Brothers, Sheila

From: Cramer, Aaron
Sent: Thursday, April 11, 2019 2:44 PM
To: Bird-Pollan, Jennifer; Brothers, Sheila; Ett-Mims, Joanie
Cc: Truszczynski, Miroslaw; Bush, Heather; Chen, Jin
Subject: NEW MS: Data Science
Attachments: DS-MS-DegPgmForm-v2-April-2019.pdf

Proposed New MS in Data Science

This is a recommendation that the University Senate approve, for submission to the Board of Trustees, the establishment of a new MS degree: Data Science, in the Department of Computer Science within the College of Engineering.

Rationale: In many areas of study, vast amounts of heterogeneous data are being continuously generated (i.e. Big Data). There is a nationwide shortage of data scientists who are able to analyze and make use of such data. The proposed program leverages UK's strengths in these areas to meet regional and national workforce demand. The proposed two-year interdisciplinary program features core coursework in computer science and statistics/biostatistics. The program is proposed to launch with a single concentration in biomedical informatics with future plans to develop concentrations in other application areas. The MS is being proposed as a non-thesis (Plan B) program with each student completing and defending a project in data science. Initial enrollment of four students growing to 26 students is anticipated.

Aaron

Aaron M. Cramer
Associate Professor, Electrical and Computer Engineering
Director of Graduate Studies, Electrical Engineering
Chair, Senate Academic Programs Committee
University of Kentucky
859-257-9113
aaron.cramer@uky.edu

1. *This form has two sections. Section A contains information required by the University Senate and Registrar's office and Section B contains information required by two external entities, the CPE (Council on Postsecondary Education) and SACS-COC (Southern Association of Colleges and Schools Commission on Colleges). Although only Section A is required for University Senate approval, every question must be answered to receive CPE approval. Please write "not applicable" wherever that is the appropriate response, leaving no area blank.*

2. *The CPE requires that a pre-proposal and full proposal be submitted. The pre-proposal is submitted after a proposed program has received college-level approval. Answers to questions identified with a * by the question number on this form should be used for the CPE's pre-proposal. Such questions are in both Section A and Section B. Please email OSPIE@l.uky.edu for more information about the CPE's pre-proposal process. The CPE's full proposal requires completion of both Sections A and B of this form and is submitted after approval by UK's Board of Trustees.*

3. *Once approved at the college level, your college will send the proposal to the appropriate Senate academic council (HCCC and/or GC) for review and approval. Once approved at the academic council level, the academic council will send your proposal to the Senate Council office for additional review via a committee and then to the Senate for approval. Once approved by the Senate, the Senate Council office will send the proposal to the appropriate entities for it to be placed on an agenda for the Board of Trustees. The contact person listed on the form will be informed when the proposal has been sent to committee and other times as appropriate.*

SECTION A – INFORMATION REQUIRED BY UNIVERSITY SENATE			
1. Basic Information: Program Background and Overview			
1a	Date of contact with Institutional Effectiveness[1]:	*9/23/2019*	
	☒ Appended to the end of this form is a PDF of the reply from Institutional Effectiveness.		
1b	Home College: *Engineering*		
1c	Home Educational Unit (school, department, college[2]): *Computer Science*		
1d*	Degree Type (Master's of Science, Master's of Business Administration, etc.): *MS*		
1e*	Program Name (Biology, Finance, etc.): *Data Science*		
1f*	CIP Code (provided by Institutional Effectiveness): *11.0802*		
1g	Is there a specialized accrediting agency related to this program?	Yes ☐	No ☒
	If "Yes," name:		
1h	Was this particular program ever previously offered at UK but subsequently suspended?	Yes ☐	No ☒
	If "Yes," describe. (*300 word limit*)		

[1] You can reach Institutional Effectiveness by phone or email (257-2873 or OSPIE@l.uky.edu).
[2] Only interdisciplinary graduate degrees may be homed at the college level.

1i*	Requested effective date:	☐ Fall semester following approval.	OR	☒ Specific Date[3]:*Fall 20 19*

1j*	Anticipated date for granting first degree(s): *Fall 2021*		
1k*	Contact person name: *Miroslaw Truszczynski* *Jin Chen* *Heather Bush*	Email: *mirek@cs.uky.edu* *chen.jin@uky.edu* *heather.bush@uky.edu*	Phone: *7-6738* *3-3162* *8-2080*

2. Program Overview

2a*	Provide a brief description of the proposed program. *(300 word limit)*
	This two-year interdisciplinary program in Data Science (DS) with concentration in Biomedical Informatics, will include fundamental courses from computer science and statistics/biostatistics. It will also include basic courses in biomedical informatics, required for concentraiton in Biomedical Informatics. Additional elective courses will come from related domains such as statistics and mathematics, and from applied domains such as biomedical and pharmaceutical sciences. The program will only offer a non-thesis option (Plan B) requiring 33 hours of graduate level coursework. The thesis option (Plan A) will not be offered. The program aims to integrate interdisciplinary data science training involving domain-specific knowledge discovery from large heterogeneous datasets generated from research or available publicly. Fifteen course hours will come from core DS courses, including three credit hours from the DS 710 Data Science Seminar course and three credit hours from DS 711 MS Project in Data Science course (the two proposals for DS 710 and DS 711 have been submitted to curriculog and are expected to be approved for Fall 2019). Six credit hours will come from foundational courses in bioinformatics, required for the concentration in Bioinformatics. The remaining twelve hours will be electives. The three sponsoring units for the MS in Data Science are: Department of Computer Science in the College of Engineering (represented by Miroslaw Truszczynski, CS DGS), Department of Biostatistics in the College of Public Health (represented by Heather Bush, Chair), and the Division of Biomedical Informatics in the College of Medicine (represented by Jin Chen, Interim Chief, and Hunter Moseley, Associate Director).
2b	*(similar to 13a)* What is the need for the proposed program? For example, is there a shortage of trained professionals or has an accrediting/professional/government body expressed a need for this type of program? Provide justification and evidence to support the need and demand for this proposed program. Include any data on student demand; career opportunities at the regional, state, and national levels; and any changes or trends in the discipline(s) that necessitate a new program. (*300 word limit*)
	In almost every major applied area of study, vast amounts of heterogeneous data are being continuously generated from a range of sources. For instance, immense amounts of heterogeneous biomedical data (known as the biomedical Big Data) are being continuously generated from sources such as Electronic Health Records, next-generation sequences, pathology labs, imaging systems, mobile sensors, and medical claims. Translation of biomedical data into predictive models and new actionable knowledge provides unprecedented opportunities for accelerating research in these areas. However, it requires high levels of expertise. There is a well-recognized national shortage of 250,000 of well-trained data scientists with advanced interdisciplinary skills to drive the discovery of new knowledge (2016 McKinsey Report). Publications in Science, Nature, and The Wall Street Journal illustrate the growing shortage of data scientists capable of handling heterogeneous big data. Also, the training of a new generation of data scientists who intersect computer science and informatics, statistics and mathematics, and applied domain areas, especially biomedical science, with the ability to address these big data challenges in a team setting is vital to the UK research enterprise. UK has tremendous amount of data resources, unique among all institutions in the Commonwealth if not the nation, ready to be capitalized for its research, education, and service missions. Such data resources include the CCTS Enterprise Data Trust, the

[3] Programs are effective the semester following approval. No program will be made effective unless all approvals, up through and including Board of Trustees and CPE approval, are received.

	Kentucky Cancer Registry, the Kentucky Health Data Trust, the Kentucky Research Data Center. The proposed graduate program will provide highly-trained, in-demand data scientists for UK, Kentucky, and the nation. Currently, Bellarmine University has a Master's of Science in Analytics, and Northern Kentucky University has a Bachelor's of Science in Data Science. Therefore, the proposed MS in Data Science complements these and provides UK a vital opportunity to establish itself as the premiere knowledge discovery DS program in the Commonwealth.
2c*	*(similar to 11a)* List the program objectives. These objectives should deal with how students will benefit from the program, both tangibly and intangibly. Give evidence that they will benefit. *(300 word limit)*
	The goal of this program is to provide students with advanced training in the areas of data science, big data, and analytics needed to support continued regional and national workforce demands. Specific program objectives, reflecting expectations for our students in the years following graduation, are that graduates of our program will: *1. Obtain employment and advance in careers appropriate to an advanced science/technical degree.* *2. Be leaders in the industrial sector, research and development, or entrepreneurship and business, or be pursuing further graduate study.* *3. Use their science, technical, and professional skills to make a positive impact on society and the world.* *4. Engage in continued professional development and life-long learning.*
2d*	List the student learning outcomes (SLOs) for the proposed program. (*300 word limit) (More detailed information will be addressed in Section A, part 5.)*
	Student Learning Outcomes (SLOs) reflect skills and abilities that students are expected to possess by the time they graduate. They include skills and abilities to: *1. Identify, analyze, and solve technical problems related to data science.* *2. Collect, organize, store, process, visualize, and analyze large data sets.* *3. Assemble computational pipelines to support data science from off-the-shelf or self-developed tools.* *4. Communicate technical concepts effectively, both orally and in writing.* *5. Demonstrate best practices as a data scientist according to policy, privacy, security, and ethical guidelines.*
2e	Provide the rationale and motivation for the program. Give reference to national context, including equivalents at benchmark institutions. *(150 word limit)*
	The national need for a fast-growing DS-trained workforce has been well publicized. This is particularly highlighted in the recent December 2016 McKinsey report, indicating very high demand for data scientists due to a 250,000 shortfall in the available workforce. Therefore, the establishment of the MS in Data Science degree program at UK is both compelling and timely. A careful examination of leading programs, including UCSD, NYU, Harvard, Columbia, and Virginia, indicates that the proposed graduate program is highly comparable to these programs and will provide equivalent or superior training. Moreover, our program has the flexibility to grow with this rapidly expanding and highly dynamic field, especially with respect to applied disciplines. While no new hires or courses are required for the initiation of the proposed program, both will naturally occur as this field of study continues to grow.
2f	Describe the proposed program's uniqueness within UK. *(250 word limit)*
	The only MS programs at UK somewhat related to the proposed MS degree in Data Science are the MS degree in Library Science with specialization in Health Informatics, and the MS in Statistics. However, these two programs are in fact quite different from the one being proposed. The unique aspects of the proposed program are its focus on fundamental algorithms for data collecting, organizing, storing, processing, and analyzing, and on development of skills to support big data processing and analysis required by research projects in a broad spectrum of areas from life sciences to natural sciences, engineering, and business.

