Development of Data Science : the Advantages and Challenges of iSchool

Zhou Xiaoying\textsuperscript{1}, Feng Xiangmei\textsuperscript{2}

\textsuperscript{1}School of Information Resources Management, Renmin University, China, xyz-ruc@qq.com
\textsuperscript{2}School of Information Resources Management, Renmin University, China, fexiangmei@163.com

ABSTRACT

It is briefly discussed the advantages and challenges of iSchools in the development of data science from three aspects, the coverage of related disciplines, the "data - information - knowledge" chain and the data literacy education. Finally it is showed that interdisciplinary, the accumulation of theory and practice and international cooperation model are the main advantages.

CCS CONCEPTS

• Computing education \textuparrow Information science education; Computing education programs \textdownarrow Computing literacy

KEYWORDS

Data Science; "data - information - knowledge" chain; data literacy; interdisciplinary

ACM Reference format:

1 INTRODUCTION

We focus on the advantages and challenges of iSchool in the course of data science development from three parts, the coverage of Disciplines, the "data - information - knowledge" chain and the data literacy education.

2 Advantages and Challenges of iSchools in the Coverage of Disciplines

As early as in 1966, "Data Science" is created by the Danish Turing Award winner Peter Naur in the terms of datalogy, and he studied the practical technical issues of automated data processing datamatics, he published "Data Science: Data and Data Processing Science and Its Role in Education" in the International Federation of Information Processing (IFIP) Conference Proceedings in 1968. In 1966, the International Council of Scientific Unions (ISCU) established the Scientific Data Council (CODATA) to facilitate the evaluation, management and use of global scientific and technical data. In 1993, Prof. Chikio Hayashi of the Institute of Statistical Mathematics of Japan, presented and discussed the concept of data science and proposed that data science is to explain the complex nature of human and social phenomena, and the underlying structure of thought using data at the fourth and fifth biennial International Federation of Taxonomies (IFCS) in Paris. Since then, the international scientific conferences on data science and systematic research have begun to emerge. Until 2002, CODATA founded the first official international academic journal on data science, Data Science Journal, marking the formal birth of Data Science\textsuperscript{[1]}. As the arrival of the era of big data, the data collection and application has become the common social phenomenon, and the research of data science is paid more and more attention by the scientific community.

From the outset, data science is an interdisciplinary research field, including data collection, analysis, modeling, application and other issues, the community large-scale data sharing and universal application is related to the supporting action of data management and service measures, such as data infrastructure, data policy, data standards, data storage, data management, data publishing and data services, in order to ensure the data use and to realize the data value.

As the data collection, analysis, modeling and application of the scientific activities popularization, there will appear the first situation, the subdivision of the aspects of the work, including collection, analysis, modeling and application, form different industries and fields; the second situation, is the data protection mechanisms need to follow up. Today, the effective combination of numerable and variation data sources, data mining technologies and tools, has enabled us to large-scale understand and study the nature and society, but these research spans multiple disciplines and involves many humanities, including data analysis, modeling and algorithm of the mathematics, statistics, computer science, data collection and transmission of the communication science, data infrastructure construction of engineering science, data policy and data Management of the humanities and social sciences, data organization, management, publishing and services of the library and information science.

The members of iSchools have a strong interdisciplinary structure, which is consistent with the strong interdisciplinary nature of the data sciences. iSchools are not established for a long time. However, unlike the simple disciplines of physics, chemistry and economics, the interdisciplinary research perspective is rooted in the teaching and research from the beginning. It is similar to solve the data problem with
the information before, interdisciplinary research is the advantage, for example, the study on the graduates of Syracuse University shows that graduates have grown into successful people based on the fusion information, information technology and management knowledge they provide[2]. The integration of disciplines or the interdisciplinary cooperation of research projects can reflect the advantages of solving information or data problems. Therefore, the experience accumulated by ischool in teaching and research can be applied to the research of data science. Multidisciplinary integration is still a way to face the data science development challenges, each subject has its own challenges to solve, but the combination of points in cross disciplinary fields to solve practical problems is a way of thinking, the result of combined may remain Union Forms such as iSchool, or may be integrated to form a new discipline or may form a new branch of the discipline.

Therefore, for iSchools, the first advantage is, the domain iSchool members contains is quite interdisciplinary with the disciplines mentioned above, such as computer science, communication science, humanities and social sciences, library and information science, So iSchool on data science research has unique conditions. The challenge is, to integrate interdisciplinary research efforts, to develop their interdisciplinary expertise in data science research and teaching, and to achieve the leading position in the data science domain at the combination point of the data-technology-human interface.

