Article

Journal of Information Science
1–19

DOI: 10.1177/0165551510000000
jis.sagepub.com
SAGE

Dialoging about data with the iSchools: exploring curricula trends

Virginia Ortiz-Repiso
Library & Information Sciences Department, Carlos III University, Spain

Jane Greenberg
Metadata Research Center, College of Computing and Informatics, Drexel University, USA

Javier Calzada-Prado
Library & Information Sciences Department, Carlos III University, Spain

Abstract
Our rapidly growing data driven culture is motivating curriculum change in nearly every discipline, including information science. This paper explores this change specifically within the iSchool community, where 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 data related education, particularly at the master's level, and that approximately 15% of the formalized degree offerings have a data focus. Overall, iSchools show a greater emphasis in data science and big data analytics, with only a few select programs providing focused curricula in the area of digital curation. Recommendation are also made for iSchools to leverage the interdisciplinary of information science, publish curriculum, 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 more interdisciplinary data education, including data curation curricula.

Keywords
Data science; big data; data analytics; digital curation; digital preservation; iSchools; information science; education.

1. Introduction

The need for a data competent workforce is impacting nearly every discipline [1]. The 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 in a position to develop and advance a distinct expertise for data-driven needs.

Clearly, the opportunity for iSchools in the growing data driven space is exciting, and information scientists find themselves working on interdisciplinary teams that previously were unimaginable. Excitement aside, the iSchool community also faces challenges in the data education space, particularly in finding to right measure to distinguish, complement, and support other data focused academic curricula, such as medical and bio-informatics, business analytics, and digital humanities. These challenges, and the 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 educating a knowledgeable, skilled, competent workforce in the data area. Here, we ask, *how are iSchools responding to the need for a data competent workforce?*

Corresponding author:
Virginia Ortiz-Repiso, Library & Information Sciences Department, Carlos III University, C/ Madrid, 126, 28903 Getafe, Madrid, Spain
Email: virginia@bib.uc3m.es

This question guided the research presented in this paper, specifically the results of a cross-institutional analysis of data related offerings in iSchools. The research took into account the full range of data oriented courses and program offerings across the iSchools. An overriding goal of this research is to gain insight into the extent of data related curricula activities within information science, particularly the iSchools, as a growing, global community.

In the sections that follow, we provide background context on iSchools and a review of relevant data-oriented disciplinary trends. Next, we present our research objectives, methodology and procedures, followed by a report of the results and integrated with discussion comments. The conclusion summarizes our key findings, makes recommendations for iSchools to collectively advance more in the data area, considers research implications, and identifies next steps.

2. The iSchool context

The iSchools, founded in 2005, represent an international consortium of schools dedicated to advancing the information field [2]. Information science is the unifying discipline across the iSchools, and draws together programs that have 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, and also include 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 a fundamental interest in the relationships among the facets of *information*, *people* and *technology*.

Today's increasingly data driven culture interconnects with these facets (information, people, and technology). 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 just a few years ago in Varvel, Bammerlin & Palmer's [4]. These 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 programs, certificate programs, and concentrations. These researchers also reported that many iSchools had an interest in revising their curricula to encompass data and that "existing digital library and archives courses are [were] contributing the most data oriented content, covering key areas such as metadata and digital preservation."

Following, in 2013, Si, Zhuang, Xing and Guo's [5] web content analysis, sheds more 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, surprising, 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 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 goals to assess the iSchools' activity in this area, it is important to first review how 'data' is understood in the iSchools/information science domain, and 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. Borgman explains we can reach more agreement looking at 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, information science introductory 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 as 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 a "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 see data science as an analytical discipline that includes the integration with business intelligence, artificial intelligence, computer science, econometrics, engineering.

Data science, as a discipline, has gained recognition in the scholarly and scientific literature, as well as the popular press. Related here 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]. To follow, 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, understand how data can be leveraged to predict outcomes.

3.1.2. *Data science and iSchool embrace*

In the iSchool community, Stanton [10], 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" as information science driven tasks, are key to the data science domain, and the ability to understand the big picture within a complex system. This viewpoint is reflects the iSchools emphasis on the information lifecycle [16]. Together, these findings help explain why iSchools are among the array of institutions supporting data science curricula; and also interconnect to the coverage of 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 the 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* [17]. The original three Vs have evolved to five Vs adding *veracity* and *value* [16, 18].

Big data is "a field dedicated to the analysis, processing, and storage of large collections of data that frequently originate from disparate sources" [18]. Big data initiatives leverage big data sources to solve business related problems. 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 [19]. '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 in the introduction, many iSchools have foundations in library and information science, although there are a number of schools with strong foundations in operations research or links 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 connections, and the fact that iSchool faculty collaborate with other disciplines, further feed the iSchool discussions about what big data is, and the skills required to make use of big data [5]. Moreover, there is also an intellectual debate about the difference between big data and data science [20]; and, in fact, the concept of 'big data science' has also emerged, with a meet up group in California [21]. Indeed, the distinction between big data, data analytics, and data science can seem blurred, although a review of literature reveals noted differences. One 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 (2015), which shows how the levels build upon each other that lead to data science (Figure 1).

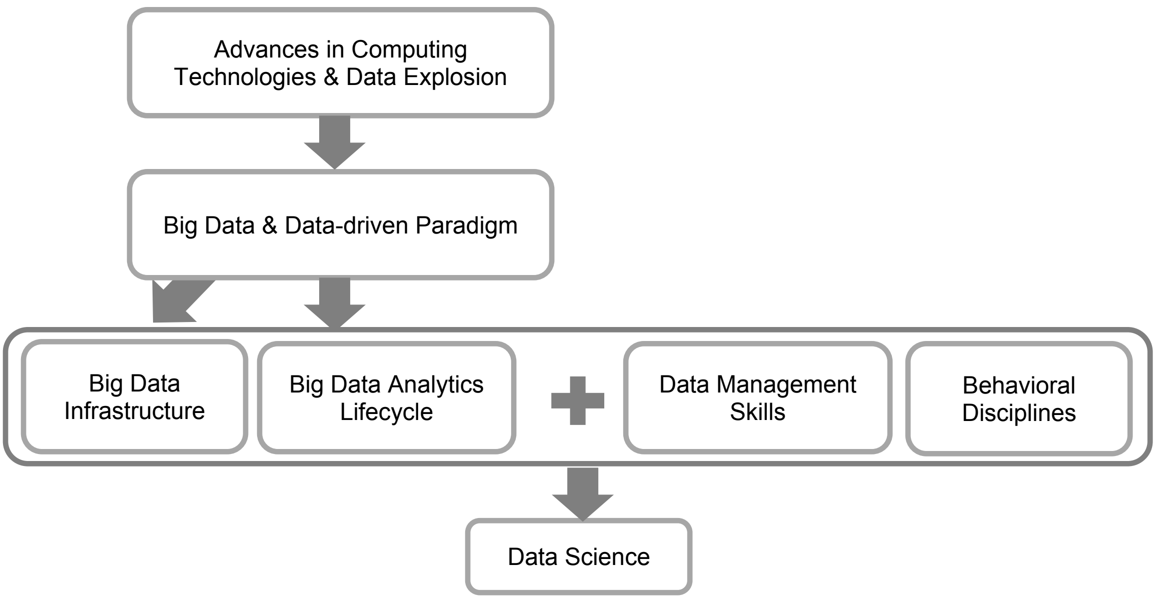

Figure 1. Emergence of data science [16].