2g	Describe the target audience. *(150 word limit)*		
	The target audience includes strong undergraduate students in STEM disciplines, especially in Computer Science, Information Systems and Science, Informatics, Engineering, Mathematics, Statistics, Biostatistics, and Life Sciences wishing to pursue graduate studies. This will include students planning after graduation to continue their graduate studies pursuing a PhD, as well as students wishing to get jobs in the government or in industrial and business organizations.		
2h*	Does the program allow for any concentrations?	Yes ☒	No ☐
	If "Yes," name the concentration(s). *(Specific course requirements will be described in Section A, part 7.)*		
	Concentration #1: *Biomedical Informatics*		
	Concentration #2:		
	Concentration #3:		
2j*	Are necessary resources available for the proposed new program? *(A more detailed answer is requested in Section A, part 4.)*	Yes ☒	No ☐
2k	Describe how the proposed program will be administered, including admissions, student advising, retention, etc. *(150 word limit)*		
	The MS in Data Science will be housed in and administered by the CS department. The day-to-day operation of the program will be the responsibility of the Data Science Gradute Committee consisting of 3-4 faculty members and chaired by the Director of Graduate Studies for the program. The Data Science Graduate Committee will be responsible for the admissions process, student advising, and planning and assessment, which will focus on ensuring student success. The DGS and members of the Data Science Graduate Committee will be selected from members of the faculty of record (see point 2n below).		
2l	Are multiple units/programs collaborating to offer this program?	Yes ☒	No ☐
	If "Yes," please discuss the resource contribution(s) from each participating unit/program. *(150 word limit) (Letters of support will be addressed in Part A, section 7.)*		
	The faculty from the departments of Computer Science and of Biostatistics and the Division of Biomedical Informatics will provide the majority of instruction and student advising. Each student will take 3-5 courses from the instructors from each unit. Faculty of the three units will also serve as primary advisors of the majority of students' projects.		
2m	Are there any UK programs, which the proposed program could be perceived as replicating?	Yes ☐	No ☒
	If "Yes," give a rationale for why this is not duplication, or is a necessary duplication. *(250 word limit)*		
	If "Yes," two pieces of supporting documentation are required. ☐ Check to confirm that appended to the end of this form is a letter of support from the unit chair/director who may perceive this program as a replicate. ☐ Check to confirm that appended to the end of this form is verification that the chair/director of the other unit has agreement from the faculty members of the unit. This typically takes the form of meeting minutes.		
2n	Will the faculty of record for the proposed new master's degree be the graduate	Yes ☐	No ☒

	faculty of the department/school offering the proposed new degree?		
	If "No," please describe the faculty of record for the proposed master's program, including: selection criteria; term of service; and method for adding/removing members. Will the existing director of graduate studies (DGS) in the department/school be the DGS for this proposed master's degree?		
	The faculty of record will consist of those faculty in the Departments of Computer Science and Biostatistics, and in the Divison of Biomedical Informatics that contribute to the program (by teaching courses or supervising students) and are interested in serving as members of the faculty of record of the program. The initial appointments witll be approved by the faculty of the respective units. Future additions to the faculty of record will require consent of the faculty of record at the time of the request. The DGS will be appointed by the Dean of the College of Engineering based on a recommendation from the Chair of the Computer Science Department, who will consult with the faculty of record of the program, and with the chair of the Biostatistics Department, the Director of the the Biomedical Informatics Division, and the heads of other units with significant presence in the faculty of record. The DGS will be selected from members of the faculty of record.		
2o	Will the program have an advisory board[4]?	Yes ☐	No ☒
	If "Yes," please describe the standards by which the faculty of record will select members of the advisory board, the duration of service on the board, and criteria for removal. *(150 word limit)*		
	If "Yes," please list below the number of each type of individual (as applicable) who will be involved in the advisory board.		
		Faculty within the college who are within the home educational unit.	
		Faculty within the college who are outside the home educational unit.	
		Faculty outside the college who are within the University.	
		Faculty outside the college and outside the University who are within the United States.	
		Faculty outside the college and outside the University who are outside the United States.	
		Students who are currently in the program.	
		Students who recently graduated from the program.	
		Members of industry.	
		Community volunteers.	
		Other. Please explain:	
		Total Number of Advisory Board Members	

3. Delivery Mode		**UK DLP and eLearning Office** [5]				
3a*	Initially, will any portion of the proposed program's core courses be offered via distance learning[6]?				Yes ☐	No ☒
	If "Yes," please indicate below the percentage of core courses that will be offered via distance learning.					
(check one)	1% - 24% ☐	25% - 49% ☐	50% - 74% ☐	75 - 99% ☐	100% ☐	

[4] An advisory board includes both faculty and non-faculty who are expected to advise the faculty of record on matters related to the program, e.g. national trends and industry expectations of graduates.

[5] For questions about alternative delivery modes, please contact UK's Distance Learning Programs and e-Learning office (http://www.uky.edu/DistanceLearning/).

[6] Per the Southern Association of Colleges and Schools Commission on Colleges (SACS) definition of distance education, distance education is a formal educational process in which the majority of the instruction (interaction between students and instructors and among students) in a course occurs when students and instructors are not in the same place. Instruction may be synchronous or asynchronous.

	NOTE: Programs in which 25% or more of the program will be offered via distance learning may need to submit a substantive change prospectus to SACS. Please contact institutionaleffectiveness@uky.edu for assistance. *The prospectus is required by SACS, but it is NOT required for Senate review.*	
3b*	If *any* percentage of the program will be offered via the alternative learning formats below, check all that apply, below.	
	☐	Distance learning.
	☐	Courses that combine various modes of interaction, such as face-to-face, videoconferencing, audio-conferencing, mail, telephone, fax, email, interactive television, or World Wide Web.
	☐	Technology-enhanced instruction.
	☐	Evening/weekend/early morning classes.
	☐	Accelerated courses.
	☐	Instruction at nontraditional locations, such as employer worksite.
	☐	Courses with multiple entry, exit, and reentry points.
	☐	Modularized courses.
3c	Give pedagogical rationale for the use of alternative delivery modes in the proposed program. Consider the aspects below and elaborate as appropriate. *(200 word limit)* • Synchronous and asynchronous components. • Balance between traditional and non-traditional aspects. • Hybrid elements.	

4. UK Resources			
4a*	Will the program's home educational unit require new or additional faculty?	Yes ☐	No ☒
	If "Yes," provide a plan to ensure that appropriate faculty resources are available, either within UK or externally, to support the program. Note whether the new and additional faculty will be part-time or full-time faculty. If "No," explain why. *(150 word limit)*		
	The three units offering the program currently have enough faculty to support the core MS in Data Science as laid out in this proposal.		
	If "Yes," when will the faculty be appointed? *(150 word limit)*		
4b*	Will the program's home educational unit require additional non-faculty resources, e.g. classroom space, lab space, or equipment?	Yes ☒	No ☐
	If "Yes," provide a brief summary of additional non-faculty resources that will be needed to implement this program over the next five (5) years. If "No," explain why. *(150 word limit)*		
	Additional administrative staff personel (half-time) to support the Director of Graduate Studies; *Administrative stipend for the DGS of the program equivalent to one summer month of support;* *Three TA lines (one in year 3, two in year 4, three in years five and beyond counting from the program inception date) to assist with instuction in core courses; these TA positions are necessary becaue of increased enrollments in the core courses expected when the new program is deployed.* *Initially the funds will come from the three units involved. As the program grows it is expected to generate sufficient tuition revenue to cover its cost (see later in the document).*		
4c	Will the program include courses from another educational unit(s)?	Yes ☒	No ☐
	If "Yes," list the courses and identify the other educational units and subunits that have approved the inclusion of their courses. *(150 word limit)*		
	Biostatistics, the Division of Biomedical Informatics, Mathematics, Statistics		
	If "Yes," two pieces of supporting documentation are required. ☒ Check to confirm that appended to the end of this form is a letter of support from the other units'		

	chair/director from which individual courses will be used. The letter must include demonstration of true collaboration between multiple units[7] and impact on the course's use on the home educational unit. ☒ Check to confirm that appended to the end of this form is verification that the chair/director of the other unit has consent from the faculty members of the unit. This typically takes the form of meeting minutes.

[7] Show evidence of detailed collaborative consultation with such units early in the process.

NEW MASTER'S DEGREE PROGRAM

4d	*(similar to question 19)* Fill out the faculty roster below for full-time and part-time faculty teaching major core courses in the proposed new master's program.

NAME List name & identify faculty member as FT (full-time) or PT (part-time).	**FACULTY CIP CODE**[8] List the applicable CIP Code for the faculty member.	**MAJOR CORE COURSES IN THE PROGRAM** List the major core courses in the program that the faculty member will teach and the frequency of the offering (e.g. "every spring")	**OTHER QUALIFICATIONS** *If applicable,* list any other qualifications and comment on how they pertain to the courses in the program the faculty member will teach. *If not applicable*, mark with "n/a."
Miroslaw Truszczynski	*11.07*	*CS 515 Algorithm Design, every Fall*	*occasionally taught by other faculty members*
Jinze Liu	*11.07*	*CS 405G Introduction to Data Bases, every Fall*	*Also taught by Licong Cui*
Licong Cui	*11.07*	*CS 626 Large Scale Data Science, every Fall*	*N/A*
Nathan Jacobs	*11.07*	*CS 460G Machine Learning, every Spring*	*Also taught by Brent Harrison*
Rada Nagarajan	*11.07*	*BMI 633 Introduction to Bioinformatics*	*Also taught by Jin Chen*
Sujin Kim	*11.07*	*BMI 730 Prncipals of Clinical Informatics*	*N/A*
Rama Kavuluru	*11.07*	*BMI 733 Biomedical Natural Language Processing*	*N/A*
Jin Chen	*11.07*	*BMI 734 Introduction to Biomedical Image Analysis*	*N/A*
Heather Bush	*27.05*	*STA/CPH 580, Biostatistics 1, every Fall; CPH 630, Biostatistics 2, every Spring*	*N/A*
Philip Westgate	*26.11*	*CPH 630, Biostatistics 2, every Spring*	*N/A*
David Fardo	*26.11*	*CPH 630, Biostatistics 2, every Sprin*	*N/A*
Emily Slade	*26.11*	*STA/CPH 580, Biostatistics 1, every Fall; CPH 630, Biostatistics 2, every Spring*	*N/A*

