upcids
Data Science for Public Policy

Proceedings of the Workshop on Closing the Digital Skills Gap for the Philippines

May 27, 2019 • 8:30 AM–5:30 PM
Institute of Environmental Science and Meteorology
University of the Philippines Diliman

upcids
Data Science for Public Policy

Proceedings of the Workshop on Closing the Digital Skills Gap for the Philippines

May 27, 2019 • 8:30 AM–5:30 PM
Institute of Environmental Science and Meteorology
University of the Philippines Diliman

PROCEEDINGS OF THE
WORKSHOP ON CLOSING THE DIGITAL SKILLS GAP FOR THE PHILIPPINES
is published annually by the
PROGRAM ON DATA SCIENCE FOR PUBLIC POLICY of the
UNIVERSITY OF THE PHILIPPINES
CENTER FOR INTEGRATIVE AND DEVELOPMENT STUDIES
Lower Ground Floor, Ang Bahay ng Alumni
Magsaysay Avenue, University of the Philippines
Diliman, Quezon City 1101
Telephone: 8981-8500 / 8426-0955 (loc. 4266 to 68)
Email: datascience.cids@up.edu.ph / cidspublications@up.edu.ph
Website: cids.up.edu.ph

ISSN 2719-0560 (print)
ISSN 2719-0579 (online)

Contents

Introduction

The **Workshop on Closing the Digital Skills Gap for the Philippines** was held last May 27, 2019 at the Institute of Environmental Science and Meteorology, University of the Philippines (UP) Diliman. The workshop was spearheaded by the UP Office of the Vice President for Academic Affairs (OVPAA), together with the Program on Data Science for Public Policy (DSPP) of the UP Center for Integrative and Development Studies (UP CIDS) as its organizer and secretariat. The workshop had the following objectives:

(1) To begin unifying efforts to close the digital skills divide by discussing curriculum, resources, plans, policy interventions, and recommendations for government institutions, particularly for the Commission on Higher Education (CHED), that may include the successful application of early education strategies; individually adaptive and low-cost artificial intelligence-based math learning platforms; workplace experience integration in high school and college science, technology, engineering, and math (STEM) courses; and lifelong learning instruction;

(2) To discuss the next steps, areas for collaboration, and the possibility of developing a loose network of academic institutions engaged in data science/analytics; and

(3) To contribute the Workshop output on academe's initiatives as input to the Analytics Association of the Philippines (AAP) panel/presentation at the Asia-Pacific Economic Cooperation (APEC) Project Data Analytics and Raising Employment (DARE) Forum to be held in July 2019.

Programme

MORNING SESSION

8:00 AM – 8:30 AM	**Registration**
8:30 AM – 8:40 AM	**Opening Remarks** **Dr. Maria Cynthia Rose B. Bautista** Vice President for Academic Affairs, University of the Philippines
8:40 AM – 9:00 AM	**Keynote Presentation on Analytics Framework** **Mr. Sherwin Pelayo** Analytics Association of the Philippines
9:00 AM – 9:20 AM	**Keynote Presentation on 21st Century Skills** **Dr. Patricia B. Arinto** Former Assistant Vice President for Academic Affairs, University of the Philippines
9:20 AM – 9:40 AM	*Break (optional)*
9:40 AM – 10:30 AM	**Industry Panel Discussion on Job Roles** *Discussion Themes* • What analytics skills are needed? • What do they need from the academe? *Resource Speakers* • **Mr. Francis del Val** CEO and President, Cobena Group • **Ms. Michelle Alarcon** President and Managing Director, Z-Lift Solutions • **Mr. JP Palpallatoc** Managing Director, Digital, Data and Analytics, Data Business Operations, Accenture Philippines *Moderator* **Mr. Sherwin Pelayo** Analytics Association of the Philippines
10:30 AM – 12:00 NN	**Sharing of Data Science Curricula** *BS Data Science* **Dr. Jaime DL. Caro** Professor, Department of Computer Science, University of the Philippines Diliman

Masters in Data Science

- **Dr. Brenda A. Quismorio**
  Program Director, Master in Applied Business Analytics,
  University of Asia and the Pacific
- **Dr. John Paul C. Vergara**
  Professor, Information Systems and Computer Science,
  Ateneo de Manila University
- **Dr. Christopher P. Monterola**
  Executive Managing Director,
  Analytics, Computing, and Complex Systems Laboratory,
  Asian Institute of Management
- **Dr. Joselito C. Magadia**
  Professor and Director for Graduate Studies,
  School of Statistics, University of the Philippines Diliman

Ph.D. Data Science

- **Dr. Johnrob Y. Bantang**
  Associate Professor, National Institute of Physics,
  University of the Philippines Diliman
- **Dr. Prospero C. Naval, Jr.**
  Professor, Department of Computer Science,
  University of the Philippines Diliman

Open Forum

12:00 NN – 1:00 PM Lunch

AFTERNOON SESSION

1:00 PM – 1:15 PM **Presentation of Ongoing Research**
Dr. Brenda A. Quismorio
Program Director, Master in Applied Business Analytics,
University of Asia and the Pacific

1:15 PM – 1:30 PM **Briefing on Breakout Sessions**
Dr. Fidel R. Nemenzo
Chancellor, University of the Philippines Diliman and
Former Convenor, UP CIDS Program on Data Science for
Public Policy

1:30 PM – 2:30 PM **Breakout Sessions**

- **Data Scientist**
  Moderator: Dr. Johnrob Y. Bantang
- **Data Engineer**
  Moderator: Mr. Sherwin Pelayo

