COVER PAGE

INSTITUTION: IUPUI

COLLEGE: School of Informatics and Computing

DEPARTMENT: Data Science

DEGREE PROGRAM TITLE: Ph.D. in Data Science

FORM OF RECOGNITION TO BE AWARDED/DEGREE CODE: Ph.D. in Data Science

SUGGESTED CIP CODE: 11.0199 (FYI: Computer and Information Sciences, Other)

LOCATION OF PROGRAM/CAMPUS CODE: IUPUI

PROJECTED DATE OF IMPLEMENTATION: January 2017

DATE PROPOSAL WAS APPROVED BY INSTITUTIONAL BOARD OF TRUSTEES:

SIGNATURE OF AUTHORIZING INSTITUITONAL OFFICER _____

DATE _____

DATE RECEIVED BY COMMISSION FOR HIGHER EDUCATION _____

COMMISSION ACTION _____

DATE _____

Program Description

**Ph.D. in Data Science To Be Offered by Indiana University at IUPUI**

## 1. Characteristics of the Program

a. Campus(es) Offering Program: IUPUI
b. Scope of Delivery (Specific Sites or Statewide): Residential
c. Mode of Delivery (Classroom, Blended, or Online): Classroom
d. Other Delivery Aspects (Co-ops, Internships, Clinicals, Practica, etc.): None
e. Academic Unit(s) Offering Program: Data Science program, School of Informatics and Computing
f. Form of Recognition to be awarded/Degree code: Ph.D.
g. Suggested CIP code: 11.0199
h. Projected Date of Implementation: January 2017

## 2. Rationale for the Program

a. Institutional Rationale (Alignment with Institutional Mission and Strengths)

- Why is the institution proposing this program?

This proposal describes a new 90-credit Ph.D. in Data Science at IUPUI with a curriculum of residential and hybrid courses that is part of a larger program initiative in Data Science in the School of Informatics and Computing at both the IUPUI and Bloomington campuses. This initiative includes the recently approved Master of Science in Informatics at IUPUI with specializations in Data Analytics, Biomedical Informatics, Knowledge and Data Management, User Experience Design, and Sports Analytics. The curriculum enjoys synergies with our school in Bloomington, which has a highly successful Master of Science and Ph.D. Minor in Data Science, with online course offerings available to students on both campuses.

On the IUPUI campus, the Ph.D. degree is to be offered and manage by the faculty affiliated with the graduate Informatics program from SOIC, which includes the Department of Human-Centered Computing, the Department of BioHealth Informatics, and the Department of Library and Information Science. As is described below, big data is becoming pervasive throughout academia and the economy, and there will be an increasing need for people who can work with big data and manage big data projects. We believe that our departments are well positioned to provide doctoral training in the knowledge, skills, and experience necessary for our students to work with and advance the boundaries of big data in academic and public and private sector settings. As can be seen in the table, Incomplete list of Data Science and cognate Doctoral programs, in Section 4.b below, at the end of 2015, there were 14 doctoral programs in Data Science offered by universities and colleges around the world, 9 of which are in the U.S.; there is no Doctoral degree offered in Data Science in the Big 10 (Pennsylvania State University has a Ph.D. in Social Data Analytics); Indiana University and the School of Informatics and Computing would be early entrants and clear leaders in this domain.

The world is awash in digital data from the web, a wide range of sensor technologies, a wider range of information and communication technologies including email and social media and an even wider range of public and private sector organizations. This is the message of the highly influential 2010 book, The Fourth Paradigm (Gray et al., 2010), published by Microsoft. Manyika et al. (2011) note that "big data has now reached every sector of the global economy. Like other essential factors of production… much of modern economic activity simply couldn't take place without it" (p. 4). An emerging grand challenge, one requiring new methodologies, techniques, and organizational understanding, is in the collection, curation, organization analysis, visualization, dissemination, and use of these heterogeneous data over the lifecycle of the data for such purposes as scientific discovery, medical advances, entrepreneurial activity, and public policy formulation. People in the public and

private sectors are taking note of this development as are academics, who are exploring ways of dealing with big data, defined by the National Science Foundation (2012) as "large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future."

The significance of big data is being recognized at the highest levels of the Federal Government's research agencies; in 2013, the Director of National Institutes of Health "announced plans to recruit a new senior scientific position, the Associate Director for Data Science" (NIH News, 2013). This follows efforts at the NIH "to build up and integrate extremely large datasets in biomedical research and to develop new informatics tools for managing and analyzing these resources" in two new initiatives: "Big Data to Knowledge (BD2K) and InfrastructurePlus" (Association of American Medical Colleges, 2012). More recently, the NSF announced program guidelines for their "BIGDATA" initiative, which "seeks novel approaches in computer science, statistics, computational science, and mathematics, along with innovative applications in domain science, including social and behavioral sciences, geosciences, education, biology, the physical sciences, and engineering that lead towards the further development of the interdisciplinary field of *data science*." (NSF, 2015)

SOIC in Bloomington is also applying for a Ph.D. in Data Science. The presence of this program on both campuses is essential to the success of joint collaborative grant applications for large NIH, NSF, and other external grants that include multiple IU stakeholders, such as the Department of Statistics, School of Public Health, Medical School, and Regenstrief Institute.

Our nation's research enterprise is positioned to see significant gains from the accelerated growth in the generation of data that are born digital, and advances in security, privacy, and ethics will be paramount to realizing the gains. The gains are furthered by the potential to link scholarly publications with the data on which the research is based and uniquely identify each dataset for discovery through a world of linked data. The economic potential of research can be realized more fully and quickly when the research can be replicated, and when data can be shared broadly and used and reused in science, citizen science, and policymaking. Currently, data sharing and reuse is severely limited, however, by a lack of adequate metadata and semantics about datasets. Without adequate metadata we find ourselves with the problem that most data are written once and never read. The data scientist, having graduated from our program with a Ph.D. in Data Science, holds the answer to this problem.

- What will completing this program prepare the students to do?

  The Ph.D. in Informatics prepares graduate who are able to define and investigate relevant research problems in Data Science. The program trains students to address such research problems with inventive and creative solutions that generate new knowledge by demonstrating a high degree of intellectual merit and the potential for broader impact. The Ph.D. curriculum trains students to produce research contributions that advance the theory and practice of Data Science.

  The program hones students' ability to (a) define, create, adapt, and apply rigorous research methods; (b) communicate research findings effectively to peers through scholarly, peer-reviewed publications that appear in international venues; (c) define, conduct, and manage a research project that involves several people and interdisciplinary expertise; (d) contribute to writing research grant proposals aimed at securing external funding to support research activities; (e) understand and address pedagogical, ethic and professional issues related to their research, including approval processes and certification for human-subject research. With these in-demand skills, the program aims at shaping graduates who can become successful researchers either in academic settings or in industrial research and development laboratories. Examples of potential occupations upon graduation in the area of specialization include Assistant Professor, Post-doctoral Research Fellow, Lecturer, Director of Research, Data Scientist, Senior Data Analyst, Strategic Innovation Manager, Big Data Consultant, and Senior Consultant.

- How is it consistent with the mission of the institution and of the school/department?

IUPUI is Indiana's premiere urban research and academic health and life sciences campus. The Ph.D. in Data Science supports this mission through the applications of Data Science to healthcare needs to improve the health and quality of life of Indiana's citizens. They also support local industry in the information technology sector. IUPUI's vision is to be one of the best urban universities, recognized locally, nationally, and internationally for its achievements. Our Ph.D. program leverages on collaborative opportunities in Indianapolis that is unique in the state.

The campus' mission is to advance the State of Indiana and the intellectual growth of its citizens to the highest levels through research and creative activity, teaching and learning, and civic engagement. Premier institutions are distinguished by the success of their Ph.D. programs. With its strong commitment to teaching and research, IUPUI promotes the educational, cultural, and economic development of central Indiana and beyond, offering a distinctive range of bachelor's, master's, professional, and Ph.D. degrees. Our proposed Ph.D. in Data science will make a substantial contribution to that portfolio. The University, through the Research Investment Fund (RIF) grant, recognizes the high productivity and impact the Ph.D. students in our existing Ph.D. in Informatics have made through their research publications and other endeavors, with our School ranking near the top of the list in terms of funding.

