Issues in Information Systems

CHARTING A NEW CURRICULUM FOR A DATA-DRIVEN WORLD

Lynn R. Heinrichs, Elon University, lheinrichs@elon.edu
Duke Hutchings, Elon University, dhutchings@elon.edu
Michele Kleckner, Elon University, mkleckne@elon.edu
Megan Squire, Elon University, msquire@elon.edu

ABSTRACT

In today's world, everything can be monitored and measured. Access to the Internet provides an endless source of data that organizations can exploit for business intelligence. The cost of storage has dwindled to pennies. Yet, having data alone is not enough; it must be mined, analyzed, and visualized in order to make sense. The explosion of data is creating demand for a new set of expertise in the workplace. The purpose of this paper is to describe a substantive curriculum shift at the authors' institution from a BS in Computer Information Systems to a BS in Information Science. The new program provides graduates with data-intensive skills that can be applied across disciplines.

Keywords: Data, Information Systems, Information Science, Curriculum

INTRODUCTION

Faculty members in the information systems field are more than accustomed to the constant challenge of delivering state-of-the-art curricula. Some changes are mostly course maintenance such as updates related to new software releases. Other changes reflect significant paradigm shifts that cause an entire upheaval of a curriculum to reflect a new direction.

The authors recently participated in a substantial effort to rechart a BS in Computer Information Systems to a BS in Information Science. The decision to undertake a dramatic change was based upon both external and internal factors. The new data-driven curriculum will be offered beginning Fall 2011 with emphasis on competencies in software, data, and statistics.

The purpose of this paper is to describe the program's rationale, purpose, design, and implementation. The curriculum has received positive support from both internal and external constituencies.

THE EXPLOSION OF DATA

According to Redman [11], information technologies are rapidly becoming commodities that are cheaply available to all. These technologies enable organizations to: (1) acquire and store vast amounts of data and information, (2) deliver data from one place to another, (3) use data for basic organizational transactions, and (4) manipulate data to create new data and information.

As the world we live in becomes more and more interconnected, the amount of data that can be collected and stored has exploded. "Whether we're talking about web server logs, tweet streams, online transaction records, "citizen science," data from sensors, government data, or some other source, the problem isn't finding data, it's figuring out what to do with it" [6]. Data is the raw material of knowledge. In today's world, everything can be monitored and measured. According to Brynjolfsson, an economist and director of the Massachusetts Institute of Technology's Center for Digital Business, "the big problem is going to be the ability of humans to use, analyze and make sense of the data" [5].

Facilitating the collection and retention of data has been the incredibly cheap cost of storage. In his book, "delete: The Virtue of Forgetting in the Digital Age," Mayer-Schonberger [8] describes how storage costs have plummeted over time:

In 1957, IBM introduced the 305, a computer with magnetic disks as storage devices that offered up to 5 megabyte of space, and which was valued at around $1 million (in 2006 terms). The cost of the storage unit alone ran to about $70,000 per megabyte in the 1950s; by 1980 that price had come down to below $500 (all in 2006 U.S. dollars), less than one percent of what it had been just two-and-a-half decades earlier. Twenty years later, in 2000, storage cost had plummeted to about 1¢, 1/50,000th of what it was in 1980. And in 2008, the cost of storage for one megabyte of information had been reduced to one hundredth of a cent. For fifty years, the cost of storage had roughly been cut in half every two years, while storage density increased 50-million fold . . . (pp. 62-63).

The Information Age is unfolding in three distinct phases [11]. The first phase involved the development of a coherent IT infrastructure by organizations; that phase is coming to an end. The second phase involves improving data accuracy by as much as two orders of magnitude. The last phase, exploiting data and information, will be full of tough issues to address.

Organizations are challenged to find ways in which their vast repositories of data can provide competitive advantage. IBM is helping organization's see the benefits and opportunities of data through its Smarter Planet Web site. The site answers the question "why data matters" by sharing insights gained from collaborations with more than 600 organizations worldwide:

> Data is being captured today as never before. It's revealing everything from large and systemic patterns—of global markets, workflows, national infrastructures and natural systems—to the location, temperature, security and condition of every item in a global supply chain. And then there's the growing torrent of information from billions of individuals using social media. They are customers, citizens, students and patients. They are telling us what they think, what they like and want, and what they're witnessing. As important, all this data is far more real-time than ever before [4].

