Creating a Fertile Environment for Innovation that Provides Researchers with the Freedom to Explore Unexpected Opportunities and Hidden Ideas

Andrew Connolly, Professor of Astronomy, University of Washington, is developing telescopes and instruments to produce digital maps of the night sky.

Toby Bradshaw, Chair, Department of Biology, University of Washington, deals with the whole range of biological phenomena at very different scales -- all the way from molecules up to ecosystems, across time scales and across all of biological diversity.

Magdalena Balazinska, Associate Professor of Computer Science & Engineering, University of Washington, is building the next generation of data management systems to give scientists access to new tools that accelerate discovery.

Bill Howe, Associate Director of the eScience Institute, University of Washington, works on Big Data infrastructure for scientists in a wide range of disciplines -- including oceanography, astronomy and biology.

Toby Bradshaw, Chair, Department of Biology, University of Washington, deals with the whole range of biological phenomena at very different scales -- all the way from molecules up to ecosystems, across time scales and across all of biological diversity.
Dear Friends –

Washington Research Foundation (WRF) grant support creates a fertile environment for innovation that provides researchers with the freedom to explore unexpected opportunities and hidden ideas as they present themselves.

We also want to encourage and stimulate the engine of innovation that runs inside academic institutions and develop those results into solutions for commercial problems.

That’s why I’m so pleased that we’re able to bring the thoughts and words of eight cutting-edge professors from the University of Washington together in this publication.

Please read on and learn, as I have, from:

- Thomas Daniel, Professor of Biology, who seeks to understand the ways in which living systems acquire information and use that to control movement, make decisions and learn.
- Bill Howe, Associate Director of the eScience Institute, who works on Big Data infrastructure for scientists in a wide range of disciplines – including oceanography, astronomy and biology.
- Magdalena Balazinska, Associate Professor of Computer Science & Engineering, who is building the next generation of data management systems to give scientists access to new tools that accelerate discovery.
- Ed Lazowska, the Bill & Melinda Gates Chair in Computer Science & Engineering, who is trying to spread data-intensive science across the entire UW campus.
- Hank Levy, Chair, Department of Computer Science & Engineering, who wants to produce technologies that will really change the world and solve the huge challenges we all face.
- Andrew Connolly, Professor of Astronomy, who is developing the telescopes and instruments that we use to produce digital maps of the night sky.
- Toby Bradshaw, Chair, Department of Biology, who deals with the whole range of biological phenomena up to ecosystems, across time scales, and across all of biological diversity.
- Ginger Armbrust, Professor of Oceanography, who studies the ocean’s microbes, which produce an abundance of oxygen.

I hope these articles nourish you and give you a true sense of our innovative future. And, if you’d like to talk about any of the ideas advanced in this publication, please be in touch.

I look forward to chatting.

Sincerely,

Ronald S. Howell
Chief Executive Officer
Washington Research Foundation
I'm a professor in the biology department at the UW. I'm also a professor in the Neurobiology and Behavioral Program, adjunct professor in bioengineering, and, to add a few titles, I'm also the director of the UW Institute of Neuroengineering and the Air Force Center of Excellence on Nature Inspired Flight Technologies.

Our work seeks to understand the ways in which living systems acquire information and use that to control movement, make decisions and learn. We use a combination of approaches from basic neuroscience, engineering, device development and computational methods to reverse engineer living neural systems. That, in turn, inspires new devices and enables a deeper ability to connect with, and assist, individuals with neural deficits. The confluence of the three domains that are transforming technology in the 21st century – devices, neuroscience and computing – heralds fertile territory for innovation.

More specifically, my research is done on all levels, from brain processing of neural information and sensory processing, all the way to the transmission of information, out to muscular dynamics and whole body dynamics. Our lab does everything from neuroscience to biomechanics, bioengineering and information processing. We deal with a lot of neural data and with motion data and things like that.

Perhaps, most importantly, though, we do all of this in a partnership with students and researchers – from high school interns to undergraduates to graduate students to postdoctoral trainees – all the way to established international leaders. Why is what we do important?

There are really two tendrils we can talk about. In studies of movement, there are myriad issues about how nervous systems acquire information. And, once acquired, the question is how is sensory information transformed into commands that control muscles and give rise to movement? This is true for all creatures. Whether it’s a flying insect or a running human or a swimming fish, they all face this general challenge of information processing for movement control.

At the same time, there are a huge number of disorders of movement in humans. These disorders include muscular dystrophy, multiple sclerosis, Parkinson’s disease, limb loss, brain injury – anything that can, in any way, affect movement is part of a problem that we all have to deal with.

In terms of the basic biology, we know that as we move around, we acquire visual information, such as the horizon or obstacles. We’re also acquiring information about gravity and about where our body parts are relative to each other. (Where are my arms? Where are my legs?)

The brain and nervous system acquires this information. And it has to make sense of this as it initiates the right motor commands. Those motor commands go to thousands of neurons, controlling myriad muscles.

