



## A cross-institutional analysis of data-related curricula in information science programs: A focused look at the iSchools

Journal:	<i>Journal of Information Science</i>
Manuscript ID	JIS-17-0192.R1
Manuscript Type:	Research Paper
Keywords:	Data science, Big data, Data analytics, Digital curation, Digital preservation, iSchools, Information science, Education
Abstract:	<p>Our rapidly growing, data-driven culture is motivating curriculum change in nearly every discipline, not the least of which is information science. This paper explores this change specifically within the iSchool community, in which information science is a major unifying discipline. A cross-institutional analysis of data related curricula was conducted across 65 iSchools. Results show that a majority of iSchools examined (37 out of 65, 56.9%) currently offer some form of data related education, particularly at the master's level, and that approximately 15% of their formal degree offerings have a data focus. Overall, iSchools have a greater emphasis on data science and big data analytics, with only a few programs providing focused curricula in the area of digital curation. Recommendations are made for iSchools to leverage the interdisciplinary nature of information science, publish curricula, and track graduate success so that iSchools may excel in educating information professionals in the data area. Future data education in iSchools may benefit from further interdisciplinary data education, including data curation curricula.</p>
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p>	
Fig_4.mht	

SCHOLARONE™  
Manuscripts

# A cross-institutional analysis of data-related curricula in information science programs: A focused look at the iSchools

Journal of Information Science

1–14

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

jis.sagepub.com



## 1. Introduction

The need for a data competent workforce is impacting nearly every discipline [1]. iSchools, a community of institutions that aim to lead education and research in information science, are embracing this challenge. The iSchool community's attention to data-focused education is not surprising, as core processes of information science (collecting, organizing, managing, accessing, and supporting the use and manipulation of information) are acutely relevant to data driven disciplinary areas such as data science. In terms of curriculum development, these core processes place information science and the iSchool community able to develop and advance distinct expertise meeting data-driven needs.

Clearly, the opportunity for iSchools in the growing data driven area is exciting, and information scientists find themselves working on interdisciplinary teams that previously were unimaginable. The iSchool community also faces challenges in the data education space, particularly in developing appropriate methods that distinguish, complement, and support other data focused academic curricula, such as medical and bio-informatics, business analytics, and digital humanities. These challenges, and the overall goals of higher education, point to a need to assess how the field of information science and the iSchool community are, in fact, contributing to producing a knowledgeable, skilled, competent workforce in the data area. Here, we ask, *how are iSchools responding to the need for a data competent workforce?*

This question guided the research presented in this paper, specifically the results of a cross-institutional analysis of data related offerings in iSchools. This work focused on iSchools as a global information science community that has unifying goals. The research considered the full range of data oriented courses and program offerings across iSchools. An overriding goal of this research is to gain insight into the extent of data related curricular activities within information science, particularly with respect to iSchools, as a growing, global community.

In sections below, we provide background context on iSchools and review relevant data-oriented disciplinary trends. Next, we present our research objectives, methodology and procedures, followed by a report of the results integrated with discussion. We conclude with a summary of key findings, make recommendations for iSchools to collectively advance in the data area, consider research implications, and identify next steps.

## 2. The iSchool context

iSchools, originating in 2005, represent an international consortium of schools dedicated to advancing the information field [2]. Information science is the unifying discipline across the iSchools, drawing together programs with strong foundations in documentation, library science, archives, informatics, and, in a number of cases, operations research and business analysis [3]. Today, iSchools encompass a broader expanse of specialties, including archival science, informatics, human and social computing, business intelligence, and computer science. Although these areas have different foci independently, as part of the iSchool consortium they share fundamental interest in the relationships among the facets of *information, people and technology*.

Today's increasingly data driven culture interconnects with these facets. Furthermore, the iSchool community's focus on the relationship between information, people, and technology helps explain why data related curricula across iSchool programs are pursuing a number of different data-related emphases. The emerging diversity of curricula in this area was reflected recently in the work of Varvel, Bammerlin & Palmer's [4]. The authors analyzed 475 courses on data curation, data science, and data management in 158 programs at 55 iSchools and other LIS schools. Although only 13.3% of programs at 17 institutions (27% of iSchools in the larger sample) qualified as "data centric", the programs covered master's degree curricula, certificate programs, and separate concentrations. These researchers also reported that many iSchools had interest in revising their curricula to encompass data and that "existing digital library and archives courses are contributing the most data oriented content, covering key areas such as metadata and digital preservation."

Si, Zhuang, Xing and Guo's [5] web content analysis sheds further light on this topic. These researchers reported data related programs and course offerings at 38 iSchools. They observed an increase, in that 25 (65.78%) of schools in their sample offered data related courses covering topics such as systems analysis related to digital content, data processing, digital storage and management, data visualization, data curation, digital preservation, among others. Song and Zhu's (2015) more recent work took a different approach, exploring beyond iSchools, analyzing 42 bachelor's and master's data science programs in the U. S.; they found that data science master's programs were more prevalent than the bachelor's programs. Perhaps surprisingly, they also observed that the largest number of these programs were offered in information science departments, followed by computer science and statistics. Tang and Sae-Lim's [6] work is similar in that they focused on data science graduate programs in the U.S. across eight disciplines, including information science as represented by iSchools; although they pursued a stratified random sample. They reported that "iSchools stood out among eight disciplines for delivering multi-level analytical skills."

The studies reviewed above show a growing interest in data education in programs grounded in information science, and prompt us to raise questions about the uniqueness and competitiveness of data education in information science, including iSchools. Given the complexities and multidisciplinary nature of data curricula, and the goal of assessing iSchools' activity in this area, it is important to first review how 'data' is understood in the iSchools/information science domain, as well as disciplinary trends.

### 3. Data and iSchool trends

Data is a simple, yet complex word. Borgman [7] emphasizes that the only agreement in defining data is that there is no single definition for data, explaining that we can reach more agreement by viewing data as "representations of observations, objects, or other entities used as evidence of phenomena for the purpose of research or scholarship." Even with this fluid definition, introductory information science courses frequently begin with the data pyramid, with data as the base level, followed by the layers of information, knowledge, and the pinnacle of understanding, and even wisdom [8, 9]. This pyramid has provided a foundation for educating information professionals for decades, and is at the core of many information science programs; although new trends in the data area point to a data-focused rubric for study that includes *data science*, *big data/data analytics*, and *digital curation and preservation*. While these trends are not mutually exclusive, they reflect the chief themes and nomenclature in iSchools, and form a useful framework for examining data driven curricula. To further aid our analysis, we review each of these areas in the subsections that follow.

#### 3.1. Data science

##### 3.1.1. Defined

Data science is generally understood to be the computational and quantitative analysis of large datasets to create information and knowledge; the science stems from the use of methodological frameworks, processes, and tools used to analyze data and derive insight. Stanton [10] defines data science as "an area of work concerned with the collection, preparation, analysis, visualization, management, and preservation of large collections of information," and explains further that it is much more than simply analyzing data. Granville [11] promotes this point, and compares data science to several overlapping, analytical disciplines such as computer science, statistics, machine learning and data mining, operations research and business intelligence. Nielsen & Hjørland [12] have a similar view, and view data science as an analytical discipline that includes integration with business intelligence, artificial intelligence, computer science, econometrics, and engineering.

Data science, as a discipline, has gained recognition in the scholarly and scientific literature, as well as the popular press. There are noted national and international reports projecting workforce needs in data science [13], identifying the data scientist's skill set, and promoting data science as one of the best, even sexiest, professional disciplines [14]. Further, the Data Science Association (DSA) [15] explains that a data scientist is someone who uses "scientific methods to liberate and create meaning from raw data - somebody who can play with data, spot trends and learn truths few others know." Data scientists are trained in forming and asking questions, and understand how data can be leveraged to predict outcomes.

##### 3.1.2. Data science and the iSchools' embrace

In the iSchool community, Stanton [16], stands as a leader advocating how information science, as a discipline, prepares the data scientists to communicate with data users. The skills that Stanton promotes underscore how tasks such as the "collection, preparation, analysis, visualization, management, and preservation of large collections of information" are

information science driven **tasks key** to the data science domain and the ability to understand the big picture within a complex system. This viewpoint reflects the iSchools' emphasis on the information lifecycle [17]. These findings help explain why iSchools are among the array of institutions supporting data science curricula; and also interconnect to Big data and data analytics, as reviewed in the next section.

