Ninth Annual
Strategic Leaders Global Summit

September 27-29, 2015
National University of Singapore
Singapore

Implications of "Big Data" for Graduate Education

We acknowledge the generous support
of our sponsors for the 2015 Strategic
Leaders Global Summit:

Table of Contents

2015 Strategic Leaders Global Summit on Graduate Education: Timed Agenda

Sunday, 27 September, 2015

Time	Details
18:30–20:30	Opening Dinner at "Edge," Level 3, Pan Pacific Singapore Hotel

Monday, 28 September, 2015

Time	Details
8:15	Assemble in Lobby of Pan Pacific Singapore Hotel
8:30–9:00	Shuttle Bus to National University of Singapore
9:00–9:15	Participant Registration, University Hall, The Nexus, Level 6
9:15–9:30	**Welcome and Introduction**
	Professor Tan Eng Chye, Deputy President (Academic Affairs) and Provost, National University of Singapore **Suzanne T. Ortega**, President, Council of Graduate Schools **Mohan Kankanhalli**, Vice-Provost, Graduate Education, National University of Singapore
9:30–11:00	**Panel 1: National Trends and Perspectives**
	Moderator: **Barbara Knuth**, Senior Vice President and Dean, Graduate School, Cornell University
	Brenda Brouwer, Vice-Provost and Dean, Graduate Studies, Queen's University
	Hans-Joachim Bungartz, Graduate Dean, Technische Universität München
	Shiyi Chen, President, South University of Science and Technology of China
	Bernadette Franco, Provost for Graduate Studies, Universidade de São Paulo
	Barbara Knuth, Senior Vice President and Dean, Graduate School, Cornell University

	Laura Poole-Warren, Pro Vice-Chancellor (Research Training) and Dean of Research, The University of New South Wales
	Nagi Wakim, Dean, College of Graduate Studies, United Arab Emirates University
11:00–11:30	Break
11:30–13:00	**Panel 2: Benefits of Big Data to Graduate Institutions**
	Moderators: **Shiyi Chen**, President, South University of Science and Technology of China, and **Laura Poole-Warren**, Pro Vice-Chancellor (Research Training) and Dean of Research, The University of New South Wales
	Karen Butler-Purry, Associate Provost for Graduate Studies, Texas A&M University
	Maggie Fu, Acting Dean, Graduate School, University of Macau
	Noreen Golfman, Provost and Vice-President (Academic), Pro-Tempore, Memorial University of Newfoundland
	Magnús Lyngdal Magnússon, Managing Director, Graduate School, University of Iceland
	James Wimbush, Dean of the University Graduate School, Indiana University
	Kate Wright, Associate Deputy Vice-Chancellor, Research Training, Graduate Research School, Curtin University
13:00–14:00	**Lunch at the Linkbridge (on Level 2)**
14:00–15:30	**Panel 3: Weighing the Costs: Resources and Privacy Issues**
	Moderator: **Hans-Joachim Bungartz**, Graduate Dean, Technische Universität München
	Mohan Kankanhalli, Vice-Provost, Graduate Education, National University of Singapore
	Julia Kent, Assistant Vice President, Communications, Advancement and Best Practices, Council of Graduate Schools
	Mary McNamara, Head, Graduate Research School, Dublin Institute of Technology
	Kevin Vessey, Dean of Graduate Studies and Research, Saint Mary's University
15:30–16:00	Break
16:00–17:30	**Panel 4: Enhancing Learning and Student Success**
	Moderator: **Brenda Brouwer**, Vice-Provost and Dean, Graduate Studies, Queen's University
	Martin Gersch, Chair of Business Administration & Head of the Competence Center E-Commerce, Freie Universität Berlin

	Shireen Motala, Director, Postgraduate Studies, University of Johannesburg
	Y. Narahari, Chair, Division of Electrical Sciences, Indian Institute of Science, Bangalore
	David Payne, Vice President and Chief Operating Officer, ETS
	Zaidatun Tasir, Dean, School of Graduate Studies, Universiti Teknologi Malaysia
	Paula Wood-Adams, Dean, Graduate Studies, Concordia University
18:00–18:15	Shuttle Bus to Kent Ridge Guild House
18:30–20:30	**Dinner at "The Scholar"**
20:30–21:00	**Shuttle Bus to Hotel**

Tuesday, 29 September, 2015

Time	Details
8:15	Assemble in Lobby of Pan Pacific Singapore Hotel
8:30–9:00	Shuttle Bus to NUS
9:00–9:30	Group Photo
9:30–11:00	**Panel 5: Preparing the Next Generation of Experts**
	Moderator: **Bernadette Franco**, Provost for Graduate Studies, Universidade de São Paulo
	Liviu Matei, Provost for Graduate Studies, Central European University
	Steve Matson, Dean, The Graduate School, University of North Carolina at Chapel Hill
	Nirmala Rao, Dean, Graduate School, University of Hong Kong
	Mark J. T. Smith, Dean, Graduate School, Purdue University
	Nicky Solomon, Dean, Graduate Research School, University of Technology, Sydney
	Wang Yaguang, Executive Vice Dean, Graduate School, Shanghai Jiaotong University
11:00–11:30	Break
11:30–13:00	**Panel 6: Enhancing Research Collaboration and Productivity**
	Moderator: **Nagi Wakim**, Dean, College of Graduate Studies, United Arab Emirates University
	Paul C. Burnett, Dean of Research and Research Training, Queensland University of Technology
	Chen Chuanfu, Dean of the Graduate School, Wuhan University
	Niels Dam, Vice President, Product Management, ProQuest

	Lucy Johnston, Dean of Postgraduate Research, University of Canterbury
	Buyinza Mukadasi, Director, Research and Graduate Training, Makerere University
13:00–14:00	**Lunch at the Linkbridge (on Level 2)**
14:00–15:30	**Practical Actions**
	Moderators: **Mohan Kankanhalli**, Vice-Provost, Graduate Education, National University of Singapore, and **Suzanne T. Ortega**, President, Council of Graduate Schools
15:30–16:00	Shuttle Bus to Hotel
17:00–19:00	Optional Tour: Singapore River Cruise

Introduction

Welcome and Introduction

Suzanne T. Ortega
President
Council of Graduate Schools

It is a great honor for the Council of Graduate Schools to co-host the Ninth Annual Strategic Leaders Global Summit on Graduate Education in partnership with the National University of Singapore (NUS). Before I begin my formal remarks, I would like to express a few words of appreciation. First, I would like to thank Mohan Kankanhalli, Vice-Provost for Graduate Education at NUS, for his commitment and collaboration throughout the process of planning this event. It was Professor Kankanhalli who first suggested the topic for the 2015 summit, and I know that our agenda has benefited from his perspective both as a leader in graduate education and as a computer scientist. It has truly been a pleasure for CGS to develop this year's program in conversation with him and his talented staff.

I would also like to give particular thanks to the two co-sponsors for this year's event, Educational Testing Service (ETS) and ProQuest. Both companies have deep commitments to graduate education, and a lot to tell us about the world of big data. They are also truly global companies that will bring an important perspective to our discussions over the next few days.

Nine Years of the Global Summit
Now that the summit is nearly a decade old, we have the opportunity to take a step back and reflect on the questions and themes that have emerged in this forum over the years. The first summit was convened in 2007, in Banff, where an international group of graduate deans—myself among them—came together to identify new opportunities to collaborate on global issues in graduate education. Since 2008, CGS has joined a remarkable group of institutions in co-hosting the summit on a variety of topics. These have included research ethics and scholarly integrity (2008), international collaborations (2009), quality assessment (2010), career pathways for graduate students (2011), "brain circulation" (2012), the promises of technology for graduate education (2013), and interdisciplinary learning (2014).

If you look back over the proceedings of previous summits, I believe you will find that three dominant themes have emerged. The first is **accountability**. As we each confront new demands for accountability in our respective national and regional contexts, how can we ensure that our accountability efforts actually improve the preparation of master's and doctoral students? How can we make sure that accountability drives stronger graduate programs instead of becoming an end in itself? A second theme is **building global partnerships**: how do we build trust and meaningful connections in a world that is getting bigger and more "global" every day? How do we make sure that our students know how to navigate such a world as researchers, innovators and professionals? Finally, a third theme is what I would call **new paradigms for knowledge creation and information sharing**. As we discover new methods and tools for learning and research, we must ask ourselves new questions: how do graduate schools productively and responsibly advance the creation of new knowledge? How can we be sure that students and future researchers are prepared to inherit and transform their fields?

I suspect that our current summit topic, "big data," will broach all three of these important themes—accountability, global partnerships, and new approaches to knowledge creation.

"Big Data" Challenges in Graduate Education

There was a time, probably not too long ago, when experiments in using large amounts of data were seen as the domain of computer science. As digital technologies have enabled the collection and processing of data relevant to every field, from the humanities to STEM to the social sciences, I think we now know that those days are over. It is also fair to say that big data questions go far beyond the purview of academics and researchers. This point was made by Alex Szalasay, an astrophysicist at Johns Hopkins University, when he was interviewed for a special report on managing information in The Economist: "People should be worried about how we train the next generation, not just of scientists, but people in government and industry" (*The Economist*, February 27, 2010, 1).

As Szalasay's remarks suggest, the institutions that train the next generation of leaders in science, humanities, government and industry have a special role to play in the world of big data. We must be aware of and responsive to increasing workforce demands for data scientists. We must be aware of the ways that fields like biology have been fundamentally changed by the need for advanced computational skills and algorithms. We must be aware of the ways in which big data have created new problems in scholarly recognition and credit. (For anyone looking for an example of these changes, I refer you to an article describing how difficult it is for the middle author on a100-author paper (often the statistician) to get the credit he or she deserves). This is the "brave new world" into which we send our students, and for which they and we must be prepared.

Our new responsibilities in a world of big data also extend, of course, to the way we process information about graduate education. This is an example of the way accountability efforts, which depend on increasingly large amounts of data, intersect with the big data topic. Developing the infrastructure to collect, analyze and store larger and larger amounts of information, establishing protocols for protecting the rights of those about whom information is collected, and finally, ensuring that we use resulting data to a positive end—improving the lives and learning of our students and faculty—are challenges that most global universities are experiencing acutely. It is enough to make those of us with a research background question whether there is such a thing as too much information. Nevertheless, I am hopeful that our discussions over the next two days will enable us to better harness and use the data that surrounds us for the greater good of our students and communities.

A New Experiment

As you look over the agenda for the 2015 summit, you will notice that the last session is titled "Practical Actions." While in past years, we used the last session to develop a set of common principles related to the summit topic, this year we will invite participants to formulate a few practical action items to take back to their own campuses. We hope that you will be willing to allow CGS to share at least some of these actions as part of the electronic proceedings we will publish at the conclusion of the meeting. These actions are intended to serve as a kind of menu of ideas for other leaders in graduate education seeking to advance conversations on big data with their colleagues and partners.

As is the case every year, our work together will benefit from the diversity of our national and institutional perspectives. This year, summit participants represent graduate institutions in 15 countries: Australia, Brazil, Canada, China, Germany, Hong Kong, Hungary, Iceland, Ireland, Macau, Malaysia, New Zealand, Singapore, South Africa, Uganda, the United Arab Emirates and the United States. As we begin this year's summit, I would like to thank each one of you for the presentations you have prepared and for the high level of engagement I know you will bring to our discussions.

References

Data, data everywhere. (2010, February 27). *The Economist*. [Special Report on Managing Information]. 1-14.

Voosen, P. (2015, July 6). Team science leaves many researchers lost in the crowd. *The Chronicle of Higher Education*. Retrieved from http://www.chronicle.com

1: National Trends and Perspectives

Super Size Me – Big Data Comes of Age

Brenda Brouwer
Vice-Provost and Dean, Graduate Studies
Queen's University (Canada)

Super size is the new regular. The New York Times announced in 2012 that we are in the 'Age of Big Data'[1] and pundits almost uniformly agree that it will change the way we live, how we work and how we learn. As academics and administrators, data-driven decision making is nothing new, but the reality is, that despite having access to considerable amounts of data we rarely delve into it beyond extracting the most rudimentary descriptive information. The potential for leveraging archival data with system-wide information and social media data to better understand student decision-making, for example, is most certainly there - providing we have the analytical expertise to mine it and the comfort to do it.

First, the question of what is big data? There are many definitions and most refer to a collection of data derived from multiple and varied sources (traditional and digital) that requires pairing with nontraditional tools and algorithms to manage and analyze humongous datasets (big data analytics) to understand complex processes, systems and behaviours to enable action. It is in the untangling of the 'big data hairball'[2] that it's true power may be unleashed to the benefit of researchers, marketers, governments and educators to name a few. People certainly see the value in such grand scale analyses when it could lead to the identification, prevention, detection and treatment of disease and health-related ailments (e.g. Genome Canada). Many of the same people however, may feel various degrees of discomfort when they find themselves receiving targeted advertisements after making an online purchase or web browsing. Perhaps the distinction in reaction lies in where the needle points on one's moral compass with respect to the analysis-based action - societal good versus corporate gain. Or maybe it's more about the relative acceptability of using one's personal information without expressed consent or in ways that were not clearly articulated a priori. These issues are non-trivial when one contemplates the potential uses of big data in graduate education.

Consider the data that universities have access to: historical and current applicant data, student records, alumni records, programs of study including instructors and graduate supervisors, financial assistance, and more. Applying analytical models to profile characteristics of prospective applicants, those likely to succeed in specific programs, and identify geographical regions of high yield can enable targeted recruitment campaigns and even match prospective students with supervisors to optimize potential for success and timely completion. In principle, the more granular the data, the more accurate the model[3]. The attraction of increasing the odds of meeting enrolment targets in a highly competitive environment and achieving successful performance outcomes in graduate education where the institutional investment is high is eminently clear, but the conscience perhaps less so. The concern relates primarily to the 'actionable' element. In Canada, each province and territory has its own Freedom of Information and Protection of Privacy Act (FIPPA) that ensures standards for the collection, use, disclosure and destruction of records in the custody of universities. Universities have an obligation to use information for the purpose for which it was collected and to inform people of the intended use of the information when it is acquired. Safeguarding such information is of paramount importance so not surprisingly, heated debates about the legitimacy of using information extracted from sources such as applications

---

[1] Lohr S. The age of big data. New York Times, Feb 11, 2012, Sunday Review Section
[2] Lisa Arthur, What's Big Data? Forbes, CMO Network 8/15/2013
[3] Goff JW, Shaffer CM. Big data's impact on college admission and recruitment Oct 24, 2014

for purposes seen to be other than admissions or a description of the intake cohort can, and likely will occur. The value of big data analytics and the competitive advantage afforded to institutions that use it[4] will drive (or pressure) others to adopt the approach along with rigorous institutional standards and policies that will assuage privacy concerns. The bigger challenge may be in finding the expertise to conduct the analytics; however, think twice about third party options as this would most certainly fan the flames of FIPPA dissension.

The shortage of data analytics expertise across sectors is good news story for universities and they have seized the opportunity to respond to it. The proliferation of new Master's programs focused on Data Analytics (Management Analytics, Bioinformatics, and other sector-relevant names) that have launched in the past 3 years is a testament to the high demand. Multi-sector projects such as the Big Data Talent Gap in Canada, a partnership among industry, government and academic leaders, aims to understand and find solutions for the advanced analytics skills gap in science & technology, business, healthcare, finance, etc… Their findings promise to offer direction to addressing the need for analytic expertise and universities, graduate schools in particular, will be instrumental in developing big data analytics talent and capacity for the future.

Big data will grow bigger adding complexity to management and analytics. Aggregate national graduate student survey data, sector-wide performance data, and labour market outcomes with student records and you're ready to deep dive into a new era of marketing, niche programming, forecasting, and transformative potential. Add to the ever-growing sources of structured data the wealth of unstructured personal data from public accessed social media profiles and now actionable analytics could indeed get personal. It is essential that care and thought accompany both the use of big data and the emergent actions stemming from the analytics, and of course, compliance with privacy laws is paramount. As big data analytics become ubiquitous, we would be wise to heed Uncle Ben's advice to Peter Parker (aka Spider-man): 'With great power, comes great responsibility' and not lose sight of our accountability for actions taken.

---

[4] See Coen S. How campuses can use predictive analytics to focus college student recruitment more strategically. http://blog.noellevitz.com/2012/03/01/campuses-predictive-analytics-focus-college-student-recruitment-strategically/ March, 2012

Big Data – A German and an Informatics Perspective

Hans-Joachim Bungartz
Graduate Dean
Technische Universität München (Germany)

Trying a Definition – the Complex Thing

Typically, the notion of "big" in Big Data is not only related to size in terms of bytes, but rather to the so-called "4 V's". Each of these represents a challenge that can turn simple data into Big Data: data volume (size), data velocity (real-time requirements, e.g.), data variety (entailing the need to combine data from different sources, such as sensors, satellites, or simulations), and data veracity (meaning safety, correctness, or reliability, e.g.). Also the idea what really is a challenge differs widely: What is huge for the needs of a small company may appear quite tiny for climate or high-energy scientists. A closer look at what makes Big Data complex reveals a variety of dimensions – from the underlying infrastructure (networks, compute power, storage and archiving facilities) via core informatics technology (data bases, data mining and exploration) and analytics abilities (machine learning, visual analytics, statistics) up to usage issues (including ethical topics). Hence, Big Data is not a simple extension to what we have been doing so far, it is a really big and new issue.

Potential Impact on Science … and more

The best way to understand the "where can we go?", also with a societal dimension, is to look at an impressive example. So take the one of the analysis of Twitter data, which have turned out to be well suited as an indicator of an imminent natural disaster such as an earthquake – just by exploiting the observations people make and communicate immediately before such an event. Concerning science, the focus on data just marks another step in what is sometimes called the computational, i.e. computation-driven, change process of research – repeating what simulation and optimization did before. Most fields of science and engineering did already undergo this change of paradigm (what are modern astrophysics or material science and engineering without computations?), the data orientation now revolutionizes the essence of research in the life sciences, the social sciences, and even the humanities. If we base our research on data, if we use data – available explicitly or implicitly – to design or improve experiments and mathematical models, new issues arise: How reliable are the data? Is the insight we believe to gain from exploring the data a really general one, or is it just valid for our specific data collection? To what extent do input data determine (and possibly falsify) our results? Can we quantify the uncertainty residing in the data available?

German View #1 – Concerns

One first German specialty is probably a certain reluctance in the public discussion, which finds its expression in a clear bias towards risks and concerns. To some extent, Big Data is considered to come with a Big Brother flavor (cf. Orwell's "1984"). This was the case with the health card discussion, it became obvious when Google started its photography campaign "Street View" in the neighborhoods, and it further increased by the recent friendly foreign secret service activities in Germany. As a result, there is a strong emphasis on data security, safety, reliability, protection, and privacy. This has to be taken into account before launching any Big Data activity.

German View #2 – Research Landscape

The data issue is present all over the research funding landscape. Similar to the NSF's regulations and other comparable schemes, the German Research Foundation (DFG) requires data plans for all data-relevant project proposals as well as the 10-year archiving of all research data for purposes of reproducibility etc. Moreover, DFG launched a nation-wide Priority Program on Big Data. The German Science Council (Wissenschaftsrat), the highest scientific advisory board of the federal and the state governments, established a "Council of Information Infrastructures", which shall coordinate infrastructure strategies for research information and data nation-wide. And the Federal Ministry for Education and Research (BMBF) recently launched two "Big Data Research Centers" as national hubs, together with a funding line for data-related research, addressing either data technology or data applications. Finally, there is an ongoing discussion in the scientific communities whether a "Data Ecosystem" should be established in Germany, consisting of several Data Centers – comparable to the ecosystem which was designed and implemented for high-performance computing about two decades ago (different tiers, pyramid structure). Nevertheless, there is an awakening or increasing feeling of not being prepared for dealing with the variety of aspects the data issue entails in a reasonable and appropriate way in many communities and institutions.

German View #3 – Education

As in other countries, there are not (yet) so many data-specific study programs at the universities in Germany. This may be astonishing, given the high relevance and the obvious lack of and demand for experts. However, the target group is often too broad – from computer scientists with a special "data flavor" to almost arbitrary domain scientists, sometimes without any informatics background, but nevertheless hopefully with a deep data understanding to be developred. The latter aspect reveals the complexity of this educational task: A topic with a strong technological component has to be taught to "non-technology-affine" communities also. At TUM, currently, two pioneer master's programs are being prepared: "Mathematics for Data Science" and "Data Engineering and Analytics", aiming at educating next-generation data technology experts with a mathematical or informatical background, respectively. In addition to that, minors or single modules are currently designed and integrated into several domain science programs (life sciences, medicine, etc.). At the PhD level, traditionally research-oriented and not course-intensive in Germany (according to the mandatory master's degree before entering doctoral research and training), the data issue is frequently limited to topics such as Good Scientific Practice or to isolated compact courses aiming at providing a mere data awareness.

TUM Graduate School View #1 – Research Training

In principle, the situation described above is also true for TUM and TUM Graduate School (there is only a small amount of mandatory credits to be earned via topical coursework). TUM GS, however, uses its university-wide structure to particularly foster cross-disciplinary research by providing courses and exchange formats in topics of high cross-disciplinary relevance, such as data analysis, data management, or descriptive statistics at an advanced level. Doctoral candidates working in Earth system science and remote sensing are being trained in interpreting data from various sources for modelling of complex processes. Researchers in data-relevant fields such as traffic control, logistics, or social sciences work in small, interdisciplinary research groups that receive start-up funding from the Graduate School. Hence, concerning research-driven training, the data issue has arrived to some extent.

TUM Graduate School View #2 – Management and Leadership

At the management and leadership levels, it is still not self-evident that German universities have sufficient data on their doctoral researchers for statistical and strategic purposes. Hence, the focus still has to be put on data acquisition; data exploration (for internal optimization purposes or for a "customer-orientation" – most of our graduates aim at a career outside academia) has not yet reached a significant level. TUM Graduate School has introduced a software-based data acquisition and CRM tool that allows for a smoother administration of all processes related to doctorates, as well as for a first-level insight into fundamental properties of the community of our doctoral candidates. Further developments are currently discussed to also include interfaces to TUM databases (research information system, alumni network) and to comply with upcoming national standards given by the Federal Office of Statistics.

The Impacts of Big Data on Graduate Education

Shiyi Chen
President
South University of Science and Technology of China

Thanks to the rapid information technology advancement, the revolution of thinking in the big-data era shows us how quantity breeds quality. In Shenzhen, China's Innovation Capital, at which the South University of Science and Technology is located, many high-tech companies begin to focus on the application of big data technology. The most famous China's social software Wechat which is created in Shenzhen and is all-in-one messaging app is so popular in the country that it has totally changed the way of communication among people and their life style as well. The total user numbers of Wechat has reached 600M in 2015 and a person without Wechat is regarded out of date. Taobao, founded in May 2003, is a focus on diversification, value and convenience of online shopping platform for Chinese consumers. My wife uses Taobao for everything, including purchasing food, drugs, clothing, furniture, etc. Information can be spread in a much more efficient and effective way, and companies such as Tencent, Baidu, and Alibaba can easily gather massive information through the analysis of its huge user group, which can be used to assess buying behavior, eating habits, personal income, education, and so on. The potential application of big data in China also includes public health and medical treatment for personalized medicine. Big data can also facilitate basic and higher education through ICTs (Information Communication Technology) and MOOCs, thus greatly avoiding the inconvenience of traditional education. Supercomputing also develops rapidly. Accordingly, lots of domestic Big Data alliances and industries have formed. Under such case, the State provides great support and the State Council issued an Outline of the Action Plan for Promoting the Development of Big Data recently.

In addition to testing, reasoning, and computing, big-data has become the fourth paradigm of scientific research method and big data education has emerged as an important future direction in graduate cultivation. At present, many universities in China have set Big Data as an independent major. For example, Peking University has established a data science major at Advanced Academy of Interdisciplinary Sciences. The University has also independently set up the first data sciences interdisciplinary doctoral program in China in 2014 and big data science committee was founded to coordinate the campus-wise activities. Tsinghua-Qingdao Institute of Data Science was founded and a big data master project was set up in April 2014. Besides, colleges and universities such as University of Chinese Academy of Sciences and University of Electronic Science and Technology of China have also established specialized organizations to start scientific research, application and education on big data. From the recent practice, it is shown that the big data education is featured with production-education-research model. A double tutorial system, with one advisor from academia and another from industrial is often seen.

The graduate cultivation in the big-data program requires collaborative innovation and interdisciplinary program. Big data talents are talents, who need to have a good knowledge of computer like database and software, as well as skills in mathematics and statistics. Such talents cannot be cultivated by a single discipline, and may need to acquire additional knowledge in finance, economics, journalism, biomedicine or management. At

data master's degree project of Tsinghua University coordinates School of Information Science and Technology, School of Economics and Management, School of Public policy and Management, School of Social Sciences, Institute for Interdisciplinary Information Sciences and PBC School of Finance, with the Graduate School staying in charge of it. Renmin University of China, Peking University, University of Chinese Academy of Sciences, Central University of Finance and Economics, Capital University of Economics and Business also group together their preponderant subjects to foster the training of big data talents by granting Master degrees in Applied Statistics.

The graduate cultivation in the big-data program also requires paying attentions to the detailed pathway, particularly it should involve with enterprises and employers. It is the applied talents who are educated in the big data master project, so we have to value the practical training and help our students solve problems in reality. In China, not only universities, but also a dozen media like People's Daily, Xinhua News Agency, CCTV, big data companies like China Mobile, China Telecommunications, China Unicom, Baidu, Alibaba, Tencent, and some cloud computing bases participate in the collaborative innovation platform founded by the five universities mentioned before. These units bring outstanding problems from real world to research.

Privacy protection is an important issue in the big-data education. Big data comes from procedural and immediate recordings of behaviors, which focus on micro-performance of individuals, so the problem of privacy protection is inevitable. We should ensure the security of individuals' privacy during data collection, disposition, analysis, usage and management. Decisions on Enhancing Protection of Network Information published by Standing Committee of the National People's Congress in December, 2012 clarifies that collecting and using citizen's personal electronic information in business should express purpose, method, scope, and get permissions from them, that is following the informed and licensing principles. Unfortunately, a detailed guidance for implementation is not available so far.

Towards Big Data Analytics at University of São Paulo

Bernadette Franco
Provost for Graduate Studies
Universidade de São Paulo (Brazil)

Agma Traina
Professor
Universidade de São Paulo (Brazil)

Introduction and Motivation

The amount of data generated by enterprises, academia and government increases in volume and complexity in a very fast pace. This imposes many challenges to the database researchers, from how to deal with such a volume of data in a timely manner to how to empower the data owners to take advantage of the information stored in the databases. Considering the usual Big Data definition with three Vs: volume, velocity, and variety, or the longer one with five Vs (adding veracity and value), at USP, the issue is not the volume, but the variety of the data generated yearly. It is a challenge to process them in innovative ways to add value and employ the technological benefits of the new state-of-the-art algorithms and methods developed by the databases and data mining research groups. At the Graduate Studies Dean´s Office of USP, we aim at getting the intrinsic relationship between the data provided by our community, and to cross the embedded information (hidden or not), in a precise and timely manner. By doing so, we aim at providing more efficiency, satisfaction and resources to the administration in the decision making processes regarding our graduate programs.

Big Data Analytics at USP

The Graduate Studies Dean's Office (GSDO) with the University of São Paulo – Brazil, is no different than other universities regarding the care of information of students, courses, disciplines, advisors and the transactions performed every day. Our goal is the keep the graduate courses as the most sought-after and prestigious in Brazil and Latin America. USP is a public university and the largest in Brazil and in South America, with campuses in eight cities in the state of São Paulo. USP has 263 graduate programs, around 25,000 graduate students and 5,500 advisors.

Our major aim regarding Big Data Analytics at USP concerns on how to identify programs, courses and other Graduate Studies entities that could get more attention from the office, not only to move forward in the top ranked programs, but also to collaborate and support other programs towards a better academic/professional level. There are several resolutions/levels of data summarization and analysis that aid the GSDO in this task. For example, one effort is to generate information about the network of advisors and students in the programs that best support other research groups considering given criteria. Also, to get the research production of a program classified following a given criterion. This kind of answer should be provided in a blink of the eye, allowing the analyst to browse in the results of the posed queries.