So, to move around in space requires the acquisition of a huge amount of information from multiple modalities, so that’s multiple inputs. It drives decisions about motor programming to multiple actuators, multiple outputs. In engineering parlance, this is called a MIMO system – multiple input; multiple output. The mathematics of that, the understanding of that, the way in which biology solves the MIMO problem, is inspiring new technologies.

It’s also a problem we need to understand for how we move, how we deal with things. When parts are broken, when I have loss of information, or lower conduction due to a demyelinating disease, how does that affect movement dynamics? The research relates not just to understanding the breadth and beauty and diversity of movement in biology, which is truly awesome, but it also deals with fundamental human-condition issues of movement and movement control.

There are even deeper delve downs as to why you would want to study an insect flying. Why would anybody want to do that? The answer is actually quite simple. Their brains acquire the same kind of information ours do, only slightly better. They actuate motions in ways that we can never accomplish. They are undergoing incredible flight maneuvers that no aircraft can perform.
Neuroscience – New Technology Innovation

Not only is this a problem of human health, not only is it a problem with just fascination with biology, but it’s inspiring new controls technologies. We’re funded, in part, by the Air Force to do work on autonomous flight and decision-making in complex environments with low information, like dark conditions. How do you control something like that? How does biology solve the problem?

I’d love to point out the fact that the living things I’ve mentioned are doing amazing computing. These living computers are acquiring information at rates that’s very difficult for us to do. They’re doing it in the lowest power ever done for any compute because biology computes more efficiently than synthetic systems, but not as fast.

You’re wondering how do people get into this? I focus on the neuroscience side of things, which deals with data coming into and out of nervous systems, often at incredible rates. I need people who are comfortable with mining giant data sets, doing hypothesis testing with giant data sets, but I also need people who know how to get the data.

How do you acquire information from a nervous system, from lots and lots of channels, thousands of neurons all at once? How do you get those data? That’s a device problem. I can’t understand natural computing unless I can get access to it. I can’t get access to it unless I understand what I need.

So what resources do I need? The dominant thing to know is that if you’re doing a really successful operation, you’re going to get great grants. Fortunately, NIH is excited about neurosciences. And, frankly, the defense department is fundamentally interested in not just autonomous distance and autonomous robotics and things that can navigate safely without hurting other people, but they’re also interested in devices that may help injured warriors. People who have suffered limb loss, who have suffered brain damage, people who have mobility disorders. These are the kind of programs that are poised to provide funding in this emerging domain of neuroscience, engineering and computing.

But our biggest need is people – people who have as much expertise in computing as they do in living systems. They could teach a computing class and they could teach a biology class. It’s not somebody who says, “I know how to communicate with a computer scientist.” It’s somebody who has significant contributions to make in both the computational and biological domains.

There’s more. We, in the university, are asking – every university is asking – what is the future of higher education and research intensive universities? What is their flavor? Are they going to continue in the same sort of hiring mode? Perhaps there will be increased attention on teams solving problems, problems of societal relevance.

The talent in the Northwest understands this. And there’s also a strong confluence of people who understand devices, computing and biology in this region. That’s extremely encouraging for me in this field, especially as we try to achieve major breakthroughs for the future.
I work on big data infrastructure expressly for scientists in all disciplines – in oceanography, astronomy, biology, and so on. I'm in the Department of Computer Science & Engineering, but my primary position is in the University of Washington eScience Institute, where we have a mission of advancing the research and practice of data-intensive science across all disciplines.

One thing that we find is that there is technology coming out of industry that is funded and developed and used commercially, but it hasn't penetrated well into the sciences. This is maybe not surprising because technologists develop tools for other technologists – the customer is an IT department. But research labs typically don't have IT departments. They have really smart and committed people, but not dedicated IT staff. And so they're solving their problems in a somewhat ad hoc and reactive manner.

What my group does is to adapt and refine, and sometimes build anew, technologies for working with large, noisy and complex data, but without losing focus on the requirements that we see coming out of science.

For example, database software is ubiquitous on the Web, and in every enterprise, but it's not ubiquitous in science. Some people believe that the fundamental assumptions made by databases are an ill fit for the requirements of science, but that hasn't been our experience. Our experience is that the delivery model is the problem. The real difficulty is trying to install a database, set it up, configure it and maintain it. So we stand up a database in the cloud, then wrap it in services to make it easy to get data in, write queries and get data out. The result is that we can use database technology in science contexts where ordinarily wouldn't work.

We have people in fishery science, for example, that have moved all their processing onto our system. We have people in oceanography who don’t write a line of code in any programming language, but are happily writing these long database queries on the Web. We have people in chemistry who have been publishing data through our system and citing the data in papers, as opposed to standing up text files on the Web. Not only can you cite it with a DOI, but it’s fully accessible, programmable and shareable through your browser. So, we’re constantly taking good ideas from the frontier of computer science research, and from what’s going on in industry, and really adapting them for use in the sciences. I’m also seeing that the tools and techniques we’re developing for the frontier of data-intensive science are positioned to drive innovation in industry applications as well, because the requirements are converging. In particular, we see the common need to empower non-specialists to extract knowledge from large, complex and noisy datasets.