### 3.2. Big data and data analytics

#### 3.2.1. Defined

Big data and data analytics, together, reflect another major data trend. The terms 'big data' and 'data analytics' are used somewhat interchangeably, with big data being the more general concept, and the analytics end focusing on actual analytical approaches. These terms interconnect with data science, in that big data is the 'raw material' or a key source for conducting data science. Big data characteristics were identified in 2001 by Laney as *volume*, *velocity*, and *variety* [18]. The original three Vs have evolved to five Vs adding *veracity* and *value* [17, 19].

Big data is "a field dedicated to the analysis, processing, and storage of large collections of data that frequently originate from disparate sources" [20]. Big data initiatives leverage big data sources to solve business related problems, among others. The Internet of Things (IoT) presents new opportunities for analysis and discovery, and pursuing market-related prediction, and explains the push for advanced analytics within industry, academia, and government [21]. 'Data analytics' is prominent in big data discussion, particularly with an emphasis on analysis.

#### 3.2.2. Big data, data analytics and iSchool embrace

As discussed above, many iSchools have foundations in library and information science, although there are **several** schools with strong foundations in operations research or linkages with business oriented information programs. Examples include Singapore Management University: School of Information Systems and Carnegie Mellon University: School of Information Systems and Management, Heinz College. These developments, and the fact that iSchool faculty collaborate with many disciplines, from the natural and social sciences to the humanities, invite iSchool discussions about big data definitions, and the skills required to make use of big data [5]. Moreover, there is an intellectual debate as to the difference between big data and data science [22]. In fact, in combining the two phrases, the concept of 'big data science' has emerged as part of a meet up group in California [23]. All of this explains why the distinction between big data, data analytics, and data science can seem blurred at time, although a review of the literature in these areas reveals noted differences. One of the most insightful views of the relationship among these concepts (**big data, data analytics, and data science**) is the diagram presented by Song and Zhu[17], reproduced in Figure 1, illustrates a top-down hierarchy with computation, technologies, and the data explosion leading to specialized topics and skills, and resulting in data science.

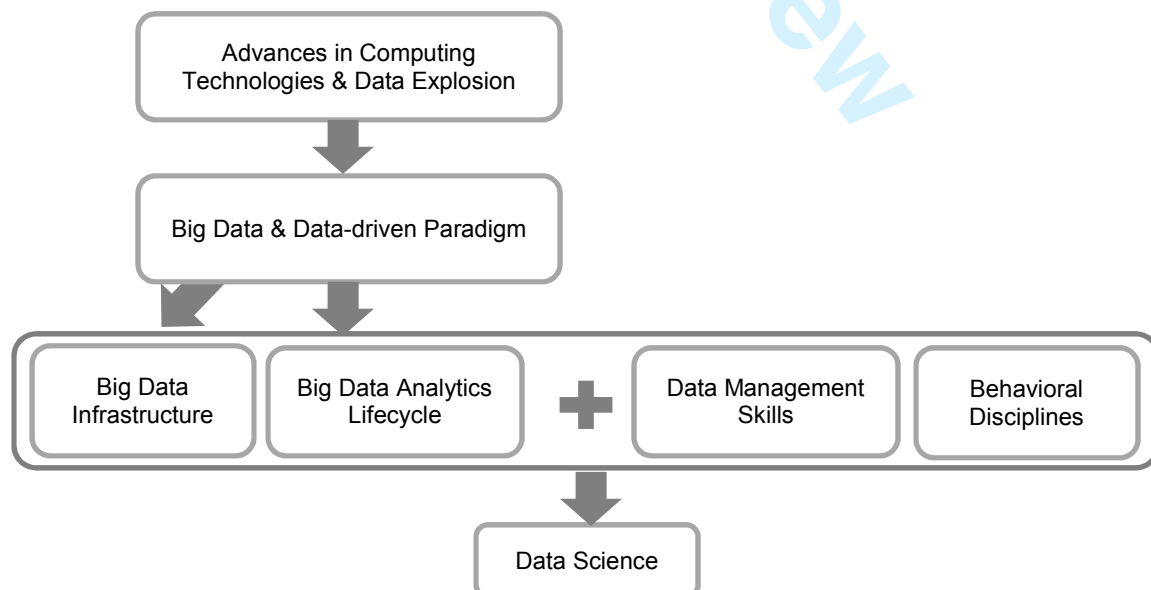


Figure 1. Emergence of data science [17].

### 3.3. Digital curation and preservation

#### 3.3.1. Defined

Digital curation, according to the Digital Curation Center (DCC) involves maintaining, preserving and adding value to digital research data throughout its lifecycle. Curation generally includes data preservation, with an emphasis on the long-term value of data and efforts to make data available for further high quality research [24, 25].

Digital curation encompasses lifecycle management of digital assets [26]. Drawing from the DCC model, the process of curation begins with conceptualization and creating or receiving digital items, and involves various processes, such as appraisal, integrate into a system, preservation, and preparing a digital resource for access, use, and reuse.

As for the terms big data, data analytics, and data science, the terms *data curation*, *digital curation*, and *digital or data preservation* are often used interchangeably. This interplay of terms reflects the fact that they are often used simultaneously. Gold [27] provides insight here, indicating that “data curation” and “digital curation” may appear to be interchangeable terms, even if these terms refer to related but distinct concepts. More precisely, “digital curation” is used for the curation of digital objects including compound digital objects; with “data curation” for the curation of records or measurements of information (“data”).

#### 3.3.2. Digital curation, preservation and iSchool embrace

Curation and preservation are a part of traditional library and archival science programs, making the digital aspect an easy add-on. The preservation emphasis connects more with programs with strong links to historiography, such as King’s College in the UK. In this area, Walters & Skinner [28] observe that different groups of scholars and their library partners have evolved some terminological differences; however, they acknowledge that the nuances can obscure the fundamental unity in libraries’ roles and services relating to digital curation and preservation. The literature also reveals a host of related descriptors to define individuals in these professions, such as *data humanist* [29] *science librarians* [30], *data librarian* [31], *librarians/science informationists* [32], *information specialists* [33], and *research informationist* [34]. This heterogeneity is reflected in job profiles, roles and responsibilities, and even within types of institutions. A case in point is the ‘data scientist’ job profile in Europe, for which there is no common definition [35]. The diversity and heterogeneity reviewed here are reflected in the different career structures across institutions and within the iSchool community.

### 3.4. Summary: iSchools and the data area

Our review presented above is structured around the themes of data science, big data, and data curation and preservation. This rubric was developed based on extensive review of the literature and of the iSchool programs in the data area. These themes not only represent the most consistent nomenclature across data focused programs in iSchools, but are part of the *lingua franca* discussion at professional conferences. Moreover, these themes were confirmed by earlier work of Song & Zhu [17]. The research that follows is, thus, framed by these areas to aid in the presentation of the analysis and discussion.

## 4. Research Objectives

The iSchool community, like many other interdisciplinary communities, has been pursuing curriculum change to address data driven workforce needs. Although this trend is apparent, reporting on the extent of curriculum change is limited.

We are at a juncture where it is timely to examine how data is being addressed in information science education, and specifically in the iSchools community. Our research is motivated by the following questions:

- What is the extent of data related iSchool curricula activities? and
- What data driven emphases and foci are found in iSchools?



## 5. Methodology

To pursue our research questions, we conducted a cross-institutional analysis of iSchools, including a cluster analysis of courses offerings. This method is appropriate for gathering data on the extent of data related curricula [4–6, 36]. Our data source was the iSchool Directory, representing a consortium of leading programs joined by the common theme of information science. The data is based on the official roster of 65 iSchools released in February 2016, and reflect active and approved curricula for the period of 2015-2017. We also targeted iSchools given the international scope and breadth of information disciplines by the members. The individual iSchool websites served as our data source.