The proposed Ph.D. in Data Science aligns well with Indiana University's mission and vision, because it will be an exemplar of "an excellent world-class, relevant, and responsive education across a wide range of disciplines in baccalaureate, graduate, and professional education" (Principles of Excellence, 2010). By addressing the demand for professionals capable of applying data science principles to specific domains, it will also provide "leadership in creative solutions for 21st century problems." In addition, the PhD in Data Science aligns well with the University's vision, because it is an innovative program with few peers of its kind in the US and abroad.

The mission of the School of Informatics and Computing is to excel and lead in education, research, and outreach spanning and integrating the full breadth of computing and information technology, including the scientific and technical core, a broad range of applications, and human and societal issues and implications. The School aims to lead the nation in creating a new, broad, and interdisciplinary view of computing and information technology, and uses this viewpoint as the foundation of its main areas of emphasis: education and research, economic development and entrepreneurship, and diversity.

The proposed Ph.D. in Data Science is consistent with the missions of the institution because it positions the School of Informatics and Computing to become an international leader in training students to move into data science careers *with one of the first Ph.D. programs in the country focused on the management and analysis of big data and its organizational and social impacts*. There are currently many efforts to define data science.[1] The breadth of scholarship covered by the faculty teaching in the degree program on both campuses is expected to make this degree superior to most offerings by other universities and to help define the scope of data science nationally. The proposed program is the integration of cutting edge developments in computing and information and communications technologies, state of the art statistical techniques and analysis, and discipline inspired research problems, leading to a broad interdisciplinary view of data science.

- How does this program fit into the institution's strategic and/or academic plan?

  The proposed Ph.D. in Data Science aligns well with Indiana University's vision because it will be an exemplar of "an excellent world-class, relevant, and responsive education across a wide range of disciplines in… graduate, and professional education" and mission, because by addressing the demand for data scientists with Ph.D. in Data Science, it will provide "leadership in creative solutions for 21st century problems." http://pres.iu.edu/vision/principles-of-excellence/index.shtml

  The Ph.D. in Data Science is consistent with IUPUI's strategic plan for increasing capacity for graduate education. It is aligned with a number of the strategic actions including the following: "2.

[1] For example, from NIST: http://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-2.pdf

Over the next seven years, expand capacity for Ph.D. education by doubling enrollments and degree conferrals in Ph.D. programs; increasing the number of research faculty and research publications by 50 percent; increasing research support that includes funding for graduate students by 30 percent. 3. Develop unique interdisciplinary Ph.D. degrees and Ph.D. degrees, including a degree program linked to the 'big and bold' research project proposed by the task force on accelerating innovation and discovery and a program linked to an active IUPUI research center."[2] The Ph.D. in Data Science also aligns with strategic actions for leveraging IUPUI's strengths in the health and life sciences[3] and with the IUPUI Academic Plan, including its second and third goals, namely, the Signature Centers Initiative and Translating Research into Practice (TRIP).[4]

- How does this program build upon the strengths of the institution?

  The Department of Human-Centered Computing, the Department of BioHealth Informatics, and the Department of Library and Information Science in the School of Informatics and Computing at IUPUI with the support of their counterparts in Bloomington—and in collaboration with the Department of Biostatistics in the School of Public Health, the Department of Statistics in the School of Science, the School of Medicine, and the Regenstrief Institute—are uniquely able to address the increasing demand for people with a Ph.D. in Data Science with a rigorous, high-quality, and competitive program. Given the combined expertise among the faculty of the departments, we will be able to offer students a broad and innovative curriculum that will enable them to gain the knowledge and experiences needed to become intellectual leaders in this rapidly developing field. The IUPUI campus and its surrounding community are home to the IU School of Medicine, Regenstrief Institute, several major hospital systems, and pharmaceutical, scientific, and technology corporations. The School of Informatics and Computing's Indianapolis location supports strategic collaborations with these organizations.

  *Describe the student population to be served?*

  The location of the program on the IUPUI campus enables nontraditional students who are employed by local industries to pursue a Ph.D. in Data Science while remaining employed at their institutions. This is a very significant need for the local community and is also one of the primary missions of the IUPUI campus.

  *Appendix 1: Institutional Rationale, Detail (This appendix should contain links to the institution's strategic and/or academic plan or the plans themselves.)*

b. State Rationale

- How does this program address state priorities as reflected in *Reaching Higher, Achieving More*?

  The proposed Ph.D. in Data Science is a student-centered program that is "Workforce-aligned, recognizing the increasing knowledge, skills and degree attainment needed for lifetime employment and ensuring Indiana's economic competitiveness." We believe that our doctoral degree will take its place alongside "high demand academic programs that are critical to Indiana's economy," because of the expected demand for trained data scientists; it will be an example of a program demonstrating "continuous efficiency." It will have high quality instruction from the faculty of the partner units meaning that interdisciplinary and interdepartmental coordination will be the norm and will produce operational efficiencies. In terms of quality, the Ph.D. in Data Science will have clearly defined and empirically measurable student learning outcomes and rigorous learning assessment procedures. (All quotes from *Reaching Higher, Achieving More*.)

[2] http://strategicplan.iupui.edu/StudentSuccess
[3] http://strategicplan.iupui.edu/HealthLifeSciences
[4] http://www.planning.iupui.edu/599.html

The proposed Ph.D. in Data Science is critical to Indiana's economy as this program provides educational, research, and training opportunities for the citizens of Indiana in the highly critical areas of information technology innovation, which can have a direct impact on local industries such as Eli Lilly & Co., Dow Agrosciences, IBM, ExactTarget, the Indiana State government (e.g., FSSA), and Indiana healthcare systems, just to name a few. The Ph.D. affords employees from local businesses the opportunity to engage with our faculty in research, which has three significant impacts: first, it helps the faculty to engage in industry-centered research problems brought in by Ph.D. students from industry; second, the students obtain a Ph.D. while continuing their employment; and, third, input from industry can shape the curriculum, which is beneficial because informatics is a continually evolving field, and it is important for informatics to address fast changing market needs.

Information technology and computer-related occupations continue to demonstrate strong growth in Indiana and, more specifically, in central Indiana.[1] Our experience with past graduates of our existing Ph.D. programs confirm that our Ph.D. students are often from the central Indiana industry base—individuals who seek the terminal credential needed to attain promotion within their industry. In a recent article, Forbes listed Indianapolis among the top ten US cities creating tech jobs[2]—citing the lower costs of cities like Indianapolis as a key factor in tech growth.

In a broader context, the proposed program aligns with the key drivers identified in *Indiana Vision 2025*, a comprehensive, multi-year initiative to provide direction, leadership, and a long-range economic development strategy for the state of Indiana to help ensure the prosperity of Hoosiers. One of its strategic goals is to increase the proportion of Indiana residents with postsecondary credentials in STEM-related fields to "Top 5" status internationally. It states, "*developing human potential through education, training and exposure to new ways of thinking and doing things is of paramount importance, especially given current demographic trends.*" The Ph.D. in Data Science takes advantage of interdisciplinary collaborations on the two campuses to train people who will enter STEM fields as researchers and academics.

This is the right time for this program because big data is becoming pervasive throughout the economy and, according to a 2013 report by Yan, "with the escalating interest in big data, chasing after talent [with] the skills to detect value from big data has become one of the top priorities for many enterprise executives."

c. Evidence of Labor Market Need

i. National, State, or Regional Need

- Is the program serving a national, state, or regional labor market need?

Royster (2013, 9) provides an overview of the national labor market need.

> A major impediment to the widespread use of big data is the lack of workers with the appropriate training and skills. Big data work can require not only knowledge of statistical analysis and computer systems, but experience in the relevant field or industry, such as healthcare or physics.