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 [22, 23].

Digital curation encompasses lifecycle management of digital assets [24]. Drawing from the DCC model, the process of curation begins with conceptualization and creating or receiving a digital items, and involves various processes, such as appraisal, ingest into a system, preservation, and preparing the digital resource for access, use, and reuse.

Similar to 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 is reflective of their often being presented together. Gold [25] 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; "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 as easy add-on. We see the preservation emphasis connected more with programs that have strong links with historiography, such as King's College in the UK. In this area, Walters & Skinner [26] observe that different groups of scholars and their library partners have evolved some terminological differences; however, they also acknowledge that the nuances can obscure the fundamental unity in libraries' roles and services relating to digital curation and preservation. Literature also reveals a host of related descriptors to define individuals in this professions, such as *data humanist* [27]; *science librarians* [28]; *data librarian* [29]; *librarians/science informationists* [30]; *information specialists* [31]; and *research informationist* [32]. This heterogeneity is reflected in job profiles, roles and responsibilities, and even within an institution type. A case in point is the 'data scientist' job profile in Europe, where there is no common definition [33]. The diversity and heterogeneity reviewed here is reflected in the different career structures across institutions and within the iSchool community.

3.4. Summary: iSchools and the data area

Data science, big data, and data curation and preservation represent the most consistent nomenclature and themes detected across data focused programs in iSchools; an observation confirmed by Song & Zhu [16]. For the immediate research in this paper, we have, therefore, selected these areas to provide framework for examining data driven curricula addressed in the next section of this paper. Finally, to further aid in our analysis, we present the following diagram (Figure 2), showing a view of how information science (including the collection of subdisciplines), interrelates to the data driven areas that forming our rubric. The result is a unique blend that the iSchools can offer.

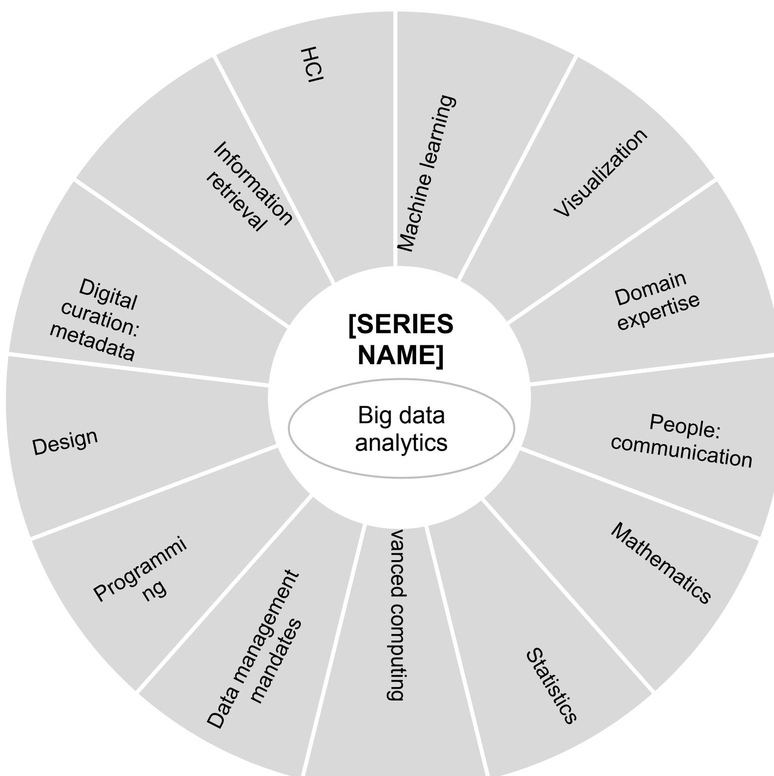

Figure 2. Data science and data trends.

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, the reporting on the extent of curricula change is limited.

We are at a juncture now where it seems 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?
- What data driven emphases and foci are found in iSchools?

5. Methodology

In order to pursue our research questions, we conducted a cross-institutional analysis of iSchools, including a cluster analysis of courses offerings. This method seems most appropriate for gathering data on the extent of data related curricula [4–6, 34]. Our data source was the iSchool Directory, representing a consortium of leading programs that are joined by the common theme of information science. The data is based on the official roster of 65 iSchools released in Feruary 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 in the represented via membership. The individual iSchool websites served as our data source.

Our procedures include the following steps:

(1) Consult the iSchools Directory [35] to identify and locate each of the iSchools' institutional websites.
(2) Identify the data-oriented degrees, specializations, and certificates for undergrads, graduate and Ph.D. in the 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 (main area covered) using a coding scheme developed and vetted by the researchers.
(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 coursed offered in non-data focused degrees. The courses were examined using a cluster analysis, that involved three processes:

    - Courses were organized using our rubric of data science, big analytics, and digital curation. We sought to respect the nomenclature reflected literature and the degree names.
    - 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 2nd-level clusters. We verified terminology by consulting the ASIST thesaurus, the Library of Congress Subject Headings, further literature review, and consultation with Drexel University researchers who have pursued research in the data area. This established our dataset for further analysis.

(5) Tally data related courses that were offered in non-data focused degrees were processed and analyzed separately.

6. Data analysis and discussion

6.1. iSchools' degree programs and disciplines

At the time of data gathering, a total of 597 degrees were found across the 65 iSchools analyzed. As we can see 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.

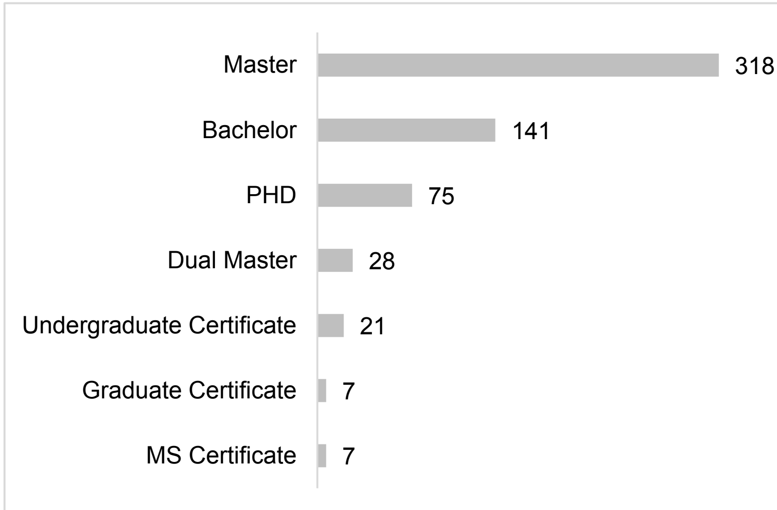

Figure 3. 597 iSchools degree programs.