	• **Analytics Manager** Moderator: Dr. Brenda A. Quismorio
3:30 PM – 4:30 PM	**Plenary Session** **Dr. Eugene Rex Jalao** Associate Professor, Department of Industrial Engineering and Operations Research, University of the Philippines Diliman
4:30 PM – 5:30 PM	**Next Steps**
5:30 PM	**Closing Remarks** **Dr. Fidel R. Nemenzo** Chancellor, University of the Philippines Diliman and Former Convenor, UP CIDS Program on Data Science for Public Policy

Master of Ceremonies
Dr. Karl Robert L. Jandoc
Fellow, UP CIDS Program on Data Science for Public Policy

Opening Remarks

MARIA CYNTHIA ROSE B. BAUTISTA, Ph.D.
Vice President for Academic Affairs,
University of the Philippines System

In her opening remarks, UP Vice President for Academic Affairs Dr. Maria Cynthia Rose B. Bautista stressed that the basis of the workshop was to consolidate the efforts of various institutions—whether in the industry or in the academe—in developing data science as a field in the Philippines.

She also elucidated that the University of the Philippines (UP) is expected to be at the forefront in the fields of omics and data science. UP is also a member of the Asia Pacific Rim Universities (APRU), which is the equivalent of the Asia Pacific Economic Cooperation (APEC) for research universities. The APRU believes that UP can convene other higher education institutions (HEIs) in the Philippines in the area of data analytics and data science. This workshop aims to provide important inputs for the presentation of UP in the next APEC Meeting in Singapore in July 2019.

KEYNOTE PRESENTATION

The Analytics Framework: Building the Analytics Ecosystem of the Philippines

Mr. SHERWIN PELAYO
Analytics Association of the Philippines

Mr. Pelayo discussed the guiding framework in developing data science as a field in the Philippines. He began with a brief history of government

efforts in putting forth a data-driven country. Especially important for the education sector is the issuance of CHED Memorandum Orders (CMOs) 11 and 12, s. 2013, which set the guidelines for a specialization track on Business Analytics for Business Administration and Information Technology programs in higher education. This continued up to the briefing of industry partners in 2019.

Mr. Pelayo then talked about the Data Science framework in relation to the development of an ecosystem that will make the Philippines a data-driven country that is globally competitive in analytics and a leading source of analytics talent for the benefit of society. The analytics ecosystem in the Philippines was described as a link between skilled supply and demand among analytics training providers, analytics service providers, higher education institutions, and businesses and organizations.

He also defined important terminologies and explained various roles within this Data Science framework. *Analytics* progresses data along the value chain as it transforms data into information, information into insight, and insight into imperatives (or actionable insights). This intends to deliver the right decision support to the right people at the right time for the greater good. A *data steward* refers to a person who develops, enforces, and maintains an organization's data governance process to ensure that data assets provide the organization with high-quality data. Meanwhile, a *data engineer* is a person who designs, constructs, tests, and maintains data infrastructures, including applications that extract, transform, and load data from transactional systems to centralized data repositories.

A *functional analyst* utilizes data and leverages on derived insights to help organizations make better decisions on a specific functional domain. Lastly, an *analytics manager* develops and guides data-driven projects from initiation to planning, execution to performance monitoring, and until their closure. Many people loosely define a data scientist based on the aforementioned roles. However, contrary to popular belief, a *data scientist* is actually the one who leverages statistical techniques and creates analytical models to derive new insights from quantitative and qualitative data.

Mr. Pelayo concluded his presentation with the discussion of the APEC Data Science and Analytics (DSA) Competencies, which includes ten components. The first of these components is *domain knowledge and application* or the application of domain-related knowledge and insights to effectively contextualize data, achieved through practical experience and exposure to emerging innovations. Its entry-level proficiency understands collected data and how they are handled and applied in the specific industry domain. A person with an intermediate proficiency in this competency is someone who develops content strategy and information architecture to support a given industry domain, while an expert makes business cases to improve domain-related procedures through data-driven decision-making.

The next competency is *data management and governance*, which enables one to develop and implement data management strategies, incorporating privacy and data security, policies and regulations, and ethical considerations. Intermediate proficiency enforces policies and procedures for data security, privacy, intellectual property, and ethics, while expert-level proficiency develops policies on data security, privacy, intellectual property, and ethics.

The third competency is *operational analytics*, which uses general and specialized business analytics/intelligence techniques for the investigation of all relevant data to derive insight for decision-making. The basic proficiency in this particular competency is performing business analysis for specified tasks and datasets. A person having intermediate proficiency level can identify business impact from trends and patterns. An expert in this competency can identify new opportunities to use historical data for organizational process optimization.

The fourth competency is *data visualization and presentation*, defined as the ability to create and communicate compelling and actionable insights from data using visualization and presentation tools and technologies. A basic proficiency in this competency is preparing data visualization reports or narratives based on provided specifications. An intermediate proficiency means being able to create infographics for effective presentation and communication of actionable

outcome. A person considered as an expert in this data visualization and presentation is able to select and develop new visualization methods used in a specific industry.

Research methods, the fifth competency, is defined as the utilization of scientific and engineering methods to discover and create new knowledge and insights. A basic proficiency for this is an understanding and the use of the four-step research model: hypothesis, research methods, artifact, and evaluation. An intermediate proficiency in research methods is developing research questions around identified issues within existing research or business process models, and an expert in this competency design experiments that include data collection (passive and active) for hypothesis testing and problem solving.