Mike Loukides [6], a senior editor for O'Reilly Media, predicts that the "future belongs to the companies who figure out how to collect and use data successfully." This means not just companies using their own data, but also data mashed up from many other sources.

The challenge of turning data into business intelligence or new products creates the need for employees with an appropriate skill set. In an interview with *McKinsey Quarterly*, Hal Varian explained:

> The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids . . . now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it [9].

Traditional information systems programs include coursework in database management and other data-based skills. However, the traditional IS program does not really provide the complete skill set required for today's data-driven environments.

Data-Driven Skills and Employment

What skills are needed to succeed in a data-driven world? Mason and Wiggins [7] have developed a taxonomy of data science which addresses this question. "Both within the academy and within tech startups, we've been hearing some similar questions lately: Where can I find a good data scientist? What do I need to learn to become a data scientist?" Mason and Wiggins describe the five primary activities of the data scientist as obtain, scrub, explore, model, and iNterpret "(or, if you like, OSEMN, which rhymes with possum)."

Flowingdata.com [2] describes the emergence of data science as a new field and the skills of the data scientist as blending expertise areas that seem sometimes disjointed:

1. Computer Science - acquire and parse data.
2. Mathematics, statistics, and data mining - filter and mine.
3. Graphic Design - represent and refine.
4. Infovis and Human-Computer Interaction (HCI) – interaction.

The job outlook for information technology positions has remained strong, even in a weak economy. Of the 2010 graduates at the authors' institution in either CIS or computer science, 100% found related employment or enrolled in graduate school.

According to the Occupational Employment Outlook for 2008-2018 [1], the job opportunities in computer-related industries will continue to remain strong:

- The Information Sector includes fast-growing, computer-related industries. The data-processing, hosting, and related services industry, are expected to grow by 53 percent; this includes establishments that provide Web and application hosting and streaming services.

- The Professional, Scientific, and Technical Services sector includes employment in computer systems design and related services. It is expected to increase by 45 percent, accounting for nearly one-fourth of all new jobs in this industry sector. Employment growth will be driven by increasing demand for the design and integration of sophisticated networks as well as Internet and intranet sites.

Getting a handle on employment growth for data-related jobs is more complex. O'Reilly Research examined job listing data for two open source projects, Hadoop and Cassandra, which are good proxies for the market as a whole in "data-based" employment [6]. A steady year-over-year increase in job listings was evident related to the two open source projects. In addition to O'Reilly's research, the McKinsey 2011 *Big Data* report [10] showed that demand for analytical talent in the US could be 60% greater than its projected supply over the next 7 years. The report goes on to explain "The United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data. (p. 7)"

Other anecdotal data such as IBM's "Why Data Matters" campaign illustrate the need for professionals with strong data skill sets. Collecting, storing, securing, analyzing, and visualizing the data will be a high priority of leaders in many industries, providing the impetus for job growth in the information and professional services sectors.

DESIGNING A DATA-DRIVEN CURRICULUM

The change from the BS in Computer Information Systems to the BS in Information Science was driven by both external and internal factors. *Externally*, increased Internet connectivity is giving organizations greater access to services and data. There is less need to "build" applications and more ability to "buy" services. Endless sources of data are available that need to be mined for business intelligence. This shift in how organizations use IT and the problems they need to solve is changing the skill set required by entry-level professionals. Instead of being "systems" specialists, graduating students need to become "data" specialists.

Internally, the department had experienced several changes that also dictated a new direction. First, the CIS program previously shared a common set of courses with the MIS concentration in business administration; the MIS program was deleted several years ago, so the need to support it no longer existed. Second, CIS faculty hired over the last decade possessed strong backgrounds in data-related competencies. They believed that a shift from the current "systems-focused" program to a "data-focused" program was consistent with their areas of expertise. Finally, faculty members believed transition to a data-driven program also would provide greater opportunities for partnerships with other disciplines throughout the university.