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

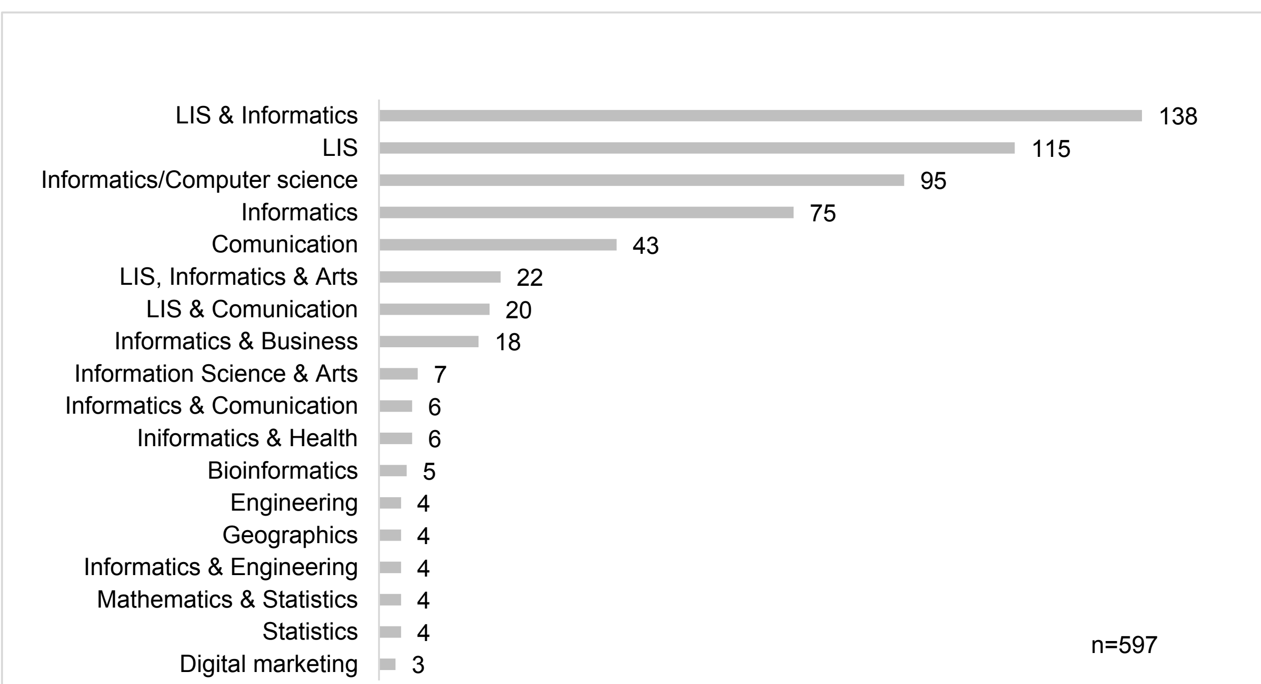

Figure 4. Most relevant disciplines involved in iSchools' degree programs.

6.2. iSchools' data related degrees

As a result of our analysis, we identified a total of 87 degrees (14.6% of the 597 degrees) that are in some way related with data, either because they are data specific –that is, they offer complete curriculum, for data or a graduate certificate, or specialization that is data focused– or because the degree include data related courses (Figure 5). To provide more context, the 87 data related degrees were offered by 37 iSchools. This means the 56.9% of the iSchools are offering some type of data related education. On the one hand, we anticipated a higher percentage of iSchools to be 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 seems appropriate, given the wide diversity of curricula that are supported in iSchools, ranging from public libraries and cultural centers to mathematics and more computationally intensive work.

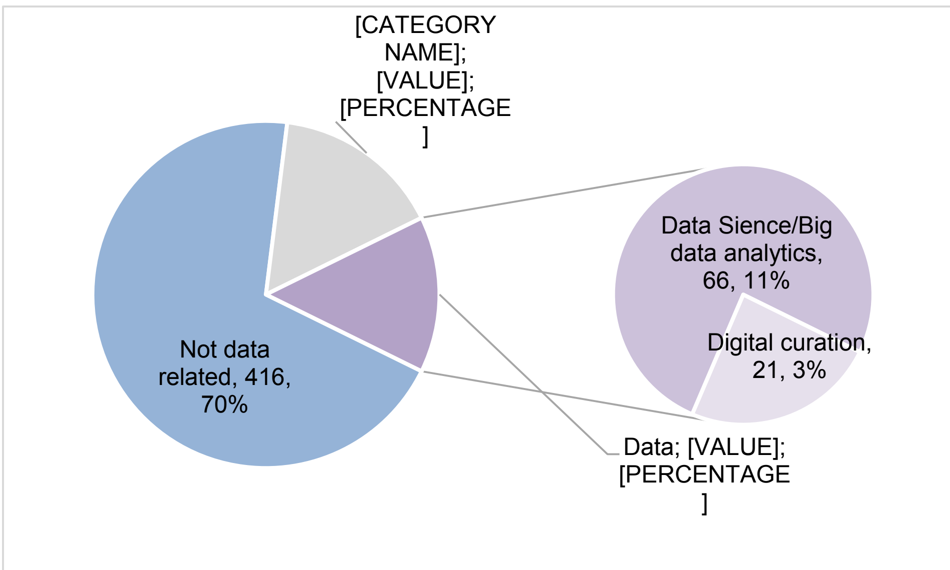

Figure 5. iSchools' degree programs and data education

A closer examination of the 87 data related degrees (Figure 6, top-portion) shows that a third of them (30%, 26 of the 87) are data specific. The other 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.

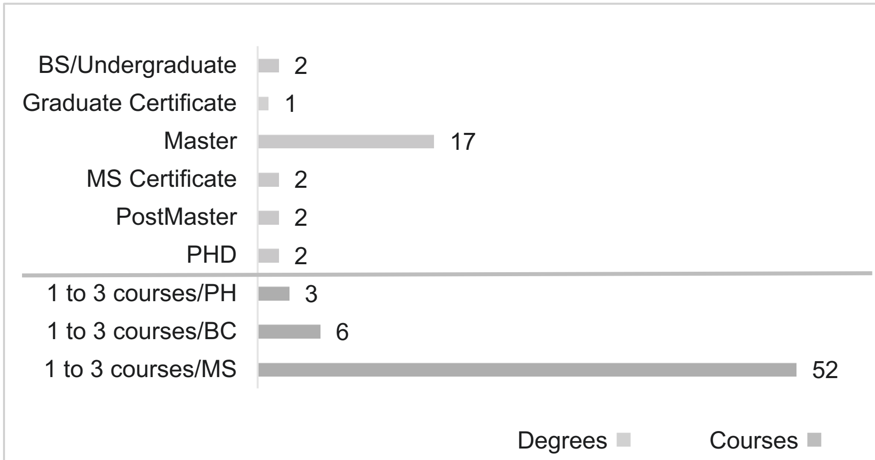

Figure 6. 87 data related degrees.