The data employed in our analytics processes are based on internal reports and personal information from the entities that integrate our databases and information systems. The anonymized information, that is, without indication of the involved people, is delivered to the next organizational level in the university hierarchy. In this way we preserve the privacy of the people who are the agents of our Graduate Programs.

When we talk about big data the main issue is scalability in all data management processes. Therefore, our data analytics systems must scale well and promptly provide the needed results. We are taking advantage of interactive visualization to allow our analysts to get the information from the data summarization and integration. This approach supports the specialists in processing the data and makes the analyses more robust.

Challenges and Achievements

One of our major challenges is integration of data from all the systems that support our Graduate Programs and the others in Brazil. A global view of the Graduate programs, or of specific elements of the Programs, allows USP to quickly see the performance of our Graduate Programs as a whole, the productivity of each program, and other relevant information, such as duration of the courses, dropouts, among others. In addition, we can compare the performance of USP with other Brazilian institutions. This is quite important, because our major assets are the high quality of the research developed in the Programs and its impact on academia and industry, and the success of students' education and professional placement.

All Brazilian Graduate Programs are evaluated by the Brazilian Ministry of Education, through the Agency for Graduate Studies (CAPES) and ranked according to their performance. USP has 5% of Brazilian Graduate Programs, but 70% of our Programs are rated as top level (the best of Brazil and comparable to the best of international ones), by CAPES standards. There is a wide diversity in the robustness and quality of Brazilian graduate programs. Thus, to count on information systems that allow not only the data integration and analysis from the inside point of view, but also from the outside, is a concern to us at USP. The grants and scholarships for master's and doctorate students, provided by CAPES, are based on results of this evaluation, done in a continuous manner. The CAPES Sucupira platform (https://sucupira.capes.gov.br/sucupira/) is in charge of gathering the Graduate Programs information required for the evaluation and USP is integrating its internal data provided with this platform, as well as with the Lattes curriculum vitae platform of advisors and students (http://lattes.cnpq.br/).

The DataUSP-PosGrad[1] and GMine[2] systems provide very useful tools in order to follow the data distribution and perform analysis of the information concerning our Graduate Programs. As usual in data analytics platforms, those systems allow interactive ways of browsing the data and computing statistics and correlations among the Graduate Programs information, such as enrollment of students considering levels, achievements as well as degrees and production associated with the Programs. It is possible to get information regarding the relationship of advisors and alumni as well. The visualization of the results is offered in several ways in order to provide to the analyst a more insightful comprehension of the data information.

Conclusions

The Graduate Studies Dean's Office at USP is aware of the opportunities of exploring the big data generated by the Graduate Programs in order to support them for effective management and quality improvement. The ability of data integration to allow fast summarization and analytics with visualization shall lead us to have a closer and more efficient management of the rich information generated by our programs. Also, we have graduate programs at USP (in computer science and data engineering) that offer disciplines for teaching and training our students on organizing, retrieving and querying such data using scalable approaches, as well as the techniques to mine relevant knowledge from big data.

---

[1] https://uspdigital.usp.br/datausp/
[2] J. F. Rodrigues Jr., H. Tong, Jia-Yu Pan, A. J. M. Traina, C. Traina Jr., Christos Faloutsos: Large Graph Analysis in the GMine System. IEEE Transactions on Knowledge and Data Engineering, 25(1): pp. 106-118 (2013)

Using Big Data to Individualize Graduate Education

Barbara Knuth
Senior Vice President and Dean, Graduate School
Cornell University (United States)

Lessons from Other Sectors

Discussions about "big data" abound. In business and marketing sectors, the focus is particularly on the context of how to better target or entice customers, personalizing the product or sale to meet the specific interests of the individual, or designing corporate policies to better support staff. Well-known firms such as Google, Dell, and Target use big data to identify improvements in policy and strategy with their customers and with their employees. One firm, for example, analyzed its big data to identify that women were less likely to self-nominate for internal promotions, resulting in women being promoted at lower rates than men with similar (or weaker) credentials, so the firm provided workshops on how to self-nominate for promotions and other opportunities.

In scientific sectors, big data discussions are especially related to discovering patterns and discerning what those patterns might predict. We see big data applications in computational biology, genomics, engineering, information sciences, sociology, and many other fields.

How might graduate education learn from these approaches to big data and make improvements in how graduate students learn, and how we, as graduate school deans, support students in reaching desired milestones and learning outcomes? How can we use the ideas of big data, whether our data are truly "big" or just moderately-sized, to discern meaningful patterns that will help us individualize, tailor, and improve the graduate education experience?

Data Volume, Velocity, Variety, Veracity

What are big data? Various academic and marketing sources have described big data as a process – an ability to meaningfully aggregate large quantities of information in useful ways, pressing the boundaries of technological capacity for analysis. Oxford Dictionaries suggests big data refers to a thing – "extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations … ."

About a decade ago, the META Group (now Gartner, Inc.) described a "3V" model characterizing big data, focusing on data volume, velocity, and variety. More recently, a 4th V has been added, veracity. Data velocity refers to the speed of data in and out, or the speed of data availability, processing, and results. In the old days, and in some current contexts, graduate schools would do "batch processing" of various information or relational data sets, with information at a specific point in time. With improvements in systems and technology, we now have near-real-time and real-time data availability and analysis, and with some systems, real-time data visualization. How do we capture and share these real-time data to make real-time, meaningful improvements in graduate education?

Data volume refers to the amount of data, measured in megabytes, gigabytes, terabytes, and more. Graduate deans may not get into the levels of "exabytes" and "yottabytes" but we have access to reams of data unimaginable to our predecessors. But we still lack access to some important data – such as, in some countries, being able to track the career paths of individual students and alumni.

Data variety refers to the various types and sources of data available to us, whether

structured or unstructured, visual, auditory, mobile, or other, from individuals directly, from sensors in devices, or as reflected in individual behavior searching the internet or engaging in other activities. As graduate deans, we regularly collect a variety of data, through student registration and enrollment records, participation lists from courses or workshops, surveys of student and alumni experiences, focus groups, alumni websites and social media, and through many other sources.

How do we pull all of this information together to make useful meaning? The newest "V," veracity, refers to the quality of the data we have available to us. Quality has many dimensions, and reflects the degree of confidence we can place in the information. Often graduate deans do not have control over the quality of the data available to us, especially as we strive to make use of the variety of available data.

Barriers to Using Big Data in Graduate Education

Aside from control over the quality of data, graduate deans face several barriers in making effective use of big data. Most graduate deans are resource-limited, lacking access to the analytical staff and the systems that will allow collection, coordination, analysis, and meaningful interpretation of large data sets. In some contexts, useful data could be collected but decisions are made not to do so in deference to potential privacy concerns of individual students.

Benefits from Big Data for Graduate Education

How might graduate deans use the technological and analytical developments of "big data" to improve the graduate experience, overall and for each individual student?

At Cornell University, we are on the path to building and integrating systems to enhance our ability to have available near-real-time data about student progress, accurate information about student engagement in academic and co-curricular experiences, and relational data reflecting student and alumni perceptions about their academic and student experiences developmentally over the course of their graduate career and beyond.

Data pulled and aggregated from central data sources, reflecting the individual progress of each student against specific academic milestones, are shared with the student, with faculty, and with staff to help guide the student's academic planning and progress, and to signal when individual students may be faltering and could benefit from focused attention and support.

As much as possible, we scan student identification cards at workshops, events, and counseling sessions, so that we can begin to see patterns of effectiveness and impact of these aggregated support activities for individual students and different groups of students, but we can also see who is not participating and making use of these various supports … and reach out to those students to encourage engagement.

We are surveying students from time of admission and matriculation through graduation, to learn from them their perspectives about their graduate student experience, and surveying alumni regarding career paths and value of their graduate education. Over time, we'll be able to aggregate these data sets to identify patterns from the student experience that may relate to career paths after graduation.

We are building and using data visualization techniques to make these data readily available, in digestible form, to our faculty directors of graduate studies and our staff graduate field assistants.

Our data analysis efforts will be further enhanced when we can aggregate these data with truly "big" data sets – not only across our own institution, but across peer and global institutions, to be able to learn more deeply from the underlying patterns that will emerge.

Big Data: Generating Big Ideas for a Big Future

Laura Poole-Warren
Pro Vice-Chancellor (Research Training) and Dean of Research
The University of New South Wales (Australia)

This paper will outline the definitions of big data in common use in Australia and discuss some of the key challenges in using big data in graduate education. Big data is clearly a topic that is at the centre of strategic thinking in all sectors, most notably business and government. The driver for this interest is the massive, largely untapped resource that currently resides in global data. Estimates of the expansion of data suggest that from 2010 to 2020, global data production will grow by over 4000%[1] to 35 zetabytes (10^{12} gigabytes). Higher education institutions in Australia, like those throughout the world, grapple with challenges involved in collection, analysis and application of new knowledge mined from big data sets.

Big data definitions
The Australian Government has developed a big data strategy that informs departments and provides a guide on the integration and use of big data in their operations. The definition of big data proposed in the Government's strategy is *"Big data refers to the vast amount of data that is now being generated and captured in a variety of formats and from a number of disparate sources"*[2].

The Governments definition also builds on Gartner's widely quoted definition of *"...high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight, decision making, and process optimization."*[3] by suggesting that two additional "Vs" be added *"..the veracity or reliability of the data, and the volatility or sensitivity of data.."*[2].

The Commonwealth Bank, one of the four largest Australian banks and the sixth largest company on the stock exchange, defines big data as being made up of the large volumes of data collected from a variety of sources such as social media, online web activity and transactional data *"..combined with your existing customer data. Analysing your big data helps you discover patterns and trends and find new business opportunities"*[4].

Legal and policy environment
Australian law is based on thirteen privacy principles, with key principles prohibiting use or disclosure of information for any other purpose except that for which it was collected, unless there is consent.[5] However, as de Zwart et al[6] noted, our current models governing privacy may not be appropriate to the rapidly moving field of big data and the associated analytics. Specifically, their paper states that *"Current technologies make surveillance and data capture a convenient byproduct of ordinary daily transactions and interactions. Data capture is so ubiquitous that it is easier to capture it all and interrogate it later. Little regard has been had to the individual privacy interests of citizens within this context and current privacy paradigms are ill-equipped to address algorithmic and predictive uses of big data."*[6].

---

[1] http://www.csc.com/insights/flxwd/78931-big_data_growth_just_beginning_to_explode
[2] *Australian Public Service Big Data Strategy*, © Commonwealth of Australia 2013.
[3] Gartner, The Importance of 'Big Data': A Definition, http://www.gartner.com/id=2057415
[4] https://www.commbank.com.au/blog/how-to-make-the-most-of-big-data.html?ei=acl_big-data
[5] http://www.oaic.gov.au/privacy/privacy-resources/privacy-fact-sheets/other/privacy-fact-sheet-17-australian-privacy-principles
[6] De Zwart M, Humphreys S, van Dissel B. Surveillance, big data and democracy: lessons for Australia from the US and UK' University of New South Wales Law Review 37(2)713-747.

The concerns of other researchers highlight the importance of maintaining confidentiality while still providing appropriate access to data. Until such challenges are overcome, barriers to longitudinal research will persist. As noted by Pugh and Foster, mechanisms for accessing matched data such as that on education with that from later-life such as employment and health outcomes are deficient. *"To the extent that educational outcomes and later-life outcomes are linked, our ability to craft evidence-based education policies in the future will be greatly enhanced by efforts to create appropriately confidential, de-identified, matched data."*[7]

Engagement in big data at university level

Universities across Australian have shown significant engagement in initiatives relating to big data. In the most part, these are focussed on coursework education, in particular graduate programs and research. Courses offered promise to produce a new generation of data savvy graduates who can support integration of big data initiatives across a range of businesses. Scholarly research in big data is also studying large and complex problems within many disciplines as well as addressing challenging interdisciplinary problems.

While universities have shown a high level of engagement with the need for better understanding of big data through research, and with producing graduates equipped to manage big data projects, there remain many challenges. These can be categorised as those relating to the legal and policy environment, which encompass matters of privacy and confidentiality as discussed above, and those relating to institutional preparedness and capacity to use big data. The latter relates to the challenges of providing better support systems and analytics and exploiting the massive data sets generated to understand trends and develop strategy.

Using big data to inform graduate education strategy

Mechanisms for collection and analysis of big data in the higher education sphere lag behind the needs of institutions. These mechanisms require robust IT systems and analytics, skilled personnel and appropriate policy as well as coordination of initiatives at a national level. Since many of these components are underdeveloped, universities have not widely engaged in using such data to generate useful information that can inform strategy.

Recent research examining data across large Australian data sets has pointed out that similar analysis to that possible in the secondary school sector is not yet possible in the higher education sector due to lack of central coordination.

"Despite the relatively small size of the entire Australian tertiary sector and of the tiers within it, no independent body in Australia has taken charge of organising multi-institutional large-N data-sharing at the tertiary level."[7]

Although quantitative data reflecting numbers in undergraduate and graduate cohorts, disciplinary breakdown of students and various other data is collected by the Australian government, there is little longitudinal data on graduate destinations and long-term outcomes. Combining social media and other data with student data to allow prediction of educational needs is a field in its infancy and there is significant opportunity to develop and implement better data collection and analytical tools for mining this valuable data. Another critical factor is the lack of people skilled in the use and development of analytical tools for data mining, a gap that graduate education is well placed to fill.

The future of graduate education will benefit from investment in the generation of knowledge mined from analysis of big data. However, there is a need to ensure an appropriate balance between the pressures from organisations and individual researchers for access to

---

[7] Pugh K, Foster G. Australia's National School Data and the 'Big Data' Revolution in Education Economics. The Australian Economic Review, vol. 47, no. 2, pp. 258–68, 2014.

better quality data and the need to protect confidentiality and individual privacy. Graduate education can clearly play a role in both of these spheres by better educating people in big data analytics, by developing and implementing more efficient and effective systems for collecting data and by addressing legal and ethical challenges. Ultimately, universities, like business and government need to engage with the use of big data analytics in driving future education strategy.

Big Data and Graduate Education in the United Arab Emirates: Challenges, Trends and Perspectives

Nagi Wakim
Dean, College of Graduate Studies
United Arab Emirates University (United Arab Emirates)

Graduate education in the United Arab Emirates (UAE) dates back to 1991, when the United Arab Emirates University launched the first graduate program, the Master of Science in Environmental Science. However, it was in the recent decade when graduate education in the UAE expanded as a result of the proliferation of colleges and universities across most of the seven Emirates reaching a total of 243 graduate programs at 35 institutions with a total enrollment of approximately 11,000 graduate students in 2013-2014.

The Ministry of Higher Education and Scientific Research (MOHESR) was established in 1992 to oversee higher education and scientific research policies in the UAE. More recently, the Center for Higher Education Data and Statistics (CHEDS), a unit within the Ministry, was established in 2012 in essence to address the issue of "big data" in the Higher Education (HE) sector of the country. CHEDS' main role is to collect data on higher education in the UAE and to use, report and interpret them in order to enhance the quality of higher education in the country. Basically CHEDS correlates the data and determines findings and trends and tries to find answers to interesting questions about the HE sector. Published reports on the data are available at CHEDS webpage (www.cheds.ae).

Big Data in Context

The data paradigm has evolved significantly in the last half century largely due to major advances in digital technologies—a significant shift on average every 15 years. Technologies of the 1960s and the seventies enabled bulk data storage in digital formats, which then led to the development of software programs for manipulating and sharing the data. In the 1990s mass storage devices led to the evolution of large data warehouses that inspired the development of specialized analytical tools for mining and understanding the data and, hence, the emergence of "big data".

In principle, "big data" refers to large, complex data that are difficult to understand using traditional data analysis techniques. In the UAE higher education sector, including graduate education, a part of the complexity is not due to the massive volume of the data but instead to the syntactic and semantic heterogeneities of the data, which are a consequence of the diversity of higher education institutions that operate in the country. There are multiple educational systems in the UAE (e.g., the British, American, Indian and French models) and each system has its own standards, conventions and definitions. Furthermore, some foreign institutions even store and manage their data in computer systems at their home institutions outside the UAE, thereby, posing additional challenges to the data acquisition process of the MOHESR and to efforts to standardize the data.

Data Acquisition and Management

The MOHESR collects data from the HE institutions on different elements, which include student profiles and outcomes, staff and faculty profiles, degree programs, course completion, student scholarships, library, facilities, and research outputs. The raw data are collected twice each year (in October and February) from all institutions, public and private. The submission

of data is voluntary; however, CHEDS works closely with all institutions to ensure complete and timely collection of their data, especially since the Ministry has no particular funding leverage over private institutions to mandate their submission of the data.

The process of homogenizing and validating the data is iterative, laborious and time consuming due to the diverse standards and differences in data syntax and semantics adopted by the institutions. The data are ingested and integrated into the CHEDS' data warehouse only after consistencies are checked and verified.

The data are maintained on secure computer servers with strict access measures to ensure privacy and confidentiality. Access to the data is limited to authorized members of the CEHDS staff and all requests for data are reviewed through an internal process prior to approval. In addition, all reporting of the data is done in aggregate forms without any identifying attributes to protect their confidentiality.

Benefits and Potential

Data are collected for sixty-two key indicators grouped into five functional, broad categories [Table 1]: (1) Institutional governance and finance, (2) research and innovation, (3) academic programs, (4) human capital and (5) students and learning environment. The indicators are used to conduct statistical analyses on, for example, financial robustness of institutions, student completion and attrition rates, quality of faculty and services, institutional performance targets, and identification of new programs to be offered.

Category of Indicators	Number of Indicators in Category	Areas Covered by Indicators in Category
Institutional Information	13	General scope; financial resources; library; labor markets
Research and Innovation	8	Research funding; research productivity; research students
Academic Programs	21	General scope; academic partnerships; admissions; student attainment; post-graduation outcomes
Human Capital	13	Workload; Recruitment and retention; evaluation and promotion; faculty quality; faculty diversity
Students and Learning Environment	7	Diversity; student to faculty and staff ratios; classes

Table 1: Categories of the UAE Higher Education Indicators.

The availability of the data has made it possible to perform a preliminary assessment of the UAE higher education sector and to compare it to that of other countries. Recent reports published by CHEDS (in 2012 and 2013) highlight some of the findings in areas of (1) Student Admission, Selectivity, Satisfaction and Success and (2) Students Enrollment and the Learning Environment, and provide a comparison of the UAE system with other systems (in Australia, UK and USA). Furthermore, the use of indicators has allowed the Ministry to develop an internal ranking of its HE institutions with an individual score assigned to each institution, which provides stakeholders with an added value on the overall performance and standing of the institutions.

There is a keen interest in the national data for the graduate education sector. In addition to assessing student success and degree completion rates, projections are now possible to assess needs for new programs and the appropriate localities for offering them, especially for professionally-oriented programs. Additionally, we are particularly interested in gauging the demand for research-based programs in order to do target-recruitment of prospective students and to ensure that the infrastructure and support for research will

meet the emerging needs, including funds for student scholarships and career-development programs for doctorate students.

Furthermore and at the institutional level, "big data" has begun to play a pivotal role in helping each institution plan ahead its five-year budget, perform target recruitment of new faculty and adjust distribution of faculty positions, better plan campus facilities and services and, more importantly, intervene timely to improve student success. At the national scene, one of the key benefits is being able to assess the country's progress towards meeting national targets for the higher education sector by tracking its performance on five key indicators: (1) graduation rates, (2) the number of national students in graduate programs, (3) the number of degree programs accredited by international organizations, (4) the ratio of national students enrolled in remedial/bridge programs and (5) the ratio of budget expended on research and innovation.

The availability of the data and the need to report different information to international bodies, like UNESCO, ALESCO, and some embassies, as well as to other organizations, such as for accreditation purposes, has brought about certain openness and more transparency in a culture where reporting everything, including not-so-positive facts or outcomes, was not a common practice. Consequently, important benefits are being reaped from sharing data and findings with the international community, and from benchmarking and comparing outcomes to more mature and robust HE systems, which leads to identifying best practices, improving models, and contributing to a universal body of information for a higher level of analysis and possibly new knowledge on trends and possibilities in the global higher education sector in general and graduate education in particular.

Sources
(1) The UAE Higher Education Factbook, 2013-2014. (2) Indicators of the UAE Higher Education Sector, Vol. 1, Nov. 2012. (3) Indicators of the Higher Education Sector: International Comparison, Jan. 2013.

2: Benefits of Big Data to Graduate Institutions

Using Big Data for Targeted Recruitment, Programming Professional Development and Developing Interventions to Improve Student Success

Karen Butler-Purry
Associate Provost for Graduate Studies
Texas A&M University (United States)

Current trends in higher education include an evolving view of appropriate use of data in decision making. The focus is shifting from primarily data-driven decision-making to data-informed leadership. Under the new perspective, data are not thought to "drive" decisions, but rather educators/administrators combine their values and insights with data analysis to reach conclusions and make decisions. In today's data-rich environment, many possibilities exist for how leaders can utilize big data (the massive amount of graduate data) to inform decisions in graduate education.

Texas A&M University assembles a large amount of data on graduate students, in addition to participating in several national surveys assembling student-level data and national comparatives. For example, the university collects data from graduate applicants such as GPAs, attendance at previous institutions and standardized test scores (e.g. GRE, GMAT, etc.). In addition, Texas A&M facilitates processing of graduate academic requirements using a custom-designed software application and paper forms which accumulates graduate student data such as degree plan, advisory committee members and completion times for matriculation milestones. Our office manually enters this data or pushes the data to the university's student information system.

Texas A&M recognizes current data use trends and sets out to improve our data collection methods and further facilitate processing of graduate academic requirements. We joined with our enterprise information systems and information technology campus units to begin developing a new graduate student portal tool. The tool will facilitate processing of graduate academic requirements and will perform other functions essential to the graduate student experience. It will accumulate a wide range of graduate level student data. The portal will draw data directly from numerous sources including student input, the university student information system, graduate student ORCID (Open Research and Contributors ID) records, etc. The new portal tool will add on-line processing of graduate academic requirements such as approval workflows for proposals, exam results, and theses and dissertations. Another task the portal tool will accomplish involves collecting individual graduate student funding data such as record of awards of university fellowships, prestigious national fellowships, graduate assistantships, and non-resident tuition waivers.

A third feature of the portal tool involves collecting data on student participation in training and professional development experiences such as:

- Responsible Conduct of Research (RCR) Training
- New Teaching Assistant Training
- English Language Proficiency (ELP) compliance for international students
- New Graduate Student Orientations
- Professional Development Activities and Interventions
  - Texas A&M G.R.A.D. Aggies (Graduate Resources and Development for Aggies) offered by the Office of Graduate and Professional Studies (OGAPS), Career

Center, University Writing Center (UWC), Student Counseling Center, TAMU@ CIRTL (Center for Integrating Research, Teaching and Learning), TAMUS E&S AGEP (Texas A&M System Energy & Sustainability Alliance for Graduate Education and the Professoriate), Academy of Future Faculty, etc.
  - Professional Development Certificates Programs [CIRTL, AGEP, G.R.A.D. Aggies, etc.]

A final component of the portal tool includes collecting data on graduation and career placement for doctoral graduates.
- Career Outcomes data – at graduation [salary, employment]; and post- graduation, such as Academic Analytics (graduates in academia); ORCID records (publications, employment, etc.); surveys describing the impact of professional development experiences on career; exit surveys of graduates

Currently, Texas A&M utilizes graduate student data to inform a number of university decisions. For instance, determining the portion of funds distributed to colleges and interdisciplinary degree programs to support graduate operations and/or new initiatives is determined based on graduate student data. In addition, Texas A&M analyzes annually the following graduate student data at the university, college, and program/major levels to assess progress of colleges toward meeting university metrics:
- Internal Comparisons
  - Graduate assistant salaries (median, min, and max);
  - Graduate Assistantships, Fellowships, Loans – number funded, plus total and average award amounts
  - Non-resident tuition waivers – who receives them
  - Applicants (Overall and new graduate student enrollment), Degrees granted – current and five year comparisons lined out by race/ethnicity and gender, degree objective
  - Completion, Retention/Attrition, Time to Degree – current, five, and ten year comparisons; lined out by race/ethnicity and gender, degree objective
- Comparison with peers
  - AAUDE – graduate assistant stipends, tuition and required fees, doctoral time to degree, doctoral completion
  - AAUDE doctoral exit survey – employment at graduation, salary

Moreover, Texas A&M has performed extensive one-time data analysis focused on a particular issue or topic. For example, during our participation in the CGS STEM Master's Completion and Attrition Project, we included all of our terminal master's programs. This allowed us to carefully examine our professional and research master's programs and identify those programs that are not graduating students in a timely manner or retaining a desirable number of their students.

Additionally, some Texas A&M graduate programs and colleges require annual evaluation of progress to degree for doctoral students, using data collected by the university. This process currently occurs using rudimentary custom applications or a time-consuming manual collection. We plan to incorporate a function into the new graduate student portal tool to facilitate and improve these annual evaluations of all doctoral students.

Another trend worthy of examination is predictive analysis. Using predictive analysis, corporations such as Civitas Learning and IBM Business Analytics synthesize many data sources and data science approaches to identify features associated with success in an educational setting. I believe that predictive analysis of big data in graduate education has

great potential. In particular, I think it could help tackle the extremely challenging national priority of broadening participation and success of underrepresented minorities in graduate programs, particularly in STEM fields. The expanded data available through using Texas A&M's new graduate student portal tool, combined with nationally available graduate survey data will provide Texas A&M with the potential to use predictive analysis to provide insight into the aforementioned issue. Other issues that will be pursued include the appropriate stage at which to provide interventions or offer experiences/programs to graduate students which will lead to successful outcomes, and identifying any distinct characteristics (such as resilience) of incoming graduate students to precisely target new students in recruitment.

The Values of Big Data to Postgraduate Education

Maggie Fu
Acting Dean, Graduate School
University of Macau (Macau)

Over the years, more and more higher education institutions are taking advantage of Big Data to assess the performance of students, faculty and curriculum against peer universities, which in turn helps to draw up the blueprint for future development. In postgraduate education, Big Data has become an indispensable tool for yielding new insight into potential for improvement. In the University of Macau, the Graduate School is established as a central body responsible for the better management, development, and evaluation of graduate education throughout the university. We work closely with the faculty and staff to provide graduate students with assistance and guidance for them to progress through their studies. To achieve this, we need different sets of data and statistics, for instance, demographic data of students, course selection data, studentship information, graduation rate, and etc. By systematically analyzing them, we can keep track of the situation and make adjustments when necessary. For us, Big Data is the quintessential tool which improves student performance and raises faculty effectiveness, and at the same time reduces administrative workload. Big Data contributes to postgraduate education in several aspects, including student recruitment, course selection and offering, and student performance and faculty effectiveness:

Student recruitment
Without doubt, student recruitment demands considerable amount of efforts, however, it is not likely for graduate schools to hit their target admission rate every single year. What's the reason behind? As the application process has been moved online, students are applying to more schools than ever. Those piled-up applications are making it harder for graduate schools to predict whether a candidate will actually accept an admissions offer or not. If an offer is turned down, it will be sent to other candidates on the waiting list. Yet, it takes extra time and administrative workload, which in turn increases the cost of the whole recruitment process. By analyzing the historical performance and demographic profiles of current and former students, graduate schools are more likely to identify prospective students and those who are most likely to enroll. Taking it further, graduate schools can mine current and prospective student's social networks to identify potential candidates for enrollment. This also helps graduate schools to make better plans for promotion. To a certain extent, the student recruitment process is revolutionized by the use of Big Data, as graduate schools can now reach out to potential students more proactively and systematically.

