

2015

## Applied Doctorate in IT: A Case for Designing Data Science Graduate Programs

Bryan Hosack  
*Illinois State University, bhosack@ilstu.edu*

Glen Sagers  
*Illinois State University, gsagers@ilstu.edu*

Follow this and additional works at: <http://aisel.aisnet.org/jmwais>

---

### Recommended Citation

Hosack, Bryan and Sagers, Glen (2015) "Applied Doctorate in IT: A Case for Designing Data Science Graduate Programs," *Journal of the Midwest Association for Information Systems (JMWAIS)*: Vol. 1 : Iss. 1 , Article 6.  
Available at: <http://aisel.aisnet.org/jmwais/vol1/iss1/6>

This material is brought to you by the Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Journal of the Midwest Association for Information Systems (JMWAIS) by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

Date: 12-5-2014

## Applied Doctorates in IT: A Case for Designing Data Science Graduate Programs

**Bryan Hosack**

*Illinois State University, bhosack@ilstu.edu*

**Glen Sagers**

*Illinois State University, gsagers@ilstu.edu*

### Abstract

There is an increasing shortage of applied Information Technology (IT) researchers which will affect many organizations. Programs at the Master's level can provide more technicians who can handle some of the workload, but we argue and propose that only an applied doctorate would help fill the gap in advanced skills. This approach follows similarly styled programs in other technical fields experiencing advanced skills shortages, e.g., nursing. While an applied doctorate could be viable in many areas of IT, the analytical skills needed for sophisticated data analysis requires the training that only a doctoral level program can provide. This article outlines a model for creating an applied doctorate in IT, using data science as an example. The program created by following this model could be self-sustaining by leveraging corporate funding for existing employees and sponsorship of new students to supplement more traditional sources of scholarships.

**Keywords:** Applied doctorate in IT, Data Science Graduate Programs, Funding.

*"This article presents the view and opinion of its authors and does not represent the position of the JMWAIS Editorial Board or the position of the Midwest Association for Information Systems (MWAIS)."*

Copyright © 2015 by Bryan Hosack and Glen Sagers

## 1. Introduction

Applied data science graduate programs can help fill a projected shortage of skilled data analysts with advanced research and problem solving skills. A recent study by the McKinsey Institute projects “that the United States needs 140,000 to 190,000 additional workers with ‘deep analytical’ expertise and 1.5 million more data-literate managers, whether retrained or hired” (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, and Hung Byers, 2011). In discussing business intelligence and “big data” with managers in Midwest U.S. companies, anecdotal evidence indicates this skill gap is already having a substantial impact and regional needs are increasing. According to a study by Simoneau (2014), business intelligence and analytics skills are ranked eighth on the top ten IT skills sought by employers; with companies indicating that demand has doubled since last year. Data scientists are now considered the “rock stars” in IT, requiring a unique skillset that makes them in high demand (Sadowsky, 2014).

Increasing job opportunities are encouraging for the limited number of current data scientists, especially since projections show no decrease in job growth and prospects in the near future. This need for more data analysts extends beyond that presented by the McKinsey report referenced above and further substantiated in discussions, such as Wladawsky-Berger (2014), that reaches across a number of industries (O’Neil, 2014). A strong indicator of the necessity for such programs and the number of job opportunities is the substantial increase in companies recruiting at relevant academic conferences. The few institutions currently offering data science programs at any level further highlight the opportunities for new programs (cf., Power, 2012).

Companies acknowledge the need for data scientists and staff with advanced analytical and problem-solving skills. The need for professionals who can interpret business context as well as conduct a statistical analysis of data will have continuing value to organizations. But how do organizations locate and/or develop these resources? This article explores creating more applied doctoral programs in data science as a potential solution. The approach has been used in other high demand professional fields, primarily medical, to fill highly skilled niche positions. Recently, academic departments at both Minnesota State University and Illinois State University have created Doctorate of Nursing Programs to deliver an advanced and applied focus for the nursing field. Such a program in data science potentially yields the necessary applied skills developed within a scientific, conceptual framework. The goal of an applied program should be to have 80% or more of graduates returning to industry and government as opposed to traditional doctoral programs in which 80% or more graduates go into university teaching or research labs. The emphasis on the research requirements for a dissertation will provide the needed and necessary experience beyond what is typically gained through a Master’s thesis. An applied doctorate program would provide the ability to address many of the conclusions provided by Wixom, Ariyachandra, Goul, Gray, Kulkarni, and Phillips-Wren (2011) on the state of business intelligence in academia, benefiting not only graduate students but undergraduate courses. Such a program could provide many synergistic opportunities to departments participating in delivering an applied doctorate.