Our procedures included the following steps:

- (1) Consult the iSchools Directory [37] to identify and locate each of the iSchools' institutional websites.
- (2) Identify the data-oriented degrees, specializations, and certificates for undergraduates, graduates, and PhDs in data areas for each iSchool, drawing from accessible digital documentation. Data was manually collected and put into an Excel PivotTable for analysis.
- (3) Classify the degree programs by country, type of degree, discipline, and concentration using a coding scheme under the rubric of data science, big data analytics, and digital curation. For this step, inter-coder reliability was independently reviewed and verified by an iSchool faculty member.
- (4) Assess data related courses associated with existing data degree programs. Two types of data related courses were identified: courses offered in data focused degrees and courses offered in non-data focused degrees. As part of this step, institutional course catalogs were examined to determine degree requirements. The courses were tallied with a cluster analysis, based on course descriptions and titles. That approach involved three processes:
  - Courses were organized according to our rubric of data science, big analytics, and digital curation. We sought to adhere to the nomenclature reflected in the literature and degree designations.
  - Subject matter of each course was assessed by examining the course title and content in syllabi and course descriptions.
  - Topical labels were appended to each course, normalizing course titles and grouping courses into 2<sup>nd</sup>-level clusters. We verified terminology by consulting the ASIST thesaurus, Library of Congress Subject Headings, further literature review, and consultation with Drexel University colleagues who have pursued research in the data area. The data was then an Excel Pivot Table, establishing our dataset for further analysis.
- (5) Tally data related courses that were offered in non-data focused degrees, for separate processing and analyzes.

## 6. Data analysis and discussion

### 6.1. iSchools' degree programs and disciplines

At the time of data gathering, a total of 597 degrees of study were found across the 65 iSchools analyzed. As shown in Figure 3, master's degrees predominate (53%), followed by bachelor's degrees (23%) and doctoral degrees (13%). These schools also offer a series of certificates.

[insert Figure 2]

Figure 2. 597 iSchools' degree programs.

The key disciplines involved in the delivery of these 597 degrees across iSchools are presented in Figure 4. The results show that LIS, LIS & informatics (combined), and informatics & computer science predominate. This result is not surprising, given that the iSchool movement was initiated by three universities in the U.S. in 1998 with foundations in library and information science and informatics [2]. Figure 4 also confirms that interdisciplinary approaches are a hallmark of the iSchools.

[insert Figure 3]

Figure 3. **Most relevant disciplines involved in iSchools' degree programs.**

### 6.2. *iSchools' data related degrees*

We identified a total of 87 degrees (14.6% of the 597 degrees) data related to data, either because they are data specific or because the degree includes data related courses. Figure 5 shows a) degrees which represent all the degrees across all areas of study, b) data degrees, c) degrees with data courses, and d) degrees without data courses. The 87 data related degrees were offered by 37 iSchools, such that 56.9% of the iSchools offer some level of data related education. On the one hand, we expected that a higher percentage of iSchools were engaged in the data area, given apparent emphases at the iSchool conference and scientific and scholarly venues in which iSchool faculty disseminate their research. On the other hand, the percentage is not surprising, given the wide diversity of **curricula in** iSchools, ranging from public libraries and cultural centers to mathematics and computationally intensive endeavors.

[insert Figure 4]

Figure 4. iSchools' degree programs and data education (597 degrees).

A closer examination of the 87 data related degrees (see Figure 6, top-portion) shows **that a third of these degrees (26 of 87, 30 %) are data specific.** The remaining roughly two-thirds of these degrees (Figure 6, lower-half) require 1 to 3 data related courses and are offered primarily at the master's level.

[insert Figure 5]

Figure 5. 87 Data related degrees.

The United States (U.S.) has the largest concentration of data focused degrees among the iSchools today (see Figure 7). These results reflect the fact that the U.S. has the **greatest** number of iSchools. The number of data courses offered by three Portuguese iSchools appears high compared to other countries.

[insert Figure 6]

Figure 6. 87 data related degrees by country.

### 6.3. *iSchools' data focused degrees*

Almost half the degrees (46.15 %, 12 of the 26 data specific degrees) address data science, while five (19.23 %) are related to big data analytics and nine (34.61 %) address digital curation (see Figure 8). Additionally, iSchools offer degrees across all levels of higher education from the undergraduate to the doctoral level, but in the three areas of our rubric, most data degrees (65%) are at the masters' level (see Figure 8).

[insert Figure 7]

**Figure 7.** 26 Data degrees by subject area.

As shown in Figure 9, English speaking countries (the United States, United Kingdom, and Australia), and particularly the United States, have the greatest number of degrees in data science and big data analytics (11 out of 17, or 64%). Interestingly, other degrees in data science and big data analytics observed in non-English speaking countries are delivered in English. Digital curation appears to be fairly equally distributed in the UK and the United States.

[insert Figure 8]

**Figure 8.** 26 Data degrees by country.

As shown in Figure 10, the majority of data education efforts are in the disciplines of LIS & informatics combined, and most degrees are interdisciplinary and include informatics, statistics, business, and management.

[insert Figure 9]

**Figure 9.** Disciplines involved in data degrees.

In order to gain more insight into the cross-disciplinary activity, we examined the curricula in greater detail, looking at data-points such as degrees level, offering opportunity (online or face-to-face), and date of implementation.

#### 6.4. Basic comparison of *iSchools'* data degrees

Given the prevalence of data in our information culture, we anticipated impacts across all degree areas. Table 1 provides an overall picture of schools with degrees in data science, big data analytics, and digital curation, for different degree levels.

Within our sample, we identified, in data science, two bachelor's degrees, seven masters', two graduate certificates/diplomas, and one PhD program. All were begun between 2014 and 2016. The duration of these degrees varies from 6 months for the graduate certificates to four years of the bachelor's. Master's degree programs range from one to two years. Only one of the master's degree programs is fully online, while the others are face to face or are considering a mixture of face-to-face and online courses.

In the big data analytics area, we identified five master's degrees, all them were initiated between 2014 and 2015. The duration of these degrees is from 1 to 2 years. Two are face to face and three use mixed methods. In digital curation, there are five master's degrees, two graduate certificates, one postmaster, and one PhD All began between 2013 and 2016. Four are offered on campus, two online, and other use both online and face to face approaches.

As anticipated, the disciplines involved are primarily LIS & Informatics. The main component in digital curation is library and archival science; although in data science and big data analytics the chief component is informatics, drawing also from computer science, business and statistics. As a result, digital curation degrees are less interdisciplinary than the others.

At the time (January 2016) of data collection, only the University of Illinois reported offering two master's degrees in both disciplines: A Master's in Big Data and a Master's in Digital Curation. Note also that several master's programs offer different specializations or use the same subjects to offer graduate certificates or diplomas. An example of this



approach is Indiana University, with a Master's in Data Science Specialization, Master of Information Science with a Data Science Specialization, and a Master of Library Science with a Data Science Specialization.

### 6.5. What is being taught?

The cluster analysis examining course offerings in the three areas of our rubric. Some overlap was found in clusters of courses related to data science degrees and big data analytics, which allowed us to more clearly identify similarities and differences among course offerings. Because no significant overlap was found between digital curation degrees and data science or big data analytics degrees, we present the results of our clustering analysis separately for this case.

#### 6.5.1. Data science and big data analytics: similarities

Degrees in data science have a total of 161 courses offerings, and big data analytics degrees have a total of 81 courses offerings. These offerings were categorized into 50 and 41 clusters, respectively. Table 2 shows the overlap found for 20 of these clusters (22 % cluster overlap). The similarities noted represent 57.14 % of the total number of courses in data science (92 courses), and 71.60 % of the total number of courses in big data analytics (58 courses) respectively.