5. Assessment – Program Assessment and Student Learning Outcomes (SLOs)	
5a	Referring to program objectives, student benefits, and the target audience (questions 2c and 2g), explain how the *program* will be assessed, which is different from assessing student learning outcomes. Include how the faculty of record will determine whether the program is a success or a failure. List the benchmarks, the assessment tools, and the plan of action if the program does not meet its objectives. *(250 word limit)*
	The program-level objectives are for students graduating from the program to have professional skills that will allows them to *1. Obtain employment and advance in careers appropriate to an advanced science/technical degree.* *2. Be leaders in the industrial sector, research and development, or entrepreneurship and business, or be pursuing further graduate study.* *3. Use their science, technical, and professional skills to make a positive impact on society and the world.* *4. Engage in continued professional development and life-long learning.* *These program-level outcomes will be assessed primarily based on data gathered from regular alumni surveys, including job placement data and self-assessed alumni satisfaction with the professional skills acquired in the program in support of objectives 1-4. The survey will be administered and analyzed by the graduate committee every three years.* *Together with the alumni survey data, every three years the graduate committee will also review secondary measures of the overall quality of the program: the appropriateness of core courses in their support of objectives 1-4; the relevance of final projects to objectives 1-4; time-to-graduation; and full-time and part-time enrollment numbers, GPA.*
5b	*(related to 2d and 14.c)* Based on the SLOs from question 2c, append a PDF of the program's curriculum map[9] to the end of this form.
5c	Append an assessment plan[10] for the SLOs to the end of this form.

6. Non-Course Requirements			
6a	Will the program require completion of a bachelor's degree from a fully accredited institution of higher learning?	Yes ☒	No ☐
	If "No," explain below. (*150 word limit*)		
6b	The Graduate School requires applicants to have an overall GPA of 2.75 on undergraduate work. Will the program have a higher undergraduate GPA requirement?	Yes ☒	No ☐
	If "Yes," describe below. (*150 word limit*)		
	GPA of 3.0. Successful applicants will come from STEM areas.		

[9] Course mapping (or "curricular mapping") is a representation of how faculty intend to approach and assess each of the student learning outcomes identified for the courses for the degree program, with an emphasis on only those courses required for all degree candidates. It is a master chart that indicates which objectives are being met, to what extent, and how often. This identifies whether an objective is "introduced," "developed," and/or "mastered" within a given course; it may be helpful also to chart any classroom-based assessment measures used to demonstrate that claim.

[10] An assessment plan is typically a tabular grid that illustrates the artifacts, rubrics, assessment team, and periods of assessment for the SLOs.

6c	Will the proposed program include requirements for testing (e.g. GRE, GMAT, TOEFL) to be considered for admission?	Yes ☒	No ☐
	If "Yes," name each test and describe the specific requirements, scores, etc. below. (*150 word limit*)		
	GRE quantitative and verbal tests are required, as well as TOEFL or IELTS for international or ESL applicants. There are no explicit minimum scores beyond any criteria established by the UK Graduate School, but the overall strength of the application package, including standardized test scores, will be considered in its entirety when evaluating candidates for admission.		
6d	Will the program have a world language requirement?	Yes ☐	No ☒
	If "Yes," describe below. (*150 word limit*)		
6e	The Graduate School allows transfer of up to nine credits or 25% of course work. Please describe transfer credit limitations below for the proposed program. (*150 word limit*)		
	The program will follow the general Graduate School policies for credit transfer.		
6f	Will the program have a thesis requirement (Plan A)? (If "Yes," explain the requirements below. If "No," proceed to question 6g)	Yes ☐	No ☒
6g	Will the program have a non-thesis requirement (Plan B)? (If "Yes," explain the requirements below. If "No," proceed to question 6h)	Yes ☒	No ☐
	If "Yes," explain the requirements below.		
	Every student will be required to complete a final project.		
6h	Provide the final examination criteria.	Yes ☒	No ☐
	As part of the required course DS 711 MS Project in Data Science, students are required to provide a written project report accompanied by an oral defense to a faculty committee established according to Graduate School policies. Committee members examine the technical competency of students at the oral defense, which acts as the program final exam.		
6i	Describe termination criteria.	Yes ☒	No ☐
	Students must meet all requirements of the Graduate School. In particular, they must maintain a GPA of 3.0 or better. Students whose GPA fals under 3.0 are placed on probation; if they do not meet conditions set in the probation letter, they are removed from the program.		

7. Course Requirements.

7a	Document the total credit hours required by level below. At least two-thirds of the minimum requirements for the master's or specialist degree must be in regular courses, and at least half of the minimum course requirements (excluding thesis, practicum, or internship credit) must be in 600- or 700-level courses.			
	400G-level: *25%*	500-level: *25%*	600-level: *25%*	700-level: *25%*

7b*	What is the total number of credit hours required for the degree?[11] (e.g. 24, 32)	*33*
	If an explanation about the total credit hours is necessary, use the space below. *(150 word limit)*	
	33 course work credits, at least 18 credits of the course work must be taken at the 600 or 700 level.	

Use the grids below to list core courses, electives, courses for a concentration, etc.
Use the course title from the Bulletin or from the most recent new/change course form.

7c* **Program Major Core Courses.** These courses are required for all students in the program and include prerequisite courses. Check the appropriate box to describe the course as either "program core" or "prerequisite."

Prefix & Number	Course Title	Type of Course	Credit Hrs	Course Status[12]
CS626	*Large Scale Data Science*	☒ Pgm Core ☐ Prerequisite	*3*	No Change
CPH580	*Biostatistics I (same as STA580)*	☒ Pgm Core ☐ Prerequisite	*3*	No Change
CPH630	*Biostatistics II (same as STA681)*	☒ Pgm Core ☐ Prerequisite	*3*	No Change
DS710	*Data Science Seminar (three credits of this course are required; will be repeated three times)*	☒ Pgm Core ☐ Prerequisite	*1*	New
DS711	*MS Project in Data Science*	☒ Pgm Core ☐ Prerequisite	*3*	New
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core ☐ Prerequisite		Select one....
		☐ Pgm Core		Select one....

[11] A non-thesis option (Plan B) requires that six or more graduate credit hours of course work be submitted in lieu of a thesis.
[12] Use the drop-down list to indicate if the course is a new course ("new"), an existing course that will change ("change"), or if the course is an existing course that will not change ("no change").

		☐ Prerequisite		
		*Total **Core Courses** Credit Hours:*	*15*	
7d	Is there any narrative about prerequisite courses for the program that should be included in the Bulletin? If "Yes," note below. (*150 word limit*)	Yes ☒	No ☐	
	The program will be of interest to students with diverse backgrounds. Therefore, the program has no specific prerequisites beyond the requirement of the BS degree in a STEM discipline and the undergraduate GPA of 3.0.			
7e	Is there any narrative about core courses for the program that should be included in the Bulletin? If "Yes," note below.	Yes ☒	No ☐	
	Every student must complete the core requirement, which consists of completing core courses earning B or better in each of them. *Upon approval of the DGS of the program, the core courses CPH/STA580 and CPH630/STA681 may be substituted by mathematically more rigorous courses of similar scope.*			
	Program Guided Electives[13] (Guided electives for all students in the program.)			
7f*	Does the program include any guided electives? (If "Yes," indicate and note the specific courses in the grid below. If "No," indicate and proceed to question 7i.)	Yes ☐	No ☒	
7g*	Using the grid provided, list the guided electives below.			

Prefix & Number	Course Title	Credit Hrs	Course Status[14]
CS405G	*Introduction to Data Bases*	*3*	No Change
CS460G	*Machine Learning*	*3*	No Change
CS515	*Algorithm Design*	*3*	No Change
			Select one....
			Select one....
			Select one....
			Select one....
			Select one....
			Select one....
			Select one....
Total Credit Hours as Guided Electives:			

7h	Is there any narrative about guided electives courses that should be included in the Bulletin? If "Yes," note below. (*150 word limit*)	Yes ☒	No ☐
	Every student must complete at least one of the three guided electives: CS 405G Introduction to Data Bases, CS 460G Machine Learning, and CS 515 Algorithm Design		
	Program Free Electives[15]**.** (Free electives for all students in the program.)		

[13] Guided electives are available to all students in the program and are organized as groups of elective courses, from which a student chooses one (or two, or three, etc.).

[14] Use the drop-down list to indicate if the course is a new course ("new"), an existing course that will change ("change"), or if the course is an existing course that will not change ("no change").

7i*	Does the program include any free electives? (If "Yes," indicate and proceed to question 7j. If "No," indicate and proceed to 7l.)	Yes ☒	No ☐
7j*	What is the total number of credit hours in free electives?	*9 credit hours*	
7k	Provide the free electives courses language that will be included in the Graduate School Bulletin. (*150 word limit*)		
	Nine credit hours of free electives will complement six credit hours required by the concentration (see below) and the fifteen credit hours of core courses and to satisfy the total 33 credit hour requirement. All electives must be approved by the DGS. At least two free electives must be at the 600 or 700 level. Moreover, free electives must include at least two courses (6 credits) with a strong data science component in the subject area of the student's project.		
	Courses for a program's concentration(s).		
	Click HERE for a template for additional concentrations[16].		
7l	Does the program include any concentrations? (If "Yes," indicate and proceed to question 7m. If "No," indicate and proceed to 7p.)	Yes ☒	No ☐
7m	Concentration name: *Biomedical Informatics*		

Prefix & Number	Course Title (Check the appropriate box to describe the course as "a core course for the concentration" or "an elective course for the concentration.")		Credit Hrs	Course Status[17]
BMI633	*Introduction to Bioinformatics*	☒ Core ☐ Elective	*3*	No Change
BMI730	*Principles of Clinical Informatics*	☐ Core ☒ Elective	*3*	No Change
BMI733	*Biomedical Natural Language Processing*	☐ Core ☒ Elective		No Change
BMI734	*Introduction to Biomedical Image Analysis*	☐ Core ☒ Elective		No Change
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core		Select one....

[15] Program free electives are available to all students in the program (regardless of any concentration(s)) and the choice of which course(s) to take is up to the student. Courses are not grouped but can be described as "student must take three courses at the 600-level or above."