The next competency is *data engineering principles* or the use of software and system engineering principles and modern computer technologies to develop data analytics applications. Having the knowledge and ability to program selected Structured Query Language (SQL) and Not Only SQL (NoSQL) platforms for data storage and access, in particular writing extract, transform, load (ETL) scripts, is the basic proficiency level in this competency. An intermediate proficiency under this competency includes designing and building relational and non-relational databases to ensure effective ETL processed for large datasets, while an expert uses modern big data technologies to process different data from multiple sources.

The seventh competency is *statistical techniques*, which involves the application of statistical concepts and methodologies to data analysis. A basic proficiency in this competency is knowing and using statistical methods such as sampling, analysis of variance (ANOVA), hypothesis testing, descriptive statistics, regression analysis, and others. In order to be considered intermediately proficient in this competency, a person must be able to select and recommend appropriate statistical methods and tools for specific tasks and data. An expert in this proficiency can identify problems with collected data and suggest corrective measures, including additional data collection, inspection, and pre-processing.

The eighth competency is *data analytics methods and algorithms*, which is the implementation and evaluation of machine learning methods and algorithms on data to derive insights for decision-making. The demonstration of understanding and performing statistical hypothesis testing and explaining statistical significance of collected data are the basic proficiencies needed in this particular competency. An intermediate proficiency means a person can apply quantitative techniques (e.g., time series analysis, optimization, simulation) to deploy appropriate models for analysis and prediction, and an expert can assess data on reliability and appropriateness, and select appropriate approaches and their impact on analysis and the quality of the results.

Computing is the ninth competency, defined as the utilization and application of information technology, computational thinking, programming languages and software, and hardware solutions for data analysis. Performing basic data manipulation, analysis, and visualization is the basic proficiency needed in this competency. Applying computational thinking to transform formal data models and process algorithms into program code are needed to be considered for intermediate proficiency level. Experts in this competency select appropriate application and statistical programming languages, and development platforms for specific processes and datasets.

The last competency is obtaining *21st-century skills*, which exhibit cross-cutting skills essential for analytics at all levels, including, but not limited to, collaboration, ethical mindset, empathy, social and societal awareness, dynamic (self) re-skilling, and entrepreneurship.

KEYNOTE PRESENTATION

Developing 21st Century Skills: Implications on the Curriculum and Pedagogy

PATRICIA B. ARINTO, Ed.D.
Dean, University of the Philippines Visayas Tacloban College and
Former Assistant Vice President for Academic Affairs,
University of the Philippines System

UP Visayas Tacloban College Dean and former UP Assistant Vice President for Academic Affairs Dr. Patricia Arinto began her talk on 21st Century Skills with a primer on the concept. According to Dr. Arinto, 21st-century skills is defined by private curriculum provider Applied Educational Systems (AES) as the abilities that today's students need in order to succeed in careers in the Information Age.

It is said that there are 12 "essential skills" needed by today's students in order to keep up with the growing demands of the industry. These are critical thinking, collaboration, information literacy, technology literacy, leadership, productivity, creativity, communication, media literacy, flexibility, initiative, and social skills.

Conversely, Dr. Arinto also expounded on the cognitive and non-cognitive skills, which are skills that can be demonstrated within core academic content areas and are important to success in education, work, and other areas of adult responsibility.[1] The three domains—the cognitive, interpersonal, and intrapersonal domains—have a taxonomy consisting of eight competency clusters. Table 1 (on the next page) shows this taxonomy.

[1] US National Research Council, 2012

TABLE 1 Taxonomy of 21st-century skills

Domain	Cluster	Terms used for 21st-century skills	O*Net skills[2]	Main ability/ personality factor
Cognitive competencies	Cognitive processes and strategies	Critical thinking, problem solving, analysis, reasoning/ argumentation, interpretation, decision making, adaptive learning, executive function	System skills, process skills, complex problem-solving skills	Main ability factor: fluid intelligence (Gf)
	Knowledge	Information literacy (research using evidence and recognizing bias in sources), information and communications technology literacy, oral and written communication, active listening	Content skills	Main ability factor: crystallized intelligence (Gc)
	Creativity	Creativity, innovation	Complex problem-solving skills (idea generation)	Main ability factor: general retrieval ability (Gr)
Intrapersonal competencies	Intellectual openness	Flexibility, adaptability, artistic and cultural appreciation,	None	Main personality factor: openness

[2] O*Net refers to the Occupational Information Network (O*NET) project that produced a large database of information on 965 occupations and on the basis of which occupations are described along several dimensions, including worker characteristics (abilities, interests, work values, and work styles) and requirements (skills, knowledge, and education)

TABLE 1 Taxonomy of 21st-century skills (continued)

Domain	Cluster	Terms used for 21st-century skills	O*Net skills[2]	Main ability/ personality factor
		personal and social responsibility (including cultural awareness and competence), appreciation for diversity, adaptability, continuous learning, intellectual interest and curiosity		
	Work ethic/ conscientious-ness	Initiative, self-direction, responsibility, perseverance, productivity, grit, Type 1 self-regulation (metacognitive skills, including forethought, performance, and self-reflection), professionalism/ ethics, integrity, citizenship, career orientation	None	Main personality factor: conscien-tiousness
	Positive core self-evaluation	Type 2 self-regulation (self-monitoring, self-evaluation, self-rein-forcement), physical and psychological health	None	Main personality factor: emotional stability

TABLE 1 Taxonomy of 21st-century skills (continued)