Course selection and offering
Whether a course will be offered or not depends on the number of students enrolled and the availability of lecturers, however, the arrangement is normally made before the actual course enrollment period, therefore, two phenomena occasionally occur: a course is offered but there is insufficient number of students; or else, too many students signed up for a course with limited quota available. To avoid such kind of misallocation of resources, Big Data could be especially useful. Graduate schools can recommend a possible curriculum and major to the prospective students by making a comparison between the new students' previous academic performance and former student profiles. Apart from this, providing additional information on majors and career options is important for students to make informed decisions that are best

for them. Furthermore, surveys can be employed to investigate the inclination of students on course selection.

On the other hand, keeping track of the admissions statistics helps graduate schools to design the offering of courses in advance more effectively. Take UM as an example, we make cross comparison between several sets of data regarding local students and non-local students, including the number of approved admissions quota, the number of applicants by first choice, the number of offers issued, and the number of offers accepted. This data provides useful information to us and faculty when we consider what courses we should offer in the coming year, and it enables us to maintain an optimal balance between the development of the graduate school and the growth in the number of students. As we know that, if the number of students outgrows the scale of the school, it is impossible to ensure that every one can be provided with adequate resources, and thus will eventually affect the quality of postgraduate education.

Student performance and faculty effectiveness

One of the benefits of Big Data is that it helps graduate schools to monitor and improve student performance and faculty effectiveness. In UM, the academic and the demographic data of students is stored in a central system known as the Student Information Web (SI Web), it is convenient for us to keep track of the completion of course credits and academic results of every single student, through this we can reach out to at-risk students in a timely fashion. Another important indicator of student performance is the graduation rate. By observing the graduation rates over the years, we can identify what courses the students have difficulty with, and allocate appropriate resources in such case, for instance, by reviewing and streamlining the course curriculum, or introducing the mentorship program among students and faculty members.

Apart from student performance, faculty effectiveness also plays a significant role in postgraduate education. It is the practice of UM to conduct the Teaching Evaluation by the end of each course, during which students are requested to fill in a questionnaire which reviews the teaching effectiveness of the lecturer. The grades and comments collected anonymously will be inputted in a database and compiled as a general report, the results will also be sent to the corresponding lecturer for reference. Such practice provides a chance for the school, the faculty and the students to reflect on the classroom activities and the supervision on coursework.

Furthermore, UM has adopted a web course tool named UM Moodle for the use of students and faculty, through which both parties can engage in discussions via private messages and public forums. Moodle offers a user-friendly channel for the faculty staff to upload teaching resources, set and grade assignments, as well as issue announcements. Most importantly, Moodle is equipped with the function of keeping activity logs, faculty can thus keep track of the learning progress of students, and devise teaching plans more effectively.

Conclusion

Big Data will surely gain more attention in the coming years, as the values that it brings to postgraduate education are not limited to student recruitment, course selection and offering, and student performance and faculty effectiveness. However, new challenges also come along with benefits, the biggest question ahead is how can we utilize the Big Data while at the same time use it in a moral and appropriate manner? This question will remain on the agenda of graduate schools and make us more determined and united to improve the future of postgraduate education.

Big Opportunities for Big Data: Business Intelligence in Graduate Education

Noreen Golfman
Provost and Vice-President (Academic)
Memorial University of Newfoundland (Canada)

Deans of Canadian graduate schools are likely to find familiar two clear trends in recent years. First, university budgets have faced deep cuts across the country. Cuts have appeared both in terms of reduced operating funding by provincial governments or program cuts by the federal government. Second, there has been a noticeable increase in the number of graduate students enrolling in Canadian universities. Graduate enrolment in Canada grew by 147 percent from 1980 to 2010. Graduate student populations have also become far more diverse, with increasing numbers of foreign students, female students, and part-time students enrolled in a growing range of academic disciplines (Universities Canada, 2011).

Graduate education is considered essential for the viability of both universities and national economies (Holdaway et al., 1995), and so for Canada, more and different types of graduate degree-earning students are a good thing. Adequately training and supporting those students against a backdrop of severe funding cuts is a serious challenge facing graduate administrators. With salaries accounting for the bulk of operational costs, proportionate increases in staffing are unrealistic.

For graduate deans, part of the solution is to think differently about how graduate education is administered. No one really trains us to do so, but necessity is the mother of ingenuity and we are increasingly compelled to think differently about what we do. In many ways, graduate education has traditionally been an intuitive enterprise, decisions having been made based on past practice, disciplinary norms, and, frighteningly enough, gut feelings. But even the most intuitive industries can benefit from becoming more analytical. In Canada, graduate universities that can understand and strategically use data are better able to deploy resources more efficiently.

It is important to note here that data, or "Big Data" as a more inclusive term, refers to data sources that can be used for predictive analytics (Waters, 2012). Furthermore, the opportunity is to apply predictive analytics in higher education to improve student outcomes. Predictive analytics is understood as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions" (Davenport & Harris, 2007). In practice, predictive analytics are generated by a business intelligence (BI) system. BI is defined as "the tools, technologies and processes required to turn data into information and information into knowledge and plans that optimize business actions" (Muntean, 2013). When contextualized to specific business operations, BI systems can support and improve planning and control, decision-making, and organizational performance (Elbashir et al., 2013). BI systems can appear as self-service platform and include data discovery dashboards that allow users to begin working with data more rapidly and adjust to changing needs (Muntean, 2013).

At Memorial University, our focus has been on leveraging predictive analytics and creating a BI system to increase the efficacy and capacity of graduate supervision. Graduate supervisors play a critical role in student development, and we know that mentoring contributes greatly to professional development and well-being of graduate students. When supervisors are experienced in their roles, graduate students have a significantly better chance at success. Conversely, inadequate faculty supervision contributes to poor completion rates

and long completion times among research students. At an institutional level, then, graduate student persistence and success are often the net sum of good and bad mentoring practices across a university. Standardization and quality control are difficult to achieve because so much of the work is based on individual relationships and qualities.

One of the ways we are trying to apply predictive analytics to enhancing supervisor effectiveness is through "people metrics." Google's Project Oxygen is an excellent example of how the right "people metrics" can accurately predict behaviour and outcomes (Bryant, 2011). In a similar way, we are examining the attributes and behaviours of effective supervisors at Memorial. This project between the School of Graduate Studies and Distance Education, Learning and Teaching Support promises to better inform training workshops and programs for new faculty members interested in taking on supervisory roles.

Another project between the School of Graduate Studies, Information Technology Services, the Faculty of Engineering and Applied Science, and the Faculty of Business Administration aims to develop a more effective BI system to monitor student progress. Through predictive analytics, one of the goals is to identify at-risk students early enough so that appropriate intervention strategies can be applied. Preliminary analysis of our attrition data tells us the most vulnerable time in a graduate student's program is the first semester. This is particularly true for distance education students and international students.

Intervention is an important consideration. To be meaningful, analytics must be coupled with the right responses and strategies. Those responses, in turn, may either be human decisions or fully automated ones (Davenport & Harris, 2007). Harnessing course data to automatically trigger interventions for online students would be one automated approach. Automation is an important goal for Memorial's BI system since much of the work of our faculty supervisors involves burdensome administrative tasks. By automating this work, faculty members have more time to focus on training, advising, and mentorship – critical to graduate student success – and to take on more students without much more effort.

Universities that pay close attention to data are better prepared to identify opportunities and develop strategies for improving the delivery and support of institutional goals. It is also important that both strategic and operational decisions are informed by data and within a "culture of evidence" (Black, 2010). One of the best opportunities for "Big Data" is with the people who contribute the most to graduate student success. Human intervention remains important, however, as do the accumulative knowledge of good, experienced supervisors. Informed by timely, accurate, and predictive data, graduate supervisors can learn faster, improve their effectiveness, and accomplish more than we ever thought possible. Or so we hope!

References

Black, J. (ed.) (2010). *Strategic Enrolment Intelligence: Canada's First Book on Strategic Enrolment Management*. London, ON: Academica Group.

Bryant, A. (2011, Mar. 12). Google's Quest to Build a Better Boss. *The New York Times*. Retrieved from http://www.nytimes.com.

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Boston, Mass: Harvard Business School Press.

Elbashir, M. Z., Collier, P. A., Sutton, S. G., Davern, M. J., & Leech, S. A. (2013). Enhancing the Business Value of Business Intelligence: The Role of Shared Knowledge and Assimilation. *Journal of Information Systems*, 27(2), 87-105.

Holdaway, E.A., Dubois, C., & Winchester, I. (1995). Supervision of Graduate Students. *Canadian Journal of Higher Education*, 25(3) 1-29.

Muntean, M., & Surcel, T. (2013). Agile BI - The Future of BI. *Informatica Economica*, 17(3), 114-124.

Universities Canada (2011). *Trends in Higher Education*. Retrieved from http://www.univcan.ca.

Waters, J. K. (2012). Big Data: What Is It and Why Does It Matter? *Campus Technology*, 26(2), 11-16.

The Promise of Taking Advantage of an Institution's Size to Affect Change in its Graduate Programs

James Wimbush
Dean of the University Graduate School
Indiana University (United States)

Indiana University is large, very large. Founded in 1820, the University is comprised of eight physical locations which are scattered throughout the State of Indiana. The University officially enrolled in fall 2014 a total of 114,370 students of which 20,172 are graduate and professional students. Earlier that year, 34,303 degrees were awarded to undergraduates, 4,556 to master's degree seeking students, and 1,529 to those earning doctoral degrees.[1] These degree earners added to the over 600,000 living Indiana University alumni worldwide, one of the largest living alumni populations of any institution in the United States.

A major research university, Indiana University is at the forefront of technology and research. Its researchers have the benefit of "Big Red II", a state-of-the-art supercomputer, which at the time of its dedication in April 2013 was the fastest university-owned supercomputer in the nation. The supercomputer uses its unparalleled high-speed to analyze massive amounts of data at the rate of one thousand trillion floating-point operations per second.[2] Needless to say, it provides Indiana University researchers with an important tool in addressing some of the most pressing and interesting research questions of our time. And this wonderful work is resulting in major breakthroughs in science, medicine, and technology.

Being its size and age also means that Indiana University has the potential to collect and analyze large amounts of data—going back decades, and for decades to come—related to student admissions, enrolled students, faculty, and alumni. And it has already begun to make use of data readily available.

For the past two years, the Bloomington campus has used outside sources to analyze data related to faculty—recruitment and hiring, productivity, scholarly impact, as well as key indicators of program or department success. Benchmarked with comparable institutions, the value of these data has been in the campus's ability to have a sense of how well a department, program, school/college, and even a faculty member performs relative to peers. For example, data can help to determine where academic units may be lacking talent for an area identified as being important for a field, as well as how recruiting such talent will impact the program's ranking among its peers. The data can even help in analyzing the potential effects on the ranking of a unit if there is turnover of a key faculty member or members. In the area of diversity, the possibilities are many, one of which being informing departments and programs of overall unit race/ethnic/gender composition as well as tenure and promotion outcomes. Faculty data have proven to be of great benefit to the academic units thus far and show continued promise for creating favorable outcomes.

Just as data have been used effectively for faculty, as part of Indiana University's 2020 Bicentennial Strategic Plan, the University Graduate School is committed to using the wealth of data available to it from its more than 300 degree programs and numerous alumni to create a comprehensive set of metrics to facilitate decision making and programming. Data are already being tracked, by campus, on key metrics such as time-to-degree, completion, diversity, funding, etc., but not on all campuses, and not consistently. Comprehensive

---

[1] University Institutional Research and Reporting, IU Fact Book, 2014. Retrieved from https://www.iu.edu/~uirr/reports/standard/factbook/.

[2] Indiana University Media Relations, April 17, 2013, "Indiana University dedicating Big Red II supercomputer April 26". Retrieved from http://newsinfo.iu.edu/news/page/normal/24119.html.

placement data which goes beyond first placement is also not currently being collected across all programs. Given the large number of graduate alumni, collecting and analyzing placement data would be of significant benefit to the Graduate School.

The goal of the Graduate School is to create an institutional database comprising data from all campuses for tracking applicant test scores, undergraduate institution and grade point average, admissions offers and yields, completion, time-to-degree, program demographics, stipends, and placement/career data. These data would enable the Graduate School to not only better monitor program outcomes, but to also design interventions for areas of challenge. Diversity, for example, is a continued issue of concern for graduate programs. Knowing more about the demographic composition of graduate programs and whether there are differential outcomes on key metrics would be useful for creating programming to enhance inclusion and equity. Another example is data on careers. Knowledge of the career trajectories of alumni will inform programs about potential long-term placement opportunities for its students as well as any needed curriculum revisions to better accommodate training for a wider variety of professions and careers. The professional development opportunities which may result from being better informed about the actual career paths of alumni offers huge potential. The alumni data will also have an obvious benefit for fundraising and alumni engagement.

The effort to develop a database will be starting soon with a goal of not only completing the project, but actively using the data by the year 2020, Indiana University's Bicentennial year. Resources are currently being garnered for this huge undertaking with the likelihood of collaborations from other units within the University, such as the Indiana University Alumni Association.

# Benefits of Big Data to Graduate Institutions

Kate Wright
Associate Deputy Vice-Chancellor, Research Training,
Graduate Research School
Curtin University (Australia)

Assessing benefits of big data for graduate institutions can be interpreted in two ways: firstly in terms of the information available on the student body, and secondly in terms of the data produced by postgraduate students themselves. In this paper, the opportunities and challenges offered by student data are summarized, followed by a brief discussion of the issues for institutions of storing big data produced through research.

All universities collect large amounts of information on their student body. This includes data on gender, age, background, academic progress and student opinion in addition to basic information on student load, retention and completion of degrees. At Curtin University, the current student data system (StudentOne) has been in use since 1999 and during that time has collected data on Higher Degree by Research (HDR) students who have applied, accepted, enrolled etc. In addition, the university holds data from surveys (Curtin Annual Student Satisfaction Survey, CASS), and Annual Progress Reports (APR). These data have been collected over a long period of time but reside in different parts of the organization and in different formats. Hence their analysis represents a major challenge.

With the right analytical tools, the above disparate data sets can be used to provide a historical perspective on how our HDR student body has evolved over time. Mapping the trends against major changes in university and government policy or the introduction of recruitment and retention initiatives, could also give an indication of the impact of such changes. For example, Figure 1 shows HDR load for a selection of Australian universities over the period 2007 – 2013. Almost all universities experienced strong growth from 2009-2012, corresponding with a major government initiative to increase the number of Australian Postgraduate Awards (APA). The data also show a flattening off of numbers in some institutions in 2013 as the APA numbers stabilized.

Figure 1. HDR student load from selected Australian universities 2007 – 2013. All major groups are represented including GO8, Australian Technology Network, Innovation Universities and non-aligned. Data from the Australian Government Department of Education and Training.
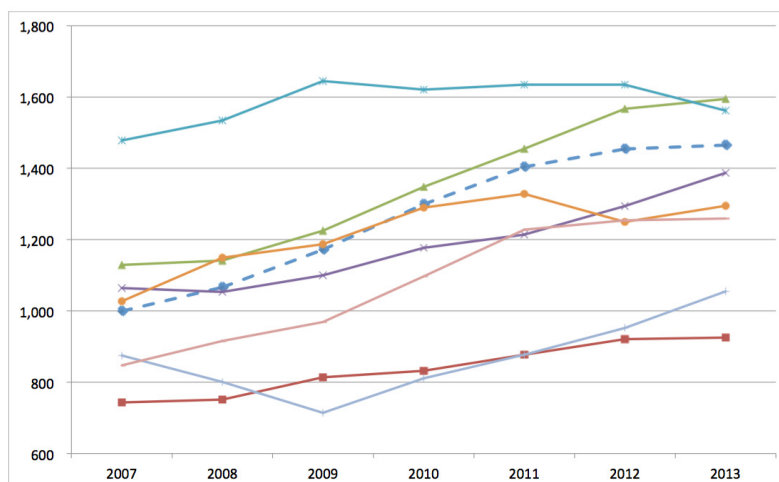
Armed with the historical perspective, we are provided with the opportunity to extrapolate and predict future trends. More importantly, our predictions can be utilized for better informed planning in terms of how we might recruit students in the future to achieve the cohort that best fits our strategic direction.

We also have the potential to predict how our students might fare as they progress through their research training journey and ask questions such as: Who is likely to withdraw? What support services will really make a difference? What factors contribute to success? What can we say about the process of becoming a researcher? Of course such analyses must be dynamic, as the data sources are continually updating and evolving. Therefore caution should be exercised when predicting out into the longer term. The life of a research student is affected not only by the internal factors relating to the university, but by external factors including financial, personal and health - in other words, life happens. Thus we should ask the question "what data do we not capture that would improve our ability to make meaningful predictions?"

Graduate research institutions also have a duty to train HDR students—there is a huge need to up-skill budding researchers in the uses and approaches to analyzing large and often complex data sets. Examples include radio astronomy projects associated with the development of the Square Kilometer Array (www.ska.gov.au), and longitudinal health projects such as the Raine study (www.rainestudy.org.au). In addition to collection and analytics, a key need for training is in the area of big data storage.

Ensuring that all data is stored and managed in such a way as to comply with requirements of governments and grant agencies represents a challenge for universities. The Australian Code for the Responsible Conduct of Research and the National Statement on Ethical Conduct in Human Research both contain clauses relating to data management and identify responsibilities of both the institution and the researchers. Added to these requirements are the increasing obligations to provide open access to research data by grant agencies. Both the Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC) have policies that require publicly funded research be made publicly available. In addition many leading journal publishers now require access to data sets that support the research findings reported. For example the Nature suite of journals requires that, where possible, supporting data be made available in public repositories (www.nature.com/authors/policies/availability.html).

The need for HDR students and institutions to manage their data securely and make it available publicly is likely to grow and institutions must provide appropriate facilities and training to ensure compliance. In summary, big data represents enormous opportunities for graduate institutions to critically examine their past performance and future possibilities. For research, access to and analytics of big data will increasingly be part of the HDR student experience. The challenges we face relate to the way in which we use big data and how we make it available to the wider community without compromising the needs of ethics and security.

3: Weighing the Costs: Resources and Privacy Issues

Big Data Privacy and Resource issues at NUS

Mohan Kankanhalli
Vice-Provost, Graduate Education
National University of Singapore (Singapore)

Singapore passed the Personal Data Protection Act (PDPA[1]) in 2012 which consists of several rules governing the collection, use, disclosure and care of personal data. It does a careful balancing act between the rights of individuals to protect their personal data, which includes rights of access as well as correction, and the needs of organizations to collect, use or disclose personal data for "legitimate and reasonable purposes". The three most important operative concepts in PDPA are:

- Consent: The individual's knowledge and consent is necessary before data can be collected, used or disclosed.
- Purpose: The individual needs to be informed of the purposes of collection, use or disclosure before it can be done.
- Reasonableness: Collection, use or disclosure of personal data is allowed only for purposes that would be considered appropriate to a reasonable person in the given circumstances.

As a direct consequence of this act, NUS established a Data Protection Office which manages the data protection policy for the university and also facilitates anybody to access their personal data records to either view or update it. It also allows for withdrawal of consent by individuals associated with the university (including students, current and former staff members, alumni, student applicants, job applicants, vendors, donors, research subjects, guests and friends). NUS also had to update its data policy and publish it online[2].

All these developments have directly impacted the handling of big data at NUS. In this paper, we will focus on issues related to graduate education. In November 2014, NUS conducted a 10-year tracer survey of all PhD alumni. This was a much more comprehensive survey than the one conducted in 2009. We wanted to understand the evolution of career pathways of our PhD alumni who had graduated during the 2004-2014 period. When compared to the 2009 survey, significantly more effort was required for the 2014 survey because of PDPA. One of the main issues which cropped up was that of consent. While we have been collecting alumni data, we had not obtained their consent to use the contact information for conducting follow-up surveys. It was not clear that such a survey would be "reasonable" since we got conflicting opinion on whether the clauses of the PDPA would be violated if we directly contacted the students. A catch-22 like situation encountered was that of whether a pre-PDPA recorded email address could be used to seek consent to use that email address for survey purposes. Since there was not sufficient clarity forthcoming on many of these issues, we erred on the side of caution and decided not to directly contact our students and alumni. Instead, we decided to contact faculty colleagues who were the advisors of the graduated students. While most of the advisors were helpful in providing the requested information, some cited PDPA and refused to provide their students' information. We had to assure these colleagues that the collected personal information would not be disclosed to anyone and that we will eventually seek the permission of the alumni for any further follow-up action.

[1] https://www.pdpc.gov.sg/legislation-and-guidelines/overview
[2] http://www.nus.edu.sg/images/resources/content/misc/pdpa.pdf

The tracer survey exercise highlighted many other big data concerns. We immediately found that mixing the results of the 2009 survey with that of 2014 yielded rich insights. However, the legacy consent issue is not very clear – whether we need to go back to the alumni to seek consent for using the pre-PDPA collected information. Since we had not sought consent of the students, we had to make sure that no identifiable personal information could be leaked out inadvertently due to the combination. This will become increasingly important when we move from descriptive and predictive analytics to prescriptive analytics.

We have had to grapple with many other privacy related issues. First is the issue of storage of big data. It is obvious that storage on cloud servers offers advantages in terms of easy access as well as back-up reliability. However, since most cloud storage providers are outside of NUS and usually outside of Singapore, one cannot be reliably assured of the security of this data. One could consider keeping the data on cloud servers in an encrypted form. However, encrypted-domain computations are still not practical and thus the data needs to be decrypted for the purpose of doing analytics. As a result, we are compelled to store data on internal data servers which is expensive and has limitations in terms of capacity.

The second issue is related to access control. Usually, login based authentication is used for managing access to sensitive data. The login credentials have different access rights based on the role in the organization. While this is useful to log who accesses the information, there is no real control if the person who is granted access downloads the data onto private storage. Once it is in private storage, it is outside the scope of system protection. Carelessness or unethical behavior can then lead to data leakage. One could partially mitigate its effects by enforcing a policy of encrypting all files in an organization. But such a policy has its own problems. Remembering multiple passwords is notoriously difficult and it can result in people physically writing down the passwords which would defeat the security purpose. NUS is currently experimenting with cloud based 2-factor authentication for many of its sensitive information services. But it is cumbersome and difficult to extend to all information services.

The third issue is related to privacy leakage when the data is analyzed for gaining insights. During the analysis and during the dissemination of analysis results, it is important to make sure that there is no unintended leakage of private data. This is particularly tricky when different facets of some sensitive data are released separately which can potentially allow somebody to piece together the private information. The usual technique for preserving privacy is the use of "k-anonymity[3]" and "l-diversity[4]" techniques. In k-anonymity, the values of sensitive items (e.g. average salary) are aggregated to the extent that data can be identified only for a group of size "k". The l-diversity technique is an extension of k-anonymity which ensures that a (non-sensitive) data field (e.g. nationality) has at least "l" different values occurring so that it cannot be used to narrow down possibilities of identifying an individual. The recent research advances are in the area of "differential privacy[5]" which advocates the use of randomization and noise to provide privacy. Ultimately, there is a trade-off between privacy and utility which needs to be carefully considered.

From the resources perspective, there is a lot of fragmentation of the efforts in data collection and storage. For graduate students and alumni, data is (often independently) collected by the office of the Provost, Alumni Affairs, Career Center as well as the various Faculties, Departments and Programs. This leads to several problems. First of all, there are duplicated efforts which cause waste of time and resources. Second, it gives rise to the problem of inconsistent data when the same field can have different values. Third, it can

---

[3] L. Sweeney. k-anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002.

[4] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam,M. l-Diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data 1 (1), 2007.

[5] C. Dwork, F. McSherry, K. Nissim, and A. Smith, Calibrating Noise to Sensitivity in Private Data Analysis, Theory of Cryptography Conference, 2006.

induce also survey fatigue among the students and alumni which can impede future data gathering efforts. Moreover, the data resides in silos in different offices, making it is very difficult to aggregate data across campus. This is exacerbated by incompatible software systems and data formats. It would be ideal to obtain seamless across the various databases but it is not easy due to different histories and different needs. At NUS, the Office of the Provost has recently started coordination efforts to overcome this fragmentation problem.

In this paper, we refer to "big data" from the somewhat narrow graduate education perspective which nevertheless is comprehensive, complex and fine-grained data. It captures the students' institutional history comprising of (a) outreach data (when the student is a prospective applicant), (b) applicant data, (c) on-campus academic, co-academic and extra-curricular activities including coursework and learning engagement information, (d) project and thesis information, (e) internships, exchange visit and experiential learning records (f) alumni information including career, salary and personal interests information. It can be seen from the earlier discussion on privacy and resources considerations that effectively managing this data can be challenging.

Legal and Ethical Issues Surrounding the Collection of Alumni Outcomes: A View from the U.S. Graduate Education Context

Julia Kent
**Assistant Vice President,
Communications, Advancement and Best Practices
Council of Graduate Schools**

Like most universities around the world, U.S. graduate institutions are working harder than ever to demonstrate educational outcomes for graduate students. A "big data" approach to demonstrating those outcomes is attractive to many universities and other entities with a stake in graduate education. Merged datasets may be greater (more valuable) than the sum of their parts, generating information that individual datasets cannot produce alone, and may also reduce the burden and costs of collecting and analyzing data.

At the same time, current big data projects to measure educational outcomes for graduate students have remained at the 10,000-foot level. The appropriate ethical and legal constraints on these big data initiatives, along with other issues, limit the degree to which U.S. graduate institutions can access, use, and share data in ways that are meaningful. This paper describes some "big data" efforts to analyze information on alumni outcomes and the legal protections that apply to them, then compares these approaches with some "bigger" or "big enough" data collection strategies that give institutions more control over how they use their own data.

Federal and State Efforts

For a number of years, the U.S. Department of Education and policymakers considered a centralized, federal system of tracking student educational information. However, Section 134 of the Higher Education Act of 2008, currently pending reauthorization, prohibits a Federal "student unit record system" that would include personally identifiable information. (It does not prevent states from creating such systems). Groups such as the Institute for Higher Education Policy have taken the position that this law limits the transparency and accountability of U.S. higher education institutions, noting that the federal government could answer "core questions" about student outcomes and allow students to compare the educational outcomes of alumni at different institutions. "Additionally," they note, "with numerous—yet often duplicative—federal, state and institutional reporting requirements, burden on institutional research offices is an ever-pressing issue" (IHEP, 2015, p. 1). In lieu of a federal system, the U.S. Department of Education has worked with most states to develop statewide longitudinal data systems (SLDSs).

Beyond the population of currently enrolled students, a number of U.S. states, and the Committee on Institutional Cooperation (CIC,) a consortium of research-intensive institutions in the U.S., have collaborated with the Census Bureau to gather data on graduate degree holders. To protect the privacy of individuals, the raw data remain within the Census Research Data Center (RDC) and are reviewed by the RDC prior to their release. Understandably, institutions cannot remove raw data from the RDC due to protections put in place by the bureau. A number of other human protections apply to a number of existing national datasets on graduate degree holders, such as the National Science Foundation's Survey of Earned Doctorates (SED) and the Survey of Doctoral Recipients (SDR).

Yet another state-level approach is to mine public databases of employment records to gather information on employment outcomes, but a disadvantage of this approach is that it cannot track alumni who find employment out of state (CGS 2014). However, the Western Interstate Commission for Higher Education has piloted data-sharing across states by employing "legally compliant data sharing agreements" that are consistent with the Family Education Rights and Privacy Act (FERPA) (Prescott, 2014).

"Bigger Data" Approaches Led by Institutions

Based on workshops with graduate deans for a CGS project on understanding PhD Career Pathways,[1] many graduate deans are interested in the potential of a big data approach to tracking student outcomes and alumni careers. However, in the context of this project and others, we have heard that CGS member institutions want more granular data on their students and programs than what current state-wide data collection efforts provide. The simultaneous desire for data at this level, alongside the desire to put those data in a larger context, has inspired CGS member institutions to participate actively in CGS "Best Practice" projects that collect aggregate institutional data for the purpose of comparison and benchmarking. Current projects such as the Doctoral Initiative on Minority Attrition and Completion (DIMAC) supported by the National Science Foundation (#1138814) and a TIAA-CREF-supported project to develop financial education projects benefiting graduate and undergraduate students are examples of initiatives that allow institutions to participate in data collection activities with dual purposes: to be used by the institution, and to be used in aggregation to inform best practices at the national level. These are not big data approaches, but like other projects organized by consortia of graduate schools, they might be considered "bigger data approaches" or "big enough data approaches." Another way that institutions have achieved the simultaneous goals of control and comparison is by contracting the services of companies that provide data dashboards on institutional outcomes that allow for comparison with peers and aspirational peers.