The following analysis and argument is structured first through a literature review of the field and domain of a data scientist and what skills a practicing data scientist needs. Then a possible curriculum and delivery framework is discussed, followed by a discussion of the cost of delivery. Finally, the next to last section reviews a sample program, prior to brief concluding remarks.

## 2. What is a Data Scientist?

Josh Wills, a Senior Director of Data Science, defined data scientist in a Twitter Tweet as “a person who is better at statistics than any software engineer and better at software engineering than any statistician.” Power (2013, 2014) defines a data scientist as a person who has the knowledge and skills to conduct sophisticated and systematic analyses of data. A data scientist extracts insights from data sets for such things as market analytics and product development, and also evaluates and identifies strategic opportunities. He/she will become the expert for digital metrics within the company. He/she will have a unique blend of thoughtful, data-driven curiosity and pragmatism, and clearly know the difference between game-changing ideas and time sucking activities.

While a lot of hype is currently associated with data science, it is here to stay (O’Neil, 2014). Companies require skilled, analytical thinkers to manage the abundance of data collected daily. This data, when properly analyzed, can potentially provide solutions to inefficient operations and increase targeting of products and services to specific customers.

A special issue of *Quanta Magazine* (2013), “Data Driven: The New Big Science,” highlights the amazing possibilities of analyzing the increasing volumes of data from many new Big Data sources. Looking at these examples from outside the realm of business and industry can help us explore new ways of studying problems in the sphere of expertise of managers and business researchers. Ever increasing circles of overlap in data problems allow us to tie disparate businesses and diverse industries together.

Imagine in an academic setting a learning management system that tells a student what the best tool(s) are for them to use in a particular course for their academic level, course format, learning style or background. Conceptually, we have the data and technology available to begin doing so today. If that is the case, what will we be able to do tomorrow to customize education? This is part of the motivation for preparing more data scientists to assist managers in organizations to make more fact-based decisions.

While many opportunities exist, what has been called “Big Data” is still in its infancy and we need to consider biases in the tools (Crawford, 2013), ethical issues around the use of new data sources, security issues, and synthesizing how all of these solutions will work together. Data scientists need to be more than skilled technicians (Harris, 2013).

Most data scientists currently come from traditional doctoral programs in engineering, science and math or are “self-taught.” While this is certainly a viable avenue to gain many of the appropriate skills, it is unlikely that the transformative potential of analyzing the vast digital data being captured will occur without new, more specialized programs that access resources from within organizations and that have students, trained in computing, working with real data on real problems in a supervised educational setting.

### 3. What Advanced Skills Does a Data Scientist Need?

A data scientist must have a solid background in technology and statistics, but equally important is an understanding of the business and context in which they practice their craft. It is this three-legged stool of knowledge that creates a challenge for an academic institution offering traditional masters or doctoral programs. The skills for analyzing data require the expertise expected in many scientific disciplines. The increasing complexity of large data sets from nontraditional sources requires expanded expertise in statistical analysis, data retrieval and management, hypothesis generation and testing, data presentation and interpretation, and report writing and storytelling. A quick review of job advertisements (cf., Power, 2014) identifies additional skill requirements: 1) Comfortable working with large, complex data sources; 2) knowledgeable of large scale data aggregation and processing; 3) Modeling and data mining; 4) Translating analytic insights into sharp, actionable business and marketing implications, 5) Advanced quantitative reasoning, 6) Ability to partner with internal and external clients, 7) Verbal reasoning and communication skills, 8) Strategic thinking and problem solving; 9) Attention to detail; 10) Project management; 11) Ability to apply skills to solve weakly structured business problems; 12) Expert knowledge of SAS, R, Stata, SPSS, Python, C++; 13) Expert skills in SQL and experience in using databases from within SAS; and 14) Experience with Windows and UNIX/Linux operating systems. Math, engineering, and most of the hard sciences do not provide adequate training in the computing and data management skills needed for a data scientist. Further, they provide little if any training in business acumen. Finally, gaining the new technical skills is easier than developing the “soft skills” desired in this list.