The development process began with an examination of programs at other institutions. Identifying comparable programs at other institutions can be difficult since they are offered under many different names. The authors believed that programs in information science and informatics would provide a good starting point. The

Communications of the ACM [3] characterizes these types of programs as focusing on problem-solving through the application of computing or computation in the context of a problem domain. It identified examples of such undergraduate programs at several premier/research institutions with varying titles (see Table 1).

Table 1. Informatics Programs

Institution(s)	Program Title
Indiana University, Bloomington University of Michigan University of Washington University of California, Irvine	Informatics
Carnegie Mellon University	Human-Computer Interaction
Georgia Tech	Interactive Computing
Rutgers	IT and Informatics
Penn State	Information Science and Technology

Other programs investigated during the initial phase of curriculum development included James Madison (Information Analysis), College of Charleston (Discovery Informatics), and Valparaiso (Information and Decision Science).

Naming the new program was one of the challenging decisions facing the faculty. After looking at several other institutions as well as related literature, the choices were narrowed down to information science, informatics, and data science. The faculty members were not comfortable with the last option feeling that it was not yet a widely-accepted title. The choice came down to information science and informatics. While there was much support for informatics, information science (ISC) was ultimately chosen. It complemented "computer science", the other program housed in the department, maintained some recognizable continuity with the former information systems program, and accurately reflected the curricular content.

The goal of the new curriculum is best captured in the course catalog description:

> The information science discipline centers on using technology to solve a wide-range of complex problems that involve capturing, analyzing, visualizing, and managing large sets of data. Everyday more than 15 petabytes (15 quadrillion bytes) of data are generated around the world. Information scientists are concerned with creating intelligence from the data that can be used to solve complex problems like mapping the shrinking polar ice sheets, tracking the spread of infectious disease, understanding customer buying patterns, and helping people manage their health. The information science program at Elon teaches students to apply state-of-the-art tools and techniques for transforming a barrage of data to consistent, trusted, and relevant information that can provide insight and support decision making. The hands-on program provides a solid foundation in programming, data, interface design, and statistics. Because of its interdisciplinary nature, information science complements a variety of majors and minors.

A comparison of course requirements between the old BS in CIS and the new BS in Information Science is summarized in Table 2. The new program is structured to meet the following objectives in mind:

- <u>To create domain independence.</u> Information science is concerned with problems in any domain. The former BS in CIS focused on business problems; majors completed four courses in the business minor as part of the "IS Environment" component. To make the new curriculum domain independent, this requirement was eliminated.

- <u>To facilitate partnerships with other areas</u>. In addition to making the revised program domain independent, the information science major encourages partnerships with other curricula by reducing total program hours from 60 to 52. Students will be positioned better to consider a second major or minor.

- <u>To enhance problem-solving and analytical skills</u>. Today's data-intensive problems require strong analytical skills. The ISC program strengthens this core component by requiring students to complete two statistics courses (STS 212 and 327) as well as CSC 130. STS 212 and CIS130 have also been added as prerequisites to selected ISC courses.

- <u>To reduce/simplify staffing loads</u>. The former CIS program created some schedule complexities and stretched faculty resources. The revised program reduces staffing requirements by eliminating two CIS courses from the major and leveraging existing computer science courses (CSC 111 and CSC 130).