How should institutions address ethical and legal issues as datasets get even bigger?

As institutions join forces with one another and with external partners to collect and merge data pertinent to graduate outcomes, how will they ensure that student and alumni rights are protected? In an essay on big data, higher education and the law, Jeffrey Sun argues that higher education institutions must "move beyond" legal and regulatory concerns involving big data to focus their attention on better data management practices. Sun's point is that legal cases should not drive university behaviors with respect to big data; rather, trends in big data should inspire universities to implement data management strategies that address in a more comprehensive way issues such as privacy, access, intellectual property, security and data breaches. Sun's specific recommendations include voluntary consensus standards, basic practices to maintain privacy, an Institutional Review Board devoted to "Big Data" concerns or the integration of big data concerns into the responsibilities of an existing board; data/record retention and destruction; and data sanitation of equipment.[2]

One way of characterizing this approach is to see it as preventive, in the same way that research integrity training for graduate students may be the best way of decreasing the likelihood of research misconduct involving graduate students. But, like research integrity training, good data management also fundamentally addresses issues of data quality. To cite only one example, a climate in which data breaches and disrespect for privacy rights are common will likely affect the willingness of individuals to voluntarily participate in data

---

[1] The current phase of this project, Understanding PhD Career Pathways for Program Improvement, is supported by the Alfred P. Sloan Foundation, the Andrew W. Mellon Foundation, and the National Science Foundation (#1534620).
[2] See Sun (2014), pp. 48-49.

collection efforts and the ways in which these individuals report information. A systematic approach should therefore prompt universities to pay close attention to the fourth "V" in the common definition of "big data": "Veracity."[3] Gathering ever larger amounts of data will only be worthwhile if we have some confidence in their integrity and reliability.

References

Sun, J.C. (2014). Legal issues associated with big data in higher education. In Lane, J.E., *Building a smarter university: Big data, innovation and analytics*. Albany: State University of New York Press.

Institute for Higher Education Policy. (2015, September). Weighing the options for Improving the national postsecondary data infrastructure. Washington, DC: IHEP.

Lane, J.E. and Finsel, B.A. (2014). Fostering smarter colleges and universities: Data, big data, and analytics. In Lane, J.E., *Building a smarter university: Big data, innovation and analytics*. Albany: State University of New York Press.

Prescott, B.T. (2014). Beyond borders: Understanding the development and mobility of human capital in an age of data-driven accountability. Boulder, CO: Western Interstate Commission for Higher Education. Retrieved from http://www.wiche.edu/news/release/100913

---

[3] Lane and Finsel's introductory chapter to Building a Smarter University provides a condensed overview of the five "V's" often used to characterize big data: Volume, Velocity, Variety, Veracity, and Visualization." See pages 6-8.

Preparation of Graduate Degree Holders for the World of Big Data

Mary McNamara
Head, Graduate Research School
Dublin Institute of Technology (Ireland)

Introduction

Big data is an evolving term, with many definitions and include, but are not limited to personal data (name, an email address, bank details, medical information etc) or climate information or digital media images or GPS signals. For the purposes of this paper, the term "big data" should be understood to mean a large set of data produced from many sources, eg people, sensors, devices, etc that is so complex that *traditional data processing techniques are inadequate*. Because most analysis today is carried out using a desk- or even lap-top PC, work with big data is not very common. Challenges include data capture, data analysis, data storage and privacy and specifically how these are resourced in publically-funded universities. For graduate schools the main question is how do we prepare the next generation of graduate degree holders to work with and manage big data?

Resourcing Developments in Ireland

Data Analytics uses statistics and advanced software to convert large amounts of data into valuable information. Globally it is estimated that the amount of data generated and collected is growing by ~40 % per annum. Industries across the world are looking for ways to extract information that will help them make better decisions and be more competitive. Many industries across many sectors including financial services, telecommunications, health, media, sustainable energy etc can use big data and it has been suggested that such companies show higher productivity, higher profitability, and increased market share.

The Irish Government's Action Plan for Jobs identified 'Big Data' as one of the areas where Ireland has distinct advantages arising from our skills base and research capability in ICT and therefore there is the potential for significant growth in terms of jobs and global expansion of the sector. Ireland has invested heavily in research capability that services Big Data needs. For example the Insight Centre for Data Analytics brings together leading academics from 5 of Ireland's research centres and was established in 2013 with funding of €75m from Government to work in key areas of priority research including Semantic Web, Sensors and the Sensor Web, Social network analysis, Decision Support and Optimisation, and Connected Health. Researchers at Insight work closely with industry partners to develop next-generation data acquisition and analytics solutions for a range of application areas.

Resourcing Developments at DIT

Establishing an industry-led data analytics technology centre to work on developing viable business tools in this emerging area is one of the major actions contained in the Government's Action Plan for Jobs. To facilitate the marriage of industrial need and academic know-how, Enterprise Ireland (Government organisation responsible for the development of Irish enterprises in world markets) and the Industrial Development Authority (Government organisation responsible for attracting FDI to Ireland) conducted an industry consultation to identify the most pressing needs in Data Analytics. This consultation resulted in a Detailed Description of Needs (DDN) and in response a team of researchers from DIT, University College Dublin and University College Cork established a new technology centre within this domain.

The Centre for Applied Data Analytics Research (CeADAR) was established in 2012 and conducts initial research into technology challenges that have been identified by industry representatives to accelerate the development, deployment and adoption of Data Analytics technology. Top-tier multinational and Irish ICT companies lead the research agenda and include eBay, Accenture, Dell, Fidelity Investments, Adaptive Mobile, HP, Moving Media, Nucleus Venture Partners, and Qumas. Research is focused on developing ways of generating business, profit and ultimately jobs from the high-growth area of data analytics.

Big Data and Privacy and Training Graduate Students
There are over 10 billion smart devices in the world and this is expected to reach 150 billion by 2020. Therefore there are issues around the privacy and ownership of data. A recent article in The Guardian suggests that 'the battle for privacy has already been fought and lost… there are few effective controls over how [data] is used or secured.' Researchers suggest that only legislation can protect the privacy of individuals.

In Europe, the current Data Protection Directive 95/46/EC does not cover globalization and technological developments like social networks and cloud computing and the European Commission are currently finalising a comprehensive reform of the rules governing data protection. The Commission plans to unify data protection within the European Union (EU) with a single law, the General Data Protection Regulation (GDPR). Across the EU there is a daily transfer of vast amounts of data across borders and therefore, common rules will be established to ensure that personal data is protected. The main objective is to give citizens control over of their personal data, and to simplify the regulatory environment for companies, thus allowing both to fully benefit from the digital economy. Under EU law, personal data can only be gathered legally under strict conditions, for a legitimate purpose. Those who collect and manage data must protect it from misuse and must respect certain rights of the data owners which are guaranteed by EU law.

The GDPR provides for stronger rights for citizens with explicit rights to be forgotten, to object to data processing and to be informed when data security is breached. In addition organisations will be required to publish transparent and easily accessible data protection policies. It is proposed that companies will deal with one single EU authority, not 28 individual country authorities, reducing bureaucracy and costs. The GDPR also includes a new 'data protection by design' principle to motivate Big Data analysts to use techniques like anonymisation, pseudonymisation, encryption, and protocols for anonymous communications. The data protection by design principle is derived from the Privacy by Design approach in engineering which takes privacy into account throughout the whole engineering process.

The PhD programme at DIT provides graduates with the knowledge and skills for productive leadership roles and global citizenship. We provide additional taught modules for discipline skills training designed so that, on completion of the PhD, our graduates are experts in their chosen subject and have the professional knowledge and capacity to practice, reflect, review and build upon disciplinary expertise and judgment. One new module for the 2015/2016 session, ICT Regulations & Professional Issues, sets out to provide graduate students with a framework of understanding of the regulations concerning information in systems and the legal issues involved in using internal and external information sources. This training also provides for discussion of best practice, opportunities and pitfalls when dealing with the acquisition and use of computer technology. The history of current legislation, challenges in compliance, and the future of related legal issues for corporations in a national and global environment are explored.

The PhD programme at DIT also provides graduates with the knowledge and skills to be civically engaged and socially responsible. The Graduate Research School promotes

good ethical research and scholarly practice and emphasises integrity and rigor and a research student, regardless of discipline, will be registered only when the research project has received approval from the Research Integrity and Ethics Committee. All first year graduate students complete a module on Ethics which introduces research students to ethical issues which may be encountered during their research project whether in the sciences and engineering or arts, humanities and social sciences. It involves tracing the development of a research project from methodology, through data collection and analysis, to publication and looks at the moral dilemmas which might arise between researcher and subject. It provides graduate students with a clear understanding of the framework for responsible ethical research practice.

Big Data Challenges at Small Universities

Kevin Vessey
Dean of Graduate Studies and Research
Saint Mary's University (Canada)

While the analysis of large volumes of data offers great opportunities for universities in recruitment, conversion of applicants to registered students, student monitoring and advising, and analysis of research performance, Big Data challenges may be even greater at smaller universities as compared to larger institutions. *Big Data* is defined here, not just by the size of the database(s), but also by the complexity of these data, and the resources need to gain insight from such data. Likewise, *Analytics* is define herein as the insight that can be gained from the systematic analysis of data by specialized software systems, whether this software is embedded within a university's information technology system, or provided externally by outside service providers.

Saint Mary's University (SMU) in Halifax, Canada, has a student body of approximately 7500 students (full- and part-time) and about 10% of these are graduate students enrolled in our 25 programs, including graduate diplomas, Master and PhD programs. Like larger universities, great insight could be gained from data held both internally to our University (such as in our student management system), and from external sources (e.g. bibliographic services, social media platforms, etc.), that could inform the graduate studies office to aid in recruitment of new students, better serve our current students, and engage alumni. However, it is primarily the limitation of resources, both human and financial, that prevent some from taking full advantage of the insight that could be obtained from these Big Data resources. These challenges and some potential solutions will be briefly addressed here.

In drilling down on such challenges, let us start with an issue which is not a resource limitation per se, but is more of a university's administrative culture, perspective and priorities. I refer to the undergraduate focus of many smaller universities. Such is the case at SMU, where 90% of the student body is undergraduate and many graduate programs are relatively new (e.g. graduate programing has increased by 50% in the last 10 years), it is understandable, even reasonable, that the first forays into Big Data and Analytics would be focussed at the undergraduate level. This is the reality at our University. For example, through an initiative of the Registrar, an analytics tools of the Education Advisory Board has been accessed for identification of undergraduate students at academic risk though analysis of our student information system. These are still early days in utilization of this service, but costs implications and focus on the undergraduate student body suggest that it may be some time before such analysis might be applied to the graduate student body.

Below are summative points on some challenges and opportunities of Big Data and Analytics from a smaller university perspective in regard to recruitment, conversion of applicants to registered students, student monitoring and advising, and analysis of research performance:

- Recruitment – Recruitment strategies and processes for graduate programs can be very challenging for smaller institutions like SMU, whose programing spans Graduate Diplomas, Master and PhD programs, research and professional programs, large-sized (i.e. hundreds of students) and small-size programs (i.e. less than 20), and domestic and international markets. Definitely "one-size does *not* fit all" when it comes to

recruitment software and analytical systems with such varied graduate programs and limited budgets. Specialized recruitment software with it built-in analytical features, such as Ellucian's Recruiter™ system, may be beyond the financial resources of graduate schools of smaller universities. However, one commonality for even diverse graduate programs is that potential applicants predominantly use graduate program websites and/or its social media sites to research the program. With such sites, web-page analytical services, such as Google Analytics, can provide useful information on visitors to the sites such as where, and even when, visitors are viewing the site. This information may be used to help direct recruitment efforts, but only in a generalized manner. To gain insight from the data accumulated from individual applicant inquiries, education-specific client/customer/constituent relationship management (CRM) software is required. Again, while these can be systems can be expensive, the graduate school at SMU has partnered with other units at our University (i.e. the business school and the alumni office) to try a find a CRM solution that is within our budget.

- Conversion of applicants to registered students – This objective has proven very difficult to achieve from a Big Data perspective at SMU. It is common for graduate applicants to SMU to also apply to other smaller universities in the Atlantic Canada region or in other regions of Canada. The competition for the best of these applicants is intense. However, without recruitment management software (such as the Ellucian Recruiter™ system mentioned above) it is virtually impossible to gain insight on why and at what point in the process, applicants decided to attend or not to attend SMU. The job of gaining such insights often falls to individuals support staff, program coordinators or faculty members who have contact with applicants. The sharing of this information is extremely limited, and there is often the risk of developing conversion strategies based upon anecdotal comments from (strong-voiced) individuals within the university that may not reflect reality.

- Student monitoring and advising – As mentioned above, the focus on using Big Data(sets) from our student information system (SIS) to aid in student monitoring and advising at SMU has been (to this point) undergraduate focused. However, I am often amazed at the insight that is gained by my Associate Dean – Student Affairs on the monitoring of students by "manually" working with reports from our SIS, consulting with graduate program managers/coordinators, and organizing/manipulating these data with simple spreadsheet software. She is able to identify students-at-risk, especially in the larger, professionally-based programs where the use of datasets on past performance to predict future performance is the most robust. However, this process is labour intensive and, since the processes and expertise around gaining this insight sit exclusively with my Associate Dean, there is risk that the knowledge may not transfer to others in the graduate studies office or her successor.

- Analysis of research performance – Analysis of research performance at SMU is an area where I feel we have made the most progress in utilizing Big Data and Analytics at a relatively low cost. Utilization of analytical options of online publication databases such as Thomson Reuters Web of Science™ and Google Scholar enables analysis of the publication performance including numbers of publications by author/research fields, by academic departments, by citation impact, and by location of co-authors, by funding agencies, etc. We also have excellent cooperation from our financial services and industry liaison/technology transfer offices in providing analysis of research funding, industry engagement, intellectual property management, etc. from across the University.

However, while these analyses are informative for evaluating research performance within our University, comparison to research performance with other universities is extremely limited. Such comparison often require (costly) bibliometric software (e.g. Thomson Reuters' *InCite™*), however open-source bibliometric mapping tools, such as SciMAT[1] are also coming online.

Significant to all topics discussed above, is the challenge of privacy and protection of personal information. This can be a particular challenge in some provinces in Canada, like Nova Scotia, where the Freedom of Information and Protection of Privacy Act does not generally allow students' personal information be held on servers outside of Canada. This means that software-as-a service (SaaS) applications must be from Canadian companies that only hold their data within Canada, or that personal data must be anonymized before it is sent to such services. Such restrictions can make integration and interpretation of data extremely cumbersome.

In closing, *Big Data* and *Analytics* present great opportunities, but also some significant challenges for graduate schools at smaller universities.

---

[1] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma and F. Herrera, SciMAT: A New Science Mapping Analysis Software Tool. Journal of the American Society for Information Science and Technology, 63:8 (2012) 1609-1630 doi: 10.1002/asi.22688.

4: Enhancing Learning and Student Success

Big Data, Digital Competencies and Educational Service Engineering

Martin Gersch
**Chair of Business Administration
& Head of the Competence Center E-Commerce
Freie Universität Berlin (Germany)**

Introduction: Freie Universität Berlin and Germany's PhD tradition in transition
The Freie Universität (FU) Berlin has approximately 29,000 students in bachelor and master's courses as well as 5,800 PhD students and offers studies in all fields of Science. In the last decade it has put its effort into internationalization, and already at the undergraduate level there are approximately 20% students from abroad. At the PhD level, the number of students from abroad is even higher.

Our Dahlem Research School: a child of the national Excellence Initiative
The Dahlem Research School is the umbrella organization for 25 (2015) doctoral programs in all fields of Natural Science, Arts, Social Sciences as well as Business and Economics at the FU Berlin. Historically, it grew out of a single graduate school in chemistry, established in 2001, and was funded by the German Academic Exchange Service (DAAD) during the first five years. As one out of six institutions, the FU Berlin successfully took part in both rounds of the so-called 'excellence initiative' of the German Government. The first round started in 2007. With a new institutional strategy the FU Berlin also acquired these highly desired grants in 2012. In order to strengthen internationally relevant research activities each round of the initiative secures a 5 years funding. The Dahlem Research School became the roof over single doctoral programs and is – beside other tasks – responsible for the development and offering of relevant topics within modules as parts of the doctoral programs` curricula. Currently, about 900 doctoral candidates are members of the Dahlem Research School, about one-third of the students coming from abroad. There are still more doctoral candidates at the FU Berlin carrying out individual PhDs, but the number of PhD students within structured programs is steadily growing.

Information rules: Big Data as a driver of digital transformation
Big Data not only refers to the *volume* of data (in giga-, tera- or exabyte), but also to its inherent *variety*, which is driven by a lack of structure and interconnections between datasets. Together with a growing *velocity* of data retrieval processes, real-time analysis and a call for timely recommendations, these attributes drive data complexity and call for new methods, competences and institutional services[1,2]. Digital transformation affects major industries and research fields but also universities and PhD programs. In the following I will focus on three examples with high impact, great potentials and diverse backgrounds.

1. Big Data fosters research
Big Data has a tremendous impact on a variety of research fields, where rich data sources be-come more and more available by new technologies, new research designs and methods. I found researchers to react in three ways: (1) Technology of client computers advances as

[1] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact.," *MIS Q.*, vol. 36, no. 4, 2012.

[2] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton, "Big Data," *Manag. revolution. Harvard Bus Rev*, vol. 90, no. 10, pp. 61–67, 2012.

complexity of integrated circuits grows and digital devices shrink. Today many researchers use virtual machines on their own PCs for testing initial prototypes of new methods and simulations or to get first impressions of their data sets (but 'Big' Data begins where single client computers usually become too small and slow). (2) Additionally, universities like the FU Berlin offer services to access highly scalable server farms, that owned by the institution. (3) Regarding the needs of single researcher or research groups FU Berlin also simplifies access to high performance clusters, owned by the institution or shared with strategic partners, like "Zuse Institute Berlin (ZIB)".

To illustrate the variety of new disciplines besides 'traditional' fields of high volume data applications, like geography, meteorology, medicine or physics, I introduce three projects. (a) Artificial honeybees are enabled to collect and analyze rich sensor data, like video streams, in self-learning algorithms within computer science[3]. (b) A variety of interlinked telecommu-nication data from Africa is used to estimate the level of literacy in economics. (c) For under-standing learning processes data on student-to-student, student-to-educator and student-to-computer-interaction is assessed within the field of learning analytics. Such approaches raise questions on handling data variety – including statistical methods, text mining and web mining – privacy, in terms of data aggregation and at least pseudonymization, as well as data security.

2. Needed digital competencies from a German and European perspective

Data privacy as well as data security are prerequisites to all Big Data projects, especially in Germany. But, is it "German Angst" or a 'tradition of reflexivity'? These topics become more and more relevant in all European cultures and especially for researchers. FU Berlin is part of a societal discussion that is concerned with closing the current gap between a mainly internationalized usage of private data on the one hand and a legitimate personal interest in informational self-determination on the other. To lower the burden on individual researchers, there is a need to develop a consent on these ethical questions by local and national privacy officers or to engage ethics councils. Besides sensitizing young researchers for such issues, we need to help them in developing digital as well as statistical competencies through specific courses within the PhD programs.

While in many fields of natural science handling huge data sets is regularly taught, the FU Berlin has just started to systematically develop and enhance relevant competencies in more fields. For this purpose, e.g. chairs on statistics offer courses on statistical programming lan-guages, like R, as well as quantitative methods. Additionally, they offer individualized coach-ings for PhD students. Furthermore technical staff at the institute of mathematics and infor-matics assists in handling Big Data, especially with regards to offering a technical infrastruc-ture and ensuring data security.

3. Data-driven improvements of teaching with Educational Service Engineering

Educational Service Engineering (ESE) integrates methods from didactics, software engineering and service management to systematically develop educational services. For this purpose services in (higher) education are modularized, data-based analyzed and (re)-engineered[4,5,6]. For testing didactical assumptions within teaching and to lift the 'fog of uncertainty'

---

[3] T. Landgraf, M. Oertel, D. Rhiel, and R. Rojas, "A biomimetic honeybee robot for the analysis of the honeybee dance communication system.," in *IROS*, 2010, pp. 3097–3102.

[4] P. Weber and A. Abuhamdieh, "Educational Service Strategy: Educational Service Platforms and E-Learning Patterns," *Instr. Technol.*, vol. 8, no. 4, pp. 3–14, 2011.

[5] R. Gabriel, M. Gersch, and P. Weber, "Service Platforms for E-Learning-supported Management Education," in *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 2007, vol. 2007, no. 1, pp. 853–859.

[6] M. Gersch, C. Lehr, and P. Weber, "Virtual collaborative Learning in international Settings - The Virtual Seminar 'Net Economy,'" *INTED2011* Proc., pp. 5078–5085, 2011.

data-based[7] Learning Analytics could be used. This field of research offers methods and procedures to collect, process and interpret data from interactions between students, educators and e-learning systems. Especially the evaluation of highly self-regulated courses calls for sophisticated methods, like e.g. network analysis as well as web and text mining. They enable educators to raise questions, like how we improve student retention and success, enhance curricula, improve student engagement and/or improve blended or online learning?

[7] G. Siemens, "Learning Analytics: The Emergence of a Discipline," *Am. Behav. Sci.*, vol. 57, no. 10, pp. 1380–1400, 2013.

Flexible Minds and Conceptual Clarity: Big Data Issues

Shireen Motala
Director, Postgraduate Research Centre
University of Johannesburg (South Africa)

What is Big Data?

Big data is a popular term used to describe the exponential growth and availability of data, both structured and unstructured. Big data has several possible meanings. From one perspective big data refers to the data itself. Industry analyst Doug Laney defined big data in 2001 in terms of the three Vs: volume, velocity and variety. Big data is thus data that can be measured in terabytes, comes from a variety of sources and in a variety of formats and streams in torrents in real time. The complexity of the data is also important, however, given the increasing need to combine it; and account must be taken of whether the data is kept centralised, decentralised or in clouds.

Another perspective is to consider big data technologies, as Davenport and Dyché (2013) do, by focusing on the availability of such technologies and their implications, such as processing speeds and costs. However, the potential of big data initiatives, the ability to conceptualise the use of big data, and the technical skills that underpin this ability, are also important. For the purpose of this discussion, I will look at big data as the ability to conceptualise the use, and analyse a variety, of data sources and types.

What can be Done with Big Data?

Big data, in terms of volume and variety of data and the big data technologies that analyse the data, currently provides many organisations with a competitive edge, in the form of improved decision making, improved operational efficiencies, cost reductions and reduced risk. In higher education and specifically in graduate institutions, while many possibilities exist, the use of big data and analytics for instructional applications is still immature (Picciano 2012).

In the South African context, too, the potential to use big data is huge but unexplored. Conventional data still tends to be used for institutional profiling and predicting, in part because the use of big data has certain system prerequisites including technical capacity, data availability and institutional readiness. The storage of large volumes of data used to be an issue, but these costs are rapidly decreasing. Instead, how to determine the relevance of specific datasets and how to use analytics have emerged as important issues. Another potential problem is that many organisations treat their big data analytics and their traditional data analytics as separate systems, with only some large companies seamlessly integrating these two sets of analytics (Davenport and Dyché 2013).

However, now that concerns around the supply of data and the analytical tools to process it no longer prohibiting the use of big data, pull versus push factors are becoming more important. One way of looking at this is by stating that value is in the eye of the beholder (Davenport and Dyché 2013), with the use of big data limited only by the imagination of potential users and managers. Based on their examination of the current status of business intelligence in academia, Wixom et al (2011, 2014) advised that developing big data skills and "business intelligence savvy" professionals should become a key focus in academia at both undergraduate and graduate levels.

Big Data at UJ – Developing Flexible Minds and Conceptual Creativity

At the University of Johannesburg the preparation of 21st century ready graduates is already a key focus, and big data readiness should strengthen this further. Although the most obvious starting points are the departments that produce IT and IS students, many other departments (including engineering, architecture, life and health sciences and management) need to be enabled. Collaboration with industry in order to gain and provide access to the analytical tools and datasets necessary to develop these skills should be a key focus. Wixom et al (2011, 2014) clearly identified vendor academic alliances as having the best potential to supply platforms for accessing and sharing the tools and datasets needed; and in addition to developing the analytical skills, an ability to conceptualise the use of big data needs to be developed.

At a graduate level, especially for master's degrees in professional areas such as engineering management, business management and health management, as well as structured interdisciplinary qualifications, an exposure to and an ability to leverage big data must become an explicit focus. Future leaders must understand how to leverage cross-functional teams that include modellers, analysts, technologists and business leaders, and should be capable of understanding problems and conceptualising needs so that the multidisciplinary teams of which they are a part can provide solutions (Wixom et al 2014).

This will require the development of skills, levels of awareness and the availability of datasets and analytical tools. However, the development of case studies and other pedagogical resources to provide both the realistic contexts and the opportunity to use the skills may be a bit more difficult to develop. These resources may be available in more developed contexts but in the African context there is very little available. The Postgraduate School at UJ must take a stronger role in advocating such preparation of our graduates and in developing the alliances and partnerships with industry that are required to make it possible.

Big Data at UJ – Graduate Research

The use of big data as empirical evidence in both master's and doctoral research is an area of potential that we have not yet explored. Although there is a focus on research into big data matters in the information systems groups, this should not be the only focus. Multi- and trans-disciplinary research are areas that may depend on the ability to use big data and the development of these areas of research beyond the structured UJ flagship programmes is a clear UJPS mandate. To make this possible the UJPS will have to take the lead in providing research capacity development opportunities to supervisors, researchers and graduate students.

Big Data at UJ – Measuring Research Output

The UJ context is one which prioritises efficient and effective data use in various forms. However, real time data which can provide predictive analysis is still to be put in place. At this stage UJ is using selected bibliometrics from the standardised 'Snowball Metrics' set, sourced from Elsevier's Scopus database and SciVal Research Intelligence tools, to provide information to academic management which can assist them in making informed research-related decisions (Towert 2015). Some of these metrics are simply tracked periodically on an institutional basis via a scorecard to monitor and measure research progress or areas requiring intervention. These metrics include Scholarly Output, Citation Counts, Average Citations per Publication, Field-Weighted Citation Impact, Outputs and Publications in Top Percentiles, Collaboration Impact, and others. Together, these metrics provide insights into the quantity, quality, impact and extent of collaboration of UJ's research endeavours, and individual bibliometric profiles can be grouped into collections of entities, including research centres, special interest groups, departments or faculties, and by so doing, allow a degree of predictive analysis to take place.

The same metrics can be applied in conjunction with more traditional methods of assessing research performance, capability and capacity, with bibliometric profiles providing objective insights into decisions to appoint faculty or nominate them for awards, promotions and other incentives. They can be particularly useful when assessing researchers from other countries and institutions, where the only other source of information is their own curriculum vitaes.

Ethical and Moral Dilemmas
It is important to keep in mind potential ethical and moral dilemmas around the use of big data. Alliances with industry around the use of datasets and tools must be managed carefully in order to avoid conflicts of interest. Using big datasets in providing empirical evidence for research may also be difficult, especially if data is used for purposes other than that for which it was gathered. The use of confidential information is very sensitive in South Africa and the POPI Act (Protection of Personal Information Act, 2013) is explicit about this. This has impacted significantly on the use of the personal data of student and staff members in South African universities. It is also of the highest importance not to marginalise or commoditize people through the use of big data.

References
Davenport, T. H. & Dyché, J. (2013). Big data in big companies. Cary, NC: International Institute for Analytics, SAS Institute.

Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20.

Towert, D. (2015). Personal Communication on the use of Big Data in measuring Research at UJ.