**Table1.** iSchools' data degrees.

Degrees	B	M	GC	GD	PhD
<b>DATA SCIENCE</b>					
Drexel University, College of Computing & Informatics: BS in Data Science	X				
Indiana University, School of Informatics and Computing: Master in Data Science		X			
Indiana University, School of Informatics and Computing: Master of Information Science. Specialization in Data Science		X			
Indiana University, School of Informatics and Computing: Master of Library Science. Specialization in Data Science		X			
University of Boras, The Swedish School of Library and Information Science: Strategic Research Program in Data Science					X
University of California-Berkeley, School of Information: Master of Information and Data Science (MIDS)		X			
University of California-Irvine, The Donald Bren School of Information and Computer Sciences: BS in Data Science	X				
University of Sheffield, Information School: MSc in Data Science		X			
University of South Australia, School of Information Technology & Mathematical Sciences: Graduate Certificate in Data Science			X		
University of South Australia, School of Information Technology & Mathematical Sciences: Graduate Diploma in Data Science				X	
University of South Australia, School of Information Technology & Mathematical Sciences: Master of Data Science		X			
University of Washington, UW Information School: Master of Science in Information Management, Data Science Specialization		X			
<b>BIG DATA ANALYTICS</b>					
Georgia Institute of Technology, College of Computing-Atlanta: MS Analytics		X			
University of Illinois, Graduate School of Library and Information Science: MS Specialization in Socio-technical Data Analytics		X			
University of Maryland, College of Information Studies: Master of Information Management Specialization in Data Analytics		X			
University of Pittsburgh, School of Information Sciences: Big Data Analytics		X			
University of Tampere: Degree Programme in Computational Big Data Analytics		X			
<b>DIGITAL CURATION</b>					
Humboldt University in cooperation with King's College: Joint Study Profile Master in "Digital Curation"		X			
Robert Gordon University, Department of Information Management of Aberdeen Business School: MSc in Digital Curation		X			
Robert Gordon University, Department of Information Management of Aberdeen Business School: PhD in Digital Curation					X

Simmons College, School of Library and Information Science: Digital Stewardship Certificate		X
University College Dublin, School of Information and Library Studies: MSc in Digital Curation	X	
University College Dublin, School of Information and Library Studies: Digital Curation Graduate Certificate (Continuing Professional Development)		X
University of Illinois, Graduate School of Library and Information Science: MS Specialization in Data Curation	X	
University of Maryland, College of Information Studies: Master of Information Management Specialization: Archives and Digital Curation	X	
University of North Carolina at Chapel Hill, School of Information and Library Science: Post-Master's Certificate in Information and Library Science (PMC)		X

Legend: B=Bachelor's; M=Master's; GC=Graduate course; GD=Graduate Diploma; PhD=Doctoral degree.

**Table 2.** Overlap analysis of data science and big data analytics courses.

Area	N. of degrees	Courses	Clusters	Cluster overlap	Course overlap
DS	12	161	50	20 (22%)	92 (57.14%)
BDA	5	81	41		58 (71.60%)

The analysis of identified clusters shown in Figure 11 suggests that, within the iSchools, data science courses are covering a larger spectrum of subjects compared to big data analytics. These results align with the findings of Song and Zhu's [17], showing that big data analytics is a part of data science, which in our case is reflected in the greater course overlap observed for big data analytics.

[insert Figure 10]

**Figure 10.** Similarities among Data Science (DS) and Big Data Analytics (BDA) degree courses.

The relationships found among courses in each cluster in a symmetrical bivariate graph, in which the ball size represents the degree of cluster overlap and the position of the balls represents the predominance of BDA or DS in the overlap. The bivariate data stem from simultaneous observation of two variables (X, Y) in a sample of 20 cluster overlaps. The X and Y axes represent the percentages of the DS and BDA respectively. The balls colors represent red for BDA, and green for DS, the blue ones show the highest coincidences.

Databases, information visualization, machine learning, and information retrieval are among the largest clusters with the most overlap in our sample; and library and information science and security technology are among the smallest clusters with overlap. Statistics, data analysis, information systems and big data are more prominent in data science; whereas data mining, probability and statistics, data analytics, and computing are predominately in big data analytics. While observing these differences, we need to consider the relatively small number of degrees examined. The similarities are not surprising, given that these areas are fundamental to both data science and big data analytics.

#### 6.5.2. Data science and big data analytics: differences

Figures 12 and 13 show clusters based upon the percentage of courses in data science and big data analytics. Within the iSchools context Data science has 29 unique clusters (68 courses that are not in big data analytics degrees), while big data analytics has 20 unique clusters (23 courses that were not in data science degrees).

[insert Figure 11]

**Figure 11.** Differences Data Science versus Big Data Analytics (Data Science degree courses only).

[insert Figure 12]

**Figure 12.** Differences between Data Science and Business Data Analytics (Business Data Analytics degree courses).

Several results stand out in these comparisons. Data science has greater variety than big data analytics, with courses such as the semantic web, metadata, data curation, cognitive science and data management, and ethics. Big data analytics is more homogenous, and has a more business oriented tilt, with courses such as risk theory and data analytics for business. It is somewhat surprising that programming and informatics are represented in data science clusters, although absent from big data analytics. This result possibly reflects the background of students pursuing these two areas of study, with data science likely attracting students from a greater diversity of disciplines [38], and big data analytics attracting a more homogenous group of students who may have already had exposure to programming. Given the novelty of these disciplinary foci in iSchools, beginning in about around 2014, it is premature to know the results of graduates entering this area.

### 6.5.3. Digital Curation

Examining course clusters in digital curation degrees provides a view of the third key data-focused area emerging in iSchools' educational programs. Figure 14 shows the distribution of course clusters for digital curation in a tree map chart, in which the area and color of each cluster is proportional to its frequency of occurrence. Digital curation and digital preservation are offered across all nine degrees in this area. Databases, digital culture, and knowledge management are offered in five of the degrees, while information management, databases, knowledge management, and metadata are offered in four. Less commonly, but still representative of three degrees, are "archives," "social media," "digital curation (technology)," and "digital curation (management)."

[insert Figure 13]

**Figure 13.** Digital curation degree courses

Unexpectedly, in our analysis only 7 iSchools (10.7 % of 65) offer data curation (9 degrees). iSchools would do well to add data curation and digital preservation to their curricula. This recommendation is based on a number of factors, including the importance of data quality and preservation, and the continued growth of open data policies and data sharing requirements mandated by federal funding agencies. Additionally, many journals have data depository requirements tied to publication as part of scholarly communication activity. These policies seek to avoid research duplication and reduce costs of data collection.

Digital curation curricula teach that data must be stored, analyzed, curated, preserved, and disseminated in order to be reusable. Libraries have expertise to perform these tasks and, as Pouchard [39] notes, college and research libraries are re-inventing themselves to respond to these necessities. Academic libraries have a history of embracing new developments supporting researchers' needs, and no-one can deny the importance of data services provided by academic libraries, particularly in light of new policies noted above. These services present a new pathway for librarians and require education specific to this need. It is clear that iSchools have an opportunity to train professionals to meet this growing need and accompanying new roles in the data world. Although many academic libraries are making efforts to prepare their staff to deal with data services, offering formal curricula is key to sustainable success in this area.

### 6.5.4. Data related courses in non-data focused degrees

As noted above, 62 data related courses are offered for degrees that are not data specific - that is, they do not offer full curricula for data or a graduate certificate or specialization. These degrees have at least one data related course and up to three courses dedicated to data. As shown in Table 3, the most popular subject is big data analytics, offered by 24 degrees. Then, data mining and data science, offered by 10 and nine degrees respectively. All of them are introductory, basic level courses. This may be indicative of the increasing importance given by iSchools to data education.

**Table 3.** Distribution of data related courses in non-data degrees.

Cluster	Number of courses
Big Data Analytics	18
Data mining	10
Digital curation	10
Data Science	9

1		
2		
3	Data Analytics	6
4	Web analytics	6
5	Digital Preservation	1
6	Research Data Management	1
7	Visualization	1
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

## 7. Conclusions and recommendations

This paper reports on a cross-institutional analysis of data related curricula supported in 65 iSchools. The study was guided by questions examining the: 1) extent of data related iSchool curricula activities, and 2) data driven emphases and foci in iSchools. The research was framed by disciplinary trends, specifically *data science*, *big data/data analytics*, and *digital curation and preservation*. The iSchools were targeted as the population to study, given that information science is a common, unifying discipline, and due to their international scope.