Another connection is that the data-intensive science community is beginning to produce a cohort of “hacker-scientist” at the Ph.D. level that sometimes does not fit traditional academic career paths. Many of these people are moving into industry, and, in some cases, they’re founding technology startups. Of course, we’re interested in keeping these people around, here on campus, but it’s a sign of how support for data-driven discovery in the sciences can fuel innovation.

Finally, I think that collaborations with scientists in the physical, life and social sciences will drive the frontier of research in the methodology fields like computer science and statistics, as much as the reverse, by exposing the important problems and the common requirements. New techniques and technologies can’t be developed in a vacuum.

Looking at the big picture, what’s happening in science is that it’s shifting from a data-poor enterprise to a data-rich enterprise. In the past, science has been hungry for acquiring new data, but that’s not really the bottleneck of discovery anymore. We are very good at acquiring data. The new bottleneck of discovery is being able to analyze all this data.

A lot of people are starting to realize that we’re not going to be making the same kind of progress as we’ve been making in the past unless we solve this problem. In a nutshell: As goes this problem, so go all fields of science.

(continued on page 11)
One of my personal research agendas is to support ad hoc, interactive science. If you have an idea, or if you have a question, you can go express your query to the system, to the data, and get an answer back in a reasonable amount of time, as opposed to the weeks and months of time it takes to engineer how you're going to answer that question, and then finally getting the answer.

I see that what's going on in industry and science is mutually reinforcing. Again, technology comes out of industry and feeds into science; but science is also producing people who have a background with a statistical rigor and a comfort level working with data — and they will go on to energize the workforce in our country.

It used to be that if you were good at math, Wall Street might hire you to be a "quant." But the new model is that if you're "good at data" then any company is probably interested. It's totally pervasive. We have an astronomer, for example. She got a Ph.D. in astronomy, and she's now working at a biotech firm. She's not a biologist, but she understands data, and she understands statistics, and so she is valuable.

We're trying to figure out how to change at an institutional level so we can reward these people properly, and so we can give them credit for the role they have in this new area of data-driven discovery.

The perfect person has a tinkerer's mindset, but this person is also willing to get his or her hands dirty in a particular domain. That's good, because sometimes computer scientists pull back a little bit, and they only want to work on the core research problems rather than deliver a fully realized system. But we can't make much of a difference in science if we only publish papers. We have to prove our good ideas by building systems that scientists will use.
Big Data - The Next Generation
By Magdalena Balazinska, Associate Professor of Computer Science, University of Washington

I’m a professor in the Computer Science & Engineering Department at the UW. My area of expertise is database management systems.

Our group is pushing the state-of-the-art when it comes to building the next generation of data management systems. By designing, building and evaluating system prototypes, giving demonstrations, publishing papers and giving talks, we influence other systems builders. By using our systems to run data science applications, we give domain scientists access to new tools that accelerate discovery.

The department is vibrant and exciting. We continuously hire new faculty in exciting areas. Recently, we hired in machine learning, natural language processing, theory, and others. It’s a fun place, a very collegial department, and people easily work together. We also have quite a significant amount of collaboration with other departments, which also makes the research very interesting.

Because I work in databases, it’s all about data management and it’s all about data management problems that real users encounter. There is no point in working on databases or data management in the absence of someone who actually needs to manage some data and is experiencing some type of pain while doing it. Because the University of Washington has a lot of top-tier departments in domain sciences, with just wonderful world experts in different areas, I got in touch with the eScience institute when I started here several years ago. This introduced me to some of the world experts in different domains who already had a lot of data and wanted to analyze that data with better tools – new tools compared to what they already had. They definitely wanted to innovate in this area.

So, I started to talk to all these experts around campus, and they were happy to talk to me, and I started a small number of collaborations, primarily with the Astronomy Department. My research in databases is really focused on managing large amounts of data. Today, it’s called Big Data Management, which means data that is either large in volume, streaming, heterogeneous, or in other ways difficult to manage with existing tools. How can we help users who experience Big Data problems by moving to the next generation of tools? As part of my research, we are building new types of data management systems. For example, we have built a new system called Myria, which is a parallel data management system, where we can put all kinds of structured data and do efficient and fast processing. Most importantly, we offer Myria as a cloud service, so users never need to install anything. They use the system directly through their browser. In the background, we can do research on how to efficiently offer data analytics as a cloud service.

Data management has always been an important problem. In the past, the focus was on managing business data. All businesses have data – inventory, customers and employees. The data has to be managed, has to have integrity, has to be updated and queried. But now we are seeing a big change. It’s not just businesses that have data; anyone can have a lot of data. And users are no longer accessing this data through custom application interfaces. Users now want to analyze their data directly. They want to ask ad hoc questions on their data. This change includes sciences – astronomers, biologists, oceanographers, they all have huge amounts of data, and they don’t want to just put that data somewhere and have someone build an interface for them to operate on it. They want direct access to the data. They want to ask questions, get answers quickly. That is really transformational in the way that we build data management systems.