Domain	Cluster	Terms used for 21st-century skills	O*Net skills[2]	Main ability/ personality factor
Interpersonal competencies	Teamwork and collaboration	Communication, collaboration, teamwork, cooperation, coordination, interpersonal skills, empathy/ perspective taking, trust, service orientation, conflict resolution, negotiation	Social skills	Main personality factor: agreeableness
	Leadership	Leadership, responsibility, assertive communication, self-presentation, social influence with others	Social skills (persuasion)	Main personality factor: extraversion

Dr. Arinto then went into discussing the need to obtain 21st-century skills. The first reason is the importance of these skills to educational success. These skills enable students to be more "engaged in the learning process and graduate better prepared to thrive in the unceasingly digitally and globally interconnected world";[3] and support "deeper learning of school subjects." Another reason to have 21st-century skills is their importance to workplace success. The emergence of computers reduced the demand for routine cognitive and manual

[3] Partnership for 21st Century Learning, "Framework for 21st Century Learning," Battelle for Kids, 2019, http://static.battelleforkids.org/documents/p21/P21_Framework_Brief.pdf.

tasks, and increased nonroutine cognitive and interactive tasks; and demand is growing for expert thinking (nonroutine problem-solving) and complex communication competencies (nonroutine interactive skills) as well as verbal and quantitative literacy. Having this particular skill set is important to improve/develop health and relationship skills (civic participation). It also serves as a prerequisite for other capabilities (digital inclusion).

The presentation then turned into how 21st-century skills were developed in the General Education (GE) curriculum of the University of the Philippines. The GE curriculum develops skills and approaches to knowledge and engages students in the intellectual work of the disciplines in a variety of fields across the arts and sciences.

The GE curriculum provides some exposure to the whole range of disciplines and areas of knowledge associated by tradition or by institutional convention like the arts and sciences. It develops a set of general critical competencies expected of all liberally educated men and women.

Dr. Arinto also expounded on the philosophy of GE as the embodiment of a liberal education, that is, holistic and integrative, non-specialist, and non-utilitarian. It also aims to developing an appreciation for the foundational disciplines, and the interconnections among them, and their continuing relevance in understanding and dealing with complex problems confronting individuals and communities, nation, and the world as a whole. GE is also an interdisciplinary program composed of core and elective courses in the humanities, social sciences and philosophy, and natural sciences and mathematics. Table 2 (on the next page) shows the learning outcomes of General Education.

Dr. Arinto then presented the initial findings from the 2018 UP GE course evaluation with the highest and lowest ratings in terms of course outcomes, and aspects of course delivery from both students and faculty members.

She also presented how 21st-century skills can achieve GE outcomes and "GE gestalt." In the case of undergraduate students,

TABLE 2 General Education (GE) learning outcomes

Program Objective	Knowledge	Skills	Attitudes/ dispositions
Broaden intellectual and cultural horizons Hone critical and creative thinking Develop a passion for learning and scholarship	Exposing students to **different knowledge traditions, perspectives, and paradigms** (range of fields and sub-fields, theories and contexts)	Training students in **various disciplines and modes of inquiry**; developing **advanced literacy skills** (textual, visual, digital) and **higher order thinking skills**, including the ability to "recognize and define problems; analyze the structure of an argument; assess the relationships between facts, assumptions, and conclusions; and perform hypothetico-deductive processes"	Cultivating in students **objectivity and open-mindedness, respect for diversity, and a global and cosmopolitan outlook**
Cultivate a high sense of intellectual and moral integrity	Providing a strong **grounding in ethics**	Developing a **capacity for critical reflection, moral reasoning, and ethical action**	Cultivating **autonomy and independence of mind**
Foster a commitment to nationalism and social justice	Exposing students to a **nationalist tradition**; **engaging with public issues**	Training students in **historical and sociological analyses**	Developing ***pagiging makabayan*, social responsibility, *pakikipagkapwa-tao*, and commitment to public service**

it is a way to discover their own interest and figure out where they want to concentrate their studies. From the student's point of view, the challenge of general education is, first and foremost, to achieve an integrated understanding of the various connections among all domains of learning (i.e., in the majors, the GE courses, in electives, and in disciplines across the arts and sciences). The other reason is for both specialized and general courses to cultivate habits of mind that will

enable a college graduate to sustain a life of intellectual curiosity and learning.

Dr. Arinto went on to discuss the case of digital humanities, which involves using digital technologies to formulate innovative research questions in the humanities and to investigate in new ways the primary materials where scholars of humanities are concerned. It is a field of engagement in which technology, textuality and humanities converge. She argued that there is a need to have good grounding in traditional humanities studies as an essential aspect for the effective practice of digital humanities.

Closing her presentation, Dr. Arinto presented some concluding thoughts on the topic at hand on the need for digital skills integration, and research-based teaching methods. Integrating digital literacy in specific educational contexts is based on the understanding that digital literacy is a nuanced and varied set of capabilities, tuned to the requirements of different roles and the practices of different subject areas. Digital literacies are expressed in specific learning, teaching, and research activities which take their meaning from the subject areas in which they are practiced.

The research-based teaching models, on the other hand, refer to the use of multiple and varied representations of concepts and tasks, such as diagrams, numerical and mathematical representations, and simulations, combined with activities and guidance that support mapping across the varied representations. Research-based teaching models put emphasis in encouraging elaboration, questioning, and explanation, and engaging learners in challenging tasks, while also supporting them with guidance, feedback, and encouragement to reflect on their own learning processes and the status of their understanding. Dr. Arinto mentioned that there is a need to pay attention to an interdisciplinary mindset, premised on a strong grounding in particular disciplines, integration, and explicit instruction and immersion.

The workshop invited members of the academe to present different Data Science programs in different colleges within the UP System, as well as from other HEIs.