According Manyika et al. (2011), "a shortage of people with the skills necessary to take advantage of the insights that large datasets generate is one of the most important constraints on an organization's ability to capture the potential from big data" (p. 103). They describe three types of skilled big data employees who will be in demand, the first two of which require PhD level skills (p. 104):

> • People with strong analytical and technical skills in statistics and machine learning who can analyze large volumes of data to derive business (and other) insights;

• Data-savvy managers and analysts who have the skills to be effective consumers of big data insights and who are capable of posing the right questions for analysis, interpreting and challenging the results, and making appropriate decisions; and

• Technology support personnel who develop, implement, and maintain the hardware and software tools needed to make and use big data.

To this list we add people who are skilled in the curation and preservation of big data and who can organize, preserve, and provide continued access to datasets and related digital materials throughout the data lifecycle. Manyika et al. (2011, pp. 104–105) estimate that by 2018, there will be some 450,000 positions for people with deep analytical skills and some 300,000 graduates with the requisite skills in the marketplace and project some 4 million positions for data-savvy managers with 2.5 million graduates with the requisite skills in the marketplace. This is part of the cloud computing explosion that has been projected (Microsoft, 2012) to create 14 million jobs by 2015 with cloud resources having 5–10 million servers (about 20% of the total server installations).

The broader context of IU and the State of Indiana also provide a strong support for this initiative. Indiana University is uniquely positioned in terms of computing facilities and cyber-infrastructure, in the Midwest and in the nation, to enable the implementation of the proposed degree. The expanding flow of health information, which includes the communication and processing of massive amounts of data, constitutes a major research domain for Data Science Ph.D.'s, and the robust presence of the School of Medicine, the School of Public Health, and other related entities in the IU system would create a synergic environment. The State of Indiana, as a major testbed for the implementation of Health Information Exchanges in the nation, also provides the right context for both the recruitment of students and employment of graduates. The IU Grand Challenges are building toward critical mass in additional interdisciplinary areas of research, and surfacing new interest in the proposed PhD. All in all, the co-presence of departmental expertise, infrastructure, and application areas lends a strong rationale for the establishment of the proposed degree.

The current and future national and international economy demands that all major universities contribute to economic development in at least three ways: producing graduates who are educated to become part of the technological and business workforce; engaging in partnerships with industry and government around technology and business issues; and contributing to the vitality of the economy through technology transfer.

In accordance with the stature of Indiana University as a major international research university, the program will serve labor market needs at all scales. The graduates of the program will be qualified to work at research institutions as well as in government and industry, including, for example, biomedical, banking and investment, defense, and pharmaceutical. Historically, graduates of similar programs have had an inclination towards entrepreneurship, a tendency that, when nurtured by affordances of Kelley School of Business and coupled with the state-level economical climate, can result in new high-tech companies in Indiana.

ii. Preparation for Graduate Programs or Other Benefits

- Does the program prepare students for graduate programs or provide other benefits to students besides preparation for entry into the labor market?

  N/A – the Ph.D. is a terminal degree

iii. Summary of Indiana DWD and/or U.S. Department of Labor Data

- Summarize the evidence of labor market demand for graduates of the program as gleaned from employment projections made by the Indiana Department of Workforce Development and/or the U.S. Department of Labor?

  There are no easily available statistics for the employment options for Ph.D.'s in Data Science. There is some difficulty using Bureau of Labor Statistics data because, as Royster (2013) explains, the agency "does not collect data specifically about data scientists [and instead] classifies these workers as statisticians or computer programmers or in other occupations." Nonetheless, according to BLS (2012), the number of positions available for computer and information research scientists, the category closest to data scientist, "is expected to grow by 19% from 2010 to 2020.… Computer and information research scientists are likely to enjoy excellent job prospects, as many companies report difficulties finding a sufficient number of these highly skilled workers." This represents an increase in employment from 28,200 positions in 2010 to between 32,800–38,000 in 2020.

  One path some of our graduates will follow is into the academy to become members of faculties in Informatics, Statistics, and related departments where training of data scientists takes place. In addition, graduates of this program will be qualified to work in research and other related settings in the private and public sectors.

  *Appendix 2: Summary of Indiana DWD and/or U.S. Department of Labor Data, Detail (This appendix should contain the detailed tables, upon which the summary of the labor market demand is based.)*

iv. National, State, or Regional Studies

- Summarize any national, state, or regional studies that address the labor market need for the program.

  The Technology Workforce Report: Employment Trends and the Demand for Computer-Related Talent in Central Indiana, the Central Indiana Corporate Partnership (CICP) found that computer-related occupations were one of only two occupational groups in Central Indiana that are growing faster than the national average—at 7.3% locally as compared with 6.4% nationally.

  Manyika et al. (2011) analyzed the demand for data scientists and summarized the labor situation as follows:

  > A significant constraint on realizing value from big data will be a shortage of talent, particularly of people with deep expertise in statistics and machine learning, and the managers and analysts who know how to operate companies by using insights from big data.…[the] demand for deep analytical positions in a big data world could exceed the supply being produced on current trends by 140,000 to 190,000 positions. (p. 10)

  By 2018, in the United States, 4 million positions will require these types of skills. However, if we add together the number of people with these skills and new graduates who will enter the market (on current trends), we reach a total of only 2.5 million people in the United States in 2018. So there is a potential shortfall of 1.5 million data-savvy managers and analysts.

  Vesset et al. (2012) project "the Big Data technology and services market to grow from $3.2 billion in 2010 to $16.9 billion in 2015. This represents a compound annual growth rate (CAGR) of 39.4% or about seven times that of the overall information and communication technology (ICT) market."

  The National Science Foundation, in a period of decreasing budgets, has disseminated a call for proposals for research and development in the domain of big data. Significantly, one

component of the long-term strategy the Foundation seeks to fund involves education and training Directorate for Computer & Information Science & Engineering (2012):

> This solicitation is one component in a long-term strategy to address national big data challenges, which include advances in core techniques and technologies; big data infrastructure projects in various science, biomedical research, health and engineering communities; education and workforce development...

*Appendix 3: National, State, or Regional Studies, Detail (This appendix should contain links to the studies cited or the studies themselves.)*

v. Surveys of Employers or Students and Analyses of Job Postings

- Summarize the results of any surveys of employers or students and analyses of job postings relevant to the program.

  According to the Bureau of Labor Statistics, the professional category that is the closest to a Data Science Ph.D. is "Computer and Research Scientists." With a median pay level in 2012 of $102,190, and 26,700 people working in this profession, job growth between 2012–2022 is expected to be 15%, faster than average.

  *Appendix 4: Surveys of Employers or Students and Analyses of Job Postings, Detail (This appendix should contain links to the surveys or analyses cited, or the documents themselves.)*

vi. Letters of Support

- Summarize, by source, the letters received in support of the program.

  *Appendix 5: Letters of Support, Detail (This appendix should contain the letters of support for the program.)*

**3. <u>Cost of and Support for the Program</u>**

a. Costs

i. Faculty and Staff

- Of the faculty and staff required to offer this program, how many are in place now and how many will need to be added (express both in terms of number of full- and part-time faculty and staff, as well as FTE faculty and staff)?

  All of the faculty are currently in place to deliver this Ph.D. degree program. Currently the school has more than 20 full-time tenured/tenure-track faculty members available to teach in the program from the Department of Human-Centered Computing, BioHealth Informatics, and Library and Information Science and more than 15 adjunct faculty members from other Schools at IUPUI and from the business community. Faculty will be eligible to serve on dissertation committees across the IUPUI and Bloomington campuses. Although we expect faculty to develop new courses in data science, we can deliver the entire degree with existing courses. We do intend to grow and develop both our data science faculty and course offerings over time.

  There will also be a Director responsible for overseeing the Ph.D. in Data Science on each campus. On the Bloomington campus, the Data Science Program has a Director selected by the Dean of the School of Informatics and Computing to serve a three-year term with a review and possible reappointment. On the IUPUI campus, administration of the curriculum in the Ph.D.

program will be managed by a curriculum committee composed of all affiliated faculty, initially from the existing units in SOIC.