The analysis confirmed that data related curriculum changes are taking place across all the iSchools, with modifications that range from supporting new data related courses offerings to implementing formalized data-focused degree programs and certificates. The trajectory of data related curriculum growth started to increase in 2013, with the degree change taking place primarily at the master's levels. More recently, change is taking place at the bachelor's level. Doctoral level opportunities are also evident, although data-science curricula are more challenging to track given the independent nature of study at this level, and the emphasis on original research. It is also important to point out that although the iSchool consortium is global, there is greater representation from U.S. and European iSchools, which is likely tied to the U.S. origins of the consortium. The results, based on the sample of iSchools, found that most of data science and big data analytics degrees are being offered by U.S. iSchools (11 out of 17, 64 %), and that digital curation degrees were found to be fairly equally represented in Europe and in the U.S. The study also showed of the 20 schools offering formal degrees in the data area, data science and big data analytics was found in more programs (13 out 20, 65 %), compared to digital curation program (7 out 20, 35 %). These results took into consideration that the distinction between data science and big data analytics in iSchool curricula is not always precise, although data science generally includes a greater diversity of courses and emphases compared to big data analytics.

The questions guiding this study are important as information science programs consider how to best respond to trends and educational needs and prepare the next generation of information professionals. The research questions pursued are reflected in recent initiatives, such as the European Commission supported EDISON Data Science Framework (EDSF) [40] and the European Data Science Academy (EDSA) [41], and well as the U.S. National Science Foundation's commissioned reports, such as, *Realizing the Potential of Data Science: Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group* [42][43] and *Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim* [44]. An in-depth comparison of these initiatives is beyond the scope of this immediate paper, although a high-level review reveals curricula recommendations similar to current iSchool activity in the data space. Common curricula topics include statistics, machine learning, data mining, predictive analytics, big data infrastructure, application design, and several other key themes. More surprising is that these commissioned initiatives fail to identify digital curation, metadata, information retrieval, and knowledge management—all common strongholds of information science; and, thus, underscore the importance of sharing the findings reported on in this paper.

Clearly, the growing ubiquity of data calls for a diversity of data skills; and iSchools along with the discipline of information science can contribute in this need through education and research. The findings reported on in this paper show that iSchools are increasing their data-focused course offerings, and that the general data oriented curricula have progressed across all the iSchool programs on some level. Perhaps, most importantly, the results help to highlight ways in which iSchools can collectively participate in the data area, and inform the following three recommendations:

1. *Recommendation 1: Leverage interdisciplinary activity.* iSchools should continue leveraging their interdisciplinary underpinnings and information science foundations to advance future data education activities. This recommendation conforms with data related professional opportunities, ranging from social and human interaction to computational needs apparent across the wide diversity data related jobs. Additionally, information professionals active in LAMs (library, archives, and museums) area increasingly becoming data savvy and technical. Data scientists and analysts not trained in iSchools can also benefit from acquiring further knowledge and skill in data curation, preservation, metadata and related areas, with through professional development options.

2. *Recommendation 2: Publish curricula.* The iSchools consortium can benefit from mandating curricula publication in the data area. The process will support more thorough analysis and better inform the iSchools and information science community as well as future employers how this sector of the data related workforce is being prepared. The publication of curricula extends to listing publications and PhD theses in the data area.

- *Recommendation 3: Track graduate success.* iSchools should track the success of graduates pursuing data related employment. Discussion focusing on data focused curricula are increasing across the iSchools. Additionally, there new job titles, such as *data research scientist*, *data services librarian*, and *research data and digital curation officer*. Tracking number the success of graduates who have earned data degrees or took data- related courses will aid iSchool consortium members and institutions offering information science degrees in preparing graduates to effectively enter the workforce in the data area.

Overall, the research reported in this paper gives insight into how iSchools are taking steps and preparing graduates to effectively address data needs in the professional workplace. The research approach and the rubric for analysis, defined by *data science*, *big data analytics*, and *digital curation*, can aid future studies. Additionally, the data gathered for this work can serve as a source for measuring future change as data education progresses. As with any development in the digital world, the change is rampant. In response, information and library science programs are increasing undergraduate programs, while facing the challenges of balancing traditional strengths, yet satisfying increasing demand for a prepared data competent workforce [45, 46]. With challenge comes opportunity. As time goes by, additional research is needed to give a fuller picture of the change. Moreover, there is a call for greater dialogue among information scientists, both educators and those in the professional workforce, as these steps will ultimately help our discipline thrive in addressing current and emerging data related challenges. Evident here is that change is underway, and information science has important expertise to offer in the data area. In conclusion, the members of the iSchool consortium are embracing data needs, and taking important steps to prepare future information professionals to successfully enter the growing, data-intensive environment.

## Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## References

- [1] Gabel TJ, Tokarski C. Big Data and Organizational Design: Key Challenges Await the Survey Research Firm. *J Organ Des* 2014; 3: 37–45.
- [2] iSchools. Origins <http://ischools.org/about/history/origins/> (2014, accessed 1 January 2016).
- [3] Wu D, He D, Jiang J, et al. The state of iSchools: an analysis of academic research and graduate education. *J Inf Sci* 2012; 38: 15–36.
- [4] Varvel VE, Bammerlin EJ, Palmer CL. Education for Data Professionals: A Study of Current Courses and Programs. In: *Proceedings of the 2012 iConference*. New York: ACM, pp. 527–529.
- [5] Si L, Zhuang X, Xing W, et al. The cultivation of scientific data specialists: Development of LIS education oriented to e-science service requirements. *Libr Hi Tech* 2013; 31: 700–724.
- [6] Tang R, Sae-Lim W. Data Science Programs in U.S. Higher Education: An Exploratory Content Analysis of Program Description, Curriculum Structure, and Course Focus. *Educ Inf* 2016; 32: 269–290.
- [7] Borgman CL. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press <https://books.google.com/books?id=pb8vBgAAQBAJ&pgis=1> (2015).
- [8] Meadow CT. *Text information retrieval systems*. San Diego: Academic Press Inc., 1992.
- [9] Rowley JE, Hartley RJ. *Organizing Knowledge: An Introduction to Managing Access to Information*. New York: Ashgate Publishing, 2008.
- [10] Stanton J. *Data Science*. Syracuse: Syracuse University <http://jsresearch.net/groups/teachdatascience> (2012).
- [11] Granville V. 16 analytic disciplines compared to data science <http://www.datasciencecentral.com/group/resources/forum/topics/16-analytic-disciplines-compared-to-data-science> (2014, accessed 1 April 2016).
- [12] Nielsen HJ, Hjørland B. Curating research data: the potential roles of libraries and information professionals. *J Doc* 2014; 70: 221–240.
- [13] McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity. *McKinsey Glob Inst* 2011; 156.
- [14] Davenport TH, Patil DJ. Data scientist. *Harv Bus Rev* 2012; 70–76.