Wixom, B., Ariyachandra, T., Goul, M., Gray, P., Kulkarni, U., & Phillips-Wren, G. (2011). The current state of business intelligence in academia. *Communications of the Association for Information Systems*, 29(16), 299-312.

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., Kulkarni, U., Mooney, B., Phillips-Wren, G. & Turetken, O. (2014). The current state of business intelligence in academia: The arrival of big data. *Communications of the Association for Information Systems*, 34(1), 1.

Incentivizing Participation in Online Educational Forums

Y. Narahari
Chair, Division of Electrical Sciences
Indian Institute of Science, Bangalore (India)

(Based on: Rohith D. Vallam, Priyanka Bhat, Debmalya Mandal, and Y. Narahari. A Stackelberg game approach for incentivizing participation in online educational forums with heterogeneous student population. Proceedings of AAAI-2015 (29th AAAI International Conference on Artificial Intelligence), Austin, Texas, USA, pp. 1043-1049.)

The question we investigate here: how do we incentivize strategically behaving students and instructors to enhance their participation levels in online learning environments?

With the explosive growth of the Internet, the area of online education has undergone a massive transformation in terms of how students and instructors interact in a classroom. Online learning environments now constitute a very important part of any academic course. Further, online education has attracted the interest of the research community due to the immense popularity of the massive open online courses (MOOCs) offered by platforms like Coursera, edX, Udacity, etc. As of January 17, 2014, Coursera students voiced themselves in 590,000 discussion threads in the education forums for a total of 343,014,912 minutes of learning across 571 courses (Coursera 2014). However, empirical studies have repeatedly shown that the dropout rates in the online courses are very high, mainly due to a lack of sustained motivation among the enrolled students. An important, but often under-utilized component of an online classroom is the online educational forum (OEF) where students and instructors discuss various administrative and technical aspects of the course. We have proposed an instructor-driven approach to orchestrate the activities of OEFs by designing optimal incentives to enhance student-instructor participation in these OEFs.

Incentive design plays an important role in encouraging participation among students in these educational forums. As part of a case study, we first analyzed the (not-so-big) data collected from two online educational forums which were part of the Game Theory (E1 254) in the Department of Computer Science and Automation, Indian Institute of Science in two separate recent offerings. The primary difference between these two offerings was that there were no incentives offered to students participating in the first offering while in the second offering, students were offered incentives (a certain percentage of marks based on the reward points accumulated by the student) to respond actively to open-ended (or discussion type) questions posted on the Piazza forum associated with the course. We observed an increased participation of students in the incentive-based course than when there were no incentives offered which is an indication of importance of appropriate incentives in driving up the participation levels in the course.

Modelling incentives for improving the participation levels has recently been studied in the literature. Online learning environments attract participation from students with heterogeneous skill levels and it is our belief that any approach to improve levels in participation should account for this heterogeneity as well. We incentivize students on a per question basis to keep up the momentum of participation in the class. Students are provided suitable incentives to post answers to the specific open-ended/discussion-style questions that are posted on the forum by the instructor. These incentives maybe in the form of free book vouchers, food coupons, or some extra grade points, as considered appropriate by the

instructor. The instructor is limited by a budget and has to make judicious use of them such that it results in higher participation levels from different types of students.

Our work first defines an interaction model in an OEF which reasonably captures the activities of the students and instructor. Once this is addressed, there is a need to understand transient behavior of the instructor and students in the time-limited online course. One of the complexities is to handle the 'continuous' nature of arrivals of the instructor and students to the OEF. Taking into consideration these factors, we model the OEF as a continuous-time Markov chain (CTMC). Using techniques from lumpability of CTMCs, we compute the transient behavior of the instructor and the students in the modelled OEF. Next, we use these computations in a more realistic game-theoretic setting where we adopt a Stackelberg game approach to model strategic behavior of students in the OEF. We believe that our problem fits naturally into a Stackelberg framework where the players of the game compete on a resource (i.e., participation time on the OEF) and the welfare maximizing leader (i.e., instructor) is in a position to exploit the first-mover advantage to trigger increased participation from the followers (i.e., student population) by designing suitable incentive schemes.

Our detailed experiments with the proposed Stackelberg model demonstrate that our approach validates several empirically/theoretically observed phenomena and also, offers utility-maximizing recommendations to the instructor as well as the different types of students on several important parameters like arrival rate, instructor bias, frequency of incentives, timing of incentives, etc.

We believe the methodology proposed could be applied in a scalable way using big data analytics to incentivize and improve participation levels of students in MOOCs and associated OEFs as well.

The Powerful Merge of "Big Data" and Assessment Data in Support of Student Success

David Payne
Vice President and Chief Operating Officer
Educational Testing Service (ETS)

Alina A. von Davier
Senior Research Director
Educational Testing Service (ETS)

Higher Education is undergoing dramatic change at all levels, with new directions in management systems, learning systems, and assessment of individuals and groups. Some of the most innovative and exciting features involve the availability of detailed student data, such as background variables, courses taken, grades, and of technology-enhanced learning and assessment tools. These changes support learning of cognitive, social, and affective skills within a common framework and allow for a detailed collection of "big data" (BD) in addition to the usual outcome data.

These days, computerized educational environments produce large amounts of data with extraordinarily high dimensionality. Extracting key features from such data is crucial, not only to make analysis computationally tractable, but also to extract relevant features of student performance. With the technological advantages of systems for the recording and capturing of multimodal data, the data from interactive learning systems and computerized assessments contain information about students' mastery of the content and their outcome results, but also their discourse, actions, gestures, speech, and body language that result in a deluge of data. One way to handle these different types of data is to make use of *computational psychometrics* (von Davier, 2015). This new discipline integrates data mining techniques and artificial intelligence/machine learning algorithms with theoretical psychometric models. This is one of the areas of expertise at Educational Testing Service (ETS)[1].

In this presentation we argue for the use of big data in education in that the availability of big data implies a more nuanced measurement and deeper knowledge about students' abilities and life circumstances, which in turn, lead to a tailored, detailed, and actionable feedback provided to students to improve their learning and engagement.

In particular, we discuss four types of BD in Higher Education that researchers at Educational Testing Service (ETS) see as promising for supporting students' success: (1) BD that resulted from the availability of ancillary information about the students; (2) BD that resulted from collecting the process data (PD), that is, the process by which a student solves a problem or interacts with a learning system; (3) BD resulted from the collaborative interactions; and (4) BD as detailed, multimodal behavioral data collected via cameras, microphones, and other sensors. In the rest of the paper we briefly discuss each of the four directions, together with examples of research conducted at ETS.

(1). ***BD from ancillary information***. These data (current and past grades, attendance, demographics, etc.) are collected by colleges' and universities' management systems. This type of BD can be used to counsel the students in their choices of courses by predicting successful paths based on the experiences gathered from other students

---

[1] http://www.ets.org/research/topics/statistics_psychometrics/next_gen

who had a similar profiles. The catch is that one needs a very large dataset to ensure that the identified profiles are representative of meaningful experiences rather than random chance (especially when profiles consist of specific combinations of data points that may be individually common but collectively rare, for example, students from low-income families in urban areas who received A's in freshman English and C's in freshman Mathematics and who don't feel very connected to the school or handle stress well. At ETS, we collaborated with Civitas Learning™[2], a learning analytics company that has access to these data through contracts with numerous universities. ETS administered *Success Navigator®*, an ETS assessment designed to help colleges reach at-risk incoming students and improve retention and completion rates. The data from the Civitas Learning™ database and the assessment data were analyzed together. The conclusion of this project is that combining different types of data can lead to more actionable insights. For example, knowing that students who have low grades in their current courses (from Civitas Learning™ database) do worse in college does not provide any actionable information. However, knowing that a combination of getting low grades and having high- stress resilience (assessment data) is related to a significantly lower likelihood of reenrollment than getting low grades and having low-stress resilience produces very actionable information. The combination of BD with assessment data allows for the design of targeted interventions to increase performance, whereas BD or assessment data alone merely predict performance.

(2). ***BD from the process data (PD)***. These data are collected during learning (in intelligent tutoring systems or training systems), or during a game-based assessment, and they usually consist of the outcome results (responses given), but also keystrokes and timing information. At ETS, we are working on several pertinent research projects and applications, such as in the National Assessment of Educational Progress (NAEP) Technology and Engineering Literacy Assessment (TEL)[3] and the Cognitively Based Assessment *of*, *for*, and *as* Learning (*CBAL™*)[4]. By collecting the PD we can better identify how students arrive at an answer, whether they have misconceptions, or whether they understood the questions. With this information we can provide feedback and develop tailored/adaptive learning and testing. For example, a careful keystroke analysis will help identify the test takers' writing style, which allows teachers to provide a more targeted instruction on writing than that just based on the final scores of the essays.

(3). ***BD from collaborative interactions***. Collaboration is one of the skills identified as the "21st-century skills" (OECD, 2013) and it receives attention among stakeholders in both higher education and the workplace. The Organisation for Economic Co-operation and Development (OECD, 2013) included a test of collaborative problem solving skills in its Programme for International Student Assessment (PISA) 2015 survey of critical skills. Collaborative assessment is also being promoted by a global initiative called Assessment and Teaching of 21st Century Skills, a partnership among Cisco®, Intel®, Microsoft® and the University of Melbourne to prepare students to live and work in Information Age societies. At ETS, we research how two or more college-level students work together remotely to solve a science problem (The Tetralogue; Hao, Liu, von Davier, & Kyllonen, 2015). We collected all the turn-by-turn interaction process data

---

from the teams, together with their responses (PD in #2 above). Examining the BD from the collaboration process, we can identify the collaboration skills that will improve the team performance, which will guide us to develop such skills in teaching practice.

(4). ***BD from multimodal behavioral data***. These data are collected from cameras, sensors, microphones, and eye trackers in conjunction with the PD in (2). These data collected during a learning session or a formative assessment may help identify the level of engagement of the student and help adapt the test based on engagement, or may prompt the student to look at parts of the screen that the student seems to have missed (based on the eye tracker).

We conclude by emphasizing not only the potential of using different sources of BD to improve student success, but also the challenges and perils around these data. The most important challenge is to identify the "right data" among the BD. The second is to validate the claims and the measurement results obtained from the application of computational psychometrics. Comprehensive research is needed to address these issues for all types of BD.

References

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. (2015). Assessing collaborative problem solving with simulation based tasks. *Proceedings for the 11th International Conference on Computer Supported Collaborative Learning*, Gothenburg, Sweden. http://www.isls.org/cscl2015/papers/MC-0297-ShortPaper-Hao.pdf

OECD (2013). *PISA 2015 draft collaborative problem solving framework*. Paris, France: Author. http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

von Davier, A. A. (2015, July). Virtual & collaborative assessments: Examples, implications, and challenges for educational measurement. Invited Talk at the Workshop on Machine Learning for Education, International Conference of Machine Learning, Lille, France http://dsp.rice.edu/ML4Ed_ICML2015

Considering Big Data and Learning Analytics in Designing Doctoral Studies Support Programmes

Zaidatun Tasir
Dean, School of Graduate Studies
Universiti Teknologi Malaysia (Malaysia)

Definition of Big Data

In this article, big data refers to Student Information System that associates students' demographic data with data of past and present learning performance, engagement with online course materials, or in-class participation levels to predict, with startling accuracy the specific outcomes such as final grades within a course.

The purpose of big data and learning analytics is for identifying actionable intelligence for efficient optimum results and influence future decisions. Big data concepts and leaning analytics are applicable to a variety of higher education administrative and instructional applications, including recruitment and admissions processing, financial planning, donor tracking, and students performance monitoring. In this article, big data concepts and learning analytics are utilized to empower Doctoral studies structure. It is used to aid Doctoral students (PhD candidates) research progress related to their learning issues such as graduating on-time, attrition and retention issues as well as the proposed solution for the respective addressed issues.

In Universiti Teknologi Malaysia (UTM), student information system for graduate studies is structurally monitored by a centralized system known as Graduate Studies Management System (GSMS). The system consists of several modules such as; Admission Module, Research Module, Viva Examination Module and Support Programme module. In specific, the Research Module serves as the platform to gather information about students' research progress in every semester where their research progress will be recorded and information is accessible by the supervisors. Upon graduation, the Research Module illustrates the research students learning journey including their participation in research activities in UTM. The recorded students' progress in the GSMS system serves as the source of information for big data and learning analytics discussed in this article towards designing rigorous Doctoral studies support programmes.

In UTM, Big Data Centre is a Centre of Excellence (CoE) that focuses on research related to Big Data. Both Big Data analytics training to educators and Professional Certificates in Data Science and Big Data Analytics for UTM's lecturers in 2015 are administered by the centre. UTM Big Data Centre is established to achieve the following objectives:

i. to intersect the core of Big Data Analytics into big data science, big data computing and big data engineering,
ii. to provide the right information at the right time, which enables data analysts, data scientist, data engineers or managers to make informed business decisions.

Doctoral Studies Support Programmes

The way forward for doctoral education should deliberate on the skills and knowledge related to the study, leveraging new knowledge and transferable skills by matching skillsets to the future career outcomes in a broader context. Development of Professional Skills Training programmes is anticipated as the appropriate support for current doctoral candidates in higher

education institution. In UTM, the PST's programmes are centralized by School of Graduate Studies and co-organized by the faculties and Postgraduate Student Societies (PGSS). In 2014, a total of 136 support programmes for PhD's students were conducted by PGSS. Based on the listed programmes, a total of 74 programmes (54%) were related to knowledge and skillsets related to students' PhD learning while 57 programmes (42%) were related to soft skills perceived as necessary for PhD students. Only 5 out of the 136 (4%) programmes were related to preparing PhD students for their potential future career upon graduation, suggesting the need to refine and enrich the Professional Skills Training programmes.

Designing Professional Skills Training Programmes based on Big Data
In 2015, School of Graduate Studies planned to design a data-driven support programmes for Doctoral students. Based on this strategy, the programmes should be customized to meet the students' needs as reported by the retrievable data in GSMS. In many previous studies, big data and learning analytics assist schools and faculties to predict learning success and failures (Arnold & Pistilli, 2012; Dietz-Uhler, & Hurn, 2013). Thus, mining big data and learning analytics on students' information about their research progress in GSMS will equipped us with the general idea of students' necessity. As a result, prediction of the suitable support programmes to assist them in their research can be formulated. This initiative will improve students' retention and raised the success rate in doctoral studies as UTM could provide support programmes in parallel with the students' needs. Findings from learning analytics narrow the existing gap between students' needs and the design of the programmes. Graduate schools improvise by being able to identify the types of support programmes correspond to every student for their doctoral research.

Matching Supervisory Styles of Supervisors with Students' Need through Big Data and Analytics
Gatfield (2005) proposed four main PhD's supervisory styles, based on a model utilising 'support' and 'structure'. The four main styles are contractual (high support, high structure), directorial (low support, high structure), laissez-faire (low support, low structure), and pastoral (high support, low structure) styles.

Big Data analytics informs supervisors about students' personal information such as students' academic background, previous courses taken, articles had been published and the socio-economic status of the doctoral students which enhance insight about the need and capability of their students. Build upon the information from learning analytics, a supervisor can identify the level of support necessary for the students. Accordingly, supervisors can prepare the suitable supervisory structure compatible to the level of support. Information obtained from analytics is also useful for supervisors to demonstrate personalized supervision styles and techniques.

Big data and learning analytics could also revamp the design of online learning environment to better suit students' online learning needs. Research reports that providing students with conducive online learning environment based on their online social learning style is important as it improves level of participation in online learning (Tu & McIsaac, 2002) in which failure to do so would affect students' satisfaction and students' perceived learning (Eom, Wen, & Ashill, 2006). Mining students' online learning experience as recorded in the online learning database informs university lecturers about students' online social learning style where predictions can be formulated about the characters of the future students and so whether pedagogies or supervisory styles have to be restored.

A small-scale research on formulating metacognitive scaffolding framework for graduate students in learning through Facebook was conducted by manipulating big data and learning analytics concept using data mining techniques (Jumaat, 2015). Technology-based instructional tools such as Facebook serve as a platform to gather a wide array of data on students' online interaction. Activity of going through big data sets to look for relevant or pertinent information; which is data mining provides new insights into which students are performing well and which students need additional scaffolding to perform better in learning. The research findings show the potential of big data in providing relevant data for predicting the types of metacognitive scaffolding through Facebook that can improve students' achievement in test. Scrutinizing information based on learning analytics with big data assist us in identification of the effective moves for improving students' retention and future success with more efficient outcome. Besides doing "trial-run" activities, big data and learning analytics save the time needed to identify appropriate support programmes for graduate students.

References

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270). ACM.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, 12(1): 17-26.

Eom, S. B., Wen, H. J., & Ashill, N. (2006). The Determinants of Students' Perceived Learning Outcomes and Satisfaction in University Online Education: An Empirical Investigation. *Decision Sciences Journal of Innovative Education*, 4(2): 215-235.

Gatfield, T. (2005). An investigation into PhD supervisory management styles: Development of a dynamic conceptual model and its managerial implications. *Journal of Higher Education and Policy Management*. 27(3): 311-325.

Jumaat, N. (2015). *A Framework of Metacognitive Scaffolding for in Learning through Facebook*. Unpublished Doctoral Thesis, Universiti Teknologi Malaysia.

Tu, C. H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American journal of distance education*, 16(3), 131-150.

Big Data: Opportunities for enhancing graduate student success in their dissertation research

Paula M. Wood-Adams
Dean of Graduate Studies
Concordia University (Montreal, Canada)

Introduction

The term "big data" as used here refers to data sets so large that traditional statistical and software tools no longer work. It also implies unstructured data of various types and the possibility of making predictions in the absence of a first-principles theory or detailed understanding of causal relationships. In particular, we consider the case of capturing of graduate student online activity with respect to academic matters and using the resulting information to enhance student success particularly in research related work.

Big data analytics combines the massive data sets, statistical analyses, and predictive modelling. This field is relevant now because the cost to collect and store massive quantities of information has markedly decreased and because of the digital interconnectedness of so many of our activities. At its best, big data analytics can be used to capture, analyse, and interpret information in real time and to provide an immediate informative or corrective response where needed. One challenge of attempting this type of approach in graduate education is the slow speed at which universities typically accept change. Even when data are available that convincingly indicate change is needed, academic bodies, at the program or institution level, are reluctant to move and certainly cannot change course at the speed required by big data driven techniques. Change can be achieved at this rate however, on the individual instructor or student level. This is the cornerstone of the current use of big data techniques in higher education. For example, an instructor can use learning analytics, a well-developed subset of big data analytics, to modify her teaching methodologies on the fly in an online or blended delivery course (Long & Siemens, 2011). An individual student can be guided towards course choices that maximize their chance of success, for example Arizona State's eAdvisor (Phillips, 2013).

Another challenge for using big data analytics in graduate education is incompatibility between the sizes of the data sets required (Manyika, et al., 2011) and the typical graduate student population at a single institution; the mismatch is on the order of a factor of hundreds of thousands. This could in theory be partially compensated for by incorporating many types of data for each student: combining confidential data from their institutional profile with complete monitoring of their online activity and accessible genealogical and other personal/professional information. It is, however, hard to imagine this being acceptable from a privacy or even an ethical standpoint. A more realistic solution might be the sharing of anonymized data on graduate student research activities between institutions and the development of national or international consortia around these efforts (Johnson, Adams Becker, Estrada, & Freeman, 2015). We believe that an interinstitutional consortium could be created to develop a big data based tool for keeping students on track with their dissertation research.

The proposal

Graduate students use many online tools during their dissertation research: abstract and citation databases, institutional repositories, digital archives, public demographic data, open access course content to name just a few examples. They also use less formal means of

gathering information including social media and search engines such as Google. This forms the basis for our proposal: a system that tracks the online research activities of a consenting graduate student and intervenes with suggestions when the student is floundering, falling behind or getting trapped in an endless cycle of fruitless searches.

We can imagine many situations in which our system might help the student be successful. At the beginning of their graduate career, the system might help the student develop the necessary expertise for tackling their research questions. For example if a student from a discipline other than physics is searching for information about a sophisticated particle theory, the system might suggest a set of more introductory content on calculus, probability theory and quantum mechanics to prepare her for tackling the more complicated subject. Or a student exploring an online archive of ancient writings might be pointed to the latest research on the relevant language or to more basic texts should her transcripts reflect a gap in her background knowledge. Later on the system might prompt the student when they have collected what appears to be enough information to begin writing the dissertation proposal. It might suggest appropriate points for consultations with her advisor or suggest names of other experts that she might contact. The system might even be able to say when the student has made a connection or discovery that does not exist in the literature. This might indicate an opportunity for further exploration or perhaps an error in the research. Conversely, the system might also caution the student who is in the process of reinventing the wheel.

As a student falls off of the program timeline a flag might be raised with the graduate school or as a student is surging ahead of expectations they might be nominated to their department for an award. Their progress towards degree completion, their research productivity, and the outcome of their defense would be monitored and (where possible) facilitated by the system. In this way, the path of each participating student would add to the knowledge base of the system creating an increasingly efficient and effective tool as time passes.

A word of caution

Big data analytics and the related corrective actions typically result in conformity; solutions are defined as those that similar people would follow in similar situations. This seems to be in contradiction to the basic creative nature of research discovery. Perhaps the use of such techniques in the training of new researchers would limit creativity and reduce the potential for genius. This issue needs to be thoroughly considered both from the perspective of cognitive and behavioural sciences and information system design. Enough flexibility must be built into the system such that an entirely unexpected research outcome is possible and that an unlimited number of unique research paths can be created by students.

Bibliography

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). NMC Horizon Report: 2015 Higher Education Edition. Austin, Texas:: The New Media Consortium.

Long, P. D., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE review, 31-37.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Phillips, E. D. (2013, January - February). Improving Advising Using Technology and Data Analytics. Change, The magazine of higher learning.

5: Preparing the Next Generation of Experts

Institutional Approaches to Big Data Analysis in a Niche University

Liviu Matei
Provost for Graduate Studies
Central European University (Hungary)

Big data and big data analysis pose new challenges to graduate education. This is true for a niche institution such as the Central European University (CEU) as well. An international university by design, CEU offers only graduate programs, and only in the social sciences, humanities, law, public policy, mathematics and its applications, and management. CEU has a an evolving, relatively well elaborated institutional approach to big data, informed by the overall mission of the University, which is to promote open society and democracy through high-level research, research-based teaching and learning, and civic commitment. This approach is articulated around several key elements: courses focusing explicitly on big data, degree and certificate programs with a big data orientation, dedicated faculty appointments, research projects, research centers with big data analysis as a core part of their activities, and dedicated public events. A particular characteristic of our approach is the systematic link between big data analysis and network science.

Several CEU departments offer courses, at the master's and doctoral level, introducing students to big data analysis. They are not stand-alone courses, dealing generically with big data, but are usually linked to particular disciplines, or rather applied to particular thematic areas. Students from other departments are encouraged to cross-list these courses. They could be part of degree or certificate programs with a major big data component, such as the M.Sc. in business analytics, the Ph.D. in network science, or the interdisciplinary certificate program in network science. Such programs are offered jointly by two or more departments or schools. The program in business analytics, for example, aims to develop skills and competences needed for creating value in a business setting from data analysis and it is offered jointly by the Business School and the Departments of Economics, Mathematics and Its Applications, and the Center for Network Science. Other big data courses are offered on an elective basis as part of degree programs that are not specifically about big data, such as the M.A. programs in Public Policy or Public Administration. In all cases attention is paid not only to technical or methodological issues, but also to legal and ethical aspects. Public policy or management programs, for example, emphasize regulatory aspects of the use of big data.

Research activities dealing with big data issues or using big data techniques are carried out in several of our academic units. As a rule, they always involve graduate students.

CEU organizes frequent public events with a big data component, for audiences of academics, practitioners, and policy makers.

These activities are supported by an institutional structure that includes traditional, but also rather new, innovative elements. At the core of this structure is the CEU Center for Network Science. Other important elements are the Department of Mathematics and Its Applications (which may look like an odd unit in a University which is otherwise primarily about social sciences and humanities), and the Center for Media, Data, and Society. CEU has one of the very few Centers for Network Science in Europe and it offers the first PhD program in this area on the continent, which is also one of the very first in the world. The Center provides an organizational platform for research and education in network science, with a special focus on applications to practical social problems. The PhD program is a pioneering, interdisciplinary program that provides tools to understand how complex

networks are structured and how they function, what are the principles governing social, political, economic, and environmental networks. Students learn ways of network modeling and explore large datasets (terabyte-sized) in areas such as health issues involving millions of individuals, financial transactions, energy delivery systems, or political party affiliation and activism. What is particular to CEU's overall approach to big data and is systematically promoted at the University is the use network science as a tool that is useful to all our disciplines. While network science and big data analysis are not the same, we have made the choice to try to link big data and network science in our educational and research activities as a way to provide a better organized and at the same time more applied framework to these activities. A particular attempt illustrating this approach is, for example, the research project of the CEU Center for Media, Data and Society on big data, bots, and the Internet of things. The projects looks at how information technologies have transformed world politics, while acknowledge that this transformation was not always for the better. More precisely, it tries to understand how technology connects us, which in turn helps to reveal the strengths and weaknesses of the Internet, including the "Internet of things". It looks, for example, at how "botnets" (a word created by combining "robot" and "network") can be used to promote political messages, or support criminal enterprises taken advantages of billions of connected devices, some of them unprotected.

A particular challenge in creating an institutional framework that would be effective for the preparation of the next generation of experts in big data analysis consists in the availability of faculty members with sufficient expertise in this area. This is even more a challenge considering that CEU does not offer programs in the sciences, but mainly in the social sciences and humanities. CEU has addressed this challenge in part by relying on the expertise of colleagues from the Department of Economics (many of whom have a strong background in Mathematics), but also by developing the Department of Mathematics and Its Applications and the Center for Network Science. The Department of Mathematics and Its Applications is the newest department of CEU and it has been created with the specific mission of promoting education and research in mathematics that is relevant for the social sciences and humanities and can help with specific research projects in these areas conducted by CEU faculty. While there have been doubts over the years that this mission could be achieved, big data proved to be an excellent examples proving how scholars and graduate students in the social sciences, humanities, and mathematics could work effectively together. Several faculty members in the Center for Network Science have a background in Mathematics, but also in Physics and Biology. Recently, CEU decided to appoint the first faculty member whose title will be "Data Scientist".

Developing Data-Literate Students

Steve Matson
Dean, The Graduate School
University of North Carolina at Chapel Hill (United States)

The term "big data" has different meanings to different people. Indeed, a quick Google search yields dozens of definitions. For the purpose of this brief article I will define this term as follows: the acquisition, management, analysis, and use of data (including an appropriate level of technical competence) with an understanding of the nature of data and its broader implications for society.

In our technology driven culture, data literacy has become essential to learning at all levels, to research and scholarship, to translational endeavors, and to future student career success. In fields from healthcare to the humanities, the acquisition, management, analysis, and use of data has become a required skillset for all college graduates and this is particularly true for students pursuing graduate degrees. A 2011 study at McKinsey Global Institute predicted that by 2018 there will be a 50% gap between demand and supply of deep analytical talent, amounting to a shortage of more than 150,000 data scientists.[1] Furthermore, the report stated, there will be an additional need for 1.5 million data literate managers and analysts with the skills to understand and make decisions based on the analysis of big data. The Labor and Economic Analysis Division of the North Carolina Department of Commerce estimates that more than 18,000 data science jobs will be created in the state between 2010 and 2020, representing 4% of all newly-created jobs. Nearly all of these new positions will require a bachelor's degree or higher.

The availability and volume of data, along with new ways of sharing information, are transforming how researchers conduct and communicate their work. This is a topic of immediate concern for scholars and scientists across disciplines. The University of North Carolina currently has several high-profile projects involving expert multidisciplinary teams that address the disruptive potential for data in various fields of research. One aspect of working with data illustrated by these projects, and critical to student development, is the ability to form and function in cross-disciplinary, often multi-institutional teams with members of diverse skillsets and backgrounds. Another aspect worth noting is how information technology is changing the paradigm of the traditional publishing model in research.