[16] Append a PDF with each concentration's courses to the end of this form.

[17] Use the drop-down list to indicate if the course is a new course ("new"), an existing course that will change ("change"), or if the course is an existing course that will not change ("no change").

		☐ Elective		
		☐ Core ☐ Elective		Select one....
7n	Provide concentration-related language that should be included in the Graduate School Bulletin. (150 word limit)			
	In order to earn the MS degree in Data Science with the concentration in Biomedical Informatics students must complete fifteen credit hours of core courses for the MS degree in Data Science, the core course for the concentration (BMI 633 - Introduction to Bioinformatics), one of three electives for the concentration (BMI 730 - Principles of Clinical Informatics, BMI 733 - Biomedical Natural Language Processing or BMI 734 - Introduction to Biomedical Image Analysis), and three free electives as specified above.			
7o	Does the program have an additional concentration? (If "Yes," indicate and proceed to question 7p. If "No," indicate and proceed to 7r.)		Yes ☐	No ☒
7p	Concentration #2 Name:			
Prefix & Number	Course Title (Check the appropriate box to describe the course as "a core course for the concentration" or "an elective course for the concentration.")		Credit Hrs	Course Status[18]
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		☐ Core ☐ Elective		Select one....
		Total Credit Hours, Concentration #2:		
7q	Provide concentration-related language that should be included in the Graduate School Bulletin for the second concentration. (150 word limit)			

[18] Use the drop-down list to indicate if the course is a new course ("new"), an existing course that will change ("change"), or if the course is an existing course that will not change ("no change").

7r	Is there anything else about the proposed program that should be mentioned? *(150 word limit)*

8. Degree Plan

8a	Create a degree plan for the proposed program by listing in the table below the courses that a typical student would take each semester. Use the spaces for "Year 3" only if necessary. If multiple concentrations are available, click HERE for a template for additional concentrations. Append a PDF with each concentration's semester-by-semester program of study to the end of this form.			
	YEAR 1 - FALL:	*BMI 633* *CS 626* *CPH 580* *DS 710 (1 credit hour)*	**YEAR 1 - SPRING:**	*BMI730 or BMI733 or BMI734* *CS405G or CS460G or CS515* *CPH630* *DS 710 (1 credit hour)*
	YEAR 2 - FALL :	*Elective* *Elective* *Elective* *DS 710 (1 credit hour)*	**YEAR 2 - SPRING:**	*DS 711*
	YEAR 3 - FALL:		**YEAR 3 - SPRING:**	

8b	With reference to the degree plan above, explain how there is progression in rigor and complexity in the courses that make up the program. *(150 word limit)*
	Students complete core course requirements in the first year. In the third semester, they progress to more advanced courses (electives), with at least two of them containing a strong data science component in the student's area of focus. As part of the DS 710 Data Science Seminar course, they identify a specific problem of interest and work through the process of problem characterization, literature review and study of prior work, hypothesized solution and detailed experimental design, implementation, data analysis, and preparation of the written report. Students work closely with their project advisor to understand and move through this process. The effort typically starts in the third semester but may start as early as in the second semester for some students. In the last semester students are enrolled in the DS 711 MS Project in Data Science and work on their final project. The program culminates with the final MS exam (project defense).

9. Approvals/Reviews

Information below does not supersede the requirement for individual letters of support from educational unit administrators and verification of faculty support (typically takes the form of meeting minutes).

	Reviewing Group Name	**Date Approved**	**Contact Person Name/Phone/Email**
9a	(Within College) *In addition to the information below, attach documentation of department and college approval. This typically takes the form of meeting minutes but may also be an email from the unit head reporting department- and college-level votes.*		
			/ /
			/ /
			/ /
			/ /

9b	(Collaborating and/or Affected Units)		
			/ /
			/ /
			/ /
			/ /
			/ /
			/ /
			/ /
			/ /
			/ /

9c	(Senate Academic Council)	**Date Approved**	**Contact Person Name**
	Health Care Colleges Council (if applicable)		
	Graduate Council	*2/21/2019*	*Roshan Nikou*

SECTION B – INFORMATION REQUIRED BY CPE AND SACS	
10. Program Overview – Program Quality and Student Success	
10a*	Highlight any distinctive qualities of the proposed program. Are any faculty nationally or internationally recognized for expertise in this field? Does this program build on the expertise of an existing locally, nationally, or internationally recognized program at UK? *(300 word limit)*
	The University began investing in biomedical informatics in 2008 when UK became one of the 60 national (and only one in Kentucky) Clinical and Translational Science Awardees (CTSA). A unique feature of this university is that it is comprehensive, the UK health colleges are located side by side with the main Lexington campus; we are one University spanning the entire spectrum of academic colleges. With the CTSA and the creation of the Center for Clinical and Translational Science (CCTS), the Division of Biomedical Informatics, the reinvigorated Center for Computational Science, and the continued NCI-designated Markey Cancer Center (the only NCI-designated Cancer Center in Kentucky), we are uniquely positioned to leverage this multidisciplinary, collaborative infrastructure with tremendous strengths in established academic departments of Computer Science and Biostatistics. UK also has significant data resources, unique among all institutions in the Commonwealth, if not the nation, ready to be capitalized for its research, education, and service missions. Such data resources include the CCTS Enterprise Data Trust, the Kentucky Cancer Registry, the Kentucky Health Data Trust, and the Kentucky Research Data Center. The proposed graduate program will provide highly-trained, in-demand data scientists for UK, Kentucky, and the nation. *The UK MS in Data Science will be distinguished in its scientific depth and breadth, its nationally recognized data science faculty, and the extensive data and computational resources listed above. The program will serve the UK, Kentucky and the nation as the premiere knowledge discovery data science program.*
10b*	*(similar to 2b)* What are the intended student learning outcomes (SLOs) of the proposed program? Address one or more of the five areas of learning – broad, integrative knowledge; specialized knowledge; intellectual skills; applied learning; and civic learning. *(300 word limit)*
	Student Learning Outcomes (SLOs) reflect skills and abilities that students are expected to possess by the time they graduate. They include skills and abilities to: *1. Identify, analyze, and solve technical problems related to data science.* *2. Collect, organize, store, process, visualize, and analyze large data sets.* *3. Assemble computational pipelines to support data science from off-the-shelf or self-developed tools.* *4. Communicate technical concepts effectively, both orally and in writing.* *5. Demonstrate best practices as a data scientist according to policy, privacy, security, and ethical guidelines.*
10c	Clearly state the student admission, retention, and completion standards designed to encourage high quality. *(300 words)*
	Admissions: Undergraduate GPA of 3.0. Successful applicants will come from STEM disciplines. *Retention: Each student enrolled in the program will have an advising committee consisting of faculty members with expertise and resources related to data science. Each student will work on a well-motivated, engaging project involving the analysis, interpretation, and application of real-world data resource(s).* *Completion: Students are to maintain a GPA of 3.0 in all core and elective courses. Students are required to provide a written project report and complete an oral project defense before their faculty committee established according to Graduate School policies (as part of the required MS Project in Data Science course). Committee members examine the technical competency of students at the oral defense, which acts as the program final exam.*
10d*	Describe how the proposed program will articulate with related programs in the state. Include the extent to which student transfer has been explored and coordinated with other institutions. Note: Convert all draft articulation agreements related to this proposed program to PDF and append to the end of this form. *(300 word limit)*
	Currently, Bellarmine University has a Master's of Science in Analytics, and Northern Kentucky University has a Bachelor's of Science in Data Science. The proposed program is different from each of them - in the scope from the first one, and in the degree type from the second one. It will complement both and will serve the UK, the Commonwealth and the nation as the premiere knowledge discovery data science MS program.

11. Mission: Centrality to the Institution's Mission and Consistency with State's Goals	
11a*	*(similar to question 2c)* List the objectives of the proposed program? These objectives should deal with the specific institutional and societal needs that the program will address. *(300 word limit)*
	The goal of this program is to provide students with advanced training in the areas of data science, big data, and analytics needed to support continued regional and national workforce demands. Specific program objectives, reflecting expectations for our students in the years following graduation, are that graduates of our program will: *1. Obtain employment and advance in careers appropriate to an advanced science/technical degree.* *2. Be leaders in the industrial sector, research and development, or entrepreneurship and business, or be pursuing further graduate study.* *3. Use their science, technical, and professional skills to make a positive impact on society and the world.* *4. Engage in continued professional development and life-long learning.*
11b*	Explain how the program objectives above in item 11a support at least two aspects of UK's institutional mission and academic strategic plan? *(150 word limit)*
	The University of Kentucky is a public, land grant university dedicated to improving people's lives through excellence in education, research and creative work, service, and health care. The proposed data science graduate program supports UK's mission in that it: *1. Facilitates learning, informed by scholarship and research;* *2. Expands knowledge through research, scholarship, and creative activity; and* *3. Serves a global community by disseminating, sharing, and applying knowledge using a data- and information-rich approach.* *It directly supports and implements UK's strategy in (1) Graduate Education and in (2) Research and Scholarship.*
11c*	How do the program objectives above in item 11a support at least two aspects of the Council on Postsecondary Education's (CPE) Strategic Agenda and the statewide implementation plan? *(300 word limit)*
	(1) A particular aspect of the CPE agenda that will be impacted by our programs is to "Increase degree and certificate completion, fill workforce shortages, and guide more graduates to a career path." There is a national shortage of Data Science graduates, and there is a lack of comprehensive training programs in Data Science in the Commonwealth. Our MS in Data Science will serve to fill this gap. *(2) Another aspect of the CPE agenda that will be impacted by the program is that "Kentucky will be stronger by training a globally competitive, entrepreneurial workforce; educating an engaged, informed citizenry; improving the health and well-being of families; and producing new research and discoveries that fuel job creation and economic growth." The University of Kentucky MS degree in Data Science contributes to this objective by offering advanced education in an area that, arguably, in the foreseeable future will remain one of the main drivers of the economy, scientific advancement and social change.*
11d*	If an approval letter from an Education Professional Standards Board (EPSB) is required, check the box below and append a PDF version of the letter to this form. ☐ (E.g. any program leading to teacher, principal, or superintendent certification, rank change, etc.)
12. Resources	
12a*	How will the program support or be supported by other programs within the institution? For example, shared faculty, shared courses, collaborative research, etc. *(300 word limit)*
	In addition to the value added to the CS and BST programs and to the research work of the BMI, the MS in Data Science will have tremendous positive impact on numerous disciplines, including the sciences, health sciences, agriculture, engineering, information technology, business and economics, and many others. Growth in this area will lead to further opportunities for collaborative research and shared curriculum with other colleges and units.
12b	What will be the projected "faculty-to-student in major" ratio? *(150 word limit)*
	Because this program is adminstered by the faculty of three units, who already support multiple undergraduate and graduate programs, and because only one new course specifically in support of the new program is proposed, the faculty-to-student ratio cannot be computed for the proposed program independently of other