Harris, Murphy, and Vaisman (2013) identify the following data science roles: Data Businessperson, Data Creative, Data Developer, and Data Researcher. These roles were determined by industry reporting how they saw themselves aligning to five skill areas: business, big data, math/operations research, programming and statistics. The scale from business person to researcher slides along these five skill spheres. The result is a diversity that goes beyond a single type of data scientist, similar to “rock” representing only a single form of music with multiple permutations as the diverse music styles blend together and share elements.

Yet, Harris, Murphy, and Vaisman (2013) see the data scientist as one who must be able to function across these five skill areas at some level with a deep focus in at least one or possibly two of the areas. This broad general knowledge combined with one specialized focus creates a “T” shaped worker, which has been used as description across the IT discipline in general. The challenge then becomes designing a curriculum to allow students develop their “T” shape in an applied doctorate.

#### 4. Curriculum Alternatives to Develop Skills

Many sources highlight the need for cross-disciplinary problem solving and inclusion of a social science perspective (Cleveland, 2001; Marchand & Peppard, 2013). Cleveland (2001), in a Bell Labs white paper, highlights skills a data science program should include:

- **Multidisciplinary Investigation** (25%) — collaboration with subject areas, such as manufacturing, business, agriculture and many others
- **Models and Methods for Data Analysis** (20%) — applied statistics and methods
- **Computing with Data** (15%) — hardware, software, and algorithms
- **Pedagogy** (15%) — how to teach the subject
- **Tool Evaluation** (5%) — keeping track of new technologies such as in-memory analytics, big data appliances, mobile solutions and others as they become available
- **Theory** (20%) — the math behind the data

Such a mix of course work builds an advanced problem solving foundation that can help meet organization's need to fill data scientist roles. To further enrich Cleveland's model curriculum, business acumen needs to be included when discussing data science. By recruiting students who currently work in a particular industry or can use an assistantship to begin to learn the business, a foundation is laid on which to build the analytics and technology skills, further reinforced via multidisciplinary coursework. The skills represented in varying proportions fit with the data science roles identified by Harris et al. (2013).

As students move through an applied program, each year should culminate in a research project drawing upon what has been learned in courses and in prior job settings. Such projects offer an opportunity for students to be mentored in solving real-world problems. As they advance through an applied data science program, students assume more responsibility as expectations increase and their knowledge of the field expands. A student's academic program should culminate with a research dissertation researching either a proprietary, business-driven or public project with faculty approval. The dissertation most likely will follow a traditional model of documenting the research problem, design, analysis and findings. Students would be encouraged to pursue multidisciplinary research develop skills to approach complicated problems from different perspectives.

#### 5. Program Delivery Alternatives

An applied doctoral program could be better suited to fill the current skill gap than a Master's program, since the focus of a doctoral program is on analytics and advanced problem solving in more applied research than traditional doctorates that are geared toward producing faculty who teach, conduct research, and provide service. Thus time typically spent in the classroom instructing in some capacity can be shifted to applying problem solving skills in an organizational setting. This type of program with its technical and practical components is an excellent fit for a team leader in data science and analytics groups. An applied program will also separate itself from traditional doctoral programs because it will focus on producing graduates to enter professional IT careers as opposed to traditional academic positions. Applied information technology data science doctoral programs could be supplemented by additional executive level programs that help address the data-literacy management gap.