Data Science Curricula

JAIME DL. CARO, Ph.D., JOHN PAUL C. VERGARA, Ph.D., JOHNROB Y. BANTANG, Ph.D., BRENDA A. QUISMORIO, Ph.D., and PROSPERO C. NAVAL, JR., Ph.D.

Dr. Jaime DL. Caro of UP Diliman presented the curriculum of the proposed Bachelor of Science in Data Science (BSDS) program of the UP Department of Computer Science (DCS). According to Dr. Caro, the program was developed to address the need for a new breed of future-ready graduates with the competencies and skills required to address the pressing demands from the industry for data scientists that can work in different domains.

The BSDS program covers the study of data and the methodologies, processes, algorithms, and systems for collecting, refining, storing, and analyzing data to arrive at useful insights and knowledge. The program envisions its graduates to have the proper skills for processing, management, and analysis of data in various forms.

According to Dr. Caro, data science benefits from developments in other disciplines such as mathematics, statistics, computer science, and business. He emphasized that the BSDS curriculum will have foundation and professional courses, including major courses in mathematics, statistics, and data science. The curriculum will cover the knowledge areas recommended in the APEC Data Science and Analytics (DSA) Competencies (2017).

Dr. Caro highlighted the requirements that academic institutions must adhere to in developing a BSDS program. First, HEIs must develop a complete set of curriculum, program outcomes, and performance indicators, as well as a program and evaluation system towards the continuous quality improvement of the program, as set in the Commission on Higher Education (CHED)'s Implementation Handbook for Outcomes-Based Education (OBE) and Institutional Sustainability Assessment (ISA) (2013). Second, faculty members who will teach in the BSDS program must have at least have a baccalaureate degree in

data science-allied fields or a graduate degree in data science or in allied fields. Third, program administrators must meet any of the following academic and professional requirements:

- A doctoral degree in data science or its allied fields; or
- A combination of a master's degree in data science or allied fields and at least one (1) year of work, consultancy, research, or tertiary-level teaching experience in data science or allied fields within the last three (3) years; or
- A combination of a doctoral degree in a STEM program and at least three (3) years of work, consultancy, research, or tertiary-level teaching experience in data science or allied fields within the last five (5) years.

For the presentation on Data Science programs at the graduate level, Dr. John Paul Vergara of the Ateneo de Manila University (ADMU) and Dr. Brenda Quismorio of the University of Asia and the Pacific (UA&P) shared their respective institution's programs.

Dr. Vergara presented the Master of Science in Data Science (MSDS) program offered by ADMU. The development of the program was a response to the growing demand for data scientists both locally and globally, who will support institutional commitment to interdisciplinarity and nation-building. The program's goal is to develop students who can use statistical and computational methods to derive insights from large datasets. Some of the expected outcomes of the program include applying a variety of data mining techniques for modeling data sets, handling incomplete and erroneous data, and applying big data processing and cloud computing techniques to address the scalability challenges of very large datasets.

ADMU's MSDS program is composed of 36 academic units of which 15 units are foundational courses, 15 units are methods and domains, and the remaining six units are for thesis writing. The university also offers BS–MS programs, such as the BS Applied Mathematics–Masters in Data Science (a five-year non-thesis program), the BS–MS Computer Science–Data Science, and the BS–MS Management Engineering–Data

Science programs. It also has a dual degree graduate program with Queen Mary University of London (QMUL) where students attend seven months of classes in QMUL. At the end of the program, students will receive both an MS Big Data Science degree from QMUL and an MS in Data Science degree from ADMU.

On the other hand, the Master in Applied Business Analytics (MABA) program of the UA&P is a two-year, part-time graduate program geared towards decision makers and working professionals. According to Dr. Quismorio, a MABA student goes through a learning journey, is geared to become an analytics manager, and is equipped with the ten APEC-recommended DSA competencies. The program aims to produce graduates capable of transforming data into actionable insights that will drive businesses and organizations for the good and of managing analytics teams. As such, the target students of the program are those with extensive experience in their respective industries. They are expected to be able to position themselves to compete for a number of fast-growing functional analyst roles, including roles as research analyst, marketing analyst, financial analyst, operations analyst, and human resources analysts.

According to Dr. Quismorio, the learning outcomes of the MABA program include:

- Leveraging data to inform strategic and operational decisions;
- Utilizing data and analytical models to inform specific functions and business decisions;
- Leveraging data analysis and modeling techniques to solve problems and glean insight across functional domains;
- Creating analytical models to derive insights from data;
- Overseeing analytical operations and communicate insights to executives; and
- Identifying, defining, and prioritizing ethical concerns related to data analytics as they pertain to persons, organizations, and society.

Dr. Johnrob Bantang from the UP National Institute of Physics and Dr. Prospero Naval, Jr. from the UP DCS presented the proposed Ph.D. program in data science under the UP Diliman College of Science. According to the proposal, the program outcomes include domain knowledge base; data science application to domain; generation of new knowledge; and collaboration and application of data science.

The program is structured to include foundation and core courses, electives, as well as research/dissertation writing. Some of the courses that will be offered by the UP DCS include Intelligent Systems, Data Mining, Machine Learning, and Reinforcement Learning. Students can also engage in active research areas, such as data analytics for healthcare, environment, and education, biomedical imaging, machine translation for Philippine languages, and brain-machine interaction.