- [15] Data Science Association. About Data Science <http://www.datascienceassn.org/about-data-science> (2016, accessed 15 July 2016).
- [16] Stanton J. *An Introduction to Data Science*. Syracuse: Syracuse University [https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1\\_1.pdf](https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf) (2012).
- [17] Song I-Y, Zhu Y. Big data and data science: what should we teach? *Expert Syst* 2015; 33: 364–373.
- [18] Laney D. 3D data management: Controlling data volume, velocity and variety. *META Gr Res Note* 2001; 6: 4.
- [19] Erl T, Khattak W, Buhler P. *Big Data Fundamentals: Concepts, Drivers & Techniques*. Boston: Prentice Hall <http://proquest.safaribooksonline.com.strauss.uc3m.es:8080/9780134291185> (2016, accessed 26 April 2016).
- [20] Erl T, Khattak W, Buhler P. *Big Data Fundamentals: Concepts, Drivers & Techniques*. Boston: Prentice Hall, 2016.
- [21] Long C. *Data science and big data analytics: Discovering, analyzing, visualizing and presenting data*. EMC Education Services, 2015.
- [22] Provost F, Fawcett T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Data Sci Big Data* 2013; 1: 51–59.
- [23] Big Data Science Meetup Group. Big Data Science <http://www.meetup.com/Big-Data-Science/> (2016, accessed 1 January 2016).
- [24] Digital Curation Centre. What is digital curation? <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation> (2016, accessed 5 May 2016).
- [25] Tenopir C, Birch B, Allard S. Academic libraries and research data services. *ACRL White Pap*. Epub ahead of print 2012. DOI: [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir\\_Birch\\_Allard.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf).
- [26] Digital Curation Centre. DCC Curation Lifecycle Model <http://www.dcc.ac.uk/resources/curation-lifecycle-model> (2008, accessed 5 December 2016).
- [27] Gold A. Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. *Library Administration* [http://digitalcommons.calpoly.edu/lib\\_dean/27](http://digitalcommons.calpoly.edu/lib_dean/27) (2010).
- [28] Walters T, Skinner K. *New Roles for New Times: Digital Curation for Preservation* [http://www.arl.org/bm~doc/nrmt\\_digital\\_curation17mar11.pdf](http://www.arl.org/bm~doc/nrmt_digital_curation17mar11.pdf) (2011).
- [29] Choudhury GS. Case Study in Data Curation at Johns Hopkins University. *Libr Trends* 2008; 57: 211–220.
- [30] Antell K, Bales Foote J, Turner J, et al. Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *Coll Res Libr* 2014; 75: 557–574.
- [31] Australian National Data Service. Information specialists and data librarian skills <http://www.ands.org.au/working-with-data/data-management/overview/data-management-skills/information-specialists-and-data-librarian-skills> (2013, accessed 15 July 2016).
- [32] Feltes C, Gibson DS, Miller H, et al. *Envisioning the future of science libraries at academic research institutions: a discussion* <http://hdl.handle.net/1912/5653> (2012, accessed 15 July 2016).
- [33] Research Libraries UK. *The value of libraries for research and researchers* [http://www.rin.ac.uk/system/files/attachments/Value\\_of\\_Libraries\\_-\\_Annexes.pdf](http://www.rin.ac.uk/system/files/attachments/Value_of_Libraries_-_Annexes.pdf) (2011).
- [34] Federer L. *Exploring New Roles for Librarians: the research informationist*. San Rafael, California: Morgan & Claypool Publishers. Epub ahead of print 2014. DOI: 10.2200/S00571ED1V01Y201403ETL001.
- [35] European Commission DG Information Society and Media Unit F3 'Géant & e-Infrastructures'. *Report on the Consultation Workshop Skills and Human Resources for e-Infrastructures within Horizon 2020* [https://www.innovationpolicyplatform.org/system/files/European\\_Commission\\_2012-Skills\\_and\\_Human\\_Resources\\_for\\_e-Infrastructures\\_within\\_Horizon\\_2020\\_5.pdf](https://www.innovationpolicyplatform.org/system/files/European_Commission_2012-Skills_and_Human_Resources_for_e-Infrastructures_within_Horizon_2020_5.pdf) (2012).
- [36] Kim J. Who is Teaching Data: Meeting the Demand for Data Professionals. *J Educ Libr Inf Sci* 2016; 57: 161–173.
- [37] iSchools. Directory <http://ischools.org/members/directory/> (2016, accessed 17 January 2016).
- [38] Dumbill E, Liddy ED, Stanton J, et al. Educating the Next Generation of Data Scientists. *Big Data* 2013; 1: 21–27.
- [39] Pouchard L. Revisiting the Data Lifecycle with Big Data Curation. *Int J Digit Curation* 2015; 10: 176–192.
- [40] Demchenko Y, Belloum A, Wiktorski T. EDISON Data Science Framework: Part 3. Data Science Model Curriculum (MC-DS).
- [41] Mikroyannidis A, Domingue J, Phethean C, et al. Designing and Delivering a Curriculum for Data Science Education across Europe.
- [42] Lazer D, Pentland A, Adamic L, et al. SOCIAL SCIENCE: Computational Social Science. *Science (80- )* 2009; 323: 721–723.
- [43] Lyon L, Brenner A. Bridging the data talent gap – Positioning the iSchool as an agent for change. *Int J Digit Curation* 2015; 10: 111–122.
- [44] National Academies of Sciences, Engineering and M. *Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report*. Washington DC: The National Academic Press, 2017. Epub ahead of print 2017. DOI:



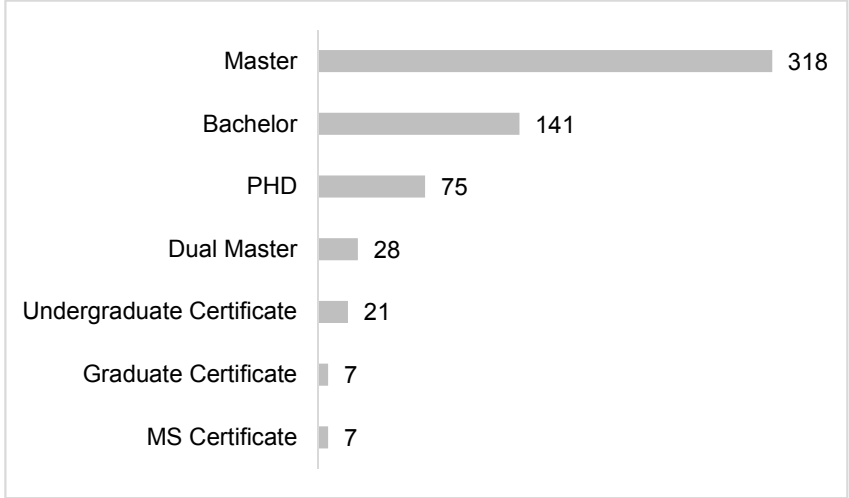
doir.org/10.17226/24886.

[45] Dority GK. *Rethinking Information Work: A Career Guide for Librarians and Other Information Professionals*. Libraries Unlimited - ABC-CLIO <http://www.abc-clio.com/ABC-CLIOCorporate/product.aspx?pc=F2685P> (2016).

[46] Information SS of. *Emerging Career Trends for Information Professionals* [http://ischool.sjsu.edu/sites/default/files/content\\_pdf/career\\_trends.pdf](http://ischool.sjsu.edu/sites/default/files/content_pdf/career_trends.pdf) (2016).

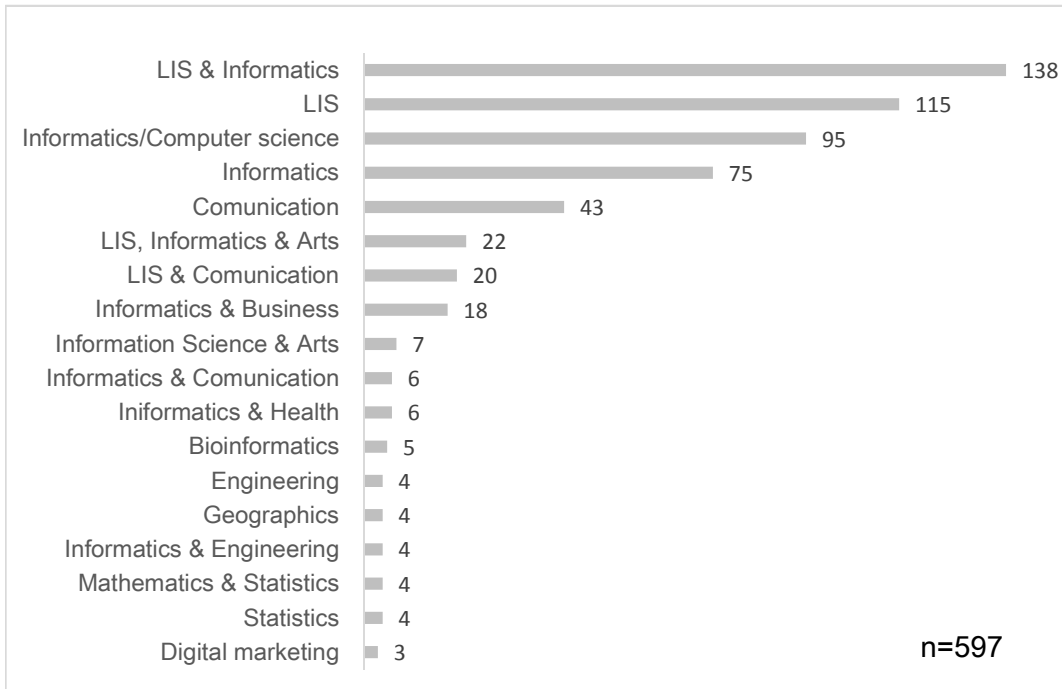
For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