An applied program can and should leverage a mix of face-to-face, blended and online delivery to provide access on a regional or national scale. Opportunities also exist to collaborate with other institutions in consortia to provide both IT and statistics course work, when applicable.

#### 6. High Cost of Program Delivery Requires the Need for Sponsorship and Support

Companies must be made aware of the expense of doctoral programs; for an applied program to work they will be required to help offset the substantial cost per student. Interested companies have multiple options for supporting such a program. Employees within an organization can be funded to go through the program with the sponsoring organization able to continue utilizing the participants' skills while they are advancing their analytics skill set. A benefit of putting current employees through such a program is leveraging their business knowledge while building their technical and statistical skills. Large companies in the Midwest have expressed interest in this model and see it as a viable option to address their long-term need providing sustainable recruitment.

Alternatively, organizations could sponsor students who are not current employees to go through the program on a co-op model working part-time (e.g., 20 hours per week) within the organization while taking classes. These assistantships would be competitive and provide a mechanism for finding students who have the prerequisite skills and dedication to be successful in the program. The assistantships will provide the ability for students to work in an organization to develop business knowledge while developing the necessary technical and statistical skills to be a successful data scientist. This allows organizations to keep key employee resources in place or move more slowly through a program while providing new resources. The unique nature of this program will provide a sustainable platform to recruit students locally, regionally, and nationally.

Finally, a successful program would also recruit students who want to attend the program using traditional funding. Students can choose to fund their own degree, and as such a program grows the potential for scholarships or other endowments could enable other students to participate.

## **7. Getting Started — Entrepreneurship in Academia**

An applied doctoral degree must combine information technology or computer science, statistics and analytics courses, potentially including business courses like marketing research and econometrics. The coursework, integrated with multiple research projects culminating in an applied research dissertation, would have a targeted time to graduation of 3-4 years. The supply gap of competent data scientists cannot be filled quickly.

Doctoral seminars in the curriculum most likely will need to be a cross-disciplinary offering. Multiple disciplines (e.g., business, agriculture, technology, education, healthcare, and others) would be assigned one or more 3-hour blocks to expose students to each of these fields. Students need to read, discuss and learn about approaches to problem solving and data analysis in that organizational area. Such courses diversify the curriculum and allow students to tailor their research interests to a particular niche while leveraging the breadth of programs offered by a research university. Faculty associated with these other disciplines would be part of a comprehensive program (e.g., running a seminar session) and would benefit from the research component of the program. By including faculty from several targeted disciplines in the research component, an applied data science program should be expected to enhance research productivity on a campus. Finally, the courses offered in a data science doctoral program could also serve other doctoral programs on a campus as electives.

Novel approaches to collaboration could be fostered through such a program and may be necessary for such a program to succeed. This could include partnerships to offer statistics or technology courses, online collaboration to deliver synchronous or asynchronous seminars to a broader student population, companies restructuring their approach to continuing education and many others. The true test for such an approach will be to develop and implement a program following such applied doctorate models as seen in healthcare, nursing and psychology.

## **8. Current State of Applied Doctorates in Academia**

There are few programs that offer applied doctorates and none within the field of IT in the United States. Currently, a program at Kennesaw State University is in the proposal phase as of late 2013 (KSU, 2013). As a field, we should look to the European model of applied doctorates or Industrial Ph.D.s (iPhD) (Ori, 2013). Since 2009, universities across Europe have begun offering programs to replace industrial researchers in their respective fields with iPhDs. This model looks to build a sustainable stream of researchers to move into industry rather than academia. The types of iPhDs or professional Ph.D.s vary in style and approach from country to country but the telling factor is that most European countries offer some form of applied doctorate or advance degree. Some of the degrees serve mid-career professionals seeking to develop skills in a certain area (Ori, 2013), similar to the model of an applied doctorate presented here. Additionally, funding follows a model similar to the one presented here with industry and organizations such as Association for International and Comparative Studies in Labour and Industrial Relations (ADAPT) in Italy (Ori, 2013), Marie Curie Actions (MCA, n.d.) or the Innovative Training Network (Euresearch, 2014) with scholarships and support from many different sectors. Furthermore, such programs include a shorter time to degree or support after receiving a degree of about three years.