We want modern data management systems to process large amounts of data and also be easy for users who are technically savvy and skilled, but not necessarily experts, not necessarily database administrators or developers. So that makes the research in database management systems very exciting because we need to build these new engines.

The shortest definition of Big Data is this – you have a Big Data problem if you can no longer manage your data with the tools you’ve been using so far.

(continued on page 15)
Fomalhaut System

Hubble Space Telescope • ACS / HRC

NASA, ESA, and P. Kalas (University of California, Berkeley)
Big Data - The Next Generation

An important question is, “How is Big Data and the new tools changing the way people do science?” Perhaps the best way to answer this is through a few examples.

I started to work with the Astronomy Department and the N-body group in particular. They had these large-scale universe simulations. The simulations produced snapshots of data. They had good tools to load a snapshot in memory, operate on it, and then load another snapshot in memory, and operate on it. But the true deep scientific questions they wanted to answer required combining data across these snapshots easily.

This is where we started to use existing and also new types of data management systems to load the simulation data and allow scientists to ask questions that combined data across as many time steps as they wanted. This also helped them to get the results fast, much faster than they used to in the past. And this really changed the way they were doing science, because they could ask a whole new set of questions and get answers much faster.

I'm also working with the Astronomy Department to process telescope images. The new telescopes are producing very large amounts of data. For example, the LSST will be producing on the order of 30 terabytes of data per night. In the past, people used to take these types of telescope images, process them, extract metadata, and query this post-processed data.

Now, what we are doing together is to take the raw pixels – putting them in a new type of data management system called SciDB that we are co-developing with MIT, Brown University, and others. By putting the raw pixels into this data management system, we allow scientists to ask questions directly on the raw pixel data, and they can ask a much broader set of questions, because they can reprocess the data however they want.

Big Data research requires a lot of computer resources. We use a lot of public clouds, so we always ask cloud providers for credits to access their clouds, and they have actually been very nice to us in granting us access to a certain extent. But research in large-scale data management systems always needs more resources than we have access to. We could always run larger-scale experiments and more experiments. We also use our own private cluster, but it's smaller and mostly good for debugging.

The second important, very important, resource is that we need ways to really interact effectively across departments. For example, we have just received an IGERT Grant for a new type of Ph.D. degree in Big Data and, as a part of that grant, we want to create cohorts of students from different departments, sitting together, taking the same classes together. These classes really belong both in computer science, in statistics, in other methods departments, and in domain sciences so they can talk to each other, so they can collaborate with each other. To support these collaborations, we need to have physical space to put these students together. We don't want them to be in separate departments. We want space for them to work together, to have seminars, to really make progress in novel ways.

I think the bottom line is that the way we work on campus with all these interdisciplinary collaborations is really accelerating. We have a lot of interactions between many departments. We have the chance to be one of the most successful universities. But we do not have that many resources. We are kind of in the trenches. We know we have to make things happen, or nothing happens. Definitely having resources to push data science to the next level will continue to be transformational to our campus. We need to make sure that we are the leaders in data science.
I’ve been a faculty member at the University of Washington since 1977. I’ve been chair of the department. Right now, I run the University of Washington East Science Institute. The goal is to spread data-intensive science across the whole campus. Computer science has evolved a lot in the past 20, 30 years. It used to be that computer science was about operating systems, networks compilers, and computer architecture. But now there are things like visualization, machine learning, large-scale cloud computing, sensors. Those are the areas in which we’ve been making investments.

The business of computer science today, and, in fact, the business of all of engineering today, is making people’s lives better, and making the world better. Science is about understanding why the world is how it is. That is critically important and inspiring – and we never forget this on a day-to-day basis at the UW.

Let’s turn to discovery in computer science, which often leads to new products and companies. There’s a long history that shows that it typically takes 15 years from the first idea to the billion-dollar industry. That hasn’t changed. But there are many times when you can assemble a collection of ideas into a billion-dollar business, and do it fairly rapidly. If you look at how we build scalable computer systems today, that’s a good example.

I think that the data-driven world is absolutely transforming scientific discovery; and I hope it is accelerating it. But faster is not all that it’s about. It’s about different as well. The way I think of the role of computer science in general is what we’re doing now, by putting the smart into everything – smart homes, smart cars, smart bodies, and smart discovery. Smart discovery is every field of discovery, because of the pervasive transformation from data poor to data rich, and because of the proliferation of sensors. Sensors are getting faster, better, cheaper, and they have higher bandwidth. The challenge used to be gathering the data; now the challenge is extracting knowledge from that data. That’s your competitive advantage. Another thing that’s going on is that almost every field is becoming an information field.

One way I describe this is that biology used to have a lot to do with taxonomy. Back in the 1950’s, Watson and Crick discovered the biochemistry of DNA. But what they really discovered was that your DNA is a digital code. You can read it, decipher it, modify it and rewrite it. From that day forward, biology was an information science. And, today, the people who do the analysis of the data are the ones who determine the next experiment.