In addition, the availability and integration of data is changing and often disrupting disciplines, organizations, and industries across sectors. Yet data alone are useless. To be of value, data must be structured and analyzed in a timely manner to gain insights that can inform decisions and improve performance. Our ability to understand the nature of the world around us will increasingly require us to be more creative in how we aggregate and analyze data.

In 2013 a group of 35 faculty and staff from 15 different disciplines was convened to explore the subject of data studies within the undergraduate and graduate curricula at UNC-Chapel Hill. This group concluded that we must prepare our graduate and undergraduate students to be data literate in order to increase their competencies for academic pursuits, to raise their competitive advantage in the marketplace of ideas and jobs, and to increase and enhance their translational skills. Each student would be equipped with a data skillset, to be customized with appropriate technical depth, that includes the following:

---

1 "Big Data: the next frontier for innovation, competition, and productivity," McKinsey Global Institute, June 2011

contextual understanding, critical and analytical thinking, problem solving, data analysis, communicating and presenting findings, decision making, and basic technical knowledge related to databases, code, storage, and security.

Importantly, the working group concluded that data studies should to be available to all students in some form and in all disciplines. The group noted that there appears to be a natural separation between students who have strong tendencies toward science/math and those who do not, at both the undergraduate and graduate levels. For example, while a course such as data analytics might be offered to all students, the curriculum should be developed with more rigorous technical requirements for science/math-intensive students, and tailored differently for those less inclined to the quantitative.

The opportunity in developing a more focused and compelling data studies curriculum will be to connect existing expertise and courses, and identify and fill gaps to achieve learning outcomes consistent with a strong data studies skillset. New courses will need to be created that provide both a structural framework for integrating technical and non-technical aspects of data studies, as well as a contextual framework that takes advantage of the University's strengths in the humanities and social sciences. These courses will need to take into account the differing mathematical and computational skills across the schools and curricula. And, finally, the new focus on data studies and the curricular offerings will need to be presented in a way that entices all populations of students.

At the graduate student level, the professional schools have the opportunity to create certificate programs in data studies or data sciences that specifically address the needs of their students. Certificates at the graduate level are often interdisciplinary and involve courses outside of one particular school. The graduate school offers support for course and certificate development in addition to providing a framework for the creation of professional science master's degree programs. A graduate certificate in qualitative data studies is already offered in the School of Education and is available to students in other graduate programs on campus. The main purpose of the certificate is to advance interdisciplinary qualitative inquiry through the knowledge, skills, and application of multiple qualitative tools and techniques. The School of Information and Library Science offers a certificate in digital curation, the main goal of which is to teach students how to plan, manage and implement practices that ensure the long term integrity and use of resources created in digital form. The certificate aims to prepare graduates for data-intensive jobs after graduation. This certificate program is currently being expanded to create a new professional science master's degree in data curation designed to train the next generation of information science professionals. The departments of Computer Science, Mathematics and Statistics/Operations Research have proposed an ambitious new professional science master's degree in Data Science designed for students with the technical skills to put big data to work in many sectors of our work force. Finally, a new interdisciplinary PhD in Health Informatics, designed to offer advanced training for careers in research, education and practice at the intersection of computing and healthcare, is in the planning stage. The program will prepare future leaders in health informatics research, innovation and policy.

The University of North Carolina at Chapel Hill is committed to having its faculty and students see the world broadly and think critically and multi-dimensionally. Therefore, data literacy should be embedded in the curriculum for all students to promote an understanding of the influence of data on individuals and society. In addition, we are committed to the creation of new degree programs at the intersections of disciplines to prepare our future data scientists while ensuring that the contextualization of data includes examining such areas as the effects of data proliferation on social constructs, communication, privacy, security, and ethical considerations.

Preparing Graduate Students for the World of Big Data

Nirmala Rao
Dean, Graduate School
The University of Hong Kong

There are numerous interpretations of the term "Big Data." For the purposes of this paper, I use the following definition, "the collection, aggregation or federation, and analysis of vast amounts of increasingly granular data" (Cate, 2014). I will consider the preparation of research students to access, collect, analyse, manage, and curate large data sets at The University of Hong Kong.

Hong Kong Context

Like all universities, teaching, research and knowledge exchange in The University of Hong Kong are affected by both local and international contexts. The University of Hong Kong is one of eight publicly funded tertiary institutions in the territory and the Hong Kong Research Grants Council funds a number of research schemes from these institutions. Among these is the Theme-based Research scheme that focuses on topics that are key to the development of Hong Kong and beyond. A new theme, Advancing Emerging Research and Innovations Important to Hong Kong,[1] was added for the 2016-2017 funding exercise and Big Data is one of the grand challenge topics under this theme. Computational and analytical approaches suitable for handling very large and potentially heterogeneous data are covered under this topic and proposals related to machine-learning research, development of distributed computational approaches, cloud-enabled implementation, to cross-domain Big Data analytics have been invited. The addition of this area signals the importance given to Big Data in the territory. This does not mean that research involving Big Data has not been occurring or been funded in the past decade. Indeed, there have been many funded research projects that have been concerned with Big Data at The University of Hong Kong and our sister institutions. For examples, Professor Victor O.K. Li of the Department of Electrical and Electronic Engineering has recently won a Microsoft Research Asia Grant for the project "Air quality monitoring with deep learning on spatio-temporal heterogeneous Big Data." There are only 15 air quality monitoring stations in Hong Kong, which are not enough to cover the whole city. By collecting urban Big Data, such as vehicular traffic, wind speed, humidity, and the urban landscape, which are available throughout the city, and by correlating such data with measured air pollution data, one is able to estimate air quality throughout Hong Kong. The Faculty of Architecture has recently opened a new HK$10M Building Information Modelling lab at its Shanghai Study Centre to analyse building and city performance (e.g., cost, energy, people, vehicle flows) using remotely sensed big urban data. It has also made an important contribution to the study of healthy city design by developing a new Big Data platform (UKBUMP) for the study of gene-environment (built, social, and natural environments) on individual health.[2]

Hong Kong has a number of reliable large databases. For example, the Hospital Authority which oversees all public hospital in Hong Kong has a number of reporting systems and the same is true for the Social Welfare Department and the Education Bureau.

1 The other themes are: Promoting good health; Developing a sustainable environment; and Enhancing Hong Kong's strategic position as a regional and international business centre.
2 http://www.tandfonline.com/doi/abs/10.1080/19475683.2015.1027791

Institutional Context.

We currently do not have taught Master's programmes that focus explicitly on Big Data but the topic is clearly emphasised in many programmes in Science, Technology, Engineering and Mathematics (STEM), library sciences and in medicine. In particular, both the Computer Science and the Electrical & Electronic Engineering Departments at The University of Hong Kong offer undergraduate and graduate classes which emphasize big data. That stated, a large number of colleagues work with Big Data and this gives our MPhil and PhD students ample opportunities to work on such projects. Colleagues work on data retrieved from government departments and on large, longitudinal family and household surveys such as the Family and the Children of 1997 Birth cohort studies. Infrastructure required includes high performance computing facilities. This is available from the HKU Computer Centre, and dedicated machines are also available in various departments. In terms of HKU funding, some work on big data is performed under the SRT on Computation and Information.

Opportunities and Challenges in preparing graduate students to handle Big Data.

There are potentially enormous societal benefits that can emanate from the analyses of Big Data. These data sets enable us to answer research questions of great importance and lead us to new conceptions of data collection and how we go about doing research. At the same time, we are also cognizant of the many challenges associated with preparing the next generation of researchers who will utilise these Big Data. Challenges are related to the nature of the data, our conceptions of research, preparing students to work with Big Data in a responsible manner, and addressing disciplinary differences.

Problems related to the nature of Big Data.

The size, complexity and heterogeneity of Big Data actually create problems when one wishes to use them to tackle research problems. What data should students use? What should be discarded? Not only do students have to be supported to learn how to organize, retrieve, mine, analyse, and model these data, they need to learn how to link data from different sources because of the research value of data integration.

Preparing students to access, collect, analyse, manage, and curate large data sets.

There are issues related to the access to Big Data and students need to be supported to navigate the hurdles. It is currently not easy to access big data sources managed by various government departments in Hong Kong as there is strict control because of privacy concerns. Students may have to access de-identified data sets but records have to be re-identified to answer research questions. For example, de-identified data from the Hospital Authority and the Educational Bureau may need to be matched to answer a research question.

Dealing with ethical and legal issues associated with managing Big Data.

All our MPhil and PhD students take a compulsory course on research ethics and have to do the CITI institutional research training programme. They may need additional courses to enhance their awareness of specific ethical and legal issues associated with managing Big Data.

Interpretation.

One of our challenges is to help ensure that students are aware of the criticisms directed towards the analysis of Big Data. In some cases, the analyses of smaller, deeper data sets may provide a more appropriate answer to a research question than a larger one. We also assume that the analyses of Big Data collected in the past will help us predict the future but this is not always the case. Further, data must be contextualised and interpreted in their social, cultural, and economic contexts. Just like we ensure all our research students do a course on Research

Ethics, perhaps we need to mandate some sessions covering what every graduate student needs to know about Big Data.

Armstrong (2014) argues that the Big Data revolution will transform how we live, work, and think. It will certainly transform the way we educate the next generation of experts.

References

Cate, F. H. (2014). Privacy, Big Data, and the Public Good. Science, 346 (6211), 818.

Armstrong, K. (2014). Big data: a revolution that will transform how we live, work, and think. Information, Communication & Society, 17(10), 1300-1302.

Big Data Analysis for Graduate Students at Purdue University

Mark J. T. Smith
Dean, Graduate School
Purdue University (United States)

As society has advanced technologically and established sophisticated networks of various types—such as cellular phone networks, data networks, social networks, and sensor networks—the volume of data available for exploitation has exploded. These data come in a wide variety of different forms, collectively providing tremendous potential to address challenges heretofore beyond our reach. Many challenges that could be tackled through big data analysis have been cited in the news media and the scholarly literature, perhaps the most exciting of which have been in the health care arena.

The term big data has been used liberally in many communities, often with different and nuanced meanings. In computer science and computer engineering circles, the term is often associated with data embodying the following characteristics: the volume or amount of data is huge; the variety of data is diverse, for example, including text, audio, images, and video from different sources such as the web, news sources, satellites, and sensors and may be of different resolutions; and the data have an associated velocity, that is, there is a rate at which the data are being collected, which can mean that the information is time sensitive. Another issue often associated with big data is accuracy or veracity. While volume, variety, velocity, and veracity are often-cited distinguishing features associated with big data, one could argue that the novelty in this emerging field lies in the use of data in new and creative ways, the gleaning of new insights from data sources that have not previously been considered, and the use of data to predict outcomes and events reliably.

Big data, as an area of study, is now the topic of conferences and workshops throughout the world. Furthermore, discussions are in progress within the standards bodies (such as the ISO/IEC) to consider appropriate new world-wide standards. These developments underscore the importance of exposing graduate students to this growing area. Big data is now a major area for education and interdisciplinary research in which Purdue University is investing. A number of faculty lines were recently created for scholars skilled in the science of managing and analyzing digital data. These new faculty members, together with the team already working in this area, will help shape the campus vision and agenda for graduate education in big data at the university.

The use of data analytics in government, industry, and academia is increasing rapidly and universities should be appropriately exposing students to and training students in data analysis methods in response to the changing times. At Purdue University, graduate students in computer science and engineering are routinely being exposed to the principles of data mining and learn about classification, clustering, and mining approaches, to mention a few. Students are also introduced to the concepts, models, and algorithms in machine learning and, more generally, the theory of statistical inference. Big data is also featured prominently in the School of Management, which has a Business Analytics MBA program. In this program, students learn the basic concepts in econometrics and multivariate analysis; are exposed to estimation, testing, and regression procedures; and are introduced to the concepts, techniques,

tools, and applications of data mining.

While we are seeing the breadth of exposure increasing within the graduate curricula at Purdue, the depth of exposure is increasing as well through an expanding research enterprise focused in this area. It is from engaging graduate students in research that we anticipate the next generation of experts will emerge. Big data activities at Purdue are capitalizing on research strengths in CERIAS (Center for Education and Research in Information Assurance and Security) which is the largest academic multidisciplinary center of its kind; building on the university's Cyber Center, which focuses on creating systems, tools, and cyberinfrastructure; and leveraging the university's Science of Information Center, which is Indiana's first and only NSF Science and Technology Center.

As part of Purdue's expansion in big data, students are provided opportunities to engage in a wide variety of research projects. For example, there are projects underway in the College of Agriculture where researchers are exploring ways to improve farming, through the analysis of mixed data. The volume of information available is massive and may include imagery, weather data, and sensor data. Faculty are discussing and developing "grow on demand" and "make what you need" strategies and engineering crops for the future.

Another example is urban city modeling and simulation. By examining mixed data that includes population information, traffic flow, mass transit, weather, and citizen behavior, Professor Dan Aliaga and colleagues have been developing models for urban city design. The models incorporate information such as temperature, humidity, sunlight, and wind patterns; pedestrian behaviors; traffic patterns; and economic conditions. These models can help an urban planner understand what modifications would be most cost effective in order to achieve a desired outcome. For example, if the outcome were to revitalize a certain neighborhood, the models could help determine appropriate changes, perhaps to roadway configurations, parks and green spaces, and pedestrian crossings. Other goals might be to lower the temperature by a few degrees, lower the humidity, or improve traffic flow at the lowest cost. Analytical models of this type can inform decisions about appropriate building materials, green space configurations, limits on the heights of buildings, the design of roadway infrastructure, and so on, to meet desired outcomes.

Genomics research is yet another area in which data analysis is promising to be transformative. An example of this is the work being explored by Professor Jun Xie, where large data sets of patients with particular diseases are being examined. Using the genome sequence data, information about patient treatments received, and treatment outcomes, Professor Xie is attempting to predict the probably of treatment success directly from genomic data. Using genomic data to predict risk of infection and optimal treatment options is an exciting area of research for many students.

Unquestionably, there is a need to educate graduate students in big data analytics, as it has the potential to enable innovation and enhance services in many industry sectors. However, organizations that employ big data analytics will need to establish policies and practices that address privacy, confidentiality, security, consumer trust and confidence, responsibility, and accountability. As universities expand programs in data science and big data analytics, the associated ethical issues that arise should be embedded in the curricula, particularly those issues related to confidentiality, privacy, and transparency. Confidentiality and privacy infringement may not always be apparent when big data analysis is applied in some industries. Similarly, transparency can be another issue deserving of attention. Big data

analysis can provide valuable information. However, when massive amounts of data about individuals are collected and shared with other agencies without the knowledge or consent of the individuals involved, ethical questions arise. People have the right to know how these data are being used and who has access.

Universities have an obligation to assure that students educated in big data analysis also have the accompanying awareness of the ethical issues and have received some level of training on sound practices for managing data. At Purdue, data analysis ethics is treated in the Responsible Conduct of Research (RCR) training workshops taken by students and is also covered in a few of the courses. As advances continue in the field of big data analytics, it will be important to update ethnics training accordingly.

Preparing the Next Generation of Experts

Nicky Solomon
Dean, Graduate Research School
University of Technology, Sydney (Australia)

As with most Australian universities, at University of Technology Sydney (UTS), big data is a big topic. Indeed it is part of the everyday language of the institution, especially when the talk is about change in the university's infrastructure, policies, research strengths, curriculum etc. It is talked about in terms of the management and analysis of the data of the institution itself, the management and analysis of data in research projects and the matching of academic expertise in big data with industry research. But also and importantly, a large focus of the talk is on student learning of skills that will help them to manage, analyse and aggregate big data.

This focus on learning is seen as central to the preparation of UTS students for future work. 'Big data' skills are understood to be core skills for all. Symptomatic of this, is the fact that UTS is seeking to embed a module, euphemistically called 'Arguments, Evidence and Intuition', in all undergraduate coursework programs. Indeed, the handbook description of this module sets the scene in relation to the need to develop 'basic' skills for using data – 'Using primary research materials, governmental reports, stories and claims drawn from current media and other sources, participants analyse and identify key features of numerical data and graphical illustrations used to support argument. By examining the ways that quantitative data can be collected, used, and abused, as evidence for supporting argument, participants have an opportunity to develop habits of mind that can be applied to the questions that should be asked, as informed citizens, of arguments and of the supporting data.' (http://handbook.uts.edu.au/subjects/36201.html)

There are an infinite number of definitions of big data but in general, there seems to be an agreement not to define it too narrowly or technically. For a start it is suggested that 'size' of the data set does not matter, as many of the storage, legal and ethical issues are relevant to all data sets, no matter how large or diverse. This is a question relevant to this symposium – Is it useful for universities to define 'big data' for institutional, research or pedagogic purposes? Does it matter to the research students in their learning?

This brief paper will focus on the preparation of graduate research students to do their research and to live in a big data world and on how UTS is responding to the challenges and the opportunities presented by the growing place of big data in research education. The paper will describe some of these responses – many of which are relevant to academics as well as research students.

There is considerable variation in the way the various disciplines engage with big data, with some seeing it as something to manage, others see its potential to ask different questions that require different kinds of aggregations and analyses, and others seeing it as a technology that signals a powerful shift in research and knowledge domains.

At times big data is considered to be only relevant to STEM research/ers and this sits alongside a stereotypical view that HASS research is qualitative as opposed to quantitative. However in the context of today's imperative for research to focus on innovation and 'real' life problem solving, there is an increased emphasis on interdisciplinary research and as interdisciplinary research connects data, knowledges and people, these disciplinary differences may became less marked. As the UTS Director of the Centre for Connected Intelligence (CIC) says of Data Science & Analytics Literacy, 'researchers in all disciplines will increasingly need a higher level of statistical and data literacy, as analytics impacts all

sectors'. Anecdotally the chat is that all academic and researchers need maths, and that to a certain extent we all need to be data experts.

At UTS the education responses go beyond the technical aspects of analysing, managing and curating large data sets, although at this stage, a key focus of investment is on providing resources for the required e-infrastructure, for data tools and for the learning of various technologies. For example:

- UTS has a partnership with Intersect www.intersect.org.au which is Australia's largest Research support agency, by providing various eResearch solutions. Experts work with UTS professional and academic staff to design and develop various technologies, platforms and training to help researchers, including research students, to manage, share, store data.
- Preparation involves various kinds of technical training including, for example, training in particular tools & practices, and making them aware of others. Intersect-run courses that touch upon data management and analysis include the ones on Excel, Google Fusion Tables and Open Refine.

Some examples of UTS specialised programs that focus on data science and/or "big data" issues are:

- Master of Data Science and Innovation – http://www.uts.edu.au/future-students/analytics-and-data-science, and a PhD in Learning Analytics. These sit within a central unit of UTS, Centre for Connected Intelligence (CIC) and prepares 'students to participate in a variety of emerging careers with the growth of data science; the data griot, data analyst, data artist, data journalist, mobile behaviour analyst, data-driven policy expert, advertising insight and online community manager to name a few.'
- The UTS Advanced Analytics Institute runs short and degree courses on data analytics, http://www.uts.edu.au/research-and-teaching/our-research/advanced-analytics-institute

Preparing graduate students to navigate the legal and ethical issues surrounding big data collection and analyses occurs in various places. For example CIC runs symposia on these issues regularly, the Graduate Research School offers modules for doctoral students and early career academics. Importantly also Research Ethics applications require careful attention to the range of legal and ethical issues that are arising from the possibility or not of de-identification or anonymisation, given the quantity of (personal) data 'out there' and the increasing sophistication of data matching algorithms. Many other ethical issues are addressed in data training, regardless of whether the data is considered to be big data. These include: the in/ability to remove/withdraw information, the apparent ongoing 'life' of data in the cloud, the potentially never-ending re-use of data for purposes other than their original project, together with the real or perceived concerns about the security of the data. Universities in Australia work with national and university codes such as the Australian Code for the Responsible Conduct of Research, Research Data Management Vice Chancellor Directive, Ethics and Research Misconduct Directives. These provide researchers with a useful framework for working with legal and ethical issues.

Perhaps an area that needs more attention is the reification of data in this Big Data and Data Science space. While big data tools offer endless opportunities for bringing together different bodies of data to produce new knowledges and solve complex problems, there is a but…. Data, without humans, won't change the world. Data requires analysis and interpretation and as students learn how to use tools and platforms they also have the challenge of working with data in ways that connect data with the social world within which it exists. It has huge potential but this potential can only be realised with a socialisation of the data.

6: Enhancing Research Collaboration and Productivity

Big Data: Analytics, Collaboration and Commercialisation

Paul C. Burnett
Dean of Research and Research Training
Queensland University of Technology (Australia)

"My name is Andreas Jones and I am enrolled in a PhD Program at the University for the Future (UF). I have two University supervisors and an Industry supervisor and am a second year student investigating the utilisation of social media postings for diagnosing Mental and Emotional Disorders, particularly Schizophrenia. I am using the postings from seven major social media databases including Facebook, Twitter, Instagram and InstantCom to develop a qualitative algorithm to diagnose Schizophrenia using social media postings over a 12 month period. I use sentence structure, word frequency and type and emotion-linguistics. The study is a collaboration between my University, Kaggle, Mental Health United Inc and Data Analytics Inc. Mental Health United has offered $1million for the best predictive qualitative algorithm and Kaggle is facilitating the competition and subsequent commercialisation of the successful outcome. My two University Supervisors are psychologists with a clinical interest in Schizophrenia and my Industry Supervisor is a specialist data scientist (working for Data Analytics Inc) who has won a Kaggle competition in the past and will share any prize received and IP generated".

Humans, through technological advancements, have the capacity to generate more information or data than storage mechanisms can handle. "Big data" is a broad term for the large amounts of quantitative and qualitative data that can be captured and stored. The critical and important issue confronting society is how to make sense of this "big data" and appropriately analyse it to address significant and complex problems and questions. It is argued that humans must rapidly shift from the "Information Age or the Age of Data Generation" into the "Age of Analytics". Data sets are now so large and complex that traditional data processing and analytic skills and applications are inadequate. Consequently, contemporary data analysts and data scientists are becoming highly valued. Doctoral and research students around the globe should prepare themselves for the future by developing specific data analysis skills and techniques. They should be dual experts in their content and data analytics.

Companies around the world are mobilising themselves in innovative ways to make sense of "big data" by developing innovative approaches towards data analytics. Kaggle and SinoTech are case examples with Australian connections and will be described further in this paper. Both companies provide insights into opportunities for future collaboration with Universities and potential to encourage research, innovation and enhance productivity through analytics.

Kaggle is a platform for predictive modelling and analytics competitions whereby companies and organizations post their data and statisticians and data miners from all over the world compete to produce best models. This "crowdsourcing" approach relies on the fact that there are countless strategies that can be applied to any predictive modelling task and it is impossible to know at the outset which technique or analytic approach will be most effective. Kaggle (www.kaggle.com) has the world's largest community of data scientists with approximately 200,000 data analysts worldwide, from fields such as computer science, statistics, economics and mathematics. They compete with each other to solve complex quantitative and qualitative data problems, and the top competitors are invited to work

on the most interesting and sensitive business problems from some of the world's biggest companies. Kaggle provides cutting-edge data science results to companies of all sizes and has a proven track-record of solving real-world problems across a diverse array of industries including life sciences, financial services, energy, information technology, and retail.

Kaggle competitions work as follows. The competition host prepares the data and a description of the problem. Data analysts experiment with different techniques and compete against each other to produce the best models/answers. For most competitions, submissions are scored immediately (based on their predictive accuracy relative to a hidden solution file) and summarized on a live leader board. After the deadline passes, the competition host pays the prize money in exchange for "a worldwide, perpetual, irrevocable and royalty free license to use the winning Entry", i.e. the algorithm, software and related intellectual property developed.

Much "big data" set are qualitative. SinoTech Group is a China-based, globally focused provider of social intelligence technology and consulting services. Established since early 2007, SinoTech helps clients leverage social media and online channels to gain greater understanding about their brands, competitors, and industry trends in China and around the world. SinoTech's social intelligence services include customized research, consulting, and syndicated reports. Clients get access to a proprietary social intelligence platform known as "SIP: Enterprise". SIP: Enterprise (short for "Social Intelligence Platform") is a leading social intelligence technology for gaining insights and determining meaning from digital Media. SIP is a powerful tool for monitoring and analysing overall online brand awareness and sentiment. Utilizing sophisticated algorithms, SIP can understand and highlight conversation and authors that are writing positive or negative information about your company, products or services, in any social network site, blog, microblog, portal, review site, wiki, and forum. SIP: Enterprise also provides insights into customer retention and loyalty through analysis of customer comments towards brands, products, services and sponsored events. Most importantly, SIP: Enterprise can provide these insights in 11 different languages.

SO………

- How might large datasets be used to guide strategic investment in research, promote collaboration, and encourage research innovation and productivity? Strategic investment should be focussed on data analytics and developing innovative cutting edge mechanisms to make sense of data and address significant and complex problems.
- What opportunities do you see for "big data" collaborations among universities? Two opportunities emerge for Universities to collaborate. Collaborative cross-nation data generation and utilising a collaborative multidisciplinary approach to developed new analytical mechanisms.
- What opportunities do you see for collaborations between universities and other institutions and organizations (i.e. government agencies, companies, social media platforms)? Companies and organisations generate massive amounts of data that they need to make sense of and use to improve productivity and efficiency. Universities can collaborate with them to make sense of their data and generate solutions and answers to relevant problems.
- How can universities navigate IP issues in the big data world? By consulting and collaborating with legal firms and companies to protect the IP generated and to commercialise the outcomes from analytic mechanisms. Universities should ensure that their IP policy specifically addresses the issues and outcomes related to "big data".

The Influence of Socioeconomic Factors on Social Media Big Data Adoption by Migrated Students: A Comparison Study

Chen Chuanfu
Dean of the Graduate School
Wuhan University (China)

Watsala Athukorala
PhD Student in Library Science
Wuhan University (China)

In recent years more and more students migrate from one country to another for the purpose of pursuing higher education. Beine, Noel and Ragot report that the globalization of higher education, due to the migration of students, has grown considerably over the past 40 years (Beine, Noel and Ragot, 2012). World Education Service predicts that that by 2020, between 4.1 million and 6.7 million students will be studying abroad (ISMFS, 2010). While sending nearly a half-million students abroad to be enrolled in undergraduate or graduate degree programs, China hosted 328,330 international students in 2012 (CSC, 2012).

Big data can be defined as a changing, expansive, and valuable environment with a large amount of data. Surveys shows that most new migrated students frequently use social media to discover knowledge, access teaching materials, make connections with faculty, and handle daily affairs. Social media has become a huge database and skill to help students adjust themselves to food, weather, and to make financial, health care, transportation arrangements and so on. It also shows that new migrated students have more social-economic concerns in using social media efficiently to handle the challenges in a new environment.

Literature Review

Some research has identified that the protection of privacy is the number one concern in the world of social media (Gruzd et al., 2012). Copyright is also cited as a major concern for scholars worried about the loss of intellectual property rights to their research (Collins & Hide, 2010). Some researchers have identified different significant factors that affect users' perceptions, namely: efficiency, entertainment, community divineness, privacy, user friendliness, and navigability (Ellahi and Bokhari, 2012). However most of the previous studies have focused on technical factors that influence migrated students in their use of social media.

Statement of the Problem and Research Method

Literature shows that little has been done to understand which socio-economic factors concern migrated students the most when using social media. The purposes of this study are to identify their major concerns when they adopt social media to explore the big data environment.

This research has selected eight factors as test indicators, including legal risk, copyright, privacy, credibility, convenience, permanency, time spent, cost, any other (please mention), and to use gender, age, discipline enrolled, country of origin, host language proficiency for correlation analysis. A survey method was used to collect data. The population sample was selected from international students studying in Wuhan University in China who

are originating from North America, Africa, Europe, and other Asian countries. The response rate was very high from both all international (n=216, 86.4%) and Chinese students (n=209, 83.6%).