Given the types of advanced degrees granted abroad, what is the current state of advanced applied or professional degrees in IT across the United States? A survey of the web, searching for programs in data science (Table 1), provides the following degree types:

Degree Type	Type of Institution (Example)	Number of Institutions Offering Degrees
Applied Doctorate	Public (Kennesaw State University)	1 (proposed)
Traditional Doctorate, Data Science	Public (University of Washington)	3
Advanced/Professional Certificate	Public and Private (Syracuse Univ)	12+
Master's Degree	Public and Private (UC Berkley, NYU, etc.)	20+

**Table 1: Web Search Survey of Graduate Data Science Programs**

The proposal at Kennesaw State University is the only applied doctorate degree that appears when searching for “applied doctorate in data science,” while only three universities and the KSU proposal appear for the search “phd in data science” in the first three pages of search results. Two of the programs were based in the United States while two are based in Great Britain. Conversely, well over 20 institutions offering Master’s degree in Data Science were found including an *Information Week* (Henschen, 2013) article highlighting the top programs in data science. Many more programs come up when searched, from South Dakota State University to Lewis College. Similarly many certificates in data science are offered in the first few pages of a Google search with programs from Indiana University to Rockhurst University. While the search is not exhaustive, it does highlight the difference between the number of advanced degrees and applied doctorates that are well below the number of data science Master’s degrees or certificates.

## 9. Ongoing Issues and Considerations

It should be pointed out in this context that many business schools may not have the in-house capability for this type of program, but that by working with the relevant disciplines, a good program can be created. A doctorate in data science could potentially be within the IS domain in business schools, but our paper suggests that an interdisciplinary approach to an applied Ph.D. program would mitigate issues with regard to expertise, resources and commitment.

Further issues include coordinating advisors from different departments or even different colleges for applied dissertation research, protecting companies who sponsor students, accreditation of the program, and addressing the diverse backgrounds of potential students. Companies and the degree will have to have sufficient flexibility to cope with these issues.

A primary concern is how long this need for advanced research skills will persist. Current indications are the trend for advanced degrees beyond a Master’s degree will continue. The challenge is creating the “rock star” that can do it all in the case of data science and meet the increasing demand (Miller, 2013; Buford, 2013).

## 10. Conclusion

The opportunity for academic institutions to innovate and diversify their approach to advanced education is upon us. The approach outlined is merely one possibility of how a program could be built to satisfy the need for advanced problem solvers in the analytics data space. Regardless of how it is structured, doctoral level programs will be necessary to impart the statistical skills needed to lead in the data analytics space. In general, North American organizations will need to consider new approaches to address the growing research skills gaps in a variety of industries, especially in the area of IT, and the approach of using applied doctorates outlined here is an opportunity that academic institutions can leverage to help both academia and industry stay competitive globally.