Part of the challenge for computer science is inventing new approaches to analyzing data, new ways to visualize it, new ways to extract knowledge from advanced and scalable machine learning, for example. Part of it is making our tools available for the non-specialist. You can’t expect every biologist, every astronomer, every chemist, every sociologist to have a master’s degree in computer science, or access to somebody who has one.

The challenge for us is producing tools that everybody can use. There’s a great company now in Seattle called Tableau Software. Tableau’s message is visual analytics for the non-specialist. That’s exactly where we have to go. What we’re really trying to do is bring these data-intensive discovery techniques to everyone.

That requires professional research scientists, it requires professional programming staff, and it requires a huge amount of support to spread this across the campus.

One thing that’s been hard to acquire, however, is funding for post docs. Post docs are really, really, really critical to moving a field forward quickly. These are the people who have finished their doctorates and want to spend another couple of years advancing their own career before moving into an independent research position, whether at a company or university. The ability to have this in the data science arena would make an enormous difference in our ability to impact the world – and that is an overriding goal for us.
I’m chair of the Computer Science and Engineering Department at the University of Washington. As an aside, I’m also a co-founder of two WRF-funded companies, Performant, which was acquired in 2003, and SkyTap, which we actually incubated in a single office in the WRF building and which now employs around 90 people.

Computer science and engineering at the UW is one of the very top departments in the country. We have outstanding faculty. We produce phenomenal students. We have a great staff that makes the department work, and we are growing at an enormous rate to try to produce students for the high-tech industry.

It’s up to our faculty to decide what areas to work in; but we try to be strategic. We’re trying to look out beyond what local industry is doing, to look out five or 10 years, maybe more, and to solve problems that people don’t even realize they have now. We’re trying to produce technologies that will really change the world and that will solve grand challenges that we all face.

When we look at computer science, more and more we see that it’s becoming interdisciplinary and is reaching out and trying to solve global challenges.

As an example of that, our Center for Game Science is trying to produce online games that help to teach children math. They’ve identified some particular math skills – understanding fractions is one of them – that if children lack, then they will never really get math. The Center for Game Science produces a highly intelligent interactive video game that helps to teach fractions. It’s actually a very sophisticated game. It gives feedback to teachers, but, more than that, the game itself collects tons of data and learns from children’s successes and failures in this game. From what it learns from the data, it can produce examples targeted to each student to help teach them in this topic area.

We’re also working on technologies to reduce energy consumption in the home. And, on the health front, we’ve demonstrated how a smart phone application can replace a spirometer, an expensive medical device used to measure airflow for asthma and other conditions, simply by using the microphone as a sensor.

Our department is involved pretty much all across the UW campus. We’re involved with the School of Medicine. We’re involved with the sciences. We’re involved with both social sciences and physical sciences. We’re involved with the law school, actually, in the new center on cyber security. Increasingly, computer science is at the center of what everyone does; and one of the reasons is that, increasingly, everyone has enormous amounts of data in their field, collected either through the Internet, or through sensors, or in other ways. We are all awash in data, but it is very difficult to know, given the enormous amounts of data, how to process it, how to store it, how to visualize it, and how to mine it for information.

One of the things we’ve done over the last couple of years is to try to take the lead. We made some incredible hires a year ago in machine learning and data visualization. These hires were big enough to get the attention of lots of major departments and, in fact, it caused The New York Times to write an article on our department that appeared on the front page of the Sunday business section, talking about our department and talking about these hires.

One of our goals is to be at the center of the data-centered world, and to provide the technologies that drive science and business.

I think that in every field we are producing the underlying technologies, but in every field people also have inordinate amounts of data. As an example, we are collecting enormous amounts of genomic data on people with disease. How do we use this data to find what causes these diseases? That’s a problem that nobody is going to solve by hand, so people in our department are using statistical techniques, such as
machine learning, to try to mine that data and provide models for understanding the relationships between genetics and disease.

Overall, we really want to be one of the main centers in the country for big data, for producing the technology that allows people to analyze big data, and for working together with scientists, with doctors, and with companies that have big data problems. Anything that allows us to become that center, to increase our expertise in this area, to appropriately be able to provide more tools for people to solve their problems, is going to help us.

We’re trying to grow significantly. We think it’s important to give more of Washington’s kids an opportunity to have the very best jobs, and we believe that the very best jobs today, the highest demand, is in information technology. That growth requires a lot of things. It requires faculty. It requires space. It requires equipment and teaching assistants and student advisors, and so on.

We’re very much focused right now on trying to grow, in order to provide the service that we think only we can provide to the state of Washington, which is educating the very best students that drive the Washington high-tech industry. And that industry consists of the largest companies, such as Microsoft, Amazon and Google, as well as new exciting high-tech start-ups. They’re all trying to hire our students. And with good reason – we represent the future.
I’m a professor in the department of astronomy at UW. We work on building the telescopes and instruments that we use to produce digital maps of the night sky. The first digital survey that was ever undertaken was the Sloan Digital Sky Survey, and the UW was one of the founding partners of that. In that survey, over a period of about 10 years, we managed to image or map out about a fifth of the night sky. Yet it was the most productive astrophysical facility ever in history. It produced over three thousand different scientific papers.