The United States (U.S.) has the largest concentration of data focused degrees among the iSchools today (Figure 7). These results are reflective of the U.S. having the largest number of iSchool members. The number of courses offered by the three Portuguese iSchools seems high compared to other countries.

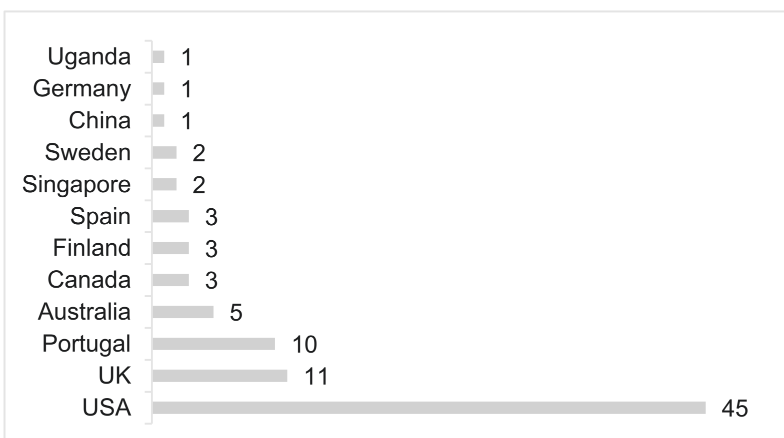

Figure 7. 87 data related degrees by country.

6.3. *iSchools' data focused degrees*

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

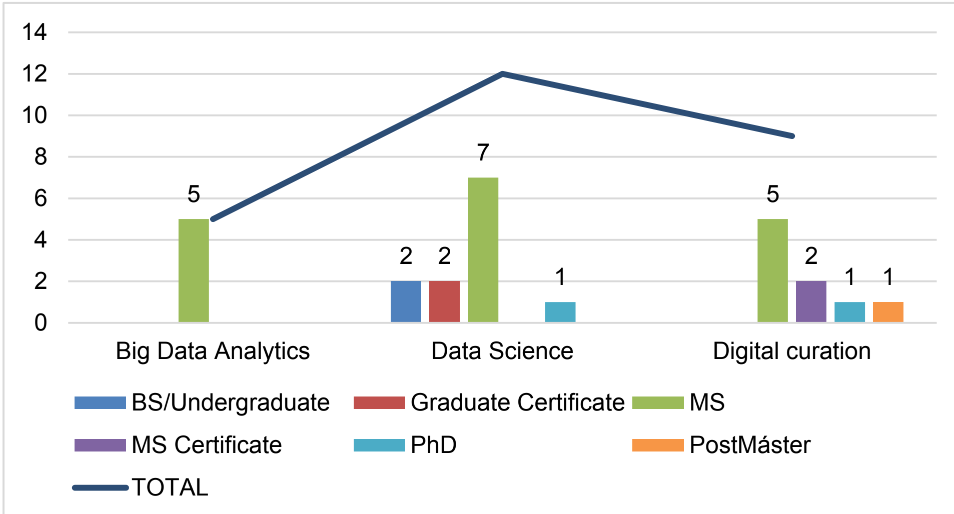

Figure 8. 26 Data degrees by subject area.

English speaking countries (United States, United Kingdom and Australia), and especially the United States (Figure 9), have the highest number of degrees in data science and big data analytics (11 out of 17, 64%). Interestingly enough, the rest of degrees in data science and big data analytics observed in non-English speaking countries are also delivered in English. Digital curation appears to be fairly equally distributed in Europe and the United States.

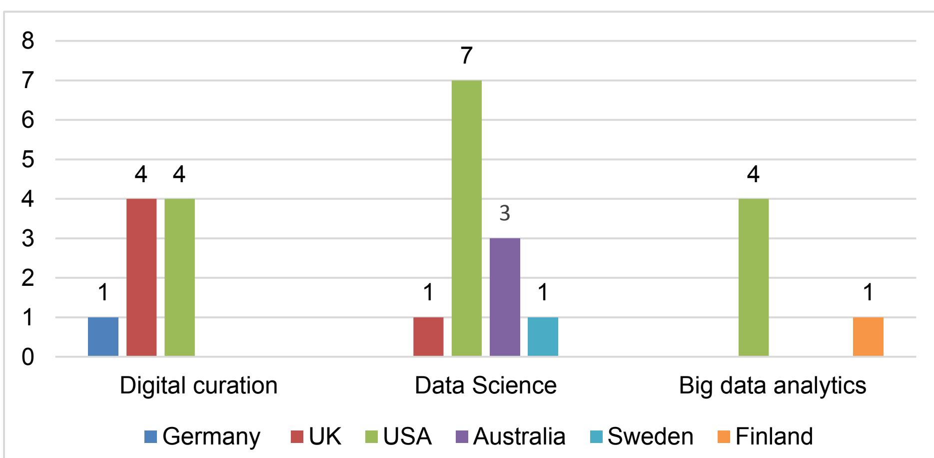

Figure 9. 26 Data degrees by country.

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

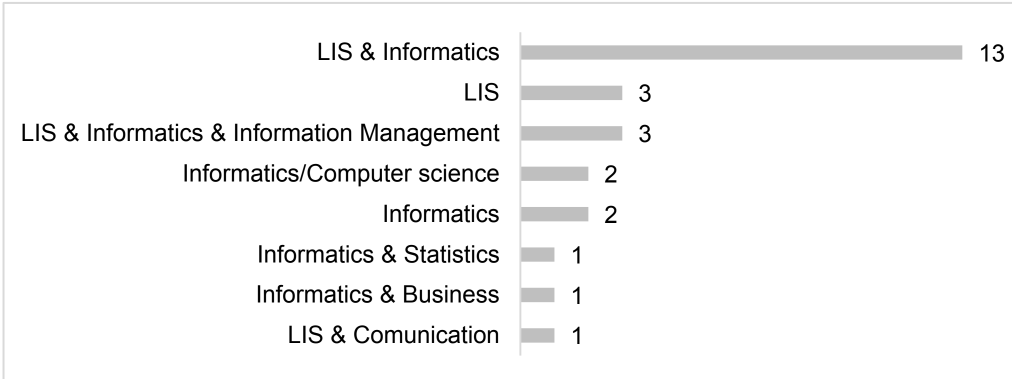

Figure 10. Disciplines involved in data degrees.

Gaining a better understanding of what is happening across these disciplines is important. In the following sections we will look in more detail at the curriculum and consider if there are interconnections.