	degree programs supported by the faculty of the three units. We anticipate 25-30 new MS students enrolled through this program, representing about 40% increase from the current total CS and BST MS enrollment.
12c	Describe the library resources available[19] to support this program. Access to the qualitative and quantitative library resources must be appropriate for the proposed program and should meet recognized standards for study at a particular level or in a particular field where such standards are available. Adequacy of electronic access, library facilities, and human resources to service the proposed program in terms of students and faculty will be considered. *(300 word limit)*
	UK library resources are already sufficient to support this program.
12d	Describe the physical facilities and instructional equipment available to support this program. Physical facilities and instructional equipment must be adequate to support a high-quality program. Address the availability of classroom, laboratory, and office space, as well as any equipment needs. *(300 word limit)*
	Physical facilities and instructional resources are already available to support this program.
13. Demand and Unnecessary Duplication	
13a*	Provide justification and evidence to support the need and demand for this proposed program. Include any data on student demand, employer demand, career opportunities at any level, or any recent trends in the discipline that necessitate a new program. *(300 word limit)* • This evidence is typically in the form of surveys of potential students, enrollments in related programs at the institution, employer surveys, and current labor market analyses. • Anecdotal evidence is insufficient. Demonstrate a systematic collection of data, thorough study of the data, and a reasonably estimated student demand for the program. • Provide evidence of student demand at state and national levels.
	The national need for a fast-growing DS-trained workforce has been well-established. This is highlighted in the recent December 2016 McKinsey report, indicating very high demand for data scientists due to a 250,000 shortfall in the available workforce. Therefore, the establishment of an MS in Data Science at UK is both compelling and timely. A careful examination of leading programs, including UCSD, NYU, Harvard, Columbia, and Virginia indicates that the proposed graduate program is highly comparable to these programs and will provide equivalent or superior training. Moreover, our program has the flexibility to grow with this fast growing and highly dynamic field, especially with respect to applied disciplines. While new hires and courses are unnecessary for the proposed program, both will naturally occur as this field of study continues to grow.
13b	Clearly state the degree completion requirements for the proposed program. *(150 word limit)*
	1. Each student completes the core requrement: earns B or better in every core course in the program *2. Each student completes at least one guided elective from the following list CS405G, CS460G and CS515* *3. Each student completes the concentration requirement (at present, BMI633 and one of BMI730, BMI733 and BMI734)* *4. Each student completes three elective courses (9 credit hours), with at least two courses having a strong data science component in the subject of the student's MS project*
13c*	Will this program replace or enhance any existing program(s) or tracks (or concentrations or specializations) within an existing program? *(300 word limit)* — Yes ☐ No ☒
	If "Yes," explain:
13d	Identify the primary feeders for the program. *(150 word limit)*
	Students graduating with degrees in STEM disciplines especially in computer science, information science, informatics and statistics, are the primary candidates for the program.
13e	Describe the student recruitment and selection process. *(300 word limit)*

[19] Please contact Institutional Effectiveness (institutionaleffectiveness@uky.edu) for more information.

	The program will be broadly advertised to students in programs listed above, in particular, to students at UK. Further, the program will have a stong web presence. *Applications (on-line applications submitted in accordance with the Graduate School Policies and including resume, relevant university transcripts, statement of purpose, letters of recommendation and standardized test scores) will be reviewd by designated members of the Data Science Graduate Committee. The recommendation will be forwarded to the DGS of the program, who will make the final admission decision.*
13f*	Specify any distinctive qualities of the proposed program. *(300 word limit)* • *Are any of your faculty nationally or internationally recognized for expertise in this field?* • *Does this program build on the expertise of an existing locally, nationally, or internationally recognized program at your institution?* • *Do you have any specialized research facilities or equipment that are uniquely suited to this program?*
	The University began investing in biomedical informatics in 2008 as part of the original CTSA proposal. A unique feature of this university is that it is comprehensive, the UK health colleges are located side by side with the main Lexington campus—we are one University spanning the entire spectrum of academic Colleges. With the CTSA award and the creation of the Center for Clinical and Translational Science (CCTS), the Institute for Biomedical Informatics, and the reinvigoratged Center for Computational Science, we are in a unique position to leverage such transdisciplinary collaborativbe infrastructure with tremendous strengths in establihed academic departments such as Computer Science and Biostatistics. We also have significant data resources, unique among all institutions in the Commonwelath, if not the nation, ready to be capitalized for EHR research, eduction and service missions. Such data resources include the CCTS Enterprise Data Trust, the Kentucky Cancer Registry, the Kentucky Health Data Trust, and the Kentucky Research Data Center. The proposed graduate program will provide highly-trained, in-demand data scientists for UK, for Kentucky, and for the nation.
13g	Provide any evidence of a projected net increase in total student enrollments to the campus as a result of the proposed program. *(300 word limit)*
	25-30
13h	Use table below to estimate student demand for the first five years following implementation.

	Academic Year	# Degrees Conferred	Majors (headcount) Fall Semester
	2021 - 2022	*0*	*4*
	2022 - 2023	*3*	*11*
	2023 - 2024	*6*	*18*
	2024 - 2025	*10*	*24*
	2025 - 2026	*12*	*26*

13i	Clearly describe all evidence justifying a new program based on changes in the academic discipline or other academic reasons. *(300 word limit)*
	Few disciplines gained prominence and entered public awareness so quickly and so forcefully as data science. It is a direct consequence of our ever-growing capacity to generate, collect and store data pertaining to individuals, groups and societies, to organizations and their processes, to natural and social phenomena, and to the vastly clear, compelling, and undisputable evidence of the value of data. The emerging field of data science, understood as an academic as well as a professional domain, aims at studying data as a phenomenon, commodity, and asset, with the objective of putting it to effective use in commercial endeavors, academic pursuits and policy setting by governmental and non-governmental organizations. Following the McKinsey Report, we define the field of data science as concerned with data generation, collection, cleaning and aggregation, analysis, interpretation, and visualization, with big data and data analytics as its subfields. As such, data science is inherently interdisciplinary, pulling together ideas and techniques from computer science (algorithms, data bases, networking, high-performance computing), artificial intelligence (knowledge representation, rule-based systems, machine learning), informatics (controlled vocabularies, ontological engineering, usability and human-data interaction, natural language processing, clinical and translational bioinformatics, -omics data and precision

	medicine), and mathematics and statistics (statistical modeling, predictive analysis, and causal inference), and seeking effective ways of applying them to data sets arising in specific applications domains including business, engineering, finance, technology innovation, medicine, scientific research, and social science.		
13j	Has the Council on Postsecondary Education identified similar programs?[20]	Yes ☒	No ☐
	If "Yes," the following questions (5h1 – 5h5) must be answered.		
(1)	Does the program differ from existing programs in terms of curriculum, focus, objectives, etc.? *(150 word limit)*	Yes ☒	No ☐
	If "Yes," explain: *Two graduate programs can be pointed to:* *(1) MS in Analytics at Bellarmine University. Our program is more broadly focused on the spectrum of topics in algorithms, big-data, data mining, machine learning, and the development of tools and pipe-lines for data processing, storing, maintaining, analyzing and visualizing. Importanty, the Bellarmine University is a private university able to reach only limited segments of the society.* *(2) Post-Baccalaureate Certificate in Data Science at the University of Louisville: That program is a certificate program not a ful-fledged degree program.*		
(2)	Does the proposed program serve a different student population (e.g., students in a different geographic area or nontraditional students) from existing programs? *(150 word limit)*	Yes ☐	No ☒
	If "Yes," explain:		
(3)	Is access to existing programs limited? *(150 word limit)*	Yes ☒	No ☐
	If "Yes," explain: *The program is limited to students with a BS degree in a STEM discipline.*		
(4)	Is there excess demand for existing programs? *(150 word limit)*	Yes ☒	No ☐
	If "Yes," explain: *The demand for data science experts is well established (cf. McKinsey Report 2016 cited also in other places in this documents).*		
(5)	Will there be collaboration between the proposed program and existing programs? *(150 word limit)*	Yes ☒	No ☐
	If "yes," explain the collaborative arrangements with existing programs. If "no," explain why there is no collaboration with existing programs.		
	The University of Kentiucky and the University of Louisville have been successfully collaborating in several educational areas. Collaboration with the post-baccalaureate certificate program at the University of Louisville is possible and will be explored.		
13k*	Are there similar programs in other Southern Regional Education Board (SREB) states in the nation?	Yes ☒	No ☐
	If "Yes," please answer the questions below to demonstrate why this proposed program is needed in addition to the one(s) currently in existence.		
13k. i*	Identify similar programs in other SREB states and in the nation.		
	North Carolina State, Graduate Program - Master Track in Data Science University of Virginia - MS in Data Science		
13k.ii*	Does the program differ from existing programs in terms of curriculum, focus, objectives, etc.?	Yes ☐	No ☒
	If "Yes," explain. *(300 word limit)*		

[20] Please contact Institutional Effectiveness (institutionaleffectiveness@uky.edu) for help with this question.