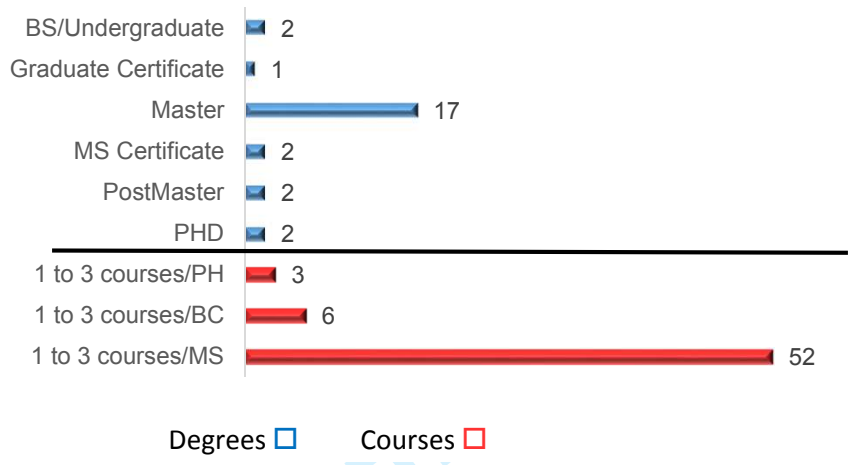
\_\_\_\_\_

Or Peer Review



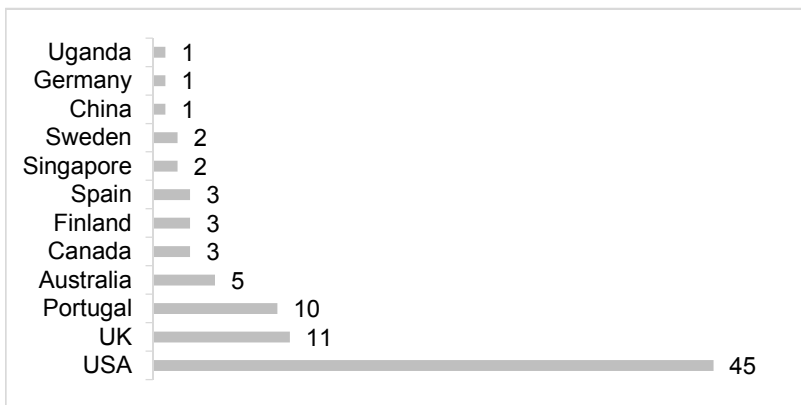
**Figure 3.** Most relevant disciplines involved in iSchools' degree programs.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



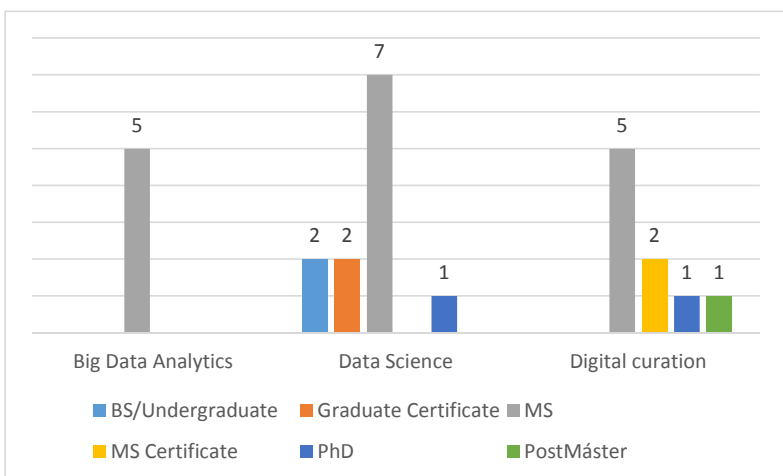
For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



For Peer Review

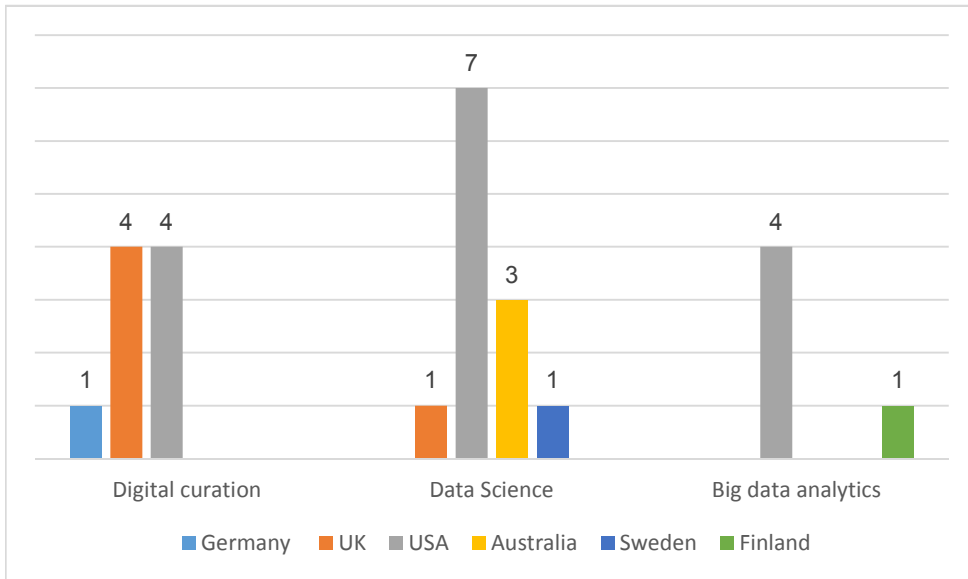
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