3 The Advantages and Challenges of iSchool in the "Data - Information - Knowledge" Chain

Many scholars have argued that a basic goal of data science is to insight from the data, and then to extract information and knowledge, and it is proved in practice. For example, US government launched, data-open service-led BD2K initiative (Big Data to Knowledge initiative), is committed to the use of biomedical data and to transform data into new knowledge, to promote the development of biomedical science[3].

In the field of information science, scholars have generally accepted the "data - information - knowledge - wisdom" theory, a chain of cognitive process. According to this study, there is information intermined from the data to the knowledge level, in fact, extracting knowledge from the data need two steps, the first is to identify the meaning of the data and to get information from it, the second is to combine the information with the prior knowledge of the human and to get new knowledge.

The "data" in Data Science was originally a member of the information family, is a special type of information. iSchool's research and teaching based on the relationship between information, technology, and people, obviously need to develop corresponding measures according to the rapid development of data science, and need to know their own strengths and expertise to develop their own strengths, to identify the location in the development process of scientific data in iSchool.

The links between people and knowledge are the basis for personal and social development, and iSchools’ key members are committed to foster information professionals who are able to establish connections between people and knowledge, as those identified by the University of Washington iSchool colleage, LIS and iSchools’ other disciplines should be committed to build an information link between people and knowledge, in order to achieve people’s access to information. The Information school of the University of Florida is committed to establish the vital connection between people and information[4]. However the information link between people and knowledge in the content and form is evolving, in the past we set up this link through the library, and later through the Internet, the future we will do through the data infrastructure, so these make disciplines deal with professional challenges.

Thus, for iSchools, the second advantage is that, from data to information, from information to knowledge, and from knowledge to intelligence, have always been researching and teaching courses on library science and information science of the iSchool members, the accumulation of theory and practice will continue to be applied and promoted in the development of data science. As early as 1945, the famous American scholar Vannevar Bush has proposed scientific research goal, which is to make full use of the growing knowledge, this is the goal of scientific development after the middle of the last century, but is also the cause of the information science born and the basic purpose of the discipline. Data is a special kind of information. One of the tasks of data science is to solve the problems of obtaining knowledge from data. The data contains useful knowledge, which is the precondition of the application value of big data. LIS disciplines advocate for all levels of society to provide professional information services can be extended to the purpose of professional data services in the era of big data. The storage, sorting and mining of knowledge, and the way of acquiring knowledge for the users through improving literacy and direct help, all reflect the knowledge complex of iSchool and the humanistic thoughts of social concern[5]. Challenges at this level are as follows, there are a number of participants in the development of data sciences, and LIS disciplines need to deepen and expand traditional knowledge from data integration and understanding to data perceptions and interactions to data learning and cognition, Such as data integration, data quality assurance, data standards, data annotation and other fields, in order to adapt to the development of data science requirements.

4 Advantages and Challenges of iSchools’ International Cooperation in Data Literacy Promotion

Information literacy has always been the hotspot of iSchool, and it is included in the nine research focuses summarized by teachers' research projects and research interests in 2007, others are information system, interdisciplinary information science, digital technology, Knowledge dissemination, information sociology, information technology, information retrieval, library services[6].

In practice, public libraries and university libraries have been engaged in the education of information literacy. Many countries, like United States, and some international organizations have developed information literacy standards. Some libraries have specialized
information literacy training programs. From the perspective of global collaboration, the library is an institution that exists in almost every city in the world. The global libraries have natural connectivity and collaboration, and have international organizations, such as the International Federation of Library Associations (IFLA). The data literacy training in the process of data science education, can make use of the accumulated experience of information literacy education and professional education, and carry out training and education of data literacy.

Therefore, the third advantage of iSchools is that they can make use of the accumulated experience and international cooperation model of information literacy education, carry out global data literacy education, enhance the society's awareness and perception of data as a whole, and can also make global collaboration in many aspects such as data technology popularize, data tools provision, data standards implementation, available data preservation and management and so on. The challenge is how to differentiate between information literacy and data literacy, how to develop a reasonable standard for data literacy, and how to develop a global data literacy education for different data needs and applications.

5 CONCLUSION

Compared to other Subject alliance, iSchools have good discipline composition, a long-term experience of extracting knowledge from data and a natural advantage in global data literacy training and international cooperation in data science research and teaching. Overcoming the above-mentioned three challenges posed by the development of data science, iSchools should be able to lead the way in interdisciplinary research in data sciences, effectively extracting knowledge from data, and data literacy training.

REFERENCES