TABLE 3 Comparison between data science and artificial intelligence

	Data Science	Artificial Intelligence
Creator domain knowledge	High	Mid to high
End user domain knowledge	High	Low to none
Purpose	Gain insight	Build system
Decision maker	Human	Human-machine (semi-automated) to machine (fully autonomous)
Autonomy	None to low	Mid to high
Math sophistication	Very high and deep	High and broad
Programming skills	Mid	High
Methods	Human interpretability to end user	Blackbox to end user
Data	Small samples/ sampling from big data	Small data to big data
Data dimensionality	Low or reduced to low	Mid to very high
Hardware	Minimal	Computer-intenstive
Realtimeness	Hours to months	Milliseconds to minutes

Simultaneous Breakout Sessions

During the afternoon session, the participants of the workshop were divided into three groups for the breakout session. The three groups were: (1) Data Science, (2) Data Engineering, and (3) Analytics Management. The participants were grouped based on their expertise and competencies. The objective of the breakout session is to discuss the development of Data Science programs of the participating universities and to identify key competencies needed to be developed at each academic level.

The moderators described the program profile for Data Science programs for each academic level. During the workshop, the moderators also reiterated the ten DSA competencies of the APEC.[4] These competencies were used as a guide for the participants in identifying learning objectives for each academic level.

Dr. Fidel R. Nemenzo, current UP Diliman Chancellor and former Convenor of the UP CIDS Program on Data Science for Public Policy, and Dr. Johnrob Bantang of the National Institute of Physics headed the breakout session on the Data Science role. The group consisted of academics from UP Diliman and UP Cebu, as well as the UA&P. Dr. Bantang noted in his discussion that data scientists leverage statistical techniques and creates analytical models to derive new insights from quantitative and qualitative data. Their expertise is usually in mathematics and statistics.

Three out of the four participants in the Data Science breakout group confirmed they are planning to create a undergraduate program in data science. However, there were no plans to create a master's or a doctoral program. The participants also laid out the program profiles for Data Science program offerings at different academic levels.

---

[4] Asia Pacific Economic Cooperation, "10 Recommended APEC Data Science & Analytics Competencies," June 2017, https://www.apec.org/Press/Infographics/10-Recommended-APEC-Data-Science-and-Analytics.

Undergraduate students are expected to perform predefined tasks and work under guidance of faculty, while master's students will be immersed into formulating and solving data analysis tasks to achieve a wide range of organization related goals, independent work on the solutions and operation. Doctoral students will have to identify new approaches, and application areas to achieve organizational goals based on existing or new collected data, architect new data models and applications, assess multiple alternative technologies and solutions based on structured analysis and experience, and propose new approaches, if necessary.

The participants also tackled the challenges of instituting Data Science programs in HEIs in the country. Some of the challenges identified were the methods used in data science, the limited human resources and facilities and equipment (such as computer laboratories and essential software), limited budget, unavailability of real-world data to be used in courses, and unreliable internet and other connectivity issues presently affecting the mode of instruction in the field.

The second breakout session was about the Data Engineering role and was moderated by Mr. Doc Ligot, the founder of Cirrolytix, a research services and social impact data analytics company. The participants of the session were academics from the UP Open University, De La Salle University, and Far Eastern University. Mr. Ligot described data engineers as the ones who clean, transform, and load data from transactional systems to centralized data repositories. Their areas of expertise are in information technology, information science, and computer science. During this session, three out of the six participants confirmed that they are planning to create a bachelor's program in Data Science. However, there were no plans to create graduate programs.

The first challenge identified by the participants of the discussion of Data Engineering roles is that big datasets are needed for the authentic learning experience of students. There is also lack of linkages with the industry, as well as the issue of scarcity of experts and academics as data engineers usually does the "dirty work" of making sure that datasets are

clean and will not end up in Type 1 or 2 errors. Similar to the earlier breakout session, there is also the challenge of limited real-world data and the difficulty of identifying experts who can help in parallel computing. Lastly, HEIs lack the necessary facilities and equipment needed for optimal learning.

Dr. Brenda Quismorio, the Program Director of UA&P's Master in Applied Business Analytics (MABA) program, moderated the session on the Analytics Management role. Academics from UP Cebu, UP Baguio, Isabela State University, and University of San Carlos were the participants in the session. According to Dr. Quismorio, analytics managers develop and guide data-driven projects—from initiation and planning, to execution and performance monitoring, up to their closure. Their expertise is project management, and they can fulfill roles such as chief data officer, project manager, data engineering manager, data science manager, and analytics translator.

For this breakout session, five out of the six participants expressed that they are planning to create a bachelor's program in Data Science. Similar to the previous sessions, there were no plans to create graduate-level programs. The challenges and issues identified by the participants during this breakout session were the extensive need for training in software usage, the lack of ample facilities and equipment in HEIs, and the need for stronger academe-industry linkages for student immersions and access to real-world data.

Discussion on the APEC DSA Competencies

The simultaneous breakout sessions discussed how the APEC DSA Competencies is expected to be developed in undergraduate and graduate program offerings in Data Science.

The participants expect undergraduate students to display competency in *domain knowledge* through understanding collected data and different handling methods, procedures, and applications

in the context of the specific research and industry domains. On the other hand, proficiency at the graduate (master's and doctoral) level with regard to this competency is concerned with understanding and developing profiles or adopting existing data analytics applications for specific research or industry domains; presenting or visualizing data in the context of given domain methods and existing practices; and developing content strategy and information architecture to support a given research or industry domain and its audience.

