

# Data Science as an Emerging Discipline: The Roles of iSchools in the Era of Big Data

X. Hu

Faculty of Education  
University of Hong Kong  
Hong Kong S.A.R., China  
xiaoxhu@hku.hk

S. K. W. Chu

Faculty of Education  
University of Hong Kong  
Hong Kong S.A.R., China  
samchu@hku.hk

J. S. Downie

School of Information Sciences  
University of Illinois  
Champaign, IL, U.S.A.  
jdownie@illinois.edu

C. W-Y. Lee

Faculty of Education  
University of Hong Kong  
Hong Kong S.A.R., China  
celinalee510@gmail.com

## ABSTRACT

Data science has been a hype in both academia and industry. It is a high time for iSchools to reflect the roles they are playing and how they can contribute in the data science space, specifically in the areas of education and research. This position paper reviews data science related teaching and research activities in tier 1 iSchools and discusses the potentials of iSchools in shaping up the future of Data science.

## CCS CONCEPTS

• **Social and Professional Topics** → **Professional topics**; *Computing Education*; • **Applied Computing** → Computers in other domains

## KEYWORDS

Data science, iSchools, education, research

## ACM Reference format:

X. Hu, S. K. W. Chu, J. S. Downie, 2017. Data Science as an Emerging Discipline: The Roles of iSchools in the Era of Big Data. In *Proceedings of Information Science to Data Science: New Directions for iSchools Workshop in iConference, Wuhan, China, March 2007*, 2 pages.  
DOI: 10.1145/1234

## 1 INTRODUCTION

With the complexity and exorbitant amount data available, information professionals are facing one of the most eminent challenges in figuring out effective and efficient ways of data processing and analysis [5]. Information professionals around the globe have started to envision addressing unresolved questions by manipulating the readily available yet highly unorganized and heterogeneous data to derive practical and valuable information and knowledge [7]. To maximize the potential of data manipulation, governments and global organizations have put forward policies and proposals that provide guidelines on accessing and using public data for research (e.g. [1,6,8]). Research libraries, especially those

affiliated with large research institutions, have taken a central and pioneer role in managing research datasets (e.g., RDAP<sup>1</sup> and HathiTrust<sup>2</sup>) as they move away from traditional data storage and retrieval service to more powerful value-adding data analysis (e.g., text analysis and visualization created by searching bibliographic information, tagclouds, timeline and map plotting, and social network relationship discovery.) Data scientists, a new form of information professionals, have evolved from its traditional role of holding and providing access to knowledge resources to being an active agent in creating knowledge [9] through discovering insightful patterns from large sets of data [5]. Having been dubbed the “sexiest job in the 21st century,” [2] and the “best job of the year,” [4] data scientist as a profession has emerged quickly and become the spotlight of a wide range of sectors which are eager to draw benefits from the new opportunities. With such fervor surrounding data science, there appears to be a lack of consensus on its definition and scope [3]. It is a high time for us, especially the iSchools, a consortium of leading institutions in the information field, to sit back from the craze of data explosion to think about how we can contribute in the data science space, specifically in the areas of education and research.

## 2 CURRENT LANDSCAPE

Among traditional academic institutions and universities, a good number of them are striving to update their existing information degree programs in the information fields and formulate new offerings. More specifically, within the iSchools community, programs or certificate courses related to data science or data curation are currently offered in many iSchools. Currently, most of the data science specific courses are offered at the Master’s level, while undergraduate programs and MOOC courses are also available in some institutions. Table 1 summarizes data science related teaching and research activities in iSchools (Tier 1 members).

While courses related to Data Science are springing up quickly, the standard, content and quality differ significantly. There is often times ambiguity of what data science entails and courses differ significantly in terms of their targeted fields. For

<sup>1</sup> <https://about.rdap.org>

<sup>2</sup> <https://www.hathitrust.org/>

example, some programs seem to focus on data curation and management while others put more emphasis on statistics, algorithms and system expertise. It is noteworthy that there are 19 data science related research centers or projects in the 27 surveyed iSchools, indicating a strong research focus in addition to education.