13k.iii*	Does the proposed program serve a different student population (e.g., students in a different geographic area and non-traditional students) from existing programs?	Yes ☒	No ☐
	If "Yes," explain. *(300 word limit)*		
	The University of Kentucky aims to serve the people of the Coommonwealth of Kentucky.		
13k.iv*	Is access to existing programs limited?	Yes ☒	No ☐
	If "Yes," explain. *(300 word limit)*		
	The program accepts candidates with the BS degree in a STEM discipline.		
13k.v*	Is there excess demand for existing similar programs?	Yes ☒	No ☐
	If "Yes," explain. *(300 word limit)*		
	The demand for data science experts is well established (cf. McKinsey Report 2016 cited also in other places in this documents).		
13k.vi*	Will there be collaboration between the proposed program and existing programs?	Yes ☐	No ☒
	If "No," explain. *(300 word limit)*		
	There will not be a collaboration with existing programs in other states		
13l	Would your institution like to make this program available through the Academic Common Market[21]?	Yes ☒	No ☐
13m	Clearly describe evidence of employer demand. Such evidence may include employer surveys, current labor market analyses, and future human resources projections. Where appropriate, evidence should demonstrate employers' preferences for graduates of the proposed program over persons having alternative existing credentials and employers' willingness to pay higher salaries to graduates of the proposed program. *(300 word limit)*		
	The national need for experts trained in data science is growing fast. This is highlighted in the McKinsey Report. The report estimates the growth in data science professionals graduating from the US colleges and universities at 7% annually and contrasts it with the demand growing even by as much as 12% annually and leading in a short term to an estimated 250,000 shortfall. The need is equally emphatically identified in the July 2017 report Data Science and Analytics Skills Shortage: Equipping the APEC Workforce with the Competencies Demand by Employers, published by the Asia-Pacific Economic Cooperation Organization (later on the APEC Report). That report estimates 370,000 new data science jobs to be created in the US in the five years from 2015 to 2020, matched with similar growth trends in other APEC countries. In Kentucky along, 50 to 100 specific Data Scientists jobs are open in recent years. Kentucky industries specifically hiring data scientists include Humana, Oracle (KY Branch), Expedia (KY Branch), UPS, SITA Global Services, The Rawlings Group, Certilytics, General Dynamics Information Technology, Latitude36, General Electric, GDIT, and Appalachian Regional Healthcare.		
13n*	Describe the types of jobs available for graduates, average wages for these jobs, and the number of anticipated openings for each type of jobs at the regional, state, and national levels.		
	Many organizations are hungry to use data to grow and improve performance—and multiple players see market opportunities in this explosion of demand. There are typically many steps between raw data and actual usage, providing multiple avenues for career development. Three main categories of jobs exist in the data ecosystem, recognizing that some jobs might fill more than one role: *- Data generation and collection, referring to the source and platform where data are initially captured;* *- Data aggregation or the processes and platforms for combining data from multiple sources; and* *- Data analysis, which provides the promise of big data through gleaning of insights from data.*		

[21] Please contact Institutional Effectiveness (institutionaleffectiveness@uky.edu) for more information.

	Such jobs exist industry (healthcare, finance, insurance, pharmaceuticals), smart cities and infrastructure, the life sciences (-omics), material sciences, and technology. *A more detailed analysis can be seen from the McKinsey 2016 Report. Additionally, recent reports indicate 1,736 job openings for data scientists with a median base salary of $116,840. (thisisstatistics.org, Jan 27, 2016*
14. Assessment and Oversight	
14a*	Describe how each program-level student learning outcome will be assessed and how assessment results will be used to improve the program. *(250 word limit)*
	(Also see attached curriculum map and assessment plan.) The program-level outcomes are for students to 1. Obtain employment and advance in careers appropriate to an advanced science/technical degree. 2. Be leaders in the industrial sector, research and development, or entrepreneurship and business, or be pursuing further graduate study. 3. Use their science, technical, and professional skills to make a positive impact on society and the world. 4. Engage in continued professional development and life-long learning. These program-level outcomes will be assessed primarily based on data gathered from regular alumni surveys. This will include job placement data and self-assessed alumni satisfaction with the professional skills acquired in the program in support of objectives 1-4. The survey will be administered and analyzed by the graduate committee every three years. Together with the survey data, every three years the graduate committee will also review secondary measures of the overall quality of the program: the currency of core courses in their support of objectives 1-4; the relevance of final projects to objectives 1-4; time-to-graduation; and full-time and part-time enrollment numbers, GPA.
14b*	Describe *program* evaluation procedures for the proposed program. These procedures may include evaluation of courses and faculty by students, administrators, and departmental personnel as appropriate. Program review procedures shall include standards and guidelines for the assessment of student outcomes implied by the program objectives and consistent with the institutional mission. *(300 word limit)*
	(See question 5a.) *Every three years in fall, the DGS will contact program alumni with the request to complete and return the alumni survey. The DGS will also compile statistics on time to degree, retention and enrollment and collect the current syllabi of all core courtses. The following spring the graduate committee for the program will meet to review the collected data, assess the program, its strngths and weaknesses and will propose corrective actions, if appropriate.*
14c	Identify both the direct and indirect methods by which the intended student learning outcomes (SLOs) will be assessed. *(300 word limit)*
	(Also see attached curriculum map and assessment plan.) *SLOs are assessed in each of the core courses through assignemtns specifically designed to measure the degree of mastery of individual SLOs, through the written report submitted at the end of the semesters 2, 3, and 4 as partial completion of DS710 requirements, and through the written and oral components of the Master's project (supporting rubric-based direct measures).* *All SLOs are assessed directly based on the written reports submitted as part of the requirements for DS710 after semesters 2, 3, and 4, using the rubric developed for this course, and based on the written project report and the final examination. A rubric with separate sub-elements for all SLOs will be filled out by committee members at the time of the final MS examination, assessing both the written report document and the oral defense as primary artifacts. The expected target performance is good or better (at least 3.0 on the numeric GPA scale).* *The DGS for the program will collect this data from course instructors and final MS examination committee members. The program graduate curriculum committee will meet annually to review this data, identify any*

	concerns or weaknesses, and recommend course or program changes to address those concerns.
14d	**Procedures for Course Mapping of SLOs *(related to question 5b)***
14d.i	Which components will be evaluated, i.e. course mapping? *(300 word limit)*
	All SLOs will be evaluated. All core courses will be evaluated for relevance to SLOs.
14d.ii	When will components be evaluated? *(150 word limit)*
	The components will be evaluated annually in early May or in the week prior to the start of the Fall semester.
14d.iii	When will the data be collected? *(150 word limit)*
	The GPA data will be collected prior to the evaluation meeting for the most recent academic year. The rubrics, which directly assess the SLOs, will be collected by the DGS after each final exam. Those collected in the most recent academic year will be used at the annual assessment meeting by the graduate curriculum committee.
14d.iv	How will the data be collected? *(150 word limit)*
	The DGS will collect the GPAs, enrollments, time to graduation statistics and job placement data. After each final exam, the members of the examination committee will complete the rubrics. The chair of the committee will deliver them to the DGS.
14d.v	What will be the benchmarks and/or targets to be achieved? *(150 word limit)*
	The expected target performance for each SLO assessed based on rubrics is good or higher.
14d.vi	What individuals or groups will be responsible for data collection? *(150 word limit)*
	The DGS and the members of the chairs of the final examination committees
14d.vii	How will the data and findings be shared with faculty? *(150 word limit)*
	The graduate committee will review and make recommendations for action items, which will be distributed to faculty in CS and BST departments and in the BMI for approval at faculty assessment meetings held each fall semester.
14d.viii	How will the data be used for making programmatic improvements? *(150 word limit)*
	The graduate committee will make recommendations for programmatic improvements, including changes in core courses, based on the assessment results.
14d.ix	What are the measures of teaching effectiveness? *(150 word limit)*
	Evaluation of teaching effectiveness is separate from program evaluation or evaluation of SLOs. *Instructors in the program will be individually assessed for teaching effectiveness using TCEs as well as peer review and other measures in accordance with the performance evaluation standards used in each instructor's home department.* *If SLO assessment indicates problems with teaching effectiveness within specific core courses, that information will be shared with the department chair for discussion and follow up with instructors as a part of the regular performance evaluation process.*
14d.x	What efforts to improve teaching effectiveness will be pursued based on these measures? *(150 word limit)*
	Efforts to improve teaching effectiveness will be pursued on a case-by-case basis with individual faculty involved with the program.

14d.xi	What are the plans to evaluate students' post-graduate success? *(150 word limit)*
	We plan to conduct an alumni survey approximately every 3-5 years. We will continually follow students' professional progress via social networks such as LinkedIn.

15. Cost and Funding of the Proposed Program[22]

15a	Will this program require additional resources?	Yes ☒	No ☐
	If "Yes," please provide a brief summary of additional resources that will be needed to implement this program over the next five years. *(300 word limit)*		
	0.5 FTE for an administrative assistant to the DGS. Given expected enrollment of 25-30 MS students, the required DGS support cannot be absorbed by the existing administrative support personnel in the CS department. *Also, three TA lines (target number for years five and beyond of the program) needed to help with the instruction in core courses (see below for funding details).*		
15b	Will this program impact existing programs and/or organizational units within your institution? *(300 word limit)*	Yes ☐	No ☒
	If "Yes, briefly describe.		
15c	Provide adequate documentation to demonstrate sufficient return on investment to the state to offset new costs and justify approval for the proposed program. *(300 word limit)*		
	See also 2e and 13a. *The costs of implementing and running the MS in Data Science can be amply met with new funds generated by the program tuition income.* *In addition there are significant non-financial benefits, as implementing this graduate program will* *- Address the emerging and critical national need for MS-level experts in Data Science.* *- Promote interdisciplinary research and team science at the University of Kentucky.*		

16.* Budget Funding Sources, by Year of Program

All the fields in number 16 are required for the CPE's pre-proposal form. Estimate the level of new and existing resources that will be required to implement and sustain the program using the spreadsheet below. Please answer in terms of dollar amounts. All narratives have a 100-word limit.

Total Resources Available from Federal Sources (Federal sources include grants, earmarks, etc.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative/Explanation:	*No known federal resources for creation of new program in this area*				

Total Resources Available from Other Non-State Sources (Non-state sources include philanthropies, foundations, individual donors, etc.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*

[22] For questions about cost and funding of the program, please contact your department chair, business officer, or associate dean for academic affairs.