For *data management and governance* proficiency, undergraduate students are expected to identify and understand data sources, structures, and types; perform data annotations; apply domain-/subject-related metadata; use common information retrieval and filtering applications for general data preparation and exploration; collect and integrate different data sources and provide them for further analysis; be familiar with data encoding standards such as Hierarchical Data Format (HDF), Network Common Data Form (NetCDF), or specialized data formats for medical imaging (such as Digital Imaging and Communications in Medicine (DICOM) and Health Level 7 (HL7)); and be aware and consistently apply policies and measures to ensure data security, privacy, intellectual property rights (IPR), and ethics during all stages of the data life cycle. Graduate students are expected to perform tasks related to data publication, accessibility, curation, and quality assurance (data curation); develop and maintain historical data repositories of analysis results (data provenance); develop and implement relevant data models; define metadata using common standards and practices for different data sources in variety of scientific and industry domains; use computer applications to manage large amounts of heterogeneous data and information; implement and manage policies and procedures for data security, privacy protection, and IPR management; and consistently handle incidents of data loss or theft.

In order to be proficient in *operational analytics*, participants expect undergraduates to use selected business analytics (BA) and business intelligence (BI) platforms commonly used in organizations;

perform business analysis for specified tasks and datasets; suggest their presentation/visualization in a form understandable by various business roles/staff; and perform tasks on user requirements gathering and user experience (UX)-related data collection, including data collected from social media.

Likewise, the participants expect master's students to deliver business-focused analysis using appropriate BA/BI methods and tools; identify business impact from trends; implement and use data warehousing technologies and data repositories to collect, store, and manage business-related data to facilitate wide range of data analysis needs; analyze needs and develop processes for user requirements gathering and analysis to enable effective UX design using appropriate data, techniques, and tools; develop and implement processes to include external data (including social media data and open data) into BA/BI applications (e.g., using sentiment analysis on social data to assess business campaigns' effectiveness); formulate and implement best practices in data presentation and visualization for organizational needs and to be able to enable sound storytelling; and identify, collect, and integrate/ingest data for a specific use case or project.

As for doctoral students, they are expected to analyze information needs, assess existing data, and suggest/identify new data required for specific business contexts to achieve organizational goals, including using social network and open data sources; analyze and identify new opportunities for the use of historical data available at the organization for organizational process optimization; operationalize fuzzy concepts to enable key performance indicators measurement and validation of business analysis; identify and assess potential challenges to provide greater insights into business performances and processes; lead in identifying required information and analytical models for interacting with the specific user groups or in the specific business sectors; understand customer-specific data footprint and use it in developing data analysis model; make business case as a result of organizational data analysis and identified trends; and provide advice on organizational processes and customer/UX improvement.

As for the discussion on *data visualization* competency, participants of the breakout sessions agreed that undergraduate students must be able to prepare data visualization reports or narratives based on provide specifications using one of popular visualization languages and tools (e.g., Excel, Tableau, D3, Processing); and be familiar with the general concepts of visualization such, e.g., Wilkinson's grammar of graphics, and their implementation in one visualization programming languages. Graduate students are likewise expected to create effective data visualization reports or narratives for various audiences that employ analytics and visualization software and strategies to support decision-making; create business dashboard, develop specialized interactive data visualization applications for selected datasets and audiences or businesses, and use animation where appropriate; use infographics for effective presentation and communication of actionable outcome from the data analysis or focused study; evaluate data visualization options for proper application in various situations, select the most appropriate visualization, and advice on visualization languages and tool selections; and select appropriate and develop new visualization methods based on approaches, conventions, and semantics used in specific research, industry, or business settings.

Conversely, on *research methods*, the participants conferred that both undergraduate and graduate students should be able to respect data as a resource and its potential applications for deriving value from data; understand and be able to use the four-step research model (hypothesis, research methods, artifact, and evaluation) for specific problems or activities; and collect data for a targeted study or research and prepare them for data analysis (enter, annotate, validate, store, and publish).

On the competency of *data engineering*, the participants mentioned that both undergraduate and graduate students must be able to have the knowledge and ability to use one or several platforms for data storage, processing, collection, and distribution; have knowledge and the ability to program selected SQL and NoSQL platforms for data storage and access (write ETL scripts in particular); understand and program parallel data processing using the MapReduce paradigm on

distributed computational platforms (e.g., Apache Hadoop, Apache Spark); be familiar and use cloud-based tools and solutions for data storage and processing; apply basic software engineering principles to write transparent, reproducible, re-usable, and auditable codes; and capture the whole process and publish in versioning control systems (e.g., GitHub, subversion, etc.).

Proficiency in *statistical techniques* at the undergraduate level includes knowledge and ability to use statistical methods for data analysis such as general statistical analysis, sampling, ANOVA, hypothesis testing, descriptive statistics, regression analysis, etc.; knowledge and ability to use at least one or several statistical programming languages and tools (e.g., R, Python, RapidMiner, SPSS) for statistical analysis, data preparation, and pre-processing; prepare and inspect data to be used with the specific statistical methods and tools; and formulate and statistically test hypothesis based on collected data. They also expect graduate students to use computer-assisted conduct of research in all of its stages—data gathering, processing, analysis, interpretation, and report writing; apply design thinking in planning and developing data analytics solutions for human consumption; and develop research questions around identified issues within existing researches or business process models. Doctoral students, on the other hand, are expected to analyze domain-related research process models, identify and analyze available data to identify research questions and/or organizational objectives and formulate sound hypothesis; design experiments which include data collection (passive and active) for hypothesis-testing and problem-solving; and guide data-driven projects using research methods for focused and explorative data analysis and suggest use of appropriate data analytics methods and big data engineering tools and facilities.