Students pursuing the Ph.D. in Data Science will apply to the School of Informatics and Computing at IUPUI and must be accepted by the Data Science Program Admissions Committee. Once admitted, the student will be assigned to a faculty member affiliated with the Data Science program who will serve as the student's initial committee Chair. The Chair will help the student select courses, ensure that he or she has the prerequisite skills to take these courses, and will sign off on the resulting program of study.

The Graduate Coordinator and Recorder at SOIC in Indianapolis will be responsible for processing the admissions and subsequent paperwork and the administrative processes associated with the degree and certificate. They will maintain the student records including the application and related correspondence, the program of study, and any other forms generated during the student's time in the program. They will also be responsible for degree audits to make sure that the student has completed degree requirements. They will be available to answer student questions about administrative matters.

*Appendix 6: Faculty and Staff, Detail (This appendix should contain a list of faculty with appointments to teach in the program and a brief description of new faculty positions yet to be filled.)*

ii. Facilities

- Summarize any impact offering this program will have on renovations of existing facilities, requests for new capital projects (including a reference to the institution's capital plan), or the leasing of new space.

  The proposed program can be delivered with existing facilities. There will be no major impact on facilities caused by this program.

  *Appendix 7: Facilities, Detail (This appendix should contain additional information on major impacts on facilities caused by this program.)*

iii. Other Capital Costs (e.g., Equipment)

- Summarize any impact offering this program will have on other capital costs, including purchase of equipment needed for the program.

  The proposed program can be delivered with existing equipment. There will be no major new capital costs associated with this program.

  *Appendix 8: Other Capital Costs, Detail (This appendix should contain additional information on other capital costs associated with the program.)*

b. Support

i. Nature of Support (New, Existing, or Reallocated)

- Summarize what reallocation of resources has taken place to support this program.

  There will be no reallocation of resources to support this program as all Data Science courses are offered to students in the existing programs.

- What programs, if any, have been eliminated or downsized in order to provide resources for this program?

None

ii. Special Fees above Baseline Tuition

- Summarize any special fees above baseline tuition that are needed to support this program.

  There will be no special fees above baseline tuition needed to support this program.

**4. <u>Similar and Related Programs</u>**

a. List of Programs and Degrees Conferred

i. Similar Programs at Other Institutions

Campuses offering (on-campus or distance education) programs that are similar:

- CHE staff will summarize data from the Commission's Program Review Database on headcount, FTE, and degrees conferred for similar programs in the public sector, as well as information on programs in the nonprofit and proprietary sectors, to the extent possible. *CHE Appendix A: Similar Programs at Other Institutions, Detail (This appendix will contain back-up tables for the summary.)*

- Institutions may want to supplement this data with supplementary contextual information, such as relevant options or specializations or whether or not programs at other institutions are accredited or lead to licensure or certification.

  There are no related programs offered at other institutions in the state.

ii. Related Programs at the Proposing Institution

- CHE staff will summarize data from the Commission's Program Review Database on headcount, FTE, and degrees conferred for related programs at the proposing institution. *CHE Appendix B: Related Programs at the Proposing Institution, Detail (This appendix will contain back-up tables for the summary.)*

  There are no related programs on the IUB campus.

b. List of Similar Programs Outside Indiana

- If relevant, institutions outside Indiana (in contiguous states, MHEC states, or the nation, depending upon the nature of the proposed program) offering (on-campus or distance education) programs that are similar:

  Peer and other institutions have begun to enter what is not yet a crowded field as can be seen in the table below**.** As of May 2015, there were nine Ph.D. programs in Data Science in the US and five outside of the country. Significantly, only two are housed in iSchools (Pennsylvania State University and University of Washington) and only three are explicitly labeled as Ph.D. in Data Science. The University of Michigan recently announced a $100 million "Data Science Initiative" that will support faculty hiring and several research centers, but the program is not yet offering a Ph.D. degree. (http://midas.umich.edu/dsi/).

**Incomplete list of Data Science Ph.D. programs**

| **School** | **Program** |
|---|---|

| | |
|---|---|
| Brown University<br>cs.brown.edu/~kraskat/phd.html | Ph.D. in Big Data: currently looking for strong students with system-building skills and out-of-the box thinking capabilities to develop the next generation of Big Data management systems. The database group is one of the world-leading groups in systems-oriented database research, known for the H-Store, C-Store, SciDB, Aurora and Borealis systems. |
| Chapman University<br>www.chapman.edu/scst/graduate/phd-computational-science.aspx | Doctorate in Computational and Data Sciences: an interdisciplinary program that emphasizes innovative research and applications of research. After completing a set of core courses, students will specialize in an area based on shared interests with one or more faculty mentors, culminating in a final dissertation of the student's choice |
| Colorado Technical University<br>http://www.coloradotech.edu/degrees/doctorates/computer-science/big-data-analytics | Doctor of Computer Science with a concentration in Big Data Analytics: designed to develop thought leaders who have mastered the tools and techniques to analyze huge amounts of distributed, unstructured data in order to produce meaningful insight and automation for their respective organizations. |
| George Mason University<br>spacs.gmu.edu/content/Ph.D.-computational-sciences-and-informatics | Computational Science and Informatics: role of computation in science, math, and engineering |
| Kennesaw State University<br>https://datascience.kennesaw.edu/ | Ph.D. in Analytics and Data Science: It is a highly applied degree, with a research component that provides candidates with a broad range of areas of application. It sits at the intersection of mathematics/statistics/programming/visualization and communication. |
| Newcastle University EPSRC Centre for Doctoral Training in Cloud Computing for Big Data<br>http://teaching.ncl.ac.uk/ccfbd-cdt/ | Ph.D. in Cloud Computing for Big Data CDT: This new multi-million pound initiative is training the next generation of experts in the analysis of "big data" using advanced statistical methodologies and the latest cloud computing technologies, addressing an acute worldwide skills shortage. |
| New York University Polytechnic School of Engineering<br>engineering.nyu.edu/academics/programs/computer-science-phd | PhD in Computer Science with Specialization in Visualization, Databases, and Big Data |
| Pennsylvania State University<br>http://bdss.psu.edu/education | Ph.D. in Big Data Social Science: Big data Social Science is a collaboration of the social, computational, statistical, and visual sciences. IGERT Trainees will engage in a full disciplinary PhD program in one of these areas, as well as a multidisciplinary curriculum that spans these areas |
| University College London<br>www.ucl.ac.uk/statistics/news/news-010212 | Ph.D. in Big Data and its Applications: Focus on big data and its applications to social network dynamics, behavioral finance, business analytics, computational retail and healthcare |
| University of East London<br>http://www.uel.ac.uk/study/courses/data-science-profdoc.htm | Professional Doctorate in Data Science: aimed at professionals who wish to enhance and/or validate data-centric, evidence-based approaches within their chosen career through a combination of taught modules and doctoral research. |
| University of Edinburgh EPSRC Centre for Doctoral Training in Data Science<br>datascience.inf.ed.ac.uk | Ph.D. in Data Science: you will take a four-year program leading to a Ph.D. The first year provides you with the background needed to carry out research. In years 2–4, you specialize to a particular area and carry out a substantial program of original research. |
| University of Southern California Marshall School of Business<br>www.marshall.usc.edu/phd/fields/iom | Ph.D. Program: Data Sciences & Operations: DSO research examines how systems and activities need to be designed to manage the flow of goods, services and knowledge assets in firms. Our doctoral program aims to train students and help them in their |

| | |
|---|---|
| | quest for original and impactful research in these areas. |
| University of Technology Sydney cfsites1.uts.edu.au/aair/courses/detail.cfm?spk_cd=C02029&spk_ver_no=4 | Ph.D. in Analytics: enables students to advance themselves in their career in computing and information technology. It offers flexibility in the choice of topic of research so it may be closely aligned with students' professional careers |
| University of Washington escience.washington.edu/blog/new-phd-tracks-big-data | Ph.D. in Big Data: Ph.D. tracks in "Big Data" through a partnership between Computer Science & Engineering and Statistics. It will be an overlay on top of departments' regular quals requirements, leading to a new certificate en route to the Ph.D. degree. Students select three out of the following four courses: Data Management, Machine Learning, Data, Visualization, Statistics |

**Sources**:
http://datascience101.wordpress.com/2012/04/09/colleges-with-data-science-degrees/
http://whatsthebigdata.com/2012/04/12/graduate-programs-in-big-data-analyticsdata-science/
http://datascience.community/colleges

c. Articulation of Associate/Baccalaureate Programs

- For each articulation agreement, indicate how many of the associate degree credits will transfer and apply toward the baccalaureate program.