Whereas some former CIS courses were dropped altogether, other courses shifted focused and were consequently renamed. Take for example the ISC course in Human-Computer Interaction, which under the former curriculum had a complete focus on basic Web page and Web site design (and in fact was called *User-Centered Web Design*). The new HCI course will still cover elements of design for the Web, but will also include key HCI areas that intersect with Information Science, including information visualization and evaluation of interactive, information-oriented products. The course well-aligns with curriculum objectives:

- <u>Domain independence</u>: HCI principles apply to software and other interactive media for any discipline or problem
- <u>Facilitate partnerships</u>: HCI itself combines elements of psychology, sociology, design, communications, computer science, and other areas. Conversely, HCI skills can be applied to a very wide variety of disciplines
- <u>Enhance problem-solving skills</u>: In HCI, the problems to be solved often require innovative solutions to address unique human needs

A list of initial course descriptions is contained in Table 3. Information science courses use a new prefix of ISC in contrast to CIS for the former program. Courses at the authors' institution are typically four semester credit hours. ISC 111, Data Science and Visualization, is a cross-listed course with computer science that can also be used to meet a general studies requirement. Of the courses shown in Table 3, ISC111 is the only one offered every semester during the first year. As the new program gains popularity, the authors hope to build economies of scale with other courses.

The ISC major will be effective for the 2011-12 academic year. The department orchestrated a marketing blitz prior to fall registration including:

- A feature article in the student newspaper and online news page.
- An informational event at the university center that included desserts and a drawing.
- Advertisements on electronic bulletin boards throughout campus.
- A Facebook page.
- A huge poster that can be used at special events along with supplemental flyers.

In addition to these activities, the curriculum passed the internal approval process just-in-time to be included in the first-year registration materials.

FUTURE PLANS

As the new program moves forward, faculty members will focus on several critical activities:

- Transitioning program offerings from the former CIS major to the new ISC.
- Revising advising materials.

- Developing program outcomes and a program assessment strategy.
- Initiating dialogues with other programs about partnership opportunities.

The department has a solid track record of partnerships with other programs on campus such as business, geography (GIS), biology (a bioinformatics course), and English (multimedia authoring). As the new program moved through the curriculum process, the authors continuously received positive feedback and encouragement from other disciplines that were either interested in pursuing partnership opportunities.

Interest among students, departments, and prospective parents has been extremely encouraging. The authors are confident that the change in direction will energize enrollments, ignite interest with prospective employers, and create new program opportunities.

REFERENCES

1. Bureau of Labor Statistics (December 2010). *Occupational outlook handbook 2010-11 edition*, U.S. Department of Labor. Available: http://www.bls.gov/oco/oco2003.htm.
2. Flowingdata.com (June 4, 2009). *The rise of the data scientist*. Available: http://flowingdata.com/2009/06/04/rise-of-the-data-scientist/.
3. Groth, D. P. and Mackie-Mason, J.K. (February 1, 2010) Why an informatics degree, *Communications of the ACM*, 2010(2) [Online]. Available: http://cacm.acm.org/magazines/2010/2/69363-why-an-informatics-degree/fulltext.
4. IBM (n.d.). Welcome to the decade of smart, *Building a Smarter Planet: 1 in a Series.* Available*:* http://www.ibm.com/smarterplanet/global/files/us__en_us__overview__decade_of_smart_011310.pdf.
5. Lohr, S. (August 5, 2009). For today's graduate, just one word: Statistics, *The New York Times* [Online]. Available: http://www.nytimes.com/2009/08/06/technology/06stats.html.
6. Loukides, M. (June 2010). *What is data science?* [Online O'Reilly Radar Report]. Available: http://radar.oreilly.com/2010/06/what-is-data-science.html.
7. Mason, H. and Wiggins, C. (September 25, 2010). *A taxonomy of data science*. Available: http://www.dataists.com/2010/09/a-taxonomy-of-data-science/.
8. Mayer-Schonberger, V. (2009). *delete: The virtue of forgetting in the digital age*, Princeton University Press, Princeton, NJ.
9. McKinsey Quarterly (January 2009). Hal Varian on how the Web challenges managers, *McKinsey Quarterly* [Online, Visitor Edition]. Available: https://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_2286.
10. McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competitiveness, and productivity*. Available: http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf
11. Redman, T. C. (2008). *Data-driven: Profiting from your most important business asset*, Harvard Business School Publishing, Boston, MA.