Narrative/Explanation:	*No non-state allocations available (or needed).*				
`					
State Resources (State sources include general fund revenue, grants, pass-thru funds, etc.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative/Explanation:	*No specific state allocations have been made.*				
Internal (The source and process of allocation and reallocation should be detailed, including an analysis of the impact of the reduction on existing programs and/or organization units.)[23]:	1st Year	2nd Year	3rd Year	4th Year	5th Year
(New) Allocated Resources	*0*	*0*	*0*	*0*	*0*
(Existing) Reallocated Resources	*0*	*0*	*0*	*0*	*0*
Narrative/Explanation:					
Student Tuition (Describe the impact of this program on enrollment, tuition, and fees.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*60000*	*168000*	*276000*	*372000*	*408000*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative/Explanation:	*The figures assume tuition at $12000/year instate, and $24000/year out-of-state. They also assume enrollments (instate/out-of-state) of 3/1, 8/3, 13/5, 17/7 and 18/8 for the first five years of the program, respectively.*				
Total Funding Sources	1st Year	2nd Year	3rd Year	4th Year	5th Year
Total New	*60000*	*168000*	*276000*	*372000*	*408000*
Total Existing	*0*	*0*	*0*	*0*	*0*
TOTAL FUNDING SOURCES	60000	168000	276000	372000	408000
17. Breakdown of Program Expenses/Requirements[4]					
(Please note – all the fields in number 17 are required for the CPE's pre-proposal form.)					
Staff: Executive, Administrative & Managerial (Include salaries and whether new hires will be part time or full time.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*39000*	*39000*	*39000*	*39000*	*39000*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative/Explanation[24]:	*A recurring commitment to 0.5 FTE for an administrative assistant, estimated at*				

[23] The source and process of allocation and reallocation should be detailed, including an analysis of the impact of the reduction on existing programs and/or organizational units.

[24] Discuss whether new hires will be full-time or part-time.

	$24000/year, and to one summer month salary for the Director of Graduate Studies, estimated at $15000. The total estimated at $39000/year				
Other Professional (Include salaries.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	0	0	0	0	0
Existing	0	0	0	0	0
Narrative/Explanation:					
Faculty (Include salaries and whether new hires will be part time or full time.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative/Explanation[25]:	*The program will not impose burden on existing resources as it is based in existing and regularly offered courses with curent enrollments at the levels allowing for growth resulting from implementing the new program.*				
Graduate Assistants (Include salaries and/or stipends.)[26]	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*30000*	*60000*	*90000*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:	*Additional TA positions (1 in year 3, 2 in year 4, 3 in year five and beyond) to support instruction in core courses addressing increased enrollments. Stipend at $18000, tuition at the instate level of $12000.*				
Student Employees (Include salaries and/or stipends.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Equipment and Instructional Materials	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Library (Include new journal subscriptions, collections, and	1st Year	2nd Year	3rd Year	4th Year	5th Year

[25] If new hires are involved, explain whether new hires will be full-time or part-time.
[26] Identify the number of assistantships/stipends to be provided; Include the level of support for each.

electronic access.)					
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Contractual Services	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Academic and/or Student Services	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Other Support Services	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Faculty Development (Include travel, conference fees, consultants, etc.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Assessment (Include personnel, software tools, data collection tools, survey administration, outside consulting services, etc.)	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Student Space and Equipment	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*

Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Other	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*0*	*0*	*0*	*0*	*0*
Existing	*0*	*0*	*0*	*0*	*0*
Narrative Explanation/Justification:					
Total Expenses/Requirements	1st Year	2nd Year	3rd Year	4th Year	5th Year
New	*39000*	*39000*	*69000*	*99000*	*129000*
Existing	*0*	*0*	*0*	*0*	*0*
TOTAL Program Budgeted Expenses/Requirements:					
GRAND TOTAL	1st Year	2nd Year	3rd Year	4th Year	5th Year
Total Funding Sources	***60000***	***168000***	***276000***	***372000***	***408000***
Total Expenses/Requirements	***39000***	***39000***	***69000***	***99000***	***129000***
TOTAL NET COST:	***-21000***	***-129000***	***-207000***	***-273000***	***-279000***

18. Course Descriptions

18a **Program Core Courses (includes pre-major and pre-professional courses)**

Prefix & Number	Course Description (from the Bulletin or the most recent new/change course form)
CS626	*LARGE SCALE DATA SCIENCE. (3) This course will offer an opportunity for students to learn big data techniques and apply them to tackle real-world data science challenges (e.g., processing, storing, querying, exploring, and mining big data). Topics include big data systems and programming models, parallel computing framework, scalable data management and processing solutions, scalable data mining techniques for large datasets, and advanced applications.*
CPH580	*BIOSTATISTICS I. (3) CPH 580 covers univariate statistical methods commonly encountered in public health studies. This includes descriptive statistics, hypothesis testing, paired and unpaired t tests, ANOVA, contingency tables, log rank test, regression and correlation.*
CPH630	*BIOSTATISTICS II. (3) Students will learn statistical methods used in public health studies. This includes receiver operator curves, multiple regression, logistic regression, confounding and stratification, the Mantel-Haenzel procedure, and the Cox proportional hazards model.*
DS710	*RESEARCH SEMINAR IN DATA SCIENCE: An independent study course that aims to expose students to a broad range of research and professional topics in data science, and enhance students' ability to do research. Students read research papers and professional articles, attend colloquium/seminar talks on topics in data science as directed by the instructor, prepare written reports on topics and make in-class presentations.*
DS711	*MASTER'S PROJECT IN DATA SCIENCE: Design and implementation of a large data science project under the supervision of a member of a faculty member.*
BMI633	*INTRODUCTION TO BIOINFORMATICS. (3) This is an introductory course aimed at a multi-disciplinary audience with an interest in applying the principles of information sciences for obtaining insight into biological processes and systems that can eventually be used to make informed decisions. (THIS COURSE*

	IS REQUIRED (CORE) for the BIOMEDICAL INFORMATICS CONCENTRATION)
18b	**Program Guided Electives Courses (for the major)**
Prefix & Number	Course Description (from the Bulletin or the most recent new/change course form)
CS405G	*INTRODUCTION TO DATABASE SYSTEMS. (3) Study of fundamental concepts behind the design, implementation and application of database systems. Brief review of entity-relationship, hierarchical and network database models and an in-depth coverage of the relational model including relational algebra and calculi, relational database theory, concepts in schema design and commercial database languages.*
CS460G	*MACHINE LEARNING. (3) Study of computational principles and techniques that enable software systems to improve their performance by learning from data. Focus on fundamental algorithms, mathematical models and programming techniques used in Machine Learning. Topics include: different learning settings (such as supervised, unsupervised and reinforcement learning), various learning algorithms (such as decision trees, neural networks, k-NN, boosting, SVM, k-means) and crosscutting issues of generalization, data representation, feature selection, model fitting and optimization. The course covers both theory and practice, including programming and written assignments that utilize concepts covered in lectures.*
CS515	*ALGORITHM DESIGN. (3) The design and analysis of efficient algorithms on data structures for problems in sorting, searching, graph theory, combinatorial optimization, computational geometry, and algebraic computation. Algorithm design techniques: divide-and-conquer, dynamic program- ming, greedy method, and randomization, approximation algorithms.*
BMI730	*BMI730 PRINCIPLES OF CLINICAL INFORMATICS. (3) This course offers an overview of Clinical informatics, which is the application of informatics principles, methods, and tools to support healthcare practice and research activities as well as business processes.*
BMI733	*BMI733 BIOMEDICAL NATURAL LANGUAGE PROCESSING. (3) This course is a technical introduction to the area of biomedical natural language processing (NLP). In the field of biomedical informatics, this focuses on the common steps in extracting information from textual data that arises from biomedical literature and clinical documents. Topics involve n-gram models, tokenization, POS tagging, and parsing.*
BMI734	*BMI734 INTRODUCTION TO BIOMEDICAL IMAGE ANALYSIS. (3) This class aims to give students a broad overview of biomedical image analysis and imaging informatics. We will introduce the state-of- the-art knowledge to understand, develop, and apply existing methods and software to handle biomedical image data to extract quantitative matrices.*
18c	**Program Free Electives Courses**
Prefix & Number	Course Description (from the Bulletin or the most recent new/change course form)
CS463G	*Introduction to Artificial Intelligence: The course covers basic techniques of artificial intelligence. The topics in this course are: search and game-playing, logic systems and automated reasoning, knowledge representation, intelligent agents, planning, reasoning under uncertainty, and declarative programming languages. The course covers both theory and practice, including programming assignments that utilize*

	concepts covered in lectures. Prereq: CS 315, CS 375, and engineering standing.
CS505	*Intermediate Topics in Data Bases: The course introduces a variety of modern techniques in database and distributed database systems. The major topics include, but are not limited to: object-oriented database systems; distributed, heterogeneous and web-based databases; knowledge based systems; physical database design; and security. The course covers a variety of methods that allow for a solution of database problems where the traditional relational database techniques are not viable or not sufficient. Prereq: CS 405 or consent of instructor.*
CS628	*Data Mining:The course will introduce the fundamental principles and main techniques in the area of data mining and its applications. The topics covered include association rule mining, clustering, classification, feature selection, similarity search, data cleaning, privacy and security issues, as well as a wide spectrum of data mining applications in the area of biomedical informatics, bioinformatics, financial market study, image processing, network monitoring and social service analysis. Prereq: CS 515 or consent of instructor.*
BMI731	*Biomedical Information Retrieval: This class is an introductory information retrieval class that is focused on biomedical information search engines. Basic IR concepts such index construction, optimization, visualization, and evaluation will be covered. In addition to core IR contexts, students will have an opportunity to learn about search engines, web crawling, and some Web 2.0 technologies based on hands-on exercises and assignments with a focus on techniques that can be used to access, retrieve, organize, and present information. Students will employ an open source indexing engine (e.g., Lemur or Lucene or something similar) to understand how back-end of retrieval engine is effectively and efficiently structured.*
BMI732	*Biomedical Ontologies and Semantic Web Techniques:This course is a conceptual introduction to biomedical ontologies and ontological modeling in biomedicine through Semantic Web techniques. Students will learn about RDF, OWL, description logics, and SPARQL and their role in designing ontologies. Biomedical terminologies such as GO, ICD-9/10, SNOMED-CT, and MeSH will be discussed as case studies. Prereq: MA 123 (or equivalent) or consent of the instructor.*
PPS710	*Techniques for Secondary Data Research:* *A successful pharmaceutical outcomes and policy researcher must have the ability to independently assess the literature in order to identify a clinically relevant research question, design a study that will address the question and analyze and present the results appropriately to the scientific community. This course will provide an introduction to the conduct of pharmaceutical outcomes and policy research through in-depth didactic and practical instruction on the development, design, and presentation of relevant research study. The course will have two components. Approximately half of the course will involve didactic instruction on specific topics related to the conduct and execution of pharmaceutical outcomes and policy research and half will be a hands-on experience in which the learner develops their own research question and hypothesis, designs a study and begins to analyze an existing healthcare dataset in order to answer a relevant pharmaceutical outcomes and policy question and present the results. Prereq: CPH 580 or equivalent.*
CPH535	*Databases and SAS Programming: Students will learn how to construct and maintain databases with applications to public health. They will also learn how to program in SAS, the leading statistical analysis system. SAS skills include report writing, MACRO writing, and Programming using SAS Intranet. Lecture, two hours; laboratory, two hours per week. Prereq: STA 291 or equivalent.*
BST655	*Introduction to Statistical Genetics: BST 655 presents an introduction to the statistical methodologies used today to investigate genetic susceptibility to complex diseases. The course focuses on linkage and association analysis with applications to real-world data. Commonly used (and freely available) software will be presented and used throughout. Because the field is constantly evolving, a focus of the material for this course will be recent statistical human genetics literature. Prereq: STA 580 or equivalent. (Same as STA 655.)*
	Additional elective courses listed in the attachment