## 11. References

- Buford, K. (2013). Data science education on the rise. *Silicon Angle*. July 17, 2013. Retrieved from <http://siliconangle.com/blog/2013/07/17/data-science-education-on-the-rise-mitiq/>
- Cleveland, W. S. (2001). Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, 69(1) 21-26.
- Crawford, K. (2013). Hidden biases in big data. *Harvard Business Review: HBR Blog*, April 1, 2013. Retrieved from <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/>
- Euresearch (2014). *ITN: Innovative Training Networks – European Industrial Doctorates*. Promotional Flier. Euresearch: Bern, Switzerland. Retrieved from [https://www.euresearch.ch/fileadmin/redacteur/Career\\_Funding/ITN\\_EID.pdf](https://www.euresearch.ch/fileadmin/redacteur/Career_Funding/ITN_EID.pdf)
- Harris, H., S. Murphy, and M. Vaisman (2013). Analyzing the analyzers: An introspective survey of data scientists and their work. O'Reilly: Sebastopol, CA.
- Harris, J. (2013). What is the philosophy of data science? *OCDQ Blog*, March 18, 2013. Retrieved from <http://www.information-management.com/blogs/what-is-the-philosophy-of-data-science-10024090-1.html>
- Henschen, D. (2013). Big data analytics master's degrees: 20 top programs. *Information Week*. January 7, 2013. Retrieved from <http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-masters-degrees-20-top-programs/d/d-id/1108042?>
- KSU (2013). Kennesaw State University applied doctorate proposal. *Unpublished*. Retrieved from [http://www.kennesaw.edu/gpcc/Year%202013-2014/proposals/2013-09-11/Final\\_Draft\\_Ph.D.\\_in\\_Analytics\\_and\\_DS.pdf](http://www.kennesaw.edu/gpcc/Year%202013-2014/proposals/2013-09-11/Final_Draft_Ph.D._in_Analytics_and_DS.pdf)
- Marchand, D. A., and J. Peppard (2013). Why IT fumbles analytics. *Harvard Business Review*, 91, (1-2), 104-112.
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute Report*. Retrieved from [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).
- MCA (no date). *Marie Curie Actions: European Industrial Doctorate*. Program Flier. Retrieved from [http://www.fp7.cz/files/dokums\\_raw/mcaeidflyer\\_1337586999.pdf](http://www.fp7.cz/files/dokums_raw/mcaeidflyer_1337586999.pdf)
- Miller, C. (2013). Data science: The numbers of our lives. *The New York Times*. April 11, 2013. Retrieved from [http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html?pagewanted=all&\\_r=0](http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html?pagewanted=all&_r=0)
- O'Neil, M. (2014). As data proliferate, so do data-related graduate programs. *Chronicle Higher Education*, February 3. Retrieved from <http://chronicle.com/article/As-Data-Proliferate-So-Do/144363/>
- Ori, M. (2013). The Rise of industrial Ph. Ds. *University World News*. December 13, 2013. Retrieved from <http://www.universityworldnews.com/article.php?story=20131210130327534>
- Power, D. J. (2012). What Universities offer Masters Degrees in analytics and data science? *Decision Support News*, 13(20). Retrieved from <http://dssresources.com/faq/index.php?action=artikel&id=250>
- Power, D. J. (2013). What is a data scientist? *Decision Support News*, 14(13). Retrieved from <http://dssresources.com/faq/index.php?action=artikel&id=265>
- Power, D. J. (2014). What are examples of data science jobs? *Decision Support News*, 15(4). Retrieved from <http://dssresources.com/faq/index.php?action=artikel&id=296>
- Quanta Magazine (2013). Data driven: The new big science (5 article collection). *Quanta Magazine*, September 30-October 9, 2013. Retrieved from <https://www.simonsfoundation.org/quanta/20130930-imagining-data-without-division/>
- Sadowsky, T. (2014). Data scientists: The new rock stars of the tech world. *Techopedia*. July 2, 2014. Retrieved from [http://www.techopedia.com/2/28526/it-business/it-careers/data-scientists-the-new-rock-stars-of-the-tech-world?utm\\_campaign=newsletter&utm\\_medium=best&utm\\_source=07092014](http://www.techopedia.com/2/28526/it-business/it-careers/data-scientists-the-new-rock-stars-of-the-tech-world?utm_campaign=newsletter&utm_medium=best&utm_source=07092014)
- Simoneau, P. (2014). Top 10 IT skills in 2014, *Global Knowledge Site*. Retrieved from <http://www.globalknowledge.com/training/generic.asp?pageid=3635&country=United+States>
- Wills, J. (2012). Tweet attributed to @josh\_wills, *Twitter*, 9:55am 3-May 12.

Wixom, B., T. Ariyachandra, M. Goul, P. Gray, U. Kulkarni, G. Phillips-Wren (2011). The current state of business intelligence in academia. *Communications of the Association for Information Systems*, 29(16) 299-312.

Wladawsky-Berger, I. (2014). Why do we need data science when we've had statistics for centuries? *Wall Street Journal*, May 2, 2014. Retrieved from <http://blogs.wsj.com/cio/2014/05/02/why-do-we-need-data-science-when-weve-had-statistics-for-centuries/>