6.4. Basic comparison of iSchool's data degrees

Given the prevalence of data in our information culture, we anticipated impact across all degree areas. Table 1 provides an overall picture of the schools that have degrees in data science, big data analytics, and digital curation, and at what degree level.

The results in data science found two bachelor degrees, seven masters, two graduate certificates/diplomas, and one PhD program. All of them started between 2014 and 2016. The duration of these degrees varies from 6 months for the graduate certificates to four years of the bachelor. Masters' degree programs range from one to two years. Only one of the master's degree programs is fully online, all the others are face to face or mixed modes contemplated.

In the big data analytics area, we identified five masters, all of them started between 2014 and 2015. The duration of these degrees is from 1 to 2 years. Two of them are face to face and three use mixed methods. In digital curation there are five masters, two graduate certificates, one postmaster, and one Ph.D. All of them started between 2013 and 2016. Four of them are offered on campus, two online, and other two hybrid methods.

The disciplines involved, as anticipated, 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 rest.

At this time (January 2016) of data collection, only the University of Illinois reported offering two masters in both disciplines: Master in Big Data and Master in Digital Curation. It should also be noted that some master's programs offer different specializations or use the same subjects to offer graduate certificates or diplomas. An example of this is Indiana University, with a Master 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 also involved looking at their course offerings in the three observed areas. Since an overlap was found in clusters of courses related to data science degrees and big data analytics, it was feasible to analyze the similarities and differences found among their course offerings. As 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 in 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 course offerings. They were categorized into 50 and 41 clusters respectively. Table 2 shows the overlap that was found on 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
DATA SCIENCE					
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			
BIG DATA ANALYTICS					
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			
DIGITAL CURATION					
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 (Figure 11) suggests that, within the iSchools, the data science courses are covering a larger spectrum of subjects compared to big data analytics. Additionally our results align with what Song and Zhu [16] found, showing that big data analytics appears to very much a part of data science, which in our case may be reflected in the higher course overlap observed for big data analytics.

The relationships found among courses in each cluster in a symmetrical bivariate graph, in which the ball size represents the degree of clusters overlap and the position of the balls represents the predominance of BDA or DS in the overlap (Figure 11).

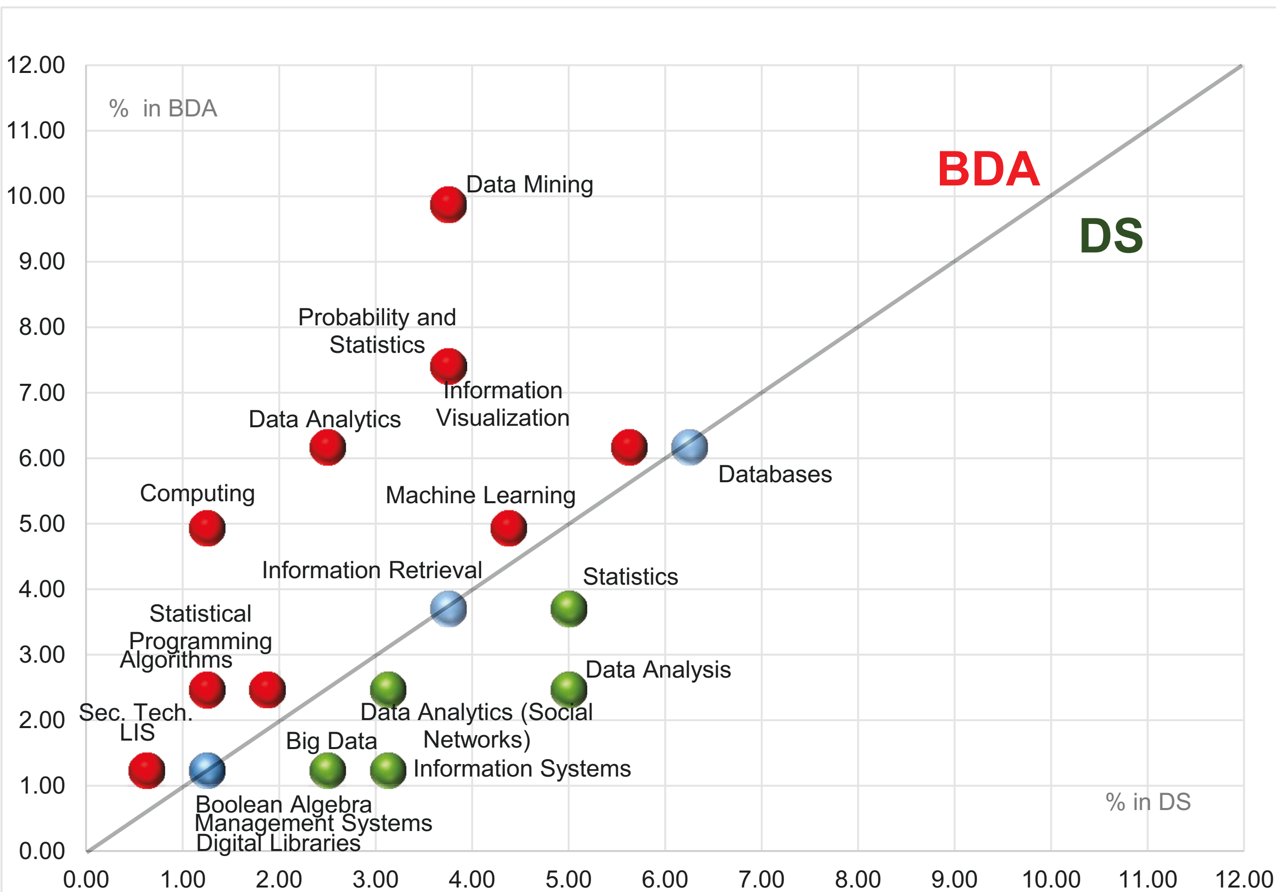

Figure 11. Similarities among DS and BDA degree courses.

Specific to our sample we found that databases, information visualization, machine learning, and information retrieval are among the largest clusters with the biggest overlap; and library and information science and security technology are among the smallest clusters that have definite overlap. In looking for different emphases, 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 more predominate in big data analytics. While observing these differences, we need to take into account the overall small number of degrees we analyzed here, within the iSchool context. The similarities are not surprising, given that they are fundamental to both data science and big data analytics.

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

In assessing an apparent, distinct iSchools context, we also looked at differences among the clusters for courses offered in data science and big data analytics degrees. Figures 12 and 13 present the clusters based on the percentage of courses in these two areas. Data science shows 29 unique clusters (68 courses that are not in big data analytics degrees); and big data analytics has 20 unique clusters (23 courses that were not in data science degrees).