Today, we work on a project called the Live Synoptic Survey Telescope – the LSST. The LSST will image half the sky every three nights, so it has a three-billion-pixel digital camera on the back of the telescope, and, over a period of 10 years, it’ll amass about 100 petabytes of data. It represents about 1,000 times more data than we’ve managed to achieve with the Sloan Survey.

So, we use this telescope to understand some of the most fundamental questions in physics today. How did the universe come about? Why is the universe accelerating? What is it about this mysterious dark energy that drives the acceleration of the universe? We do this on very large scales to understand cosmology. We do it on smaller scales, looking at the distribution of stars within our galaxy to map out why or how our galaxy formed and evolved. We look at the most energetic events, such as supernovae and hypernovae, and we also look at very small objects. We look at our own solar system, and get a census of the asteroids within our solar system, including those asteroids that might potentially impact the earth. We also go from dark energy, to understand the nature of dark matter, to understanding whether we have to modify the theories of gravity or general relatively, and whether we need to change how we explain the way our universe evolves. Do we have to change the underlying physics of the universe?

An example of this is that when we study the early universe, the formation of our universe, or the reason that it came into being, we’re looking at energies and densities that are much greater than we’ll ever be able to achieve in labs, or in accelerators on the earth. So we can probe the fundamental physics. The other reason this is so important is because we’re addressing fundamental questions such as how does life arise on distant planets? What gives rise to the formation of life? Why would life be different on one planet, and how can we detect it on these nearby cells, or around these nearby cells?

I think astronomy has the potential to impact society on many different levels.

There are the fundamental questions about why the universe formed. What’s the ultimate fate of the universe? Why we are here, right through to the technology questions. The data that we achieve, the hundreds of petabytes of images that we have to process in order to extract out very subtle features within the data, will tell us fundamentally new things about physics and astrophysics. This requires new techniques; and the techniques that we develop are the techniques that will impact all of this inquiry.

The digital camera that we use in astronomy is the same technology that is now used in the digital cameras you go to your local store and buy. So, the technology that we build is the technology that goes into society over time.

As a related example, there was a project called the Galaxy Zoo Project, where we asked the general public to go out and classify different types of galaxies. Within two weeks, over 100,000 people had signed up to work on these data, because there was something fundamental about working on scientific data that was open. This had the ability to really draw in the community. I think it was one of the most successful citizen science projects ever undertaken.

Turning to the astronomy department at the UW, it’s been around from almost the start of the university. We have researchers who work on the Hubble Space Telescope and images from that. We have the virtual planetary lab that tries to simulate how you would detect life on a distant planet. We have people working on simulations of the formation of the universe, and groups working on the experiments, the surveys, to try and map out what the digital sky looks like, or what the sky looks like.

We have an active and very vibrant graduate population. I think one of the really remarkable things about the astronomy department is that over 50 percent of our graduate students

(continued on page 25)
are female, which is almost unheard of in a physics and astronomy community. We work hard at that so we have a very diverse community in terms of the students and the research that’s going on. We have very active outreach programs to the local community. We have many students coming in, high school students, elementary school students, to show them the planetarium and things like this.

We also collaborate with people in the statistics and computer science areas at the UW.

Today, when we start having data sets that are hundreds of terabytes, or hundreds of petabytes in size, it means that we can ask any question of the data. Instead of having to go out to the telescope, we effectively have the telescope sitting on a computer. To manage that data means we have to work with computer scientists, we have to work with statisticians. The collaboration works both ways, so it’s not just that we’re being served by computer science, or by statistics; the techniques that we develop help them, and the techniques that they develop help us.

Data was so precious prior to the digital revolution that people held onto it and didn’t want to share it. But today, if we just held onto the data ourselves, there would be far fewer publications or scientific results. The fact that there is an abundance of data, and the data is made available, means that instead of having to wait 12 months to get access to a telescope to answer the question that I need to answer, I can go directly to the data. It makes science much more efficient. And, by making the data available to the public, it means that anyone can ask those questions. So it could be a computer scientist, or it could be a biologist who has a different way of thinking about patterns within data, who can make the next fundamental discovery.

In addition to having the ability to access new data storage capacity, what would make a fundamental difference to the program here at the UW would be to bring in new post docs, new graduate students, and be able to fund them to work on research that is looking five years into the future, as opposed to research where we’re just looking at the next step. The ability to do the fundamental discovery, and the ability to support those people to do these long-term blue-skies research, is something that’s important to us – especially as we move deeper and deeper with our discoveries.
I’m an evolutionary biologist, but I’m also the Chair of the UW Department of Biology.

The department is a large one on campus, and we are well-known because we deal with the whole range of biological phenomena at very different scales, all the way from molecules up to ecosystems, across time scales, from nanoseconds to millions of years, and across all of biological diversity, viruses, bacteria, plants, animals, algae, fungi, everything. We cover the whole spectrum of biodiversity.