  N/A

  *Appendix 9: Articulation of Associate/Baccalaureate Programs, Detail (This appendix should contain the actual articulation agreements relevant to the proposed program.)*

d. Collaboration with Similar or Related Programs on Other Campuses

- Indicate any collaborative arrangements in place to support the program.

  N/A

## 5. Quality and Other Aspects of the Program

a. Credit Hours Required/Time To Completion

- Credit hours required for the program and how long a full-time student will need to complete the program

  A total of 90 credit hours shall be required for this degree. No more than 30 of those hours would be counted from a master's degree taken at Indiana University or a graduate program at another university. The 90 credit hours shall consist of courses in the informatics core areas, theory and methodology, specializations, minor, and dissertation work. Ph.D. students must take at least 24 credit hours of courses in Data Science at or above the 500 level. Students will be required with the help of an advisor to take courses that sample the breadth of data science.

  Minor Area Requirement
  All students shall be required to have an appropriate minor from another department in the school or from outside the school. Minors shall be selected with the advisor's recommendation. The selected minor should be appropriate for the student's choice of specialization within informatics and might include another specialization in another department in the school. Appropriate minors outside the school would include biology, chemistry, cognitive psychology, computer science, public health, business management, healthcare administration, and biostatistics. In all cases the number of hours to

be included in the minor shall be consistent with the requirements of the unit granting the minor. Some of the courses included in the minor may also be counted towards the student's Theory and Methodology requirement or other requirements. Although a minor area of emphasis within the schools may be completed in 12 credit hours, external minors are more typically 15 credit hours. The Ph.D. in Data Science provides three options:

External Minor. A student may complete an external minor awarded by another IUPUI department or graduate program that is approved by the Data Science Program. A student opting for an external minor must follow the requirements of the external department.

Internal Minor. A student may complete at least 9 Data Science credits, in courses other than reading and research, and in an area other than the student's specialization. The area and the courses must be approved by the student's advisory committee.

Individualized Minor. To complete an individualized interdisciplinary minor a student must take at least 12 credits spanning at least two Indiana University departments/degree programs, to be recommended by the student's advisory committee and approved by the Data Science Program in advance of any coursework.

A grade average of B+ (3.3) is required for the 24 credit hours of required Data Science courses. This is in addition to the University's Graduate School requirement of a B (3.0) average for all courses taken.

*Appendix 10: Credit Hours Required/Time To Completion, Detail (This appendix should contain the semester-by-semester, course-level detail on the program curriculum, including how long it will take to complete the program, assuming full-time study.)*

**Sample Data Science Ph.D. Timetable and Checklist**

The timetable is designed to guide the students to finish their Ph.D. study in 5–6 years.

Students will be evaluated every semester in the first two years and then annually to determine whether satisfactory progress is being made.

| Year 1 | Task |
|---|---|
| Semester 1 | Take courses towards fulfilling major and minor course requirements<br>Transfer credits if applicable<br>Identify research interests |
| Semester 2 | Continue enrollment towards fulfilling major and minor course requirements<br>Start research involvement via Independent Study or RAship<br>Identify faculty advisor<br>Identify minor area<br>Start discussing with faculty advisor about forming Advisory Committee |
| Milestone | Form advisory committee |
| Year 2 | |
| Semester 1 | Continue enrollment towards fulfilling major and minor course requirements<br>Conduct research under the guidance of Advisory Committee |
| Semester 2 | Final coursework to fulfill major course requirements<br>Schedule and take Qualifying Exam<br>Finish all major courses before taking Qualifying Exam<br>Courses seven years old and older must be revalidated |
| Milestone | Fulfill major course requirements<br>Pass Qualifying Exam<br>Submit research papers |
| Year 3 | |

| | |
|---|---|
| Semester 1 | Final coursework to fulfill minor course requirements<br>Apply for Ph.D. Candidacy |
| Semester 2 | Continue Enrollment (Y790)<br>From Research Committee |
| Milestone | Acquire Ph.D. candidacy<br>Form Research Committee<br>Publish research papers and present papers at conferences |
| Year 4 | |
| | Continue Enrollment (Y890)<br>You can register for Y890 after you acquire Ph.D. Candidacy<br>Define research topic and start Ph.D. Research under the guidance of Research Committee<br>Continue Ph.D. research<br>Keep publishing |
| Milestone | Research publications and conference presentations |
| Year 5 | |
| | Continue Enrollment<br>You can register for Y901 after you have completed 90 credits<br>You can register for Y901 for up to six semesters before you need to return to Y890<br>Thesis proposal |
| Milestone | Defend thesis proposal |
| Year 6 | |
| | Dissertation defense<br>At least six months after thesis proposal<br>At least 30 days after dissertation announcement<br>If dissertation defense is held more than seven years after passing Qualifying Exam, the Qualifying Exam needs to be revalidated<br>Submission of the dissertation<br>Apply for degree |
| Milestone | Pass dissertation defense |

b. Exceeding the Standard Expectation of Credit Hours

- If the associate or baccalaureate degree program exceeds 60 or 120 semester credit hours, respectively, summarize the reason for exceeding this standard expectation.

  N/A

  *Appendix 11: Exceeding the Standard Expectation of Credit Hours, Detail (This appendix should contain detailed information on why it is necessary to exceed the standard credit hour expectation, such as links to relevant licensure and/or accreditation standards the standards themselves.)*

c. Program Competencies or Learning Outcomes

- List the significant competencies or learning outcomes that students completing this program are expected to master.

  As a scholar, well prepared doctoral students in Data Science:
  • Critically evaluate the published scholarly record
  • Critically apply the theories and methodologies of data science to new research in their primary area of study
  • Apply appropriate principles, frameworks, and models to evaluate and interpret the frontiers of knowledge in their primary area of study
  • Demonstrates expository and oral communication skills appropriate to a Ph.D., publishing and presenting work in their field

• Knowledgeable in the ethical and privacy issues of data

Specific learning outcomes include
• Demonstrates advanced competency in data science tools and techniques, applied statistical analysis, and a domain area relevant to their area of concentration
• Develops a record of relevant scholarship
• Demonstrates ability to conduct independent, original research with a depth of knowledge in the chosen area of concentration

d. Assessment

- Summarize how the institution intends to assess students with respect to mastery of program competencies or learning outcomes.

Ph.D. students are assessed in formal and informal ways. Program competencies will be assessed as follows:

Learning outcome 1: Advanced competency in data science tools and techniques, applied statistical analysis, and a domain area relevant to their area of concentration.