In the *data analytics methods and algorithms* competency, the participants expect undergraduate students to perform data preparation and pre-processing using appropriate tools and software (e.g., summarization, quality checks, cleaning, validation, sampling, etc.); understand and be able to select an approach to analyzing selected datasets, including big- or web-scale datasets; demonstrate

understanding and perform statistical hypothesis testing; explain statistical significance; be aware of ethical issues in data analysis (in particular due to possible biased collected data); and handle discovered information responsibly. Graduate students are expected to effectively use a variety of data analytics techniques, such as machine learning (including supervised, unsupervised, and semi-supervised learning), data mining, and prescriptive and predictive analytics, for complex data analysis through the whole data life cycle; apply designated quantitative techniques, including statistics, time series analysis, optimization, and simulation, to deploy appropriate models for analysis and prediction; identify, extract, and pull together available and pertinent heterogeneous data; use SQL to design complex queries on relational data and integrate them into the data processing and analysis workflow; visualize results of data analysis; design dashboard and use storytelling methods; use different performance and accuracy metrics for model validation in analytics projects, hypothesis testing, and information retrieval; integrate data analytics and processing applications into organization workflow and business processes to enable agile decision making; and apply optimization and simulation techniques to suggest improvements to organization workflows and business processes.

Doctoral students, on the other hand, are expected to develop and plan required data analytics for organizational tasks, including evaluating requirements and specifications of problems to recommend possible analytics-based solutions; identifying processes and mechanisms to effectively retrieve, assess, enrich, manipulate, re-engineer, and amalgamate data and apply them to practical data processing procedures; understanding and using modern data sources such as social media data, open data, and governmental data, and use them for blending business analytics, sentiments analysis, targeted advertisements, and services delivery; assessing data in terms of reliability and appropriateness to the possible solutions; selecting appropriate data analytic approaches (e.g., descriptive vs. diagnostic vs. predictive vs. descriptive analytics); and assessing how they impact performance of data analysis and quality of the results.

In the *computing* competency, undergraduates are expected to demonstrate computational thinking in approaching data handling and using a variety of applications to work with data; use a variety of common office tools for text and data processing, manipulation, transformation, presentation and of communication tools (e.g., Microsoft Excel, PowerPoint, or similar software from other commercial or open source vendors) to prepare and exchange data between team members and applications; write data analysis codes/processes using popular statistical or data analysis languages and tools (e.g., R, Python, RapidMiner, SPSS); use SQL queries to work with relational databases; perform basic data manipulation, analysis, and visualization using selected programming languages, tools, or software; and be familiar with the common practices in data versioning management, apply them in own and organizational workspaces.

The participants, moreover, expect master's students to use spreadsheet tools (e.g., Microsoft Excel) to build complex graphs and reports and further combine them into dynamic dashboards; use the PowerPivot functionality to build databases and data models; create visualized data; use online and cloud-based data and web analytics tools such as GoogleDocs, SurveyMonkey, GoogleAnalytics, OpenRefine, etc.; employ crowdsourcing methods where applicable; develop full data analysis applications according to specified requirements using one or several statistical and data analytics languages (e.g., R, Python, Julia, KNIME, SAS, WEKA) and tools (e.g., R-Studio, Anaconda, SPSS, RapidMiner, Microsoft Visual Studio); select appropriate language, tool, and development approach based on knowledge and experience with corresponding tools; apply computational thinking to transform formal data models and process algorithms into program code, prototype and deploy for practical use; and automate data analysis processes.

Finally, the participants presume doctoral students will use advanced spreadsheet processing functions (such as pivot tables, INDEX, VLOOKUP, and data cleansing), embedded advanced programming, statistical and query functionality/formulas (for example, those available in Microsoft Excel or Access) to accomplish business data analysis and visualization tasks; have experience in working with

cloud-based platforms for data analysis, storage, and processing and using them in data analysis applications; assess and select appropriate programming languages, software tools, and development platforms for specific organizational processes and datasets and be able to integrate applications/solutions from different platforms; and advise on the right selection and correct use of different applications.

A Unified Call on 21st Century Skills

Across the three breakout sessions, all participants came to an understanding that both undergraduate and graduate Data Science students must be able to acquire 21st-century skills and exhibit cross-cutting skills essential for analytics at all levels. The include, but not limited to, collaboration, communication and storytelling, ethical mindset, organizational awareness, critical thinking, planning and organizing, problem solving, decision making, customer focus, flexibility, business fundamentals, cross cultural awareness, social and societal awareness, dynamic (self) re-skilling, professional networking, and entrepreneurship.

Next Steps and Closing Remarks

FIDEL R. NEMENZO, D.Sc.
Chancellor, University of the Philippines Diliman and
Former Convenor, UP CIDS Program on Data Science for Public Policy

Former UP CIDS Program on Data Science for Public Policy (DSPP) Convenor and incumbent UP Diliman Chancellor Dr. Fidel R. Nemenzo delivered the closing remarks for the program. In his remarks, Dr. Nemenzo said that this gathering of a loose network of data scientists is not the end of the line, but rather, the beginning of more discussions in the future as the country advances and equips its human capital the

necessary skills to keep up with the changing demands of industry. He hopes that this workshop will evolve into regular meetings of those at the forefront of the development of data science from both the academe and industry.

ACKNOWLEDGMENTS

Office of the Vice President for Academic Affairs (OVPAA), University of the Philippines
Analytics Association of the Philippines (AAP)
Cobena ▪ Z-Lift Solutions ▪ Accenture Philippines

ISSN 2719-0560 (PRINT)
ISSN 2719-0579 (ONLINE)