Results

Based on their importance, the factors that influence migrated students in using social media are listed in the following order: privacy, convenience, time spent, cost, legal risk, credibility, permanency and copyright. For domestic students, these factors are in the following order: privacy, convenience, credibility, time spent, legal risk, cost, permanency, and copyright. The three factors of "cost, stability, copyright" are ranked higher among migrated students than domestic students.

Average values obtained after assignment to eight factors
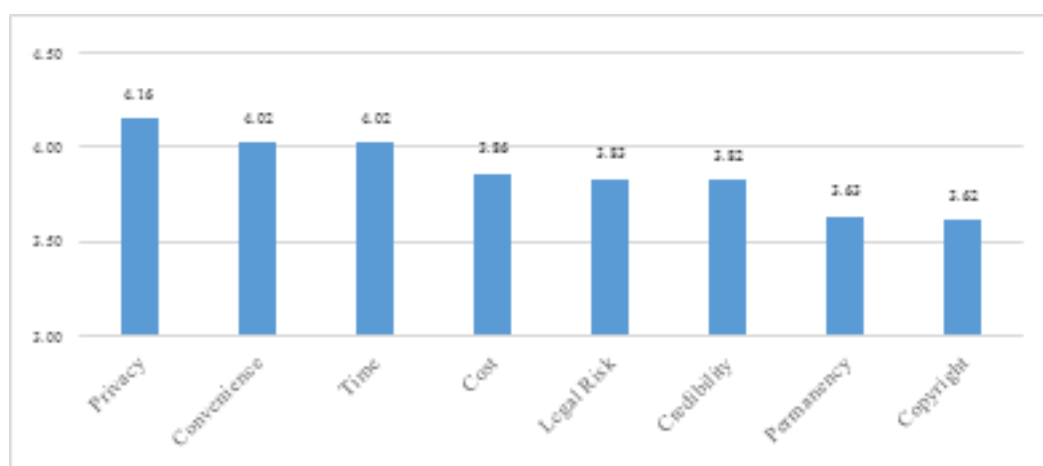
1. Privacy and convenience are always listed as top concerns, regardless of the influence of variables; privacy and convenience are always the most important influencing factors for migrated students. The factors of cost, permanency and copyright are ranked higher among migrated students than domestic ones;

2. Students of different genders have different concerns. Except for the copyright factor, which weighs more among female students than male students, in general, all other factors have more influence on male students than female students;

3. The influencing factors are different among students of different ages. Older students pay more attention to privacy, convenience and legal risk than younger students. As the length of study abroad grows, students are more concerned about credibility and permanency of social media;

4. The influencing factors are different among students with different disciplinary backgrounds. Humanities and social sciences students rank the importance of privacy, credibility and convenience higher than STEM students. However, factors of legal risk, cost, time spent, and stability are more important to STEM students than humanities and social sciences students;

5. Students from different continents (countries) have different concerns. While students from all the continents represented consider privacy as the most important influencing factors, their opinions vary widely in other factors such as legal risk, credibility, and copyright;

6. Host language proficiency is a contributing factor to influence students' concerns. In general, these concerns are ranked higher among students with low host language proficiency.

Conclusion
For international students, the use of social media to discover useful digital resources is an effective way to adjust themselves to the new learning and living environment. The major contributing factors for their use of social media are privacy, convenience and time spent. Based on this finding, social media providers should make improvements to protect user privacy. As migrated students depend heavily on social media for their learning, research needs as well as their daily life in today's digital environment, it is important for university administrators to integrate media literacy education into academic training and professional development programs for both migrated students and faculty. In addition, when media literacy education is provided, we also need to tailor it to people needs by paying attention to their differences in gender, age, discipline enrolled, country of origin, host language proficiency, and to help migrated students take advantage of the social media and to better protect their interests.

Remove Barriers, Build Trust & Empower Research

Niels Dam
Vice President, Product Management
ProQuest

As an information provider that is home to billions of documents, ProQuest knows Big Data. If the first Homo Sapien from the Cenozoic period were still alive and had read one ProQuest document each day of its life, he or she would still have not completed reading through the ProQuest data vault. Making content sets of this size useful to contemporary users requires sophisticated server architecture, information retrieval technologies, and rights management programs. The challenges and opportunities posed by the large amount of content cannot be solved by individual research groups, governments, companies or community associations alone, but is a collaborative effort where each actor can provide a positive contribution.

These massive computing resources offer scholars within many fields newfound opportunities to discover previously hidden patterns in the world. By mining billions of medical records, for example, the nineteen-year-old Brittany Wenger built a neural network classifier that could detect breast cancer in patients. At Yale University, research using machine imaging software to overlay the covers for each decade of Vogue Magazine from the ProQuest data vault, revealed that in many decades, the covers featured hauntingly similar depictions of women. Others, such as researchers in linguistics and the humanities, have used natural language processing techniques to uncover plagiarism buried for years within doctoral dissertations and scientific publications. These discoveries were made possible by the rise of powerful computing systems and the enormous collections of digitized data housed within organizations like ProQuest.

Today, too often, a researcher's time is spent on administrative tasks collecting and preparing datasets instead of analyzing and developing new insights. They build their own data-infrastructure, they in isolation negotiate and secure intellectual property rights for each dataset, and they clean and prepare each of these datasets for analysis. This is both inefficient and limits the research to be performed by large specialized groups. It is up to all stakeholders to create change that will empower students and researchers alike to unlock this new frontier of research.

Enabling Big Data analysis means fundamental changes to how content is created, delivered and distributed. Online content publishing systems are designed to enable humans to publish, find and read individual articles, books and reports. They are not designed to provide access to datasets, have machines systematically read them or provide clear guidance on what information is substantive instead of structural components that support the human discovery and reading experience. As an information provider, we are undertaking the significant investments needed to change this. These changes are not trivial nor will they be effectual without a collaborative effort to build trust among content creators and researchers using the information for analysis. Content creators are rightfully concerned that enabling massive local installations of their information has the potential to undermine their intellectual property rights, but we also have to realize that we as a community of creators, information providers and researchers need to collaboratively find a way to enable these new forms of research. We need to take steps to rebuild the trust that content providers have legitimate concerns and researchers obviously are not out to misuse information.

Learning from initiatives like the Wharton Research Data Services (WRDS) group, out of Wharton Business School, can provide a sustainable and efficient way forward to

enable researchers more broadly to take advantage of the new opportunities. WRDS provides a secure research and business intelligence platform that provides a set of highly relevant and complimentary (as well as fee-based) databases while enabling researchers to use a flexible set of tools to do and share their analysis. This alleviates the content creators' concerns for their intellectual property rights and saves researchers time with administration and data clean up, while having negligible effects on researchers' ability to perform their analysis. Having a shared infrastructure will not only provide efficiencies in the research process, but will also give researchers a new avenue for collaboration and sharing of results across traditional subject boundaries.

All available evidence suggests that this form of collaborative research will only become increasingly ubiquitous in the future and not be limited to primarily financial and medical data, which has lead the way in recent years. Researchers of ancient texts can now work with statisticians and leaders in natural language processing to break ancient ciphers in political and sacred texts. Mathematicians can now help professors of drama to discover new Shakespearean texts. Research teams are finding that the most interesting applications of Big Data revolve around the combined talents of many scholarly communities.

Big Data is, however, more than enabling data-mining of text or numerical databases. As audio-visual information explodes and video become ubiquitous, the need to unlock the information hidden within is increasingly urgent. Having launched a video curation service for universities for the purpose of getting historical video out of storage boxes and into the hands of students and researchers, we at ProQuest have direct experience with the value of combining the voice of individuals with the written sources in understanding history. Funding and research is needed to revitalize speech recognition of audio-visual material, so that it goes beyond the command recognition we use in our phones and cars today and become as efficient as optical character recognition is for historical texts.

Because ProQuest wishes to continue helping scholars achieve their goals, we are eager to enable researchers to data mine the millions of unique documents and audiovisual material we index. We routinely ship large data sets to universities around the world, and are eager to continue developing new solutions that address the needs of the research community. We aim to provide the resources with which the next generation of scholars can continue advancing the steady progress of knowledge.

Using Big Data in Doctoral Research: Is More Necessarily Better?

Lucy Johnston
Dean of Postgraduate Research
University of Canterbury (New Zealand)

Whakatikatikahia koe mō te Ao hurihuri nei
Prepare yourself for the new world

By analyzing clusters of search terms Google was able to predict influenza outbreaks faster than could be achieved through hospital records. A Harvard assistant professor of government's statistical model, based on six crude variables from previous cases, could outperform the judgements of 87 law professors in predicting the outcome of Supreme Court cases. And credit card companies found people who buy anti-scuff pads for furniture are highly unlikely to default on their card payments. The availability of big data appears, then, to be providing insights across a wide variety of domains.

The ease, and reduction in cost, of storing large quantities of data has led to a previously inconceivable increase in the availability of data which has had, and will continue to have, profound impacts on research, including doctoral research. Importantly, doctoral candidates are no longer restricted to using data which they are able to collect in the limited timeframe, and with the limited resources, of their doctoral candidature. They will have access to national and international datasets and networks of researchers, which will facilitate not only access to *more* data but also comparisons across time and across location. Data from doctoral candidate's research will also add to the data available to others such that the impact and influence of that data may be wider than that which might be achieved by the doctoral candidate in isolation. Doctoral candidates will become both providers and users of big data. This greater access to data is then combined with huge increased power in desktop computers allowing for rapid data analysis to be conducted, in contrast to years past when researchers spent long periods of time specifying the code for a single analysis and waited, often overnight, for results that now take a matter of seconds.

The exciting research opportunities afforded by such big data must be grasped. At the same time, however, some caution must be exercised to ensure appropriate and meaningful use of these data. Others at this summit will highlight the challenges that big data poses in terms of data protection, privacy and IP protection and so I will not comment on these issues here, except to note that doctoral candidates – those at the beginning of their research and leadership careers – are an important group for whom to ensure protection and not exploitation through the provision of, and access to, big data. The issues that I wish to highlight relate to ensuring that big data is used to produce high quality research findings. Big data can, but does not necessarily, lead to good research! Below I consider 3 crucial factors that we, as leaders of doctoral research programmes, must consider when guiding the use of big datasets by doctoral candidates.

What data analytic skills are needed? Fundamentally different computational skills are required for the analysis of big datasets than for the analysis of small datasets. We are not talking about adding a few more data points to a conventional analysis of variance or regression analysis. Analysis of big datasets requires data mining – the identification of relevant variables and their inter-relationships – and the ability to develop and work with appropriate algorithms to enact effective searches of the datasets. Computers blindly follow

the programmed algorithms, even if those algorithms are developed by people unfamiliar with the research questions being asked (such that the wrong questions are answered) or by people who lack appropriate computational expertise to develop appropriate search algorithms. Notably, and highlighting the risks that arise here, the areas where the availability of big data might have the most profound effects – in medicine and public health and in social sciences – are those areas in which researchers, including both doctoral candidates and their supervisors, do not typically have the relevant computational skills to access and analyze big data. Institutions must ensure the availability of such training to enable effective and meaningful use of big datasets. Given the lack of computation skills amongst social science researchers, there is a tendency for those with the computational skills (e.g., engineers, physicists) to apply these skills to new research areas. A cautionary tale from an adjunct engineering Harvard professor illustrates the risks. Based in Rwanda the researcher developed, from mobile phone data, models of the daily and weekly commuting patterns of villagers. Noticing patterns in the data it was hypothesized that commuting patterns were related to the onset of communicable diseases. Linking his data set with that from the ministry of health revealed a powerful effect – villager movement dropped prior to cholera outbreaks such that the magnitude of an outbreak could be predicted from the decrease in movement. It turned out, however, that the model was not predicting cholera outbreaks at all, but rather pinpointing the location of flooding (which limits movement and increases risk of illness outbreaks): "Ultimately, all this analysis with supercomputers was identifying where there was flooding – data that, frankly, you can get in a lot of other ways." Without knowledge of the context in which data were collected and the research questions being asked, highly sophisticated data analysis on huge quantities of data can be fundamentally flawed. What is needed is a combination of researcher skills (asking the right questions and interpreting the findings) and what the computer is good at (computation and statistics). One alone is not sufficient. It is also important to recognize the human superiority in seeing patterns in data – expert researchers must be looking at and interpreting the data patterns revealed through sophisticated computational modelling to glean the most from those data.

What data is available? Big datasets can be structured or unstructured. Structured datasets have an *a priori* agreement as to the nature of the data to be collected by all contributing to the dataset. For example, the interRAI project looking at the lifestyle and health of the elderly of which New Zealand is a partner has a standard 250 question survey administered to all participants in over 30 participating countries. Researchers can have confidence about the data contained within this project, and especially its consistency, allowing meaningful, for example cross-national, comparisons. Unstructured datasets in contrast have no prior agreements about the structure of the data or the manner of collection. Such datasets may hold very useful information, but greater caution must be exercised in interpreting findings from such datasets. Further, the biggest gains may not be in larger datasets per se but rather in the linkages between different datasets; doing so will, of course, raise more issues about the quality of the data collected and its compatibility to be combined for analysis.

What research questions are being asked? My final question is perhaps the most profound: will the availability of big data change the nature of how we conceive of research and research questions and if so with what consequence? The traditional approach to research, including doctoral research, is to identify a pertinent research question (often derived from a theoretical basis or framework) and consider how it might be tested and then, as necessary, collect relevant data. The availability of big datasets might invert this process with researchers starting with a data set and asking what they might find in it. Rather than looking at data to answer a research question the researchers are looking at data to find the question to ask, or indeed the question that the discovered data patterns answer. What impact

might this shift from a curiosity-driven to a data-driven research agenda have? History is littered with serendipitous research findings – penicillin, the microwave oven, post-it notes – but society needs curious minds as research, policy, education and business leaders. As leaders in doctoral research education we must develop approaches, including computational skill development, that enable doctoral students to take advantage of the huge opportunities provided by big data without losing the basics of fundamental scientific enquiry.

Biographical Sketches of Participants

Professor Brenda Brouwer

As Vice-Provost and Dean, Dr. Brouwer promotes and supports the graduate mission of the university, providing both academic and administrative leadership. She represents the interests and needs of graduate programs and graduate students and collaborates with faculty, deans and graduate coordinators to develop and support excellence in all aspects of graduate education including academic training and professional development. The expansion of graduate credentials, enrolment management, maintenance of high academic standards, and the establishment of policies and best practices that support graduate students academically and financially are part of her portfolio. She currently serves as President of the Canadian Association of Graduate Studies.

Dr. Brouwer joined Queen's in 1990 after completing her PhD in Neuroscience at the University of Toronto. She holds a B.Sc. in Kinesiology (University of Waterloo) and an M.Sc in Biomechanics (McGill University). She served as an Associate Dean in the School of Graduate Studies from 2005 to 2010 before moving into the role of Vice-Provost and Dean. She maintains an externally funded research program focused primarily on quantifying the biomechanical, neuromuscular and metabolic demands of mobility in healthy aging and stroke. She has supervised over 32 research master's and doctoral students and post-doctoral fellows in the area of disordered movement control and physical function.

Professor Hans-Joachim Bungartz

Hans-Joachim Bungartz is a full professor of informatics and mathematics at TUM, where he holds the *Scientific Computing* chair in the Informatics Department.

Dr. Bungartz earned degrees in mathematics and informatics and a PhD as well as his habilitation in informatics, all from TUM. He became associate professor of mathematics at Universität Augsburg, full professor of informatics at Universität Stuttgart, and returned to TUM in 2005. Since 2008, he has also been affiliated with the Department of Mechanical Engineering at University of Belgrade, Serbia. Since 2013, Dr. Bungartz has served as Dean of Informatics and as TUM Graduate Dean, heading TUM Graduate School and being in charge of doctoral education TUM-wide. In both functions, he is member of TUM's Extended Board of Management.

Dr. Bungartz has served or serves on several editorial boards, and he was a member of the scientific directorate of *Leibniz Institute for Informatics Schloss Dagstuhl*. He is involved in various national and international review and advisory board activities. In 2011, he was elected chairman of the *German National Research and Educational Network* (DFN). Furthermore, Dr. Bungartz is a board member of *Leibniz Supercomputing Centre*, one of three national supercomputing centres.

His research interests are where computational engineering, scientific computing, and supercomputing meet. He works on parallel numerical algorithms, hardware-aware numerics, high-dimensional problems, data analytics, and aspects of HPC software, with fields of application such as computational fluid dynamics. Most of his projects have been interdisciplinary ones. As an example, he coordinates DFG's Priority Program *Software for Exascale Computing*.

Professor Paul C. Burnett

Paul C. Burnett (DipT, BEdSt, MEdSt, DipAppPsych, PhD) is Dean of Research and Research Training at Queensland University of Technology (QUT). He rejoined QUT in

May 2009 as a Research Capacity Building Professor in the Division of Research and Commercialisation and was appointed to the Dean's position in March 2011. Prior to this he was at Charles Sturt University where he was Deputy-Vice-Chancellor (Research) for eight years. During that time he was responsible for the strategic direction, policy formulation, planning activities, and the effective management of research, research training, commercialisation, intellectual property and outside professional activity. He was the Presiding Officer of the University's Research Management Committee, the Board of Graduate Studies, and the Intellectual Property and Outside Professional Activities Committee and served on the University's Senior Executive, University Budget Committee, Academic Senate, Vice-Chancellor's Forum, University Course Planning Committee, and Promotions Committee. He served on the Executive of the National DVC/PVC (Research) Group, served as an AUQA Auditor for four years and chaired the audit of Notre Dame University and in 2007 was nominated by the Federal Minister to Chair the RQF Assessment Panel for Psychology and Psychiatry.

Dr. Karen Butler-Purry

Karen Butler-Purry is Associate Provost for Graduate and Professional Studies (APGPS) at Texas A&M University (TAMU), a position she has held since 2010. In addition, Butler-Purry is a professor in the department of electrical and computer engineering, having served at all faculty levels beginning with an initial appointment as visiting assistant professor of electrical engineering in 1994. Butler-Purry has vast experiences in graduate education as a faculty member, administrator, researcher and program leader. From 2001 to 2004, she served as Assistant Dean for Graduate Programs in the College of Engineering, and served as Associate Department Head in the Electrical and Computer Engineering Department from 2008 to 2010. Further, Butler-Purry has directed several fellowship and education projects promoting recruitment, retention and advancement of graduate students in STEM fields. Additionally she has served in many capacities on committees for the college, university, and professional societies. Dr. Butler-Purry developed a successful research program with funding from federal agencies such as NSF and ONR, and industry funding from electric utility companies. She has supervised and funded over 40 graduate and 65 undergraduate research students.

During her inaugural year as APGPS in response to economic reductions, Butler-Purry led a campus discussion and review of distribution policies regarding university graduate student support funds which resulted in a plan that prioritized providing sufficient support to best attract the brightest doctoral students while at the same time allowing individual colleges to better align the funds with their specific strategic priorities. Also under Butler-Purry's leadership, the TAMU Office of Graduate and Professional Studies (OGAPS) added a new university initiative to promote and support graduate student participation in professional development opportunities aligned closely with the university's novel Quality Enhancement Plan, *Aggies Commit to Learning for a Lifetime*.

Professor Chen Chuanfu

Chuanfu Chen received his MA Degree in Library Science and SJD in Intellectual Property from Wuhan University, China. He acted as the Dean of the School of Information Management with a joint appointment of Director of Research Center for Intellectual Property. He is now the Dean of the Graduate School of Wuhan University. He is member of the seventh discipline assessment groups of the academic degree commission of the state council.

He teaches Research Methodology, Legal and Intellectual Property Issues for Information Professionals. His research interests include copyright, library and information education, and online information assessment. He was major investigator of research projects funded by the Natural Science Foundation, Social Science Foundation, and EU-China Higher Education Cooperation Program.

He was a visiting Scholar at UC Berkeley (2006), Centre d'Etudes et de Recherche en Droit de l'Immateriel, Faculté Jean Monnet, Université de Paris Sud, France (1998), Law School of University of East Anglia, Norwich, England (1998), and Graduate School of Library and Information Science, University of Washington, Seattle, USA (1995-96). He acted as the member of IFLA Standing Committee on Library Education and Training. He was a member of iCaucus for iSchools.

He is the author of publications including "Impacts of government website information on social sciences and humanities in China: A citation analysis" (Chen, Chuanfu; Wang, Ping; Liu, Yaqi; et al.). *Government Information Quarterly*, 2013, 30(4): 450-463; "International Students' Perception of Library Services and Information Resources in Chinese Academic Libraries" (Nzivo, Charles N.; Chen Chuanfu.) *Journal of Academic Librarianship*, 2013, 39(2): 129-137; "Assessing the authority of free online scholarly information" (Chen, Chuanfu; Yu, Yuan; Tang, Qiong; et al.) *Scientometrics*, 2012, 90(2): 543-560; "The attitudes of LIS chairs toward the iSchools movement in China" (Chen, Chuanfu; Wang, Ping; Wu, Dan; et al.). *ASLIB Proceedings*, 2012, 64(6): 591-614.

Professor Shiyi Chen

Shiyi Chen graduated from Peking University and got his Ph.D. degree in 1987. From July 1999, Dr. Chen served as the Professor, Department of Mechanical Engineering, Johns Hopkins University (JHU). In 2005 he became the founding Dean of the College of Engineering of Peking University. In 2011 he was appointed as the Dean of the Graduate School and then, in 2013, the Vice President of Peking University. He is now the President of South University of Science and Technology of China.

Shiyi Chen is one of the first recipients of "National Thousand Talents Plan" and an elected member of the Chinese Academy of Sciences (2013). He has published more than 170 scientific papers, edited 3 books and has SCI citations in excess of 10,000 times. His research contributions include the invention and development of the lattice Boltzmann method (LBM) for computational fluid dynamics.

Dr. Niels Dam

Niels Dam is the Vice President for Product Management at ProQuest Information Solutions. He oversees the global expansion of the longstanding Dissertation and Thesis dissemination program (PQDT Global) with more than 3.8 million graduate works, of which 1.8 million are in full text. The program reached a milestone in 2014 for the first time, adding more than 100,000 new dissertations and theses. Moreover ProQuest Information Solutions is designated as an official offsite repository for Dissertations by the U.S. Library of Congress and has over the years worked with universities in the USA, Canada, United Kingdom and Australia on more than 150 retrospective scanning projects for Dissertations and Thesis through the Digital Archiving and Access Program. In addition to the Dissertation and Thesis program, Niels oversees a broad set of STM A&I and full text databases and leads the development of new digital services in video curation and text and data mining. Niels has a diverse experience from a number of strategic, sales and product management roles within

the STM industry, at first Elsevier and most recently at ProQuest Information Solutions. Niels has a background as a research scientist with a PhD from University of Aarhus, Denmark, in Chemical Physics and as a Post-Doctoral fellow at the Nanomaterial and Nanomanufacturing Research Center in Tampa, Florida, before obtaining an MBA with distinction from London Business School.

Professor Bernadette Franco

Bernadette Franco is Professor at the University of São Paulo (USP), Faculty of Pharmaceutical Sciences, Department of Food and Experimental Nutrition and Provost for Graduate Studies. Professor Franco graduated in Pharmacy and Biochemistry, with an MSc in Microbiology and Immunology and a PhD in Food Sciences. Her research, teaching and extension activities are related to food security, safety and quality and she serves as Coordinator of the Food Research Center (FoRC) at USP. Franco has published more than 150 peer-reviewed papers, books, and book chapters. She has supervised approximately 70 MSc and PhD students and post-docs. She is a Board member of the International Commission on Microbiological Specifications for Foods (ICMSF) and International Commission on Food Microbiology and Hygiene (ICFMH) and has served in CNPq, CAPES and FAPESP as coordinator of the Food Science and Technology area. Franco was President of the Brazilian Society of Microbiology for two terms and Editor-in-chief of the Brazilian Journal of Microbiology for fifteen years. She is a member of the editorial boards of several journals in food science.

Professor Xiaoqing (Maggie) Fu

Xiaoqing (Maggie) Fu became Acting Dean of Graduate School, University of Macau in October 2010. She graduated with a PhD in Finance from Cass Business School, City University, UK in 2004. Then she joined the University of Macau as an Assistant Professor of Finance. In 2010 she became Associate Professor of Finance. Prior to joining the University of Macau, she had been working in the banking industry for several years.

She currently serves as an editorial board member of three international peer-reviewed journals. In addition, she is the Director of the GARP (Global Association for Risk Professionals) University Chapter at the University of Macau as well as a member of GRE Asia Advisory Council. Her research interests are in banking and finance.

Professor Martin Gersch

Degree in Business Administration and Economics. 1991-1996: PhD thesis on the subject of "Cross-linked Relationships". 1997-1999: managerial functions in international business development with the Tengelmann Group (one of Europe's biggest retailers), including projects in Germany and Hungary. 2000-2006: Founder and head of the Competence-Center E-Commerce and habilitation on "Flexibility Traps". Received venia legendi for "Business Administration" and "Business Informatics". 2007: Full professor of Business Administration and head of Competence Center E-Commerce at the Freie Universität Berlin. 2008 (Senior) Researcher at the DFG Pfadkolleg Research Center on "Organizational Paths" at the Freie Universität Berlin (German Research Foundation DFG). 2011 Department of Information Systems at the Freie Universität Berlin; 2012 Visiting Professorial Fellow (UNSW, UoS, QUT, Australia). 2013 member of the steering group of the Focus Area "Disease in Human Aging" (Excellence Initiative). Board member and responsible for "Entrepreneurship

Education" at "(ENU) Entrepreneurial Network University" of the Freie Universität Berlin and the Charité – Universitätsmedizin Berlin. 2014 successful funding of a Junior Research Group "Health-IT and Business Model Innovation".

Current key research areas: Technology-driven change and transformation processes, information management, business process management, service engineering, e-business and e-commerce, business model analyses, entrepreneurship, e-health, management and economic theory, innovative teaching and learning concepts (blended learning/learning service engineering).

Dr. Noreen Golfman

Noreen Golfman is Provost and Vice-President (Academic) of Memorial University. She served as Dean of Graduate Studies at Memorial from June 2008 to September 2014, when she was appointed Provost and Vice-President (Academic) Pro Tempore. Under Dr. Golfman's leadership, the School of Graduate Studies experienced tremendous growth. In the last decade, the number of graduate students has doubled to a total of 3,565 graduate students.

Dr. Golfman is a professor of English and holds a PhD from the University of Western Ontario. She recently served two terms as president of the Canadian Federation of Social Sciences and Humanities, a national education advocacy group, and she is past president of both the Canadian Association of Graduate Studies and the Northeastern Association of Graduate Schools. She is currently a member of the advisory committee to the Canadian studies program at the Hebrew University of Jerusalem.

Active in Canadian cultural issues and experienced with the media, Dr. Golfman is the founding director and chair of the St. John's International Women's Film Festival, vice-chair of the Newfoundland and Labrador Film Development Corporation and chair of the board of the Friends of Canadian Broadcasting. Since 2011 she has co-chaired the board of directors of Business and the Arts NL, an organization that brokers relationships and funding between the arts and the corporate/private sectors.

In addition to publications in scholarly journals, Dr. Golfman writes on the arts and culture in popular venues, and she has worked as a commentator, reviewer and performer for CBC radio and television.

Professor Lucy Johnston

Lucy is Dean of Postgraduate Research and Professor of Psychology at the University of Canterbury, convenor of the NZ Deans and Directors of Graduate Studies (DDOGS) and chair of the Universities NZ Scholarships Committee.

Lucy graduated with a BA (Hons) from the University of Oxford and PhD from the University of Bristol, taught at the University of Cardiff before joining the Psychology Department at Canterbury in 1994. She was on the inaugural management group of the New Zealand Institute of Language, Brain and Behaviour. In 2011 she became Dean of Postgraduate Research.

As Dean Lucy has oversight for all policy and strategy for HDR candidates, skills training for HDR candidates and supervisors. She has been responsible for the introduction of Researchplus to provide doctoral graduates with both generic and disciplinary skills. She developed a course for new thesis supervisors and a series of workshops for experienced supervisors.