The Department of Biology at the University of Washington plays a huge role in teaching undergraduates and graduate students. Biology is the single largest major on campus. About 40 percent of all freshmen that enter the University of Washington will take the biology introductory course series. Biology at the University of Washington is the largest stem major in the entire state of Washington. And this has a huge downstream impact on all of the jobs that are related to life sciences, in agriculture, in aerospace, in the environment and in the health sciences. And, of course, the health sciences are the fastest growing job sector in the state of Washington.

So biology contributes tremendously to the future employment of our students, and, of course, students are interested in biology. The 21st century is the century of biology. Students have figured that out and they are very keen to get the latest in research information from our teaching faculty. And about half of all our undergraduates – we have 1,700 biology majors at the University of Washington – are engaged in research.

I think that our reputation globally has less to do with our disciplinary expertise than with the culture in our department. It’s a very collaborative, inter-disciplinary culture. We tend to hire faculty and graduate students and post-docs who are interested in a wide variety of biological phenomena and who work on multiple problems simultaneously.

Basic research in biology is essential for advancements in medicine, agriculture, bioenergy, and the conservation of biodiversity. Fundamental discoveries about the living world made by curiosity-driven UW biologists have yielded novel methods for making vaccines, genetically engineering crops, transforming biomass into liquid fuels, and tracking the source of illegal elephant ivory. But by far the greatest contribution of basic biological research is the wondrous realization that humans are related by shared ancestry to every living thing on Earth, and that our future depends upon the function of the global ecosystem of which we are a part. Our brains make us the first species in 4 billion years to be aware of this. What will we do with that knowledge?

But you can’t learn everything about the living world, which is, after all, the most complicated thing we know of in the universe. Living systems are the most complicated thing we know about in the entire universe. They have to be studied with these multi-disciplinary approaches. They can’t be studied in isolation. And, in biology, we take this holistic approach in order to understand complex systems. And that’s one of the reasons we collaborate.

Some examples of the kind of research that we do would be collaborations to study the effects of climate change on plant communities. So that might be a collaboration between an ecologist who works on Mount Rainier and a climate scientist who looks at the feedbacks between plants and the climate using mathematical and computational models. We also have faculty who have cross appointments in the School of Medicine, in the College of Engineering, in the College of the Environment. So we’re much broader than just our own department. We have collaborations all over the world.

We’re also fairly data-centric.

A good example of how large data sets have really changed the way that we do science is in the area of genome sequencing. When the human genome sequence was completed in the early 1980s or the mid-1980s, it was $10 billion. Now the whole thing can be done for a few hundred dollars. The ability to gather large amounts of DNA sequence data has, for a low
cost, changed the way that we do science. We used to have genome sequences for a tiny handful of organisms. Now, we can have genome sequences for any organism that we want in a matter of days. So lots and lots of people in my department now have complete sequences for whatever their organism of interest is. And that allows them to have a complete inventory of all the parts that are in the organism.

The challenge then becomes: “How do all the parts fit together? How do they work?” And that’s also a big-data problem. It’s a systems biology problem. How do genetic networks form, what analogies are there between genetic networks and other kinds of networks? And so we have information theorists who are on the faculty in our department; and they work purely on information theory.

It’s important to remember that access to large data sets doesn’t just change the speed with which science is done. It changes the kinds of questions that you can ask and answer. So, for example, our information theorists have partnered with our paleontologists to look at when all the land masses on Earth were condensed into one continent, Pangea. The supposition is that the animals in Pangea would be pretty uniform across the continent because, basically, you could walk anywhere on Earth on land.

But, when you use information and network theory to analyze patterns of fossils that are found across the world now, you see that organisms were just as specialized, even on this one supercontinent, as they are now, divided up into multiple continents. Those kinds of data sets were just unavailable and unanalyzable until the modern world. So it really makes things possible that weren’t possible before.

Science is very unpredictable, in terms of where it will go. So we don’t hire people who specialize in narrow areas of biology. We hire people who, at one time, would be called polymaths, people who are multi-talented experts in multiple areas, which includes biology, but not only biology. So they often do mathematical modeling or computational modeling in addition to the experimental work, and they often do theory as well.

It’s not just brilliant people. Brilliant people are scarce, but not incredibly scarce. What is scarce is brilliant people who like to collaborate in teams, who like to solve problems as a group, who like to work in interdisciplinary environments where very challenging problems that can only be solved by groups are solvable. And that’s the kind of person we’re looking for.

So we don’t look for a cell biologist or a physiologist or a botanist or a zoologist. We look for the person who is interested in a wide variety of biological phenomena, comes with a broad background and a tool kit of methods that everyone in our department can appreciate and understand. That’s the way we’ll help create breakthroughs in the years ahead.
I’m an oceanographer and I work in the School of Oceanography at the UW. We have oceanographers that are involved in the physics of the ocean, the chemistry of the ocean, the geology, and, like me, the biology of the ocean.