**Table 1: Data Science Related Teaching and Research Activities in iSchools (Tier 1 members)<sup>3</sup>**

	U.S.	Canada	Europe	Asia	Total
Number of institutions	20	1	4	2	27
Graduate programs	9	0	1	1	11
Graduate certificates	4	0	0	0	4
Undergrad programs	7	0	0	1	8
Courses (incl. MOOC)	3	1	1	0	5
Research centers or projects	13	1	3	2	19

### 3 ISCHOOL IN DATA SCIENCE EDUCATION

Comments from course takers revealed that some existing courses were too elementary in level, failing to provide adequate training for career preparation<sup>4</sup>. Indeed, with such great arrays of training programs available, it is difficult to identify the perfect programs suiting their needs and expectations. It is therefore important that education institutions and organizations such as iSchools to take the initiative to clarify, organize and optimize the existing education opportunities. As established academic institutions in the information field, iSchools should contribute based on our traditional and current strengths on data science-related areas, such as information organization and access, information management, and information preservation, ethics of information.

Working together, experts could discuss and establish a clearer definition, scopes and streams of what data science entails. It is significant that institutions specify the focus of their courses to potential course takers to avoid confusions. Further, a list of key principles, skills and toolsets necessary for data scientists can be compiled to help current and future information professionals assess their proficiency and locate courses and

programs suitable to their levels, interests and career plans. To improve the quality and effectiveness of programs, institutions should seek to bring together faculty staff from different disciplines specific to different knowledge and skills data science entails (e.g., statistics, computer science, information science, business and marketing, metadata, digital libraries, institutional repositories, preservation, reference).

### 4 GLOBAL COLLABORATION

To maximize the potential of iSchools in leading data science education and research, international collaboration is essential. In addition to MOOC courses, such as the first MOOC-based Master's degree jointly offered by University of Illinois at Urbana-Champaign and Coursera<sup>5</sup>, which make data science education accessible to aspiring data scientists from all over the world, more formal and informal meetings, seminars, workshops and conferences can be arranged among data science specialists in iSchools, to disseminate and share most up-to-date information and experience in the field. An example of such kind of international collaboration is the joint training institute co-organized by the iSchools in the University of Illinois and the University of Hong Kong in 2015<sup>6</sup> and 2016<sup>7</sup>. Drawing upon the expertise from both iSchools, the Institute aims to meet the increasing need of academic librarians in Hong Kong and neighboring regions in learning theories, techniques and best practices in data curation and data-driven scholarship services. The participants highly the vision and quality of the institute and would look forward to more similar events on data science coming in the near future.

### REFERENCES

- [1] Bakowki, P. (2014). Big data: opportunities and privacy concerns. European Parliament Research Service.
- [2] Davenport, T.H. & Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review, October 2012..
- [3] Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., ... & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66(8), 1523-1545. .
- [4] Glassdoor (2016). 25 Best Jobs in America. Retrieved from [https://www.glassdoor.com/List/Best-Jobs-in-America-LST\\_KQ0,20.htm](https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm).
- [5] Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and techniques* (3rd ed.). San Francisco: Morgan Kaufmann.
- [6] High Level Expert Group on Scientific Data. (2010). *Riding the wave—How Europe can gain from the rising tide of scientific data*.
- [7] Mcneely, C. L., & Hahm, J. (2014). The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research*, 31(4), 304-310. doi:10.1111/ropr.1208
- [8] Organisation for Economic Cooperation and Development (OECD). (2007). *OECD principles and guidelines for access to research data from public funding*. Paris. Retrieved from <https://www.oecd.org/dataoecd/9/61/38500813.pdf>
- [9] Tammaro, A. M., Ross, S., & Casarosa, V. (2014). Research data curator: The competencies gap. *BOBCATSSS 2014 Proceedings*, 1(1), 95-100.

<sup>3</sup> <http://ischools.org/members/directory/>

<sup>4</sup> <https://www.linkedin.com/pulse/why-i-left-my-masters-program-charles-pensig-1>

<sup>5</sup> <http://online.illinois.edu/mcs-ds>

<sup>6</sup> <http://library.ust.hk/events/staff-workshops/data-curation/>

<sup>7</sup> <https://lib.hku.hk/ddsal/index.html>