Assessment of advanced level competency:
• Successful completion of graduate courses comprising the program of study
• Successful completion of qualifying examinations assessing depth and breadth of knowledge
• Annual review by faculty of student progress with close advising and mentoring
• Self-reported student learning outcomes as indicated in exit surveys
• Placement in positions and careers that require ability and scholarship in this field

Learning outcome 2: Record of relevant scholarship

Assessment of substantial scholarship:
• Taking elective courses outside the core discipline
• Submission and acceptance of peer-reviewed journal articles and conference papers based on dissertation work
• Achievement of student as evidenced by selection for conference presentations, peer-reviewed publications, individual grant attainment, and professional placement

Learning outcome 3: demonstrated ability to conduct independent, original research with a depth of knowledge in the chosen area of concentration

Assessment of depth of knowledge and ability to conduct independent research:
• Successfully demonstrates ability to conduct an independent research project from initial planning to fruition
• Preparation of and defense of Ph.D. dissertation proposal
• Assessment of quality of Ph.D. dissertation, including public defense
• Critical reading of dissertation by committee of graduate faculty members and a committee member from outside of the specific IU graduate program
• Review evidence of papers presented, publications, and professional networking

In addition to assessment of learning outcomes at the level of individual courses, there are three major program assessments:

• Qualifying Examination—Written: All students shall take a written qualifying examination that aims to assess the student's knowledge and readiness to carry out successful research in the discipline. The exam covers the core Data Science courses. The examination shall be set by a group of faculty who are familiar with the content of the core courses. Questions shall cover the material pertinent to the selected core courses. Individual specializations may have additional requirements, such as the preparation of a research paper or proposal. Examinations shall be offered after three or

four regular (fall or spring) semesters. A retake examination shall be offered within one month to those who have failed to pass at 80% in all subject areas. Examinations must be completed by the beginning of the student's third year in the program but can be completed before that time once the core courses have been completed. Students who do not successfully complete the examination can retake the exam a second time but only for one part, either written or oral.

• Qualifying Examination—Oral: The oral qualifying examination covers in-depth knowledge of the student's primary research area or any part of the written exam. This examination is administered by a faculty committee within the Data Science program. That committee consists of the advisor, a representative from the student's minor, a faculty member representing the student's primary research area (who can be, as appropriate, from outside the School of Informatics and Computing), and other faculty involved in preparing the written exam. The examination shall normally be completed within one month of passing the written exam. The student must pass this examination before passing on to candidacy. Only two attempts to pass this examination shall be allowed, if the student was not required to retake the written examination.

• Dissertation Proposal: The research proposal for the dissertation must be approved by the student's research committee. That committee may have the same membership as the advisory committee or the student may choose different members. The committee shall consist of at least four faculty members and the majority of the voting members should be from within the School of Informatics and Computing. The chair for the dissertation shall be a faculty member in the School of Informatics and Computing and a member of the Graduate Faculty. The chair may not be the student's primary research advisor. At least one of the other members of the committee should have a primary appointment outside of the school. The student shall defend the proposal at a public colloquium in the school.

• Dissertation Defense: A written elaboration of significant original research must be successfully presented to the Research Committee in a defense of dissertation as described in the Graduate School Bulletin. An oral defense meeting, open to the public, is required.

The Ph.D. program shall be reviewed and modified each year by the Program Committee on each campus. The program, its specializations, and individual courses shall be assessed based on their respective student learning outcomes by direct and indirect measures and with reference to the Principles of Graduate and Professional Learning. Two formal and external evaluations of the doctorate shall take place during year three and again during year five. The third year review shall be a small one to two-day review that includes an external person. The fifth year review shall be a systematic three-day review that includes three external people. These reviews are not school reviews, but rather examine the strengths and weaknesses of the Ph.D. program. In both reviews, a written set of recommendations would be delivered to the University Dean of the School of Informatics and Computing and to the Dean of the Graduate School. Prior to each of these reviews, procedures for the review process shall be established consistent with similar reviews at Indiana University and at comparable institutions.

Assessment of postgraduation outcomes will also involve an exit survey, to find out what types of employment (academic, industry, public sector) graduates have secured, and a follow-up survey three years after graduation, to find out how their careers have developed. These data will be compared with the Survey of Doctorate Recipients, a "longitudinal biennial survey conducted since 1973 that provides demographic and career history information about individuals with a research doctoral degree in a science, engineering, or health (SEH) field from a U.S. academic institution" (NSF, 2015).

Informal means of assessment include submitting and presenting conference papers, successfully publishing articles, engaging in research projects with faculty, and designing and completing independent research projects.

Admission Requirements

Candidates should have a master's degree in a related social science, health discipline, health science, or computer science with a GPA of 3.5 or higher (4.0 scale). The applicants are required to take the Graduate Record Examination (GRE) General Test (Quantitative, Verbal, and Analytical Writing). While we do not expect to institute a fixed minimum requirement, students shall be advised that successful candidates typically have scores above the 70th percentile in the verbal, quantitative, and analytic writing sections. Notably, the school will develop and implement data-driven guidelines for assessing the likely success of prospective applicants, both in the program and after degree completion.

For those applicants whose native language is not English, IUPUI requires a 79 on the Internet-based TOEFL or 550 on the paper-based TOEFL or a 6.5 on the IELTS or a G011 or higher on the IUPUI EAP Placement Exam taken from within the last two years.[5] However, because of the importance of writing skills on a program with a dissertation requirement, for the Ph.D. in Informatics, applicants should typically score above the 70th percentile (i.e., 94 on the Internet-based TOEFL). Final decisions on admission shall be made by the Data Science program's Admissions Committee.

Beyond these measures for admission, the applicants shall submit a written statement of purpose for entering the Ph.D. program, three letters of recommendation from individuals in professional positions able to judge success (two of whom would generally be tenured or tenure-track faculty), original transcripts, and a curriculum vitae.

Anticipated student clientele: The population expected to be served by this program includes graduates of bachelor's or master's degree programs in computer science, biology, chemistry, physics, informatics, information systems/science, human-computer interaction, media arts and sciences, cognitive science, psychology, education, sociology, and business, and of clinical degree programs in medicine, dentistry, nursing, pharmacy, public health, and other related fields. While the program on the Bloomington campus is designed to meet the needs of traditional students, on the Indianapolis campus, students are more likely to be nontraditional and employed in a part-time or full-time position.

e. Licensure and Certification

Graduates of this program will be prepared to earn the following:

- State License: N/A
- National Professional Certifications (including the bodies issuing the certification): N/A
- Third-Party Industry Certifications (including the bodies issuing the certification): N/A

f. Placement of Graduates

- Please describe the principal occupations and industries in which the majority of graduates are expected to find employment.

Our graduates will be qualified to compete for jobs as Data Science faculty and as data scientists: a technically-savvy specialists who can gather, organize, analyze, and interpret big data to inform decision makers, drive successful operations, and shape technology and resource investments. Related occupational titles include data analyst, quantitative developer, risk analytics lead, predictive modeling analyst, information analytics specialist and big data platform engineer. We expect our graduates to be employed in the academy, the private sector, sectors including health care, retail, pharmaceuticals, insurance, finance, manufacturing, education, service industries, and the public and nonprofit sectors.

- If the program is primarily a feeder for graduate programs, please describe the principle kinds of graduate programs, in which the majority of graduates are expected to be admitted.

[5] http://iapply.iupui.edu/graduate/english/

N/A

g. Accreditation

- Accrediting body from which accreditation will be sought and the timetable for achieving accreditation.

  N/A

- Reason for seeking accreditation.

  N/A

**6. <u>Projected Headcount and FTE Enrollments and Degrees Conferred</u>**

- Report headcount and FTE enrollment and degrees conferred data in a manner consistent with the Commission's Student Information System

  See table below

- Report a table for each campus or off-campus location at which the program will be offered

  N/A

- If the program is offered at more than one campus or off-campus location, a summary table, which reports the total headcount and FTE enrollments and degrees conferred across all locations, should be provided.

  N/A

- Round the FTE enrollments to the nearest whole number

  60

- If the program will take more than five years to be fully implemented and to reach steady state, report additional years of projections.

  The program can be completed in four to six years, depending on whether the student has a relevant master's degree at the time of enrollment.

*Appendix 1: Institutional Rationale, Detail (This appendix should contain links to the institution's strategic and/or academic plan or the plans themselves.)*

Indiana University Strategic plan
http://pres.iu.edu/vision/principles-of-excellence/index.shtml

*Appendix 2: Summary of Indiana DWD and/or U.S. Department of Labor Data, Detail (This appendix should contain the detailed tables, upon which the summary of the labor market demand is based.)*

Bureau of Labor Statistics. (2015). Occupational Outlook Handbook. Computer and Information Research Scientists. http://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011). Big data: The next frontier for innovation, competitiveness and productivity. McKinsey Global Institute.