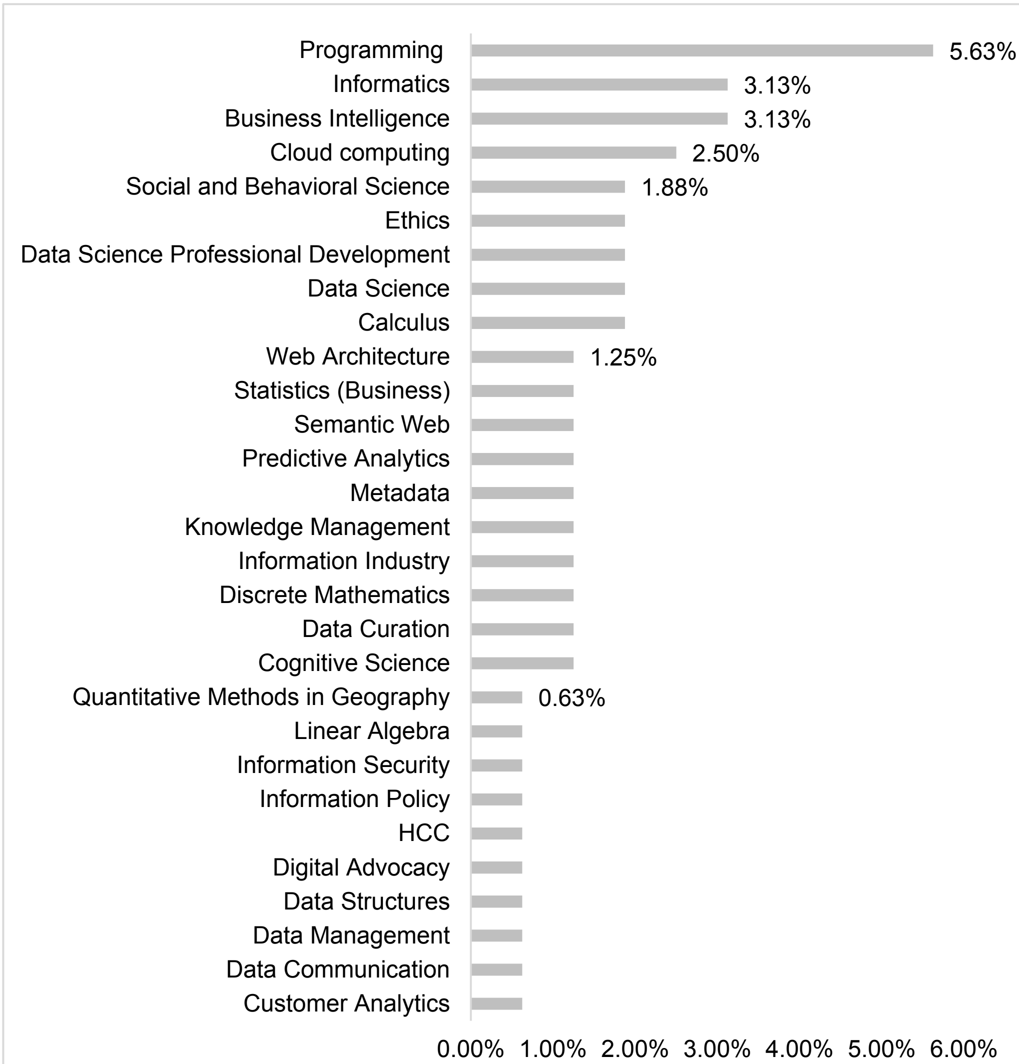

Figure 12. Differences DS versus BDA. Only DS degree courses.

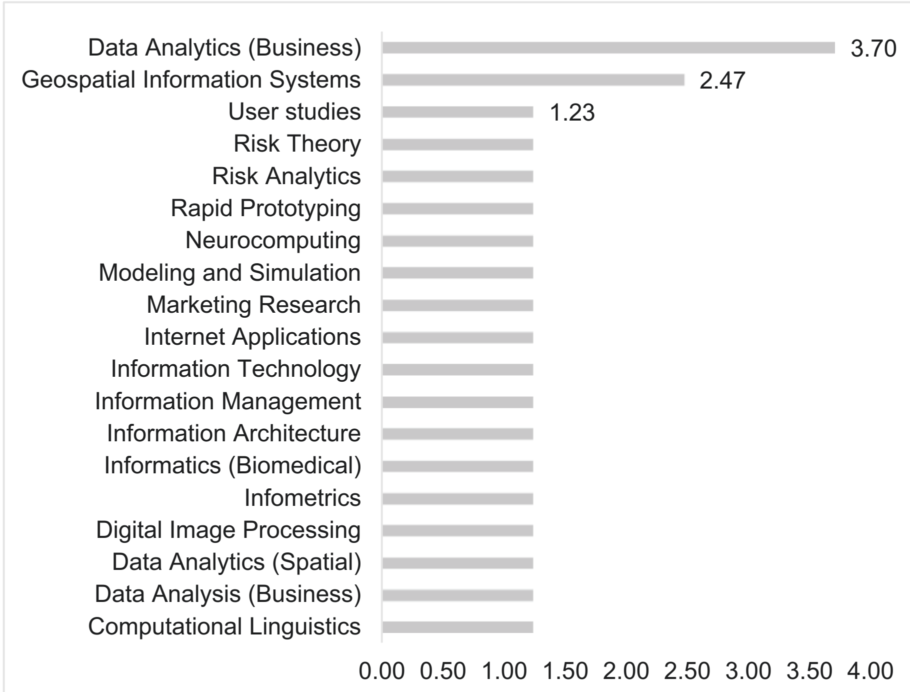

Figure 13. Differences DS versus BDA. Only BDA degree courses.

A few results stand out in our comparisons. Data science has more variety than big data analytics, with courses like 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 course like risk theory and data analytics for business. Somewhat surprising is that programming and informatics is represented in data science clusters, although absent from big data analytics. This is very likely a reflection of the background of the students pursuing these two areas of study, with data science likely attracting students from a greater diversity of disciplines [36], and big data analytics attracting a more homogenous group of students that may have already had exposure to programming. Given the novelty of these disciplinary foci in iSchools, starting, primarily around 2014, it is still a bit early to have substantial data on the types of students being attracted to this area, overall.

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. In looking at the course clusters, digital curation and digital preservation are offered across all of the (9) degrees in this area. Databases, digital culture, and knowledge management are offered in five of the degrees. Information management, databases, knowledge management, and metadata are offered in four of them. Less common, but still representing of three degrees are “archives”, “social media”, “digital curation (technology)”, and “digital curation (management).” Figure 14 shows the distribution of clusters in a treemap chart, in which the area size and color of each cluster give an idea of its frequency of occurrence.