Lucy's research expertise, with over 80 international peer-reviewed publications, is in social perception and nonverbal communication. She received a University Teaching Award

in 2008 and in 2004 held a Distinguished Visiting Professor position at the University of Connecticut.

Lucy also has an MSc in Sport and Exercise Psychology. She was awarded Oxford Blues and full colours at the University of Bristol for basketball and played on the British Universities. She rowed for her Oxford College and City of Bristol and played soccer for the University of Bristol. She recently retired from 10 seasons completing in road cycling and triathlons.

Dr. Mohan Kankanhalli

Mohan Kankanhalli is the Vice Provost (Graduate Education) at the National University of Singapore (NUS). He is also a Professor of Computer Science at the NUS School of Computing. He was the Associate Provost (Graduate Education) during 2011-2013. Earlier, he was the Vice-Dean for Academic Affairs & Graduate Studies at the NUS School of Computing during 2008-2010 and Vice-Dean for Research during 2001-2007. Mohan obtained his BTech (Electrical Engineering) from the Indian Institute of Technology, Kharagpur, in 1986 and his MS and PhD (Computer and Systems Engineering) from the Rensselaer Polytechnic Institute in 1988 and 1990, respectively.

Mohan's research interests are in Multimedia Systems, Digital Video Processing and Multimedia Security. He has made many contributions in the area of multimedia content processing – image & video retrieval, data fusion, visual saliency, computational media aesthetics as well as in multimedia security – multi-camera surveillance, content authentication & privacy. He is actively involved in the Multimedia Systems research community and was the Director of Conferences for ACM SIG Multimedia during 2009-2013. He is on the editorial boards of several journals including the ACM Transactions on Multimedia Computing, Communications, and Applications, Springer Multimedia Systems Journal, Pattern Recognition Journal, and Springer Journal on Big Data.

Mohan is a Fellow of IEEE.

Dr. Julia Kent

Julia Kent is Assistant Vice President, Communications, Advancement and Best Practices at the Council of Graduate Schools (CGS). As Director of Communications, she leads CGS's efforts to engage press and media outlets in issues important to graduate education. In the advancement area, Kent promotes the exchange of expertise between CGS and its corporate partners and supports the development of collaborations that benefit the graduate community.

Julia's work in communications and advancement is informed by seven years of research experience in the Best Practice division at CGS. She has conducted research on a broad range of topics in graduate education, including PhD career pathways; graduate admissions processes; international collaborations; quality and accountability in graduate education; research ethics and integrity; and the preparation of future faculty. Currently she serves as Co-Principal Investigator for a project co-funded by the Alfred P. Sloan and Andrew W. Mellon Foundations, "Understanding PhD Career Pathways for Program Improvement," and directs a Hobsons-supported project on holistic review in graduate admissions processes. She has overseen CGS's Global Summit program since 2009. Julia holds a PhD in British literature from Johns Hopkins University and a maîtrise de lettres modernes from the Université de Paris VII.

Dr. Barbara A. Knuth

Barbara A. Knuth was appointed Vice Provost and Dean of the Graduate School at Cornell University in 2010. She was promoted to Senior Vice Provost in October 2014. She oversees more than 90 graduate fields that include approximately 1,800 graduate faculty across ten colleges and schools, 5,100 graduate and professional students, and 600 post-docs. She served previously as Senior Associate Dean of the College of Agriculture and Life Sciences at Cornell (2007-2010), and Chair of the Department of Natural Resources (2002-2007). Under Dean Knuth's leadership, the Graduate School offers strong professional development programs focusing on core competencies of leadership, communication, personal development, teaching, and career development, emphasizing transferrable skills relevant to career paths in academia, business, government, and non-profit sectors. Flagship programs focus on writing skills and the creation of writing communities on campus, faculty-led diversity recruitment efforts, future faculty development, student financial literacy, career development, and data visualization for monitoring student milestones and evaluating and improving student experiences. The Graduate School partners with the Cornell Center for Teaching Excellence, Career Services, and the Knight Institute for Writing in the Disciplines, and is part of the multi-institution Center for the Integration of Research, Teaching, and Learning (CIRTL). Under her leadership, the Graduate School restructured its staff, launched a new web site, improved its information technology, increased its media presence, and expanded its assessment efforts. Knuth served on the Ocean Studies Board of the U.S. National Academies and is a past president of the American Fisheries Society. She was appointed Chair of the Council of Graduate Schools Board of Directors at their December 2014 meeting.

Dr. Magnús Lyngdal Magnússon

Magnús Lyngdal Magnússon joined University of Iceland's Graduate School in 2013, having earlier been Deputy Director of RANNIS (Icelandic Centre for Research), responsible for competitive funding for basic research, and Senior Advisor at the Ministry for Education, Science and Research. Magnús took part in setting up and running a Quality Enhancement Program for Icelandic Higher Education in 2009 and has appeared on numerous panels and committees on higher education in Iceland. Magnús trained in medieval church history and has published on 14th century canon law in Iceland and is an avid opera fan. Magnús is currently senior advisor to the Rector/President of University of Iceland.

Dr. Liviu Matei

Liviu Matei is Central European University's Provost and Pro-Rector, and Professor at the Department of Public Policy. He served as Senior Vice President and Chief Operating Officer from 2008 to 2014, and as Academic Secretary of CEU from 1999 to 2008.

He studied philosophy and psychology at Babeş-Bolyai University Cluj, and Sociology at Bucharest University, Romania. He received his PhD from the latter. He benefited from fellowships at the Institut Supérieur de Formation Sociale et de Communication, Bruxelles, The New School for Social Research, Université Paris X Nanterre, Université de Savoie, and the Salzburg Seminar. He taught at universities in Romania, Hungary and the U.S., and consulted extensively in the area of higher education policy for UNESCO, OSCE, the Council of Europe, the European Commission, and other

international organizations and national governments in Europe and Asia. He is a member of the Board of Trustees of the American University of Central Asia and serves on the editorial board of the Journal of the European Higher Education Area. Prior to joining CEU, he worked as a Director General for International Relations at the Romanian Ministry of Education, served as co-chair of the Working Group for Higher Education of the Stability Pact for South East Europe and on the steering committee of the UNESCO-EU Project on Management of Higher Education in South-East Europe.

Dr. Steve Matson

Steve Matson was appointed Dean of The Graduate School at The University of North Carolina at Chapel Hill in 2008. Prior to his appointment, he served as Assistant Dean for Academic Advising and Chair of the Department of Biology.

As dean, Matson has engaged the graduate community to envision and support graduate programs of the highest quality and to ensure that graduate students are provided with academic, professional and financial resources for degree completion and career success. Matson has raised nearly $6 million for graduate education to support implementation of the Summer Research Fellowship Program; creation of five-year Chancellor's Fellowships; increased professional development activities; the Bland Fellowship Professional Pathways Program and endowed professorships supporting faculty mentors and their students. He is leading an effort to enhance program review through the use of faculty productivity data and student exit surveys.

Matson is engaged in advancing opinion leaders' understanding of the importance of graduate education through his work with colleagues across North Carolina to implement Graduate Education Day at the Capitol. Matson served as chair of the CGS Government Relations Advisory Committee, serves on the CGS Board of Directors, is a member of the Board of Directors of the Graduate Record Exam and the TOEFL Board of Directors. Matson holds a bachelor's degree in mathematics from Colgate University and a doctorate in biochemistry from the University of Rochester. His research focuses on DNA repair as it relates to cancer and genome stability, and the transfer of antibiotic resistance in bacteria.

Professor Mary McNamara

As Head of the Graduate Research School at Dublin Institute of Technology (DIT), Professor Mary McNamara manages the educational programmes and events to enhance the graduate student learning experience and promotes the achievements of the postgraduate research community. DIT is the largest provider of higher education in Ireland with over 22,000 students and has been an integral part of Irish education for over 125 years. With autonomous degree-awarding authority, DIT combines the excellence of a traditional university with professional career-oriented learning and prepares graduates for productive leadership roles in the public and private sectors. The Graduate Research School at DIT defines the community of graduate students and their supervisors and provides a forum of information and consultation. It is designed as the over-arching entity for the quality assurance of graduate research and all graduate students are, in addition to being registered in their own faculty, members of the Graduate Research School. Prof McNamara is the guarantor of graduate research quality assurance and also works with the schools and colleges at DIT in the development of PhD programmes. She helps to develop and implement strategies, policies and procedures for growing research across DIT.

Prof McNamara graduated with a honours degree from the Royal Society of Chemistry and a PhD in Physical Inorganic Chemistry from University College Dublin. She is also a fellow of the Royal Society of Chemistry and an active researcher in the development of novel drug delivery systems.

Professor Shireen Motala

Professor Shireen Motala held the position of the Director of the Education Policy Unit, University of the Witwatersrand, from 1999 to February 2010. Her academic qualifications are BA(UDW), B Social Science Honours (UCT), MA (University of Warwick), PGCE (University of London) and PhD (Wits). In March 2010, she was appointed as Director of the Postgraduate Research Centre: Research and Innovation at the University of Johannesburg. Her responsibilities include leading the university-wide strategy for improving enrolment at a postgraduate level, providing research support for postgraduate students and ensuring that throughput improves in the institution. She was Chairperson of the Education Policy Consortium which brings together policy research entities nationally from 2006 to 2010. She was also the Chairperson of the UNESCO South African Commission from 2002 to 2006. In 2010 she was appointed by the Minister of Higher Education and Training to serve on the Council of Higher Education (CHE) and in 2013 to serve on the Ministerial Task Team to review the national Senior Certificate examination. She was the first inaugural president of the South African Research Association (SAERA) from 2013 to 2014. She has worked extensively in research and has provided leadership for regional and international partnerships, which have led to collaborations with universities across Africa and with northern partners. Her research record is substantial and includes books, editorship of local and international journals and chapters in books. Her research interests and expertise have been in the areas of education financing and school reform, access and equity and education quality.

Professor Buyinza Mukadasi

Buyinza Mukadasi, is a Professor of Forestry Resource Economics and Director of Research and Graduate Training at Makerere University, Uganda. He has grown though the university ranks from a Tutorial Assistant (1993–1998) at Gajah Mada University, Indonesia to becoming a Lecturer (1999); Senior Lecturer (2004); Assoc. Prof (2007) and Full Professor (Oct. 2010) at Makerere University. He has successfully supervised over 50 MSc. and 3 PhD students in the fields of forestry, environment and natural resources. On his record is 20 year experience of university research and teaching in the field of Forestry and Natural Resource Economics.

He has been principal investigator (PI) of five research projects. He has tremendous capacity for resource mobilization; currently he is managing, as PI, three projects worth about USD two million including the Intro-ACP Africa (Mwalimu Nyerere) and the Caribbean & Pacific (Euro two million). Effective 1st October 2010, he was appointed for a 4-year term as Deputy Director (Graduate Studies), Directorate of Research and Graduate Training, Makerere University. As a member of the Teaching & Learning Committee, he participated in the restructuring process of the Makerere University and in handling of the transition process. He has spearheaded several national and international research and development projects in the fields of forestry, environment and natural resource management. He is among the Ugandan Team Leaders for the "Protected Area Management, Climate Change and Poverty in Africa" Project, 2007–2011. He is also involved in the ongoing World Bank

Research and Development Project focusing on strengthening Forests, Poverty and Climate Change linkages in Africa, Magadascar, and Asia. He has served as the Chairperson, Board of Directors of the National Forestry Authority since 2010. His research interests include forest resources economics, ecosystem health and community livelihoods. He has published widely especially in the field of forest management approaches, economics of biodiversity conservation, and environmental impact assessment studies. Has a wealth of consultancy experience with special focus on Environmental Incomes and Rural Income (Funded by Cochran Fellowship Programme, USA); natural resource management for rural livelihood enhancement; and decentralized forest service delivery for rural development in the Uganda (FAO Funding).

Professor Y. Narahari

Y. Narahari is currently a Professor at the Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India. He is also Chair, Division of Electrical Sciences. The focus of his research in the last decade has been to explore problems at the interface of computer science and game theory. In particular, he is interested in applying mechanism design to problems in auctions, market design, online education, Internet advertising, social network analysis, and crowdsourcing. He has just completed a textbook entitled "Game Theory and Mechanism Design" brought out by the IISc Press and the World Scientific Publishing Company. He is a fellow of the IEEE and a J.C. Bose Fellow of the Department of Science and Technology, Government of India. More details at: http://lcm.csa.iisc.ernet.in/hari/

Dr. Suzanne T. Ortega

Suzanne Ortega became the sixth President of the Council of Graduate Schools on July 1, 2014. Prior to assuming her current position, she served as the University of North Carolina Senior Vice President for Academic Affairs (2011-14). Previous appointments include Executive Vice President and Provost at the University of New Mexico, and Vice Provost and Graduate Dean at the University of Washington and at the University of Missouri. Dr. Ortega's master's and doctoral degrees in sociology were completed at Vanderbilt University.

Dr. Ortega serves or has served on a number of professional association boards and committees, including the executive boards of the Council of Graduate Schools, the Graduate Record Exam (GRE), the National Academies of Science Committee on the Assessment of the Research Doctorate, the National Science Foundation's Human Resources Expert Panel, the North Carolina E-learning Commission, the North Carolina Public School Forum, the UNC TV Foundation, and the UNC Press Board of Governors.

Dr. David Payne

David Payne is the Vice President and COO of the Global Education Division at ETS. David heads the GRE® and TOEFL® programs, as well as higher education assessments such as the ETS® Major Field Tests, ETS® Proficiency Profile and SuccessNavigator®. He oversaw the launch of the GRE® revised General Test on August 1, 2011. With new questions, a new score scale and a new test-taker friendly design, the launch marked the biggest change to the assessment in 60 years. He also led efforts to create the SuccessNavigator assessment (which is designed to help colleges reach at-risk students and improve first-year retention rates) and the HEIghten™ assessment (which provides six student learning outcome measures).

Payne works closely with the GRE and TOEFL Boards, undergraduate and graduate education organizations and colleges, universities and public education systems. He also helps to identify assessment needs in the higher education and professional markets—both domestic and international—and to develop external relationships.

Prior to joining ETS in 2003, Payne was a full professor of psychology and Vice Provost and Dean of the Graduate School at SUNY Binghamton.

Payne holds bachelor's and master's degrees in experimental psychology from SUNY Cortland and a Ph.D. in cognitive psychology from Purdue University. He is a Fellow of the American Psychological Association and the Association for Psychological Science, and has received numerous grants, fellowships and awards. He has published five books, nine book chapters and more than 100 articles, technical reports and papers.

Professor Laura Poole-Warren

Laura Poole-Warren is the Pro-Vice Chancellor (Research Training) & Dean of Graduate Research at UNSW Australia. In these roles she has responsibility for graduate research education, researcher development and early career and postgraduate researchers. She is also a Professor of Biomedical Engineering and leads a research group in the field of biomaterials and tissue engineering.

Up to 2013, Laura was a decade-long member of the Australian Commonwealth statutory Advisory Committee on Medical Devices. In this role she advised the Australian Therapeutic Goods Administration on safety of medical devices. Laura also spent 2 years as a preclinical scientist in the biomedical industry from 1999 to 2001 working on development of implantable devices including wound dressings and embolic agents for cancer treatment.

Laura is current Convenor of the Australian Council for Deans and Directors of Graduate Research, is Vice-Chair of the International Federation of Medical and Biological Engineering Women in BME Committee and is immediate past Chair of the Go8 Deans of Graduate Research. Her other roles include being an Associate Editor of the journal Biomaterials and an advisory board member of the ARC Centre of Excellence for Climate Systems Science and the Bionic Vision Australia Consortium.

Professor Nirmala Rao

Nirmala Rao is Dean, Graduate School, The University of Hong Kong. She is also Serena H C Yang Professor in Early Childhood Development and Education and Professor, Faculty of Education.

A Developmental and Chartered (Educational) Psychologist by training, she has been recognised internationally for her research on early childhood development and education in Asian cultural contexts. Professor Rao has published widely on early childhood development and education, child development and educational policy, and educational psychology. She has been the recipient of numerous research grants and has participated in international meetings as an expert/specialist and undertaken consultancies for UNICEF, UNESCO, and the World Bank. Professor Rao serves on the editorial board for several journals, is the Associate Editor of Child Development and a member of the Steering Committee for the upcoming Lancet Series on Early Child Development. Professor Rao is also actively involved in professional organisations that aim to promote the well-being of children through research and advocacy efforts.

During her tenure with The University of Hong Kong, she has been Deputy Head of the Department of Education (1999-2002); Associate Dean, Faculty of Education (2002-

2005); Acting Dean, Faculty of Education (2006-2007); Acting Chairperson of the Board of the Faculty of Education (2008-2009); Associate Dean of the Graduate School (2010 – 2013) and has been Chairperson of the Board of Graduate Studies since 2010.

Dr. Mark J. T. Smith

Mark J. T. Smith received the B.S. degree from MIT and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, all in electrical engineering. He joined the electrical and computer engineering (ECE) faculty at Georgia Tech in 1984, where he remained for the next 18 years. While working primarily on the Atlanta campus, he spent several terms in 1991-93 on the Institute's European campus in Metz, France. Five years later he served a four-year term as Executive Assistant to the President of Georgia Tech. In January, 2003, he joined the faculty at Purdue University as head of the ECE School. From 2005-2009, Smith served as a member of the Board of Directors of the national ECE Department Heads Association and was president in 2007.

Smith was appointed Dean of the Purdue University Graduate School in 2009 and, shortly thereafter, became actively involved with the Council of Graduate Schools (CGS). Presently, he is chair elect of the CGS Board of Directors and a member of the GRE Board of Directors.

Dean Smith's scholarly interests are in the area of digital signal processing (DSP). He is a Fellow of the IEEE, and is a former IEEE Distinguished Lecturer. He has authored many technical papers, six international standards publications, three textbooks, and two edited books, the most recent of which is GPS for Graduate School—Students Share Their Stories. In addition to his professional activities, Dr. Smith's past includes Olympic competition and U.S. national gold medals in the sport of fencing.

Professor Nicky Solomon

As Dean of the Graduate Research School, Professor Solomon has overall leadership and management of research degree students across the university. Her role focuses on improving the quality of doctoral education so that research graduates make significant contributions to knowledge in their disciplinary areas, but also so that graduates have well developed research skills to help them with their career choices.

Professor Solomon's research spans a number of areas, including workplace learning, as well as on changing professional and pedagogical practices through the ongoing influence of digital developments. Her current research project focuses on the changing practices of health professionals in primary health care settings.

She has published books, chapters in edited books, journal articles and refereed conference papers.

Professor Zaidatun Tasir

Zaidatun Tasir is a Professor of Educational Technology at the Department of Educational Science, Maths and Creative Multimedia, Faculty of Education, Universiti Teknologi Malaysia. She is also the Dean of School of Graduate Studies, UTM and the research group leader of Creative and Innovative Technology in Education (CITE) under Smart Digital Community research alliance. Prior to that, she was the Deputy Dean (Social Science) of School of Graduate Studies (2009 - 2010), Deputy Dean (Postgraduate Studies & Research) (2008 - 2009), Head of Department of Postgraduate Studies (2007 - 2008), and Information

Technology Manager (2004 - 2007), Faculty of Education. She involved as a panel in preparing documentation for 11th Malaysia Plan 2016-2020 - Thrust 3: Harnessing Talent Effectiveness of Soft Programmes under Ministry of Education. She obtained her first degree, B. Sc. Comp. with Edu. (Math) (Hons.) from UTM (1995), M. Ed. (Educational Media Computers) from Arizona State University, USA (1998), and Ph.D (Educational Technology) from Universiti Teknologi Malaysia (2002). Her research interests and expertise include Design and Development of Computer and web-based Instructions, Multiple Intelligence through computer-based instruction, Problem-based learning through technology, Social Networking Tools in Education, and Online Social Learning Model.

Professor Agma Traina

Agma J. Machado Traina received the B.Sc. and the M.Sc. degrees in Computer Science from the Mathematics and Computer Science Institute at the University of São Paulo, in 1983 and 1987 respectively, and the Ph.D. in Computational Physics from the Physics Institute of São Carlos, also at University of São Paulo, Brazil, in 1991. She spent two years as a Visiting Scholar at Carnegie Mellon University, USA from 1998 to 2000 working on multimedia databases, metric access methods and selectivity estimation of similarity queries.

Traina is a full professor with the Computer Science Department of the University of São Paulo at São Carlos, Brazil. Her research interests include image and complex databases, image mining, indexing and retrieval by content of complex data, data analytics, information visualization, similarity queries in medicine and climatology. She has supervised 40 graduate students in these areas so far. She is a member of the IEEE Computer Society, ACM, SIGKDD, and the Brazilian Computer Society.

Dr. Kevin Vessey

J. Kevin Vessey is the Dean of Graduate Studies and Research, and a Professor of Biology at Saint Mary's University, Halifax, Canada. Dr. Vessey received his BSc and MSc in Biology from Dalhousie University and his PhD in Biology from Queen's University, Kingston. Aside from his 16 years as a Professor of Plant Science at the University of Manitoba, he has also been a Researcher Associate at North Carolina State University, and a Visiting Scientist at the Institut National de la Recherche Agronomique, France. Dr. Vessey's teaching and research area is plant physiology, particularly the functional interactions between crop plants and beneficial micro-organisms, and the optimizing of crops as biodiesel and bioethanol feedstocks. He has published over 80 peer-reviewed scholarly articles and book chapters and has co-edited one book. He has directly supervised over 50 undergraduate/graduate students, postdoctoral fellows, and research associates. He has been awarded several regional and national research awards including the C.D. Nelson Award by the Canadian Society of Plant Physiologists for outstanding contributions by a young scientist to plant physiology in Canada. Positions in which Dr. Vessey has recently served, or is currently serving, include Grant Selection Committees of the Natural Sciences and Engineering Research Council; Advisory Committees for the Maritime Provinces Higher Education Commission, the Nova Scotia Health Research Foundation, the Nova Scotia Research and Innovation Trust, Petroleum Research Atlantic Canada, and TRIUMF (UBC); and the boards of directors for the Canadian Association of Graduate Studies, the Offshore Energy Research Association, Plant Inoculants Canada, and Springboard Atlantic Inc.

Professor Nagi Wakim

Prof. Wakim has over two decades of experience in teaching, research and development, and administration in government and higher education, which include appointments as a department chair, dean, and vice president at several universities in the USA, Lebanon and the United Arab Emirates—where he is now the founding Dean of the College of Graduate Studies. He earned a Bachelor of Science in Computer and Engineering Sciences from the City University of New York and a Master of Science and a PhD in Computer Science from Polytechnic University in New York (now the Polytechnic School of Engineering of New York University).

He began his career at NASA's Goddard Space Flight Center where he worked on research projects in distributed, heterogeneous data systems (working with "big" science data) and information ecology—exploring ways to incorporate Artificial Intelligence in space flight operation and control. He was a tenured, full professor at Bowie State of the University System of Maryland where he worked for 15 years while continuing his collaborative projects with the federal government. He participated in numerous panels and program reviews on educational reform and capacity development especially in the STEM (Science, Technology, Engineering and Mathematics) disciplines for NASA and the National Science Foundation.

Prof. Wakim has a fond interest in human capital development and the application of information technology to enhance institutional efficiency and effectiveness. He has published numerous papers and articles and has presented his work to national and international audiences. He is a lifetime member of the Association for the Advancement of Artificial Intelligence (AAAI) since 1995.

Professor Wang Yaguang

Wang Yaguang is Chair Professor in Mathematics and Deputy Dean of the Graduate School at Shanghai Jiao Tong University. He obtained his PhD degree in the Department of Mathematics at Fudan University, Shanghai in July 1992. Since then he has been working in the Department of Mathematics at Shanghai Jiao Tong University, where he became a full professor in 1998. He was the Lise-Meitner Postdoc at Innsbruck University of Austria from October 1995 to December 1996. As guest professor, he has visited more than 20 universities in the USA, France, Germany, Austria, Switzerland, Japan, Korea and Hong Kong, and was a guest professor at Northwestern University, USA from September 2008 to April 2009. His research mainly focuses on analysis of partial differential equations and applications. From October 2009 to September 2014, he has served as the deputy chair in the Department of Mathematics, and in June 2014 he was appointed as the deputy dean of the Graduate School at SJTU.

Ms. Xiaoyue Wang

Ms. Xiaoyue Wang graduated from Peking University with a BA in Chinese Literature in 1987 and got her MA in Modern Chinese in 1995. She worked in the Graduate School of Peking University from July 1987. She served as Director of the Provost Office at Peking University from December 2004 to September 2010. She is now the Director of the Secretariat of Association of Chinese Graduate Schools (ACGS).

Dr. James Wimbush

Dr. James C. Wimbush is the Vice President for Diversity, Equity, and Multicultural Affairs, Dean of the University Graduate School, and Professor of Business Administration at Indiana University.

As Vice President for Diversity, Equity, and Multicultural Affairs, his efforts focus on three areas of priority: recruitment and retention of a diverse faculty, staff, and student population including timely completion for students; campus climate; and, advocacy and outreach.

Dr. Wimbush serves as dean of The University Graduate School, where he oversees academic masters and doctoral programs. Wimbush serves nationally as the President of the AAU's Association of Graduate School's Executive Committee, as past chair of the Board of Directors of the Council of Graduate Schools, former chair of the Board of Directors of the Graduate Record Exam, and was a member of Educational Testing Service's and the Council of Graduate Schools' joint Commission on the Future of Graduate Education.

Wimbush has received multiple awards for his teaching of management and leadership, and formerly chaired various units in the Indiana University Kelley School of Business, including the Department of Management and Entrepreneurship, Doctoral Programs in Business, and the Full-time Residential MBA Program.

An acknowledged national authority, he has published numerous articles on business ethics as they relate to human resources practices. Wimbush earned a doctorate in management and a master's degree in human resources management and industrial and labor relations from Virginia Polytechnic Institute and State University (Virginia Tech) in Blacksburg, Virginia.

Dr. Paula Wood-Adams

Paula Wood-Adams, a Professor of Mechanical Engineering, was appointed Dean of Graduate Studies of Concordia University in September 2013. While widely recognized for her academic achievements, she is also known as a strong leader, supervising over 25 graduate students and postdoctoral researchers. As an administrator, she has provided strategic and operational direction to various functions of graduate studies at Concordia, focusing her attention mainly on new curriculum, professional development and program support, and recruiting new, promising, graduate students.

She has been an active member of provincial and federal funding agency peer-review committees including the National Sciences and Engineering Council of Canada (NSERC) and the Fonds de recherche du Québec (FRQ). She has also contributed numerous articles to scientific journals on her research focus: viscoelasticity, polymer science and rheology, and has received over $2.5M in grants to support her research.

Professor Wood-Adams joined Concordia in 2001 as an assistant professor in Mechanical Engineering and in that same year received the NSERC University Faculty Award which was renewed in 2004. She has also received a Canada Foundation for Innovation New Opportunities infrastructure grant which allowed her to set up the Laboratory for the Physics of Advanced Materials. In 2006, she became the Graduate Program Director of Mechanical and Industrial Engineering, was awarded a Concordia University Research Chair, and became Visiting Professor of Chemical Engineering at Kasetsart University, Bangkok, Thailand.

She obtained her BSc in Chemical Engineering from the University of Alberta and her MEng and PhD in Chemical Engineering from McGill University.

Professor Kate Wright

Prof. Kate Wright has been actively engaged in research for more than 25 years. She obtained an Honours Degree in Geology and then a PhD in Mineral Physics from University College London, UK. After holding a series of high profile research fellowships in Manchester and London, Kate moved to Australia and Curtin University in 2004. She was appointed Professor of Mineral Chemistry in 2006, becoming Dean of Research for the Faculty of Science and Engineering in 2010. Kate was appointed to the role of Associate Deputy Vice-Chancellor Research Training in 2011 where she is responsible for developing strategy and policy for research training and ensuring quality in the University's Higher Degree by Research programs.

Kate's research interests centre around the study of microscopic defects in crystalline materials and their influence on macroscopic behavior. She still finds time to do research, but only just.