Unique Contributions and Opportunities of iSchools in Data Science Education

X.Han
School of Information Management,
Nanjing University
P.O. Box 210023
PRC
hanx015@163.com

QH.Zhu
School of Information Management,
Nanjing University
P.O. Box 210023
PRC
qhzhu@nju.edu.cn

ABSTRACT
Based on the review of different definitions of data science and data scientist, we discuss the unique contribution of iSchools for data science from four aspects: data thinking, data management, absorptive capacity of information technology and interdisciplinary attributes. The authors propose that iSchools need to take in the idea of data science, acknowledging what we can offer and admitting what we lack. Finally some suggestions are presented for the reform of information science education in data era.

CCS CONCEPTS
• Social and professional topics → Information science education
• Information systems → Information integration

KEYWORDS
Data science, big data, information science, data scientist, education

ACM Reference format:

1 INTRODUCTION
Big data and data science have both attracted researchers’ and practitioners’ attention in recent years. According to research by McKinsey, big data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus. United States could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions [1]. Also Harvard Business Review declared data scientist the “sexiest job of the 21st century” [2]. Up to now, many disciplines such as statistics, computer science and information science have shown their great interest on data science. Data science represents a set of disciplines necessary to solve big data challenges. Although there are still contradictions and no wildly recognized conclusions, many colleges have been promoting various data science programs based on their own disciplines [3]. A qualified data scientist requires many skills related to different disciplines. For information science, which focuses on data, information and knowledge, iSchools’ unique contribution to data science is worthy of discussion. The answer to this question will help us to promote the development of information science in the big data era and set up proper curriculum to train data scientist for industry.

Based on the diverse definitions of data science, data scientist, and the relationship between data science and information science, we discuss unique contribution of iSchools to data science from four aspects: data thinking, data management, absorptive capacity of information technology and interdisciplinary attributes.

2. DATA SCIENCE AND DATA SCIENTIST
Clear understanding of the definition of data science and data scientist is beneficial to find the relationship between information science and data science. In general, data science is the science (or study) of data [4]. Many scholars have given their own definitions of data science. Dhar (2013) addresses the heart of data science that ‘data science is the study of the extraction of knowledge from data’ in a narrow sense [5]. Stanton (2012) focused on all aspects of big data life-cycle and defined data science as ‘an emerging area of work concerned with the collection, preparation, analysis, visualization, management, and preservation of large collections of information’[6]. Provost and Fawcett (2013) think that ‘data science involves principles, processes, and techniques for understanding phenomena via the (especially automated) analysis of data’ [7]. Yao (2015) claims that data science has been often used to emphasize the scientific principles of data representation, collection, storage, integration, analysis and understanding. Data science is not a new term although it might has been given a new meaning [8]. Cao (2016) defined data science from different aspects. For example, from process perspective, data science is a systematic approach to ‘thinking with wisdom,’ ‘managing data,’ ‘computing with data,’ ‘mining on knowledge,’ ‘communicating with stakeholders,’ ‘delivering products,’ and ‘acting on insights.’ And from disciplinary perspective, data science is a new interdisciplinary field in which to study data and generate data products with ‘data-knowledge-wisdom’ thinking. Data science combines traditionally data-oriented disciplines together, such as statistics, informatics, and computing [4]. Data science employs techniques and
3 UNIQUE CONTRIBUTION AND OPPORTUNITIES OF ISCHOOLS FOR DATA SCIENCE EDUCATION

Based on different definitions of data science above, we think data science is a multidisciplinary field. It is directly influenced by big data phenomena and the increasing emphasis on data-driven decisions. Although data science is not a new research era, the integration of latest research of information science, computer science, and statistics has brought new idea to this domain in the big data environment. The three disciplines mentioned above have different contributions to the development of data science. Meanwhile, iSchools mainly focus on information storage, organization, management, analysis, use and dissemination in the past. So there is no doubt that iSchool’s research era intersect with data science. Apart from the more advanced information technology (especially unstructured data processing and machine learning) and statistical tools, the main task of data science corresponds with the research interests of information science. Therefore, considering the definition of data science and responsibilities of data scientists, the unique contribution and opportunities of information science to data science are mainly reflected in the following four aspects.

3.1 Data thinking

The biggest traditional advantages of information professionals are their strong information sensitivity, information literacy and information thinking. Since information is based on data, this advantage becomes data thinking and is still important and effective in big data era and data-driven environments. Data thinking is the generic mental pattern observed during the processes of picking a subject to start with, identifying its parts or components, organizing and describing them in an informative fashion that is relevant to what motivated and initiated the whole processes[14]. Although information technology is important for the daily work of data scientists, data thinking, raising data question and solving practical problems based on data are even more vital. Data, technologies, and people are the three pillars of data science, and it is obvious that these three components are not at the equal line. Data are ubiquitous, while technologies are being developed in order to cope with various big data problems. What lagged far behind is ‘people’ who can critically think about big data problems [11]. Graduates of information science are more inclined to data thinking, while graduates of computer science may be good at information technology but not at data thinking. Data thinking seems simple, but it is difficult to form in a short time just like culture. This is a great discipline culture advantage for information science which is specialized in extracting knowledge and wisdom from data and information.