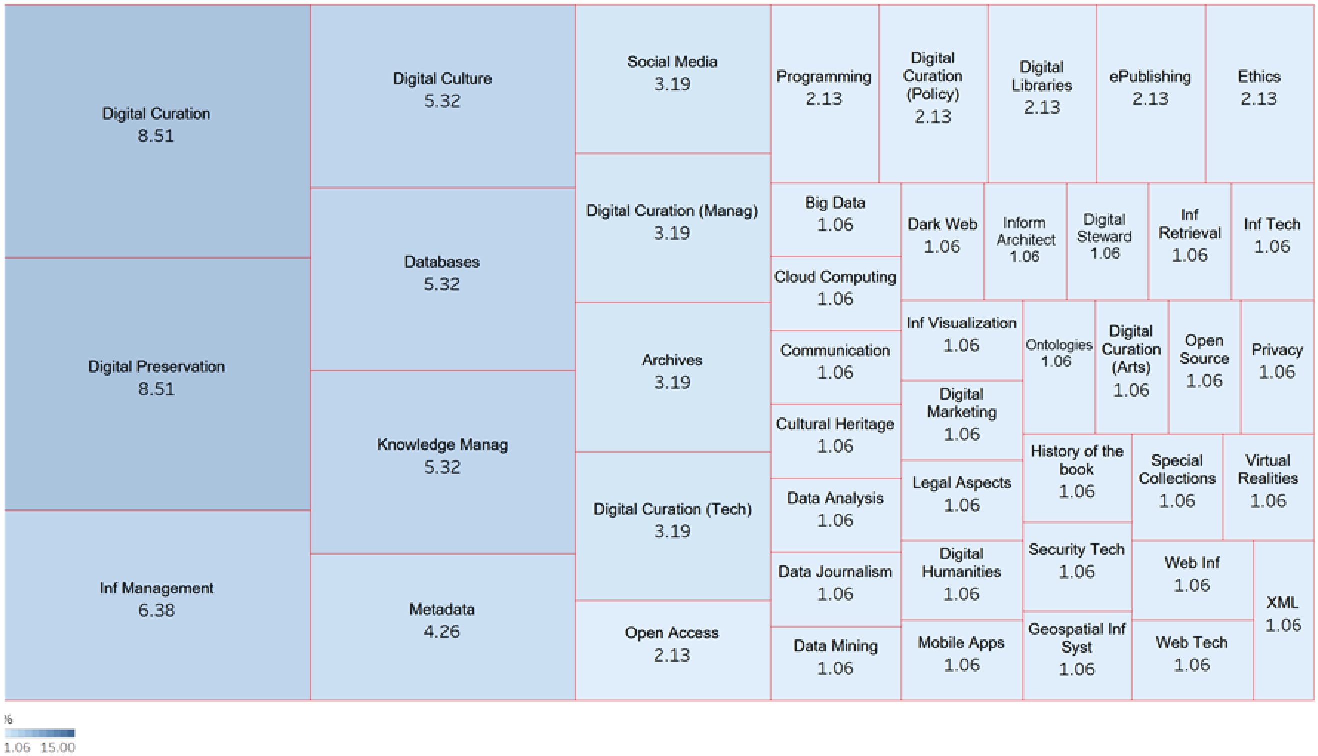

Figure 14. Digital curation degree courses

Harris-Pierce & Quan's [37] study examined data curation in 52 LIS school's in North America on data curation education reported that the most important courses dealt with metadata, collection development, hardware and software platforms or technologies, relational databases, data types, and digital preservation. These researchers also reported that the following courses were relevant and important: data reposit ories, data quality, project management, provenance, researcher practices and needs, digital scholarship in the humanities, digitalization, libraries and archives, risk management, roles of curators, sustainability, data analysis, data management, records management, records management, history, open access, stakeholders, subject knowledge, and XML.

Perhaps most unexpected with this analysis is that only 7 iSchools (10.7% of 65) offer data curation (9 degrees). It seems that iSchools would do well to add data curation and digital preservation to their curriculum. 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 deposition requirements tied to publication as part of the scholarly communication activity. These policies seek to avoid research duplication and reduce costs of data collection.

Digital curation curriculum teaches that data must be stored, analyzed, curated, preserved, and disseminated for being reused. Libraries are experts on that but the growing demand and the new panorama of big data; and, as Pouchard [38] points out, college and research libraries are re-inventing themselves to respond to that necessity. Related here, is the fact that academic libraries have always embraced new developments supporting researchers' needs, and nobody can deny the importance of data services in academic libraries, particularly in light of new policies. Data 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 new roles in the data world. Although many academic libraries are making efforts to prepare their staff to deal with data services, the offering of a formal curriculum is key to sustainable success in this area.

6.5.4. Data related courses in non-data focused degrees

As previously indicated, we found 62 data related courses offered in degrees that are not data specific - that is, they do not offer complete curriculum 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. The most popular subject is big data analytics, offered by 24 degrees (Table 3). 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
Data Analytics	6
Web analytics	6
Digital Preservation	1
Research Data Management	1
Visualization	1

7. Conclusions and recommendations

This paper reports on a cross-institutional analysis, including a cluster analysis of courses offerings, of data related curricula was conducted across 65 iSchools. As part of our assessment we first reviewed how 'data' is understood in the iSchools/information science domain, and disciplinary trends, reviewing data science, big data/data analytics, and digital curation and preservation. We then examined, specifically, 1) the extent of data related iSchool curricula activities, and 2) driven emphases and foci across iSchools. Our sample included full roster of 65 iSchools consortium members in January 2016. The iSchools were selected as leading institutions where information science is a common, unifying discipline. They were also selected because of the international scope and breadth of information disciplines.

The analysis confirmed that data related curricula change is taking place across iSchools, with the inclusion of data degrees and data related courses. The data related curricula represent important steps taken by iSchools members toward helping to educate a data competent workforce. Key findings are summarized here:

- The variety of disciplines across the iSchools (information and library science, archives, computer science, social informatics, human centered computing, etc.) has resulted in a different levels of data education, with information and library science and informatics programs leading curriculum changes to accommodate data education (Figure 4).
- Most data science and big data analytics degrees are being offered by American iSchools (11 out of 17, 64%), while digital curation degrees are fairly equally represented in Europe and in the U.S.
- Data degrees seem to proliferate starting in 2013 and are mainly offered at the master's level, which is the most common degree offered by iSchools. Although few of the iSchools offer a complete data focused degree, a nearly all of these institutions offer degrees with data related courses, in their attempt to respond to the professional trends and educational needs of future professionals in the area.
- The distinction between data science and bid data analytics in iSchools' curricula is not always precise, although the data science generally includes a greater diversity of courses and issues compared to bid data analytics (Figure 14).
- Digital curation degrees are only offered by a small percentage of iSchools (7 out 20, 35%) o compared with the number of schools that offer degrees in data science and big data analytics (13 out 20, 65%).