For Peer Review

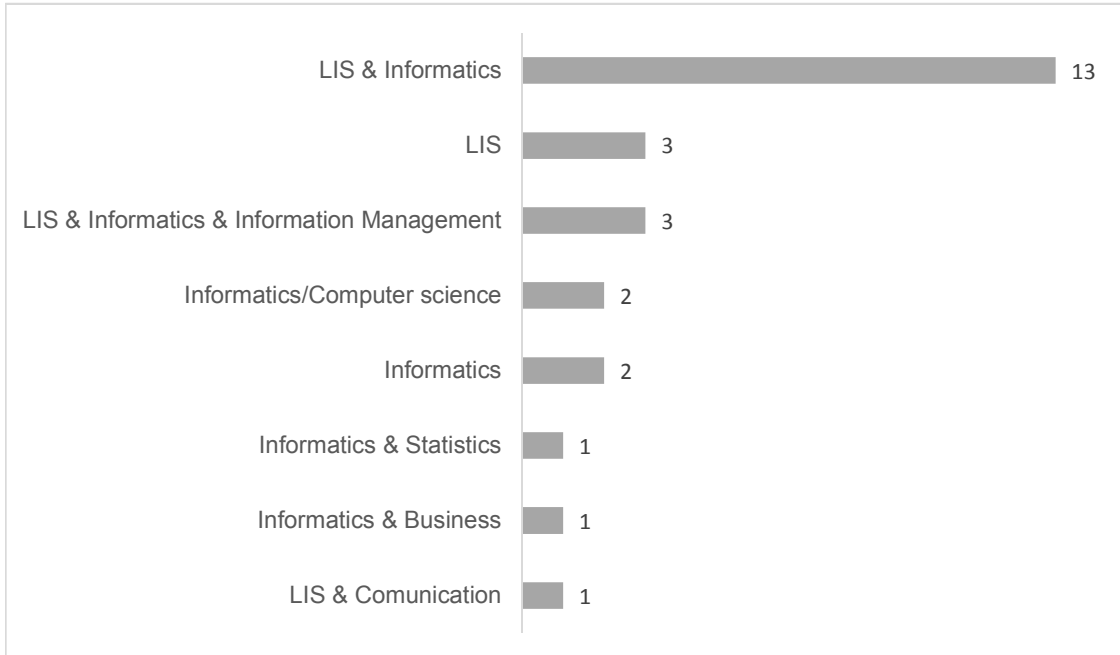


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

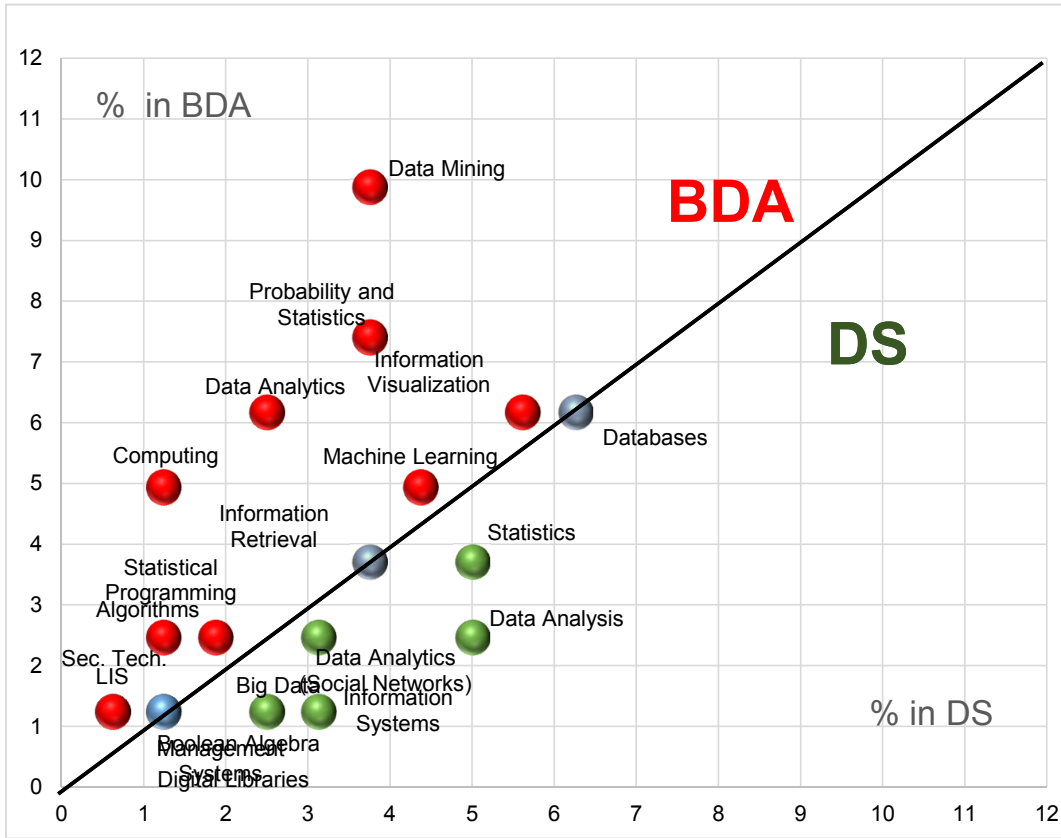


Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



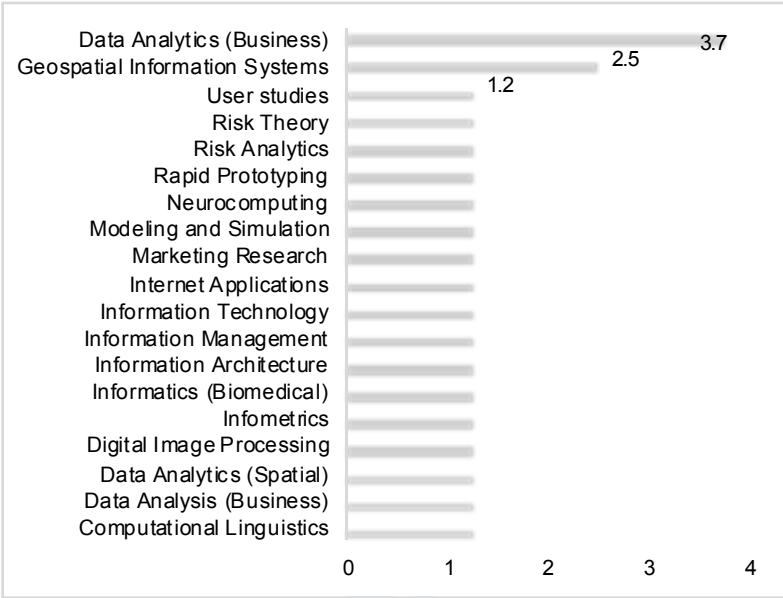
Peer Review



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



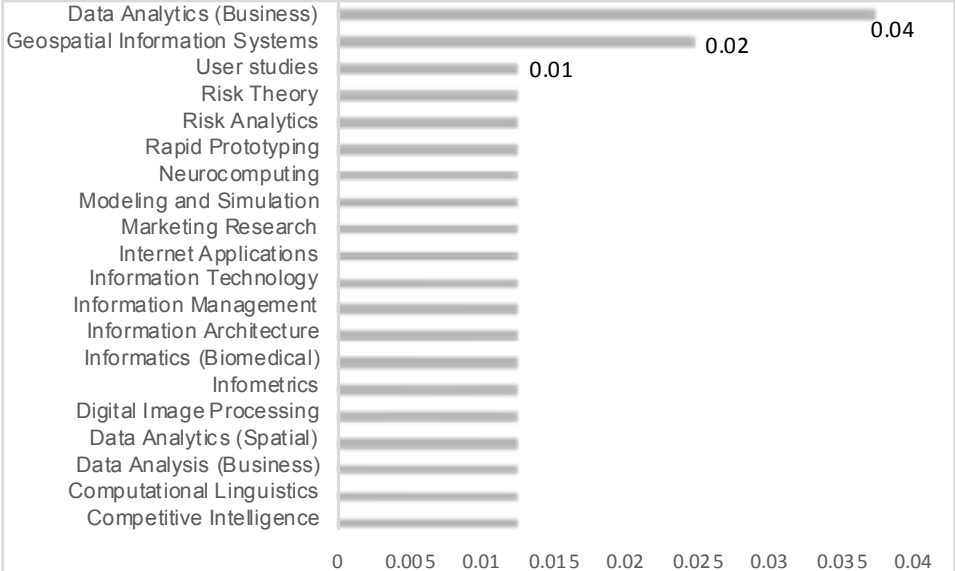
For Peer Review



Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Peer Review

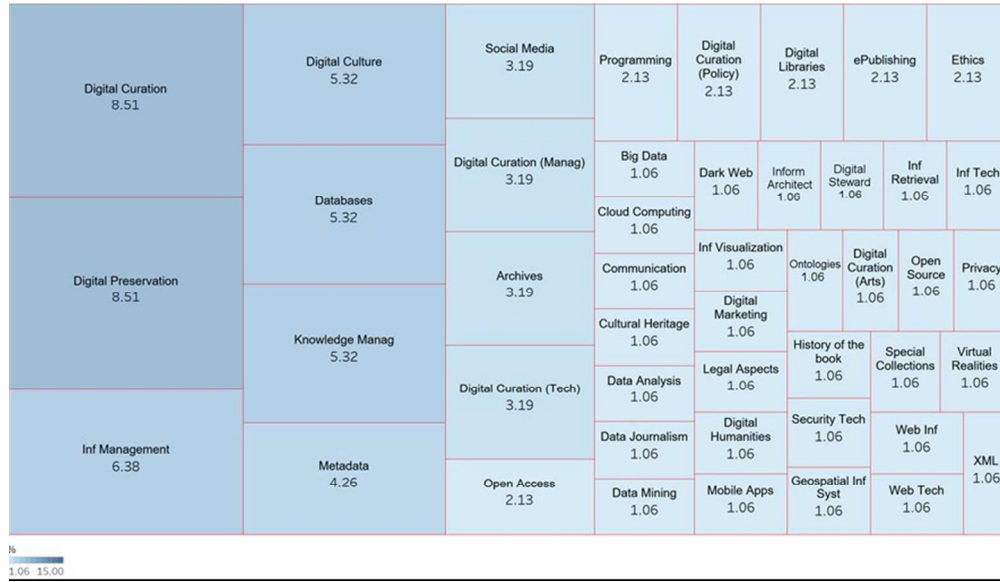


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

.....

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



150x86mm (150 x 150 DPI)

Review