3.2 Data Management

Data science requires a variety of computer technologies and statistical methods (such as machine learning and unstructured data processing), which are some important components of data science. Data science is not just advanced information technology or programming. From a more macroscopic perspective, the activity throughout data life-cycle is data management (see Fig. 1), including data collection, preservation, organization, analysis and the dissemination of results for data-based decision making. Data management is a key component of data practices besides information technology. As shown in the review the definition of information science and data science in Wikipedia, we iSchools members will be more confident of strengthening the relationship between data science and information science. Information science is an interdisciplinary field primarily concerned with the analysis, collection, classification, manipulation, storage, retrieval, movement, dissemination, and protection of information [15]. From the definition of information science above we can conclude that information science have abundant research experience in data management and data-cycle related activities, which can be seen as a unique contribution and opportunity for information science professionals to promote data science.

![Figure 1: Life cycle of data management.](image-url)
vious. At the same time information science is famous for its absorptive capacity of latest information technology. Although information science cannot surpass the computer science on information technology frontier research and development, it is effective in absorbing information technology and applying it into daily practice. Unlike other disciplines, information science has shown a close relationship with information technology since its inception. Information science is often mistakenly considered a branch of computer science, actually they are different. At present, there are no other disciplines that pay more attention to the application of information technology to solve problem than information science. The unique advantage of information science in absorbing information technology and applying it into practice makes it possible to better integrate information technology with big data problems.

### 3.4 Interdisciplinary Attributes

Data science requires not only information technology, data management capabilities and statistical methods, but also interdisciplinary knowledge to solve various questions in different backgrounds. It is a broad and interdisciplinary field [16]. A data scientist needs to handle many other matters in addition to programming and statistics. Among the existing data science education programs in America, more than half are offered by joint departments, which reflects the inter-disciplinary characteristics of data science. Besides close relationship with computer science, information science predates computer science and is a broad, interdisciplinary field (see Fig. 2), incorporating not only computer science and statistics, but also diverse fields such as archival science, cognitive science, commerce, communications, law, library science, museology, management, mathematics, philosophy, public policy, and other social sciences [17]. IS research has been alternately praised and criticized for being too interdisciplinary, but we believe this is a strength rather than a weakness in today’s data rich environment. IS scholars have invoked theories from economics, sociology, psychology, and political science, and studied phenomena such as electronic markets, consumer behavior, crowdsourcing, information security, and online retailing from a diversity of perspectives. This interdisciplinary nature of Information science makes it uniquely to exploit opportunities in the big data era [18].

![Figure 2: Interdisciplinary field of Information Science.](image)

### 4 CONCLUSIONS AND SUGGESTIONS

Big data and data science put forward new practical problems for scholars and practitioners to solve out. There is no single discipline can answer all the questions raised by data science, and each of the disciplines mentioned above needs to make their own unique contribution and provide new knowledge or technology to solve practical data problems.

For iSchools, a clear understanding of our unique contributions and deficiency is important in the data era. Information science starts off with a comparative advantage towards big data. Data thinking, data management, absorptive capacity of information technology and interdisciplinary attributes are the unique contributions and opportunities for information science to data science. Information science is one solid foundation from which big data science evolved and data science is not the same thing as information science [16]. The increasing popularity of data science must present us with opportunity. We should take in the idea of data science, acknowledging what we can offer and admitting what we lack.

It is difficult to predict future research and innovation of data science at this early stage, but the reform of information science education towards data science education is necessary and urgent because we are just what we teach. The knowledge map for data science includes data mining, machine learning, statistics, data management, knowledge discovery, computing, programming, system analysis and design, project management, communication and decision support. Even though information science has unique opportunities, we still have much to do. We need to overview the current curriculum of information science and update them to introduce more topics such as machine learning, programming, predictive modeling and statistics methods and tools that are currently closely associated with data science and the key capability for data scientist. We should focus on data analytics and conduct different kinds of programs to train specialists who have a deep understanding of data processing related analytics tasks.

For iSchools, the relationship between data science and information science is beyond question. The unique contributions and shortages are also clear. What we need to do next is taking action in our education reform in order to grab more chances than other disciplines in the data era.

### ACKNOWLEDGMENTS

This work is partially supported by the National Social Science Fund Major Project of China (grant number 15ZDB126)

### REFERENCES


[8] Yao X (2015) Big Data Analytics, Data Science and the CIS. IEEE Computational Intelligence Magazine. 10: 4-5. DOI:http://dx.doi.org/10.1109/MCI.2014.2369875