National Institutes of Health (2013). NIH to recruit Associate Director for Data Science. *NIH News*, January 10, 2013. http://www.nih.gov/news/health/jan2013/od-10a.htm

National Science Foundation. (2015). Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA)

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767

*Appendix 3: National, State, or Regional Studies, Detail (This appendix should contain links to the studies cited or the studies themselves.)*

Directorate for Computer & Information Science & Engineering. (2012). Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA). National Science Foundation

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Royster, S. (2013). Working with Big Data. Occupational Outlook Quarterly. Bureau of Labor Statistics.

Vesset, D., Woo, B., Morris, H.D. Villars, R.L., Little, G., Bozman, J.S., Borovick, L., Olofson, C.W., Feldman, S., Conway, S., Eastwood, M., Yezhkova, N. (2012). *Market Analysis: Worldwide Big Data Technology and Services 2012–2015 Forecast*. IDC.

Yan, J. (2013). Big Data, Bigger Opportunities - Data.gov's roles: Promote, lead, contribute, and collaborate in the era of big data. President Management Council Inter-agency Rotation Program.

*Appendix 4: Surveys of Employers or Students and Analyses of Job Postings, Detail (This appendix should contain links to the surveys or analyses cited, or the documents themselves.)*

Computing Research Association. (2012). Taulbee Survey May 2012
http://cra.org/uploads/documents/resources/taulbee/CRA_Taulbee_2011-2012_Results.pdf

*Appendix 5: Letters of Support, Detail (This appendix should contain the letters of support for the program.)*

*Appendix 6: Faculty and Staff, Detail (This appendix should contain a list of faculty with appointments to teach in the program and a brief description of new faculty positions yet to be filled.)*

Faculty currently affiliated with the Data Science Program

Rachel Applegate, Ph.D. — rapplega@iupui.edu
Chair, Department of Library and Information Science — IT 561
Associate Professor, Library and Information Science — 317-278-2395

Paul Biondich, M.D. — pbiondich@regenstrief.org
Adjunct Associate Professor, Health Informatics

David Bodenhamer, Ph.D. — intu100@iupui.edu
Adjunct Professor, Library and Information Science

Davide Bolchini, Ph.D. — dbolchin@iupui.edu
Chair, Department of Human-Centered Computing — IT 595
Associate Professor, Human-Centered Computing — 317-278-5144

Malaz Boustani, MD, MPH — mboustan@iu.edu
Adjunct Professor, Bioinformatics
Professor of Medicine

Karim Boustany, PhD — karboust@iupui.edu
Adjunct Assistant Professor, Health Informatics

| | |
|---|---|
| Erin Brady, Ph.D.<br>Assistant Professor, Human-Computer Interaction | brady@iupui.edu<br>IT 591<br>317-278-7672 |
| Francesco Cafaro, Ph.D.<br>Assistant Professor, Human-Computer Interaction | fcafaro@iu.edu<br>IT 579 |
| Jake Yue Chen, Ph.D.<br>Associate Professor, Bioinformatics<br>Director, Indiana Center for Systems Biology & Personalized Medicine | jakechen@iupui.edu<br>WK 311<br>317-278-7604 |
| Andrea Copeland, Ph.D.<br>Assistant Professor, Library and Information Science | ajapzon@iupui.edu<br>IT 565<br>317-274-0114 |
| Brian Dixon, MPA, PhD, FHIMSS<br>Adjunct Assistant Professor, Health Informatics<br>Assistant Professor, Department of Epidemiology, Richard M. Fairbanks School of Public Health | bedixon@iu.edu |
| Brad Doebbeling, M.D.<br>Adjunct Professor, Health Informatics | doebbeling@asu.edu |
| Thompson N. Doman, Ph.D.<br>Adjunct Professor, Bioinformatics | domantn@lilly.com |
| Lynn Dombrowski, Ph.D.<br>Assistant Professor, Human-Computer Interaction | lsdombro@iupui.edu<br>IT 583<br>317-278-7673 |
| Stephen M. Downs, M.D.<br>Adjunct Professor of Informatics | stmdowns@iupui.edu |
| Jon D. Duke, M.D.<br>Adjunct Professor, Informatics | jonduke@iupui.edu |
| Garland C. Elmore, Ph.D.<br>Emeritus Associate Professor, Informatics | elmore@iupui.edu<br>IT 500<br>317-274-4507 |
| Anthony Faiola, Ph.D., M.F.A.<br>Associate Professor, Human-Computer Interaction | faiola@iupui.edu<br>IT 585<br>317-278-4141 |
| David Haggstrom, M.D.<br>Adjunct Associate Professor, Health Informatics | dahaggst@iupui.edu |
| Richard J. Holden, Ph.D.<br>Assistant Professor, Health Informatics | rjholden@iupui.edu<br>WK 118<br>317-278-5323 |
| Sara Anne Hook, M.B.A., J.D.<br>Director, Informatics Undergraduate Program<br>Professor, Informatics<br>Adjunct Professor, Department of American Studies, IUPUI School of Liberal Arts; Adjunct Professor, Center for Intellectual Property Law and Innovation, Robert H. McKinney School of Law; Professor, University College | sahook@iupui.edu<br>IT 589<br>317-278-7690 |
| Edgar Huang, Ph.D., M.F.A.<br>Associate Professor, Media Arts and Science | ehuang@iupui.edu<br>IT 471<br>317-278-4108 |
| Marilyn M. Irwin, Ph.D.<br>Emeritus Associate Professor, Library and Information Science | irwinm@iupui.edu |
| Sarath Chandra Janga, Ph.D. | scjanga@iupui.edu |

| | |
|---|---|
| Assistant Professor, Bioinformatics | WK 309<br>317-278-4147 |
| Josette Jones, Ph.D.<br>Director, Health Informatics<br>Associate Professor, Health Informatics & Nursing<br>Affiliate Scientist, Regenstrief Institute | jofjones@iupui.edu<br>WK 120<br>317-274-8059 |
| Benjamin Keele, M.L.S., J.D.<br>Adjunct Lecturer, Library and Information Science | bkeele@indiana.edu |
| Neil Kirby, Ph.D.<br>Adjunct Professor, Bioinformatics | nkirby@dow.com |
| Mark Larew, Ph.D.<br>Adjunct Assistant Professor, Human-Computer Interaction | mlarew@indiana.edu |
| Xiaowen Liu, Ph.D.<br>Assistant Professor, Bioinformatics | xwliu@iupui.edu<br>WK 304<br>317-278-7613 |
| Yunlong Liu, Ph.D.<br>Adjunct Associate Professor, Bioinformatics<br>Assistant Professor of Medical & Molecular Genetics<br>Adjunct Assistant Professor of Medicine (Biostatistics) | yunliu@iupui.edu |
| Karl F. MacDorman, Ph.D.<br>Associate Dean, Academic Affairs<br>Associate Professor, Human-Computer Interaction<br>Adjunct Professor, Purdue University School of Engineering and Technology, Department of Electrical and Computer Engineering | kmacdorm@iupui.edu<br>IT 475J<br>317-278-8169 |
| Steve Mannheimer, M.F.A.<br>Associate Dean, Faculty Affairs<br>Professor, Media Arts and Science | smannhei@iupui.edu<br>WK 308<br>317-278-4913 |
| Samy Meroueh, Ph.D.<br>Adjunct Associate Professor, Bioinformatics<br>Associate Professor of Biochemistry and Molecular Biology | smeroueh@iu.edu |
| Michael Mirro, MD<br>Adjunct Professor, Health Informatics | michaelmirro@gmail.com |
| Mathew Palakal, Ph.D.<br>Executive Associate Dean<br>Professor, Informatics | mpalakal@iupui.edu<br>IT 475H<br>317-278-7689 |
| Sandra Petronio, Ph.D.<br>Adjunct Professor, Informatics | petronio@iupui.edu |
| Meeta Pradhan, Ph.D.<br>Assistant Research Professor | mpradhan@iupui.edu<br>WK 306<br>317-278-0148 |
| Saptarshi Purkayastha, Ph.D.<br>Visiting Assistant Professor, Health Informatics | saptpurk@iupui.edu<br>WK 119<br>317-274-0439 |
| David Russomanno, Ph.D.<br>Adjunct Professor, Informatics | drussoma@iupui.edu |
| Katherine Schilling, Ed.D.<br>Associate Professor, Health Informatics | katschil@iupui.edu<br>WK 115<br>317-278-2372 |
| Li Shen, Ph.D.<br>Adjunct Associate Professor, Bioinformatics<br>Associate Professor of Radiology and Imaging Sciences | shenli@iupui.edu |