I study the microbes that live in the ocean. The reason that I study them is because I think of them as the planet’s great recyclers. About 50 percent of all oxygen is produced by these organisms in the planet’s oceans. When you take every other breath, you should thank one of these microbes. They also produce about as much organic matter to fuel life in the ocean as all the terrestrial plants. These are really important organisms for the health of our planet. I study them, both in the lab and in the field.

Our program at the UW is one of the top programs in the country. From an educational point of view, we are one of the premier undergraduate programs, because we have experiential learning activities for all our students. All of our undergraduates have the opportunity to go on a global class vessel, which means the vessel can go into the open ocean and stay there for a real research cruise; the cruise could last for up to six weeks. Our undergraduates get the opportunity to conduct research on a real research vessel. Our graduate students are some of the premier graduate students in the country, and they will become the leaders in our field. So, at this level, we are one of the major oceanographic programs in the country, too.

The first question to ask is: “How does my work impact innovation?” My research strives to understand the role of microbial processes in regulating ecosystem function and large-scale biogeochemical cycles. Our challenge is to scale from the micrometer-sized world of microbes to their global-scale impacts.

Innovation plays a key role in our research as we develop new approaches that allow us to ask what this unseen majority of organisms are able to do. We use new DNA-based approaches, starting with an entire community of micro-organisms, to identify which organisms are present and what they are capable of doing. We develop new instrumentation that we operate from moving ships to map the distribution, abundance and types of microbes across vast expanses of the ocean. Innovation requires taking calculated risks and creating an environment where researchers with different areas of expertise can work together. Innovation often comes from the intersection of different research disciplines.

It’s also important to understand that the oceans really dictate many of the large-scale features that happen on our planet. The oceans are incredibly important for climate. Think how big the oceans are. They’re enormous. They are a giant heat sink. There are huge amounts of heat; there are huge amounts of chemistry; and they are where huge amounts of fresh water are stored. The oceans have this major impact on our climate, on our biology, on the major elements, the recycling of the major elements – nitrogen, phosphorus, sulfur. That’s what’s going on in the oceans. It’s a major source of protein for much of our planet. It’s our conduit between countries.

You can think about this in a number of ways. First, though, you can think about the power of basic research and understanding how the natural world works. I think that has an impact, because we’re increasing our knowledge of how the planet works, which is incredibly important as we consider that our planet is changing rapidly. The oceans are changing rapidly, too.

But understanding how the oceans work on a smaller scale is important for people’s livelihoods. We have a number of people that are dependent on the fisheries. We need to maintain the health of our fisheries. We have a number of people that live close to the coast. That’s where most people want to live. Understanding ocean dynamics has a direct impact on those people. And we have many people that use the ocean as a way of migrating between countries. Understanding our oceans is also key to the future, because, like it or not, our oceans are going to become more industrialized. That is what’s going to happen. We are going to turn to our oceans more and more for resources.
That said, being able to train the next generation that understands how to mitigate the potentially negative impacts from engineering the ocean is really important.

Collaboration is just a given if you want to understand the oceans. My department is composed of physicists, it’s composed of expert biologists, it’s composed of chemists, and it’s composed of geologists. I don’t think there are very many departments on a university campus where you would have that kind of internal collaboration, where I may not understand all the details of what my physicist colleagues do, and they may not understand all the details of what my biology colleagues do. That’s the art of collaboration.

And we have collaboration that stretches across the campus. We have collaborations with atmospheric scientists, for example; we’ve had collaborations with terrestrial geologists; and we have collaborations with genome scientists, because we try and understand what the organisms are telling us about the environment. We have collaborations that also go across to the fisheries, and then we have collaborations with public health scientists, computer scientists, and statisticians.

In the past, oceanography was really dominated by seagoing activities. You would go on this trip and post the data from that particular cruise.

Over the years, however, we have moved into a world where we have sensors on the sea floor, and we have autonomous vehicles that can move through the ocean without a person; they actually beam data back to land. In my group, we build instruments that we run and operate continuously on ships, collecting huge amounts of data. More and more in my field, we’re also doing genomics of entire communities of organisms. Each of these approaches creates enormous amounts of data. We have moved from a data-poor field into a data-rich environment, an environment in which we integrate a host of different and multi-dimensional data sets.

At the end of the day, though, I love to go to sea, and I need access to the sea. And, in addition to integrating lots of data sets to answer lots of questions, we need to keep developing new instruments that we can deploy in the oceans to advance our inquiry and understanding as the 21st century takes hold.
Hank Levy, Chair, Department of Computer Science & Engineering, University of Washington, wants to produce technologies that will really change the world and solve the huge challenges we face.

Ginger Armbrust, Professor of Oceanography, University of Washington, studies the ocean’s microbes which produce an abundance of oxygen.

Ed Lazowska, the Bill & Melinda Gates Chair in Computer Science & Engineering, University of Washington, is trying to spread data-intensive science across the entire UW campus.

Thomas Daniel, Professor of Biology, University of Washington, seeks to understand the ways in which living systems acquire information and use that to control movement, make decisions and learn.