Table 2. Comparison of Current BS in CIS and BS in ISC

Category	Former BS in CIS		New BS in ISC	
Required Courses	Course Title		Course Title	
	CIS 216 Programming in a Visual Environment	4 sh	CSC/ISC 111 Data Science and Visualization	4 sh
	CIS 245 Operating Systems and Networks	4 sh	ISC 245 Fundamentals of Data	4 sh
	CIS 301 Database Management and Analysis	4 sh	ISC 301 Database Management and Analysis	4 sh
	CIS 330 Systems Analysis	4 sh	ISC 310 Human Computer Interaction	4 sh
	CIS 430 Project Implementation and Management	4 sh	ISC 320 Data Mining and Analytics	4 sh
			ISC 325 Data Driven Web Development	4 sh
	Twelve hours from one of the following concentrations:		ISC 345 Information Security	4 sh
	A. Web Development :	12 sh	ISC 430 Senior ISC Experience	4 sh
	CSC 130 Computer Science I			
	CIS 310 User-Centered Web Design			
	CIS 325 Web Development			
	B. Enterprise Systems:			
	CIS 211 Management Information Systems			
	CIS 320 Building Collaborative Environments			
	CIS 345 Network Design and Security			
Electives	4 hours of upper-level CIS or CSC230	4 sh	4 hours of upper-level ISC or CSC230	4 sh
Math	MTH 206 Discrete Structures	8 sh	MTH 206 Discrete Structures	4 sh
	One course beyond core math		STS 212 Statistics in Application	4 sh
Cognate area	IS Environment:	16 sh	Computing:	
	Four courses from the Business minor		STS 327 Statistical Computing	4 sh
			CSC 130 Computer Science I	4 sh
TOTALS	Total	60 sh	Total	52 sh

Table 3. Course Descriptions for Information Science Program

CSC/ISC 111.   Data Science and Visualization
The Internet is full of rich data sources that anyone can use to answer questions and solve problems. How can we process this data to uncover interesting patterns? How can we visualize this data to reveal trends or to spur additional questions? This course teaches students how to access online data, write programs to analyze the data, and use visualization tools to describe the patterns we find in a compelling way. Students of any major are welcome. No prerequisites.

ISC 245.   Fundamentals of Data
An introduction to the storage, organization, and management of data resources. Topics include data representation, data formats, data files, data storage, and data integrity. Prerequisites: ISC111 or CSC130 or instructor permission.

ISC 301.   Database Management and Analysis
This course focuses on designing, implementing and using database systems, with emphasis on relational and object-relational models. Students design and deploy relational database models using commercial database management tools. Students will learn SQL and will be able to design complex reports and queries to answer domain problems. Pre-requisites: ISC 111 or CSC 130; ISC 245.

ISC 310.   Human-Computer Interaction
Students learn standard tools and methodologies to build usable systems, software, and other interactive user experiences. This course emphasizes understanding end-user abilities and limitations, important tasks and task properties, other aspects of requirements gathering, design principles (especially for visual interfaces), usability evaluation, and emerging trends in human-computer interaction. Pre-requisite: ISC 111 **or** CSC 130 **or** instructor permission.

ISC 320.   Data Mining and Analytics
This course provides an introduction to the concepts of data analysis and data mining using descriptive statistics, SQL, and digital visualization techniques. Students will be introduced to the many steps in the data mining process including: collection, cleaning, aggregation, transformation, mining, evaluation, presentation. Pre-requisites: ISC 301, MATH 212.

ISC 325.   Data-Driven Web Development
 This course provides a complete overview of the Web site development process. Students will create complex, interactive, data-driven Web sites using client- and server-side technologies to manage display, processing, and storage of data. Pre-requisites: ISC 301, CSC 130.

ISC 345.   Information Security
The course focuses on how organizations ensure the confidentiality, integrity, and availability of information by addressing technical security issues, people security issues, policy issues, privacy, and ethics. Pre-requisites: ISC 245, MATH 206.

ISC 430.   Senior ISC Experience
This integrative experience centers around a semester-long project in which students will be evaluated on dimensions of data design, usability, and security. Key elements of project management will be included early in the course, giving the opportunity for students to implement project management best practices in conducting their project work. Prerequisite: senior standing **or** permission of instructor.