The analysis undertaken and reported in this paper involved extensive review and discussion, and, therefore present a fairly cohesive view of the iSchool community. We acknowledge challenges, as website data varied in completeness—particularly for doctoral programs, not every website is in English. A small portion of iSchools may have changed course offering, although curricula evolve very slowly and asynchronously. Despite these limitations, the overall findings still hold. Moreover, they findings provide a valuable pointer to the future direction of travel for the iSchools. Here, we offer three recommendations for iSchools to continue to excel in the data area:

- Leverage interdisciplinary. The iSchools should continue leverage their interdisciplinary, drawing from the foundation in information science, for future efforts in data education. This recommendation integrates with opportunities across data related professions, ranging from social and human interaction to computational issue echoes the problem that can also be found in the data related professions. Here, we already see librarians and professionals in the LAMs area becoming increasing technical, and can recommended that the data scientist can benefit from graining more knowledge and skill in data curation, preservation, metadata and related areas. A greater effort from each profession working in the data area needs to be made, since neither of them have a future without mutual support.
- Publish curriculum. The iSchools consortium should mandate publication of curriculum in the data area, more consistently, not only to support additional substantive analysis, but also to help inform the iSchool and information community more clearly, and understand the unique ways in which the iSchoools contribute to educating and preparing a data related workforce needs. The publication of curriculum extends to including the listing of finished PhD theses in the data-related area, to show research contributions.
- Track graduate success. iSchools need to track the success of graduates pursuing data related work. Discussion focusing on data related curriculum is increasing across the iSchools. There is knowledge of new job titles like *data research scientist*, *data services librarian*, and *research data* and *digital curation officer*. More research about the success of data degrees and data courses in terms of number of students enrolled, educational background, graduation and employment rates, etc., can help not only individual iSchools, but the iSchool consortium as a whole.

The results presented here serve as measure in time, when the iSchool consortium continues to grow, and there is increasingly reports noting an increased need for a data competent workforce. There are questions about declining enrollments in library science, and also positive reports of new opportunities [39, 40]. There are also global communities, such as the Research Data Alliance [41, 42], bringing together a diversity of information professionals, research scientists, and industry partners who seeks to collectively address data challenges in areas ranging from data sharing and publication to specific community needs such as biodiversity or wheat interoperability. To this end, the research like that presented here, and role of iSchools to meeting future challenges, needs to continue not only within the growing iSchool community, but other disciplines. Further, the work presented here, including our rubric of data science, big data analytics, and digital curation can aid future studies, and our data may serve as a source for which to measure change. In conclusion, the finding presented in this paper and our methodology will help iSchools contribute to the larger goal of educating the next generation of information professionals to successfully embrace data challenges.

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. *Journal of Organization Design* 2014; 3: 37–45.
[2] iSchools. Originshttp://ischools.org/about/history/origins/ (2014, accessed January 1, 2016).
[3] Wu D, He D, Jiang J, et al. The state of iSchools: an analysis of academic research and graduate education. *Journal of Information Science* 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. *Library 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. *Education for Information* 2016; 32: 269–290.
[7] Borgman CL. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, 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. *An Introduction to Data Science*. Syracuse: Syracuse Universityhttps://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf (2012).
[11] Granville V. 16 analytic disciplines compared to data sciencehttp://www.datasciencecentral.com/profiles/blogs/17-analytic-disciplines-compared (2014, accessed April 13, 2016).

[12] Nielsen HJ, Hjørland B. Curating research data: the potential roles of libraries and information professionals. *Journal of Documentation* 2014; 70: 221–240.
[13] McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute* 2011; 156.
[14] Davenport TH, Patil DJ. Data scientist. *Harvard business review* 2012; 70–76.
[15] Data Science Association. About Data Sciencehttp://www.datascienceassn.org/about-data-science (2016, accessed July 15, 2016).
[16] Song I-Y, Zhu Y. Big data and data science: what should we teach? *Expert Systems* 2015; 33: 364–373.
[17] Laney D. 3D data management: Controlling data volume, velocity and variety. *META Group Research Note* 2001; 6: 4.
[18] Erl T, Khattak W, Buhler P. *Big Data Fundamentals: Concepts, Drivers & Techniques*. Boston: Prentice Hall, 2016.
[19] Long C. *Data science and big data analytics: Discovering, analyzing, visualizing and presenting data*. EMC Education Services, 2015.
[20] Provost F, Fawcett T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Data Science and Big Data* 2013; 1: 51–59.
[21] Big Data Science Meetup Group. Big Data Sciencehttp://www.meetup.com/Big-Data-Science/ (2016, accessed January 1, 2016).
[22] Digital Curation Centre. What is digital curation?http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation (2016, accessed May 5, 2016).
[23] Tenopir C, Birch B, Allard S. Academic libraries and research data services. *ACRL White Paper*. Epub ahead of print 2012. DOI: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf.
[24] Digital Curation Centre. DCC Curation Lifecycle Modelhttp://www.dcc.ac.uk/resources/curation-lifecycle-model (2008, accessed December 5, 2016).
[25] Gold A. Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. *Library Administration*http://digitalcommons.calpoly.edu/lib_dean/27 (2010).
[26] Walters T, Skinner K. *New Roles for New Times: Digital Curation for Preservation*http://www.arl.org/bm~doc/nrnt_digital_curation17mar11.pdf (2011).
[27] Choudhury GS. Case Study in Data Curation at Johns Hopkins University. *Library Trends* 2008; 57: 211–220.
[28] Antell K, Bales Foote J, Turner J, et al. Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries* 2014; 75: 557–574.
[29] Australian National Data Service. Information specialists and data librarian skillshttp://www.ands.org.au/working-with-data/data-management/overview/data-management-skills/information-specialists-and-data-librarian-skills (2013, accessed July 15, 2016).
[30] 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 July 15, 2016).
[31] Research Libraries UK. *The value of libraries for research and researchers*http://www.rin.ac.uk/system/files/attachments/Value_of_Libraries_-_Annexes.pdf (2011).
[32] 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.
[33] 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 (2012).
[34] Kim J. Who is Teaching Data: Meeting the Demand for Data Professionals. *Journal of Education for Library & Information Science* 2016; 57: 161–173.
[35] iSchools. Directoryhttp://ischools.org/members/directory/ (2016, accessed January 17, 2016).
[36] Dumbill E, Liddy ED, Stanton J, et al. Educating the Next Generation of Data Scientists. *Big Data* 2013; 1: 21–27.
[37] Harris-Pierce RL. Is data curation education at library and information science schools in North America adequate? *New Library World* 2012; 113: 598–613.
[38] Pouchard L. Revisiting the Data Lifecycle with Big Data Curation. *International Journal of Digital Curation* 2015; 10: 176–192.
[39] Dority GK. *Rethinking Information Work: A Career Guide for Librarians and Other Information Professionals*. Libraries Unlimited - ABC-CLIOhttp://www.abc-clio.com/ABC-CLIOCorporate/product.aspx?pc=F2685P (2016).
[40] Information SS of. *Emerging Career Trends for Information Professionals*http://ischool.sjsu.edu/sites/default/files/content_pdf/career_trends.pdf (2016).
[41] RDA. Research Data Sharing without barriershttps://www.rd-alliance.org/ (accessed February 13, 2017).
[42] Berman F, Wilkinson R, Wood J. Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance. *D-Lib Magazine*; 20. Epub ahead of print January 2014. DOI: 10.1045/january2014-berman.