| | |
|---|---|
| Member of Center for Neuroimaging and Center for Computational Biology and Bioinformatics | |
| James Stevens, Ph.D.<br>Adjunct Professor, Bioinformatics | stevens_james_l@lilly.com |
| Thankam Thyvalikakath, D.M.D., M.D.S., Ph.D.<br>Adjunct Associate Professor of Informatics | tpt@iu.edu |
| Tammy Toscos, Ph.D.<br>Visiting Assistant Research Professor, Health Informatics | ttoscos@iupui.edu |
| Yang Wang, M.D.<br>Adjunct Associate Professor of Informatics | ywang1@iupui.edu |
| Huanmei Wu, Ph.D.<br>Interim Chair, Department of BioHealth Informatics<br>Associate Professor, Bioinformatics<br>Associate Professor, Computer Information Technology | hw9@iupui.edu<br>WK 303<br>317-278-1692 |
| Ayoung Yoon, Ph.D.<br>Assistant Professor, Library and Information Science | ayyoon@iupui.edu<br>IT 563<br>317-278-2388 |
| Yaoqi Zhou, Ph.D.<br>Adjunct Professor, Bioinformatics | yaoqi.zhou@griffith.edu.au |

*Appendix 7: Facilities, Detail (This appendix should contain additional information on major impacts on facilities caused by this program.)*

N/A

*Appendix 8: Other Capital Costs, Detail (This appendix should contain additional information on other capital costs associated with the program.)*

N/A

*Appendix 9: Articulation of Associate/Baccalaureate Programs, Detail (This appendix should contain the actual articulation agreements relevant to the proposed program.)*

N/A

*Appendix 10: Credit Hours Required/Time To Completion, Detail (This appendix should contain the semester-by-semester, course-level detail on the program curriculum, including how long it will take to complete the program, assuming full-time study.)*

**Curriculum:** The following courses have been identified as contributing to the Data Science program. The list is not exhaustive; courses will be deleted and new courses added as departments make changes to their curricula, and with the approval of the Data Science Curriculum Committee. All of the courses will be offered on a regular schedule, although topics may change. Students will be required to take at least one course from each cluster, including at least one course from a domain area.

The Department of Biostatistics in the School of Public Health and the Department of Statistics in the School of Science are partners because their courses are part of the methods required for the degree.

Courses

Required Core Courses (24 cr.)

- INFO I501 Introduction to Informatics (3 cr.)
- LIS S517 Web Programming (3 cr.)

- NEWM N510 Web Database Concepts (3 cr.)
- INFO H515 Introduction to Data Analytics (3 cr.)
- INFO H516 Applied Cloud Computing for Data Intensive Sciences (3 cr.)
- INFO I600 Professionalism and Pedagogy in Informatics (3 cr.)
- INFO I790 Independent Study/Rotation (3 cr.)
- INFO I575 Informatics Research Design (3 cr.)

Methods Courses (18 cr.)

- NURS-L 650 Data Analysis for Clinical and Administrative Decision-Making (3 cr.)
- NURS-R 612 Interpretive Data Analysis (2 cr.)
- PBHL-B 515 Biostatistics Practicum (3 cr.)
- PBHL-B 527 Introduction to Clinical Trials (3 cr.)
- PBHL-B 546 Applied Longitudinal Data Analysis (3 cr.)
- PBHL-B 621 Advanced Statistical Computing (3 cr.)
- PBHL-B 636 Advanced Survival Analysis (3 cr.)
- PBHL-B 646 Advanced Generalized Linear Models (3 cr.)
- PSY 60000 Statistical Inference (3 cr.)
- PSY 60100 Experimental Design (3 cr.)
- PSY 60800 Measurement Theory and Interpret Data (3 cr.)
- PSY 64000 Survey of Social Psychology I (3 cr.)
- PSY-I 643 Field Methods & Experiments (3 cr.)
- SOC-R 551 Quantitative Methods (3 cr.)
- SOC-R 559 Intermediate Soc. Statistics (3 cr.)
- STAT 51100 Statistical Methods 1 (3 cr.)
- STAT 51200 Applied Regression Analysis (3 cr.)
- STAT 51600 Basic Probability Applications (3 cr.)
- STAT 51900 Introduction to Probability (3 cr.)
- STAT 52100 Statistical Computing (3 cr.)
- STAT 52200 Sampling and Survey Techniques
- STAT 52400 Applied Multivariate Analysis (3 cr.)
- STAT 52500 Generalized Linear Model (3 cr.)
- STAT 52800 Mathematical Statistics I (3 cr.)
- STAT 52900 Applied Decision Theory and Bayesian Statistics (3 cr.)
- STAT 53600 Introduction to Survival Analysis (3 cr.)
- STAT 61900 Probability Theory (3 cr.)
- STAT 62800 Advanced Statistical Inference (3 cr.)

Specialization (18 cr.)

- Disciplinary Affinities (0–6 cr.)
- Minor (12–18 cr.)

The student must complete a minor within a domain appropriate to the chosen specialization and/or research area. All courses must be graduate-level and outside the School of Informatics and Computing.

Qualifying Examinations

- Written Exam – The student must successfully complete a written qualifying examination by the end of the program's second year. The exam is established by Data Science faculty and covers subject matter taken in the program's core courses. The exam may be retaken once.
- Oral Exam – An oral examination takes place within weeks after successful completion of the written exam. The student must pass both the written and oral exam before passing on to Ph.D. candidacy. The oral exam is based on the student's response to the written exam and core course material. The exam may be retaken once.

Dissertation (30 cr.)

A dissertation is a written elaboration of original research that makes creative contributions to the student's chosen area of specialization. The student will enroll multiple times in INFO I890 Thesis Readings and Research (1-12 cr.) while completing the dissertation. All requirements must be completed within seven years of passing the qualifying exams. The dissertation process includes the following components:

- Proposal – This is an in-depth oral review undertaken by students who have made significant progress in their research. The proposal will be defended at a public colloquium. The student must complete the proposal within one year of passing the qualifying exams.
- Defense – The student must defend his or her dissertation in an open seminar scheduled when doctoral research is almost complete.

*Appendix 11: Exceeding the Standard Expectation of Credit Hours, Detail (This appendix should contain detailed information on why it is necessary to exceed the standard credit hour expectation, such as links to relevant licensure and/or accreditation standards the standards themselves.)*

N/A

## 6. Projected Headcount and FTE Enrollments and Degrees Conferred

Institution/Location: IU School of Informatics and Computing, Bloomington

Program: Master of Science in Data Science

| | Year 1 FY2017 | Year 2 FY2018 | Year 3 FY2019 | Year 4 FY2020 | Year 5 FY2021 |
|---|---|---|---|---|---|
| Enrollment Projections (Headcount) | | | | | |
| Full-Time | 2 | 3 | 5 | 7 | 8 |
| Part-Time | 0 | 1 | 1 | 2 | 2 |
| Total | 2 | 4 | 6 | 9 | 10 |
| Enrollment Projections (FTE) | | | | | |
| Full-Time | 2 | 3 | 5 | 7 | 8 |
| Part-Time | 0 | 1 | 1 | 2 | 2 |
| Total | 2 | 4 | 6 | 9 | 10 |
| Degrees Conferred Projections | 0 | 0 | 0 | 0 | 10 |

CHE Code: 12-XX

Campus Code: XXXX

County: Marion

Degree Level: PhD

CIP Code: Federal -30.3001; State - 000000