18d	**Courses for a Track.** (If multiple tracks are available, click HERE for a template for additional tracks. Append a PDF to the end of this form with each track's courses and descriptions.	
Prefix & Number	Course Type	Course Description (from the Bulletin or the most recent new/change course form)
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	
	☐ Track Core ☐ Track Elective	

19. Specific faculty involved in the degree program.

(similar to question 4d) Fill out the SACS[27]-required faculty roster below, for full-time and part-time faculty teaching in the program. Abbreviations for the NAME and COURSES TAUGHT columns are below the table. *Please contact Institutional Effectiveness (institutionaleffectiveness@uky.edu) for help with this question.*

NAME List name & Identify faculty member as F or P.	**COURSES TAUGHT** Include term; course prefix, number and title; & credit hours. (D, UN, UT, G)	**ACADEMIC DEGREES AND COURSEWORK** List relevant courses taught, including institution and major. List specific graduate coursework, if needed	**OTHER QUALIFICATIONS AND COMMENTS** Note qualifications and comments as they pertain to course taught.	**NEW COURSES** Include course prefix, number, and title.
Miroslaw Truszczynski, FT	*Fall, CS 515 Algorithm Design; 3, G and UN*	*PhD*		*DS 711 MS Project in Data Science*
Jinze Liu, FT	*Fall, CS 405G Intro to Data Bases; 3, UN and G*	*PhD*		*CS 628 Data Mining*
Licong Cui, FT	*Fall, CS Large Scale Data Science; 3, G*	*PhD*		
Nathan Jacobs, FT	*Spring, CS 460G Machine Learning; 3, UN and G*	*PhD*		
Rada Nagarajan, FT	*Fall, BMI 633 Introduction to Bioinformatics; 3, G*	*PhD*		
Suji Kim, FT	*BMI 730 Principles of Clinical Informatics; 3, G*	*PhD*		
Rama Kavaluru, FT	*BMI 733 Biomedical Natural Language Processing; 3, G*	*PhD*		
Jin Chen, FT	*BMI 734 Intro to Biomedical Image Analysis*	*PhD*		
Heather Bush, FT	*Fall, CPH 580 Biostatistics 1; 3, G*	*PhD*		

[27] Southern Association of Colleges and Schools Commission on Colleges (SACS).

	Spring, CPH 630 Biostatistics 2; 3, G			
Philip Westgate, FT	*Spring, CPH 630 Biostatistics 2; 3, G*	*PhD*		
David Fardo, FT	*Spring, CPH 630 Biostatistics 2; 3, G*	*PhD*		
Emily Slade, FT	*Fall, CPH 580 Biostatistics 1; 3, G* *Spring, CPH 630 Biostatistics 2; 3, G*	*PhD*		
FT = full time PT= part time	D = developmental UN = undergraduate nontransferable		UT = undergraduate transferable G = graduate	

Courses

The following courses, many of which are currently offered by the UK faculty, are part of the proposed curriculum of the MS in Data Science. The core courses and the required courses are marked as such. All other courses are electives that require an approval by the DGS. Students may also select as electives other related graduate courses based on the advice of their faculty mentors and in consultation with the DGS.

Letters from non-CS units allowing DS students to take the courses as core or as electives in the program are attached. They came from Dr. GQ Zhang (BMI courses), Dr. Heather Bush (BST and one CPH course offered by the Biostatistics Department), Dr. Arnold Stromberg (STA courses), Dr. Daniele Moga (CPH 713), Dr. Jeffrey Talbert (PPS 710), and Dr. Uwe Nagel (MA courses).

CS 405G Introduction to Data Base Systems

Study of fundamental concepts behind the design, implementation and application of database systems. Brief review of entity-relationship, hierarchical and network database models and an in-depth coverage of the relational model including relational algebra and calculi, relational database theory, concepts in schema design and commercial database languages. Prereq: CS 315 and graduate or engineering standing.

CS 460G Machine Learning

Study of computational principles and techniques that enable software systems to improve their performance by learning from data. Focus on fundamental algorithms, mathematical models and programming techniques used in Machine Learning. Topics include: different learning settings (such as supervised, unsupervised and reinforcement learning), various learning algorithms (such as decision trees, neural networks, k-NN, boosting, SVM, k-means) and crosscutting issues of generalization, data representation, feature selection, model fitting and optimization. The course covers both theory and practice, including programming and written assignments that utilize concepts covered in lectures. Prereq: Strong programming ability (CS 315), basic probability and statistics (STA 281), and basic concepts of linear algebra (MA/CS 321 or MA/CS 322), or instructor's consent.

CS 463G Introduction to Artificial Intelligence

The course covers basic techniques of artificial intelligence. The topics in this course are: search and game-playing, logic systems and automated reasoning, knowledge representation, intelligent agents, planning, reasoning under uncertainty, and declarative programming languages. The course covers both theory and practice, including programming assignments that utilize concepts covered in lectures. Prereq: CS 315, CS 375, and engineering standing.

CS 505 Intermediate Topics in Data Bases

The course introduces a variety of modern techniques in database and distributed database systems. The major topics include, but are not limited to: object-oriented database systems; distributed, heterogeneous and web-based databases; knowledge based systems; physical database design; and security. The course covers a variety of methods that allow for a solution of database problems where the traditional relational database techniques are not viable or not sufficient. Prereq: CS 405 or consent of instructor.

CS 515 Algorithm Design

The design and analysis of efficient algorithms on data structures for problems in sorting, searching, graph theory, combinatorial optimization, computational geometry, and algebraic computation. Algorithm design techniques: divide-and-conquer, dynamic program- ming, greedy method, and randomization, approximation algorithms. Prereq: CS 315 and engineering standing.

CS 521 Computational Science

Study of computer science techniques and tools that support computational sciences and engineering. Emphasis on visualization, performance evaluation, parallel computing, and distributed computing. Prereq: Either EGR 102 or CS 115, and CS 380/EE 380 and engineering standing.

CS 537 Numerical Analysis

Floating point arithmetic. Direct methods for the solution of systems of linear algebraic equations. Polynomial and piecewise polynomial approximation, orthogonal polynomials. Numerical integration: Newton Cotes formulas and Gaussian quadrature. Basic methods for initial value problems for ordinary differential equations. The emphasis throughout is on the understanding and use of software packages for the solution of commonly occurring problems in science and engineering. Prereq: CS/MA 321 or equivalent or graduate standing or consent of instructor. Knowledge of a procedural computer language is required. (Same as EGR/MA 537.)

CS 626 Large scale data science (core)

This course will offer an opportunity for students to learn big data techniques and apply them to tackle real-world data science challenges (e.g., processing, storing, querying, exploring, and mining big data). Topics include big data systems and programming models, parallel computing framework, scalable data management and processing solutions, scalable data mining techniques for large datasets, and advanced applications.

CS 628 Data Mining (submitted for approvals)

The course will introduce the fundamental principles and main techniques in the area of data mining and its applications. The topics covered include association rule mining, clustering, classification, feature selection, similarity search, data cleaning, privacy and security issues, as well as a wide spectrum of data mining applications in the area of biomedical informatics, bioinformatics, financial market study, image processing, network monitoring and social service analysis. Prereq: CS 515 or consent of instructor.

BMI 633 Introduction to Bioinformatics (required for the Bioinformatics concentration)

This is an introductory course aimed at a multi-disciplinary audience with an interest in applying the principles of information sciences for obtaining insight into biological processes and systems that can eventually be used to make informed decisions.

BMI 730 Principles of Clinical Informatics

This course offers an overview of Clinical informatics, which is the application of informatics principles, methods, and tools to support healthcare practice and research activities as well as business processes.

BMI 731 Biomedical Information Retrieval

This class is an introductory information retrieval class that is focused on biomedical information search engines. Basic IR concepts such index construction, optimization, visualization, and evaluation will be covered. In addition to core IR contexts, students will have an opportunity to learn about search engines, web crawling, and some Web 2.0 technologies based on hands-on exercises and assignments with a focus on techniques that can be used to access, retrieve, organize, and present information. Students will employ an open source indexing engine (e.g., Lemur or Lucene or something similar) to understand how back-end of retrieval engine is effectively and efficiently structured.

BMI 732 Biomedical Ontologies and Semantic Web Techniques

This course is a conceptual introduction to biomedical ontologies and ontological modeling in biomedicine through Semantic Web techniques. Students will learn about RDF, OWL, description logics, and SPARQL and their role in designing ontologies. Biomedical terminologies such as GO, ICD-9/10, SNOMED-CT, and MeSH will be discussed as case studies. Prereq: MA 123 (or equivalent) or consent of the instructor.

BMI 733 Biomedical Natural Language Processing

This course is a technical introduction to the area of biomedical natural language processing (NLP). In the field of biomedical informatics, this focuses on the common steps in extracting information from textual data that arises from biomedical literature and clinical documents. Topics involve n-gram models, tokenization, POS tagging, and parsing. Prereq: MA 123 (or equivalent) or consent of instructor.

BMI 734 Introduction to Biomedical Image Analysis

This class aims to give students a broad overview of biomedical image analysis and imaging informatics. We will

introduce the state-of- the-art knowledge to understand, develop, and apply existing methods and software to handle biomedical image data to extract quantitative matrices.

BMI 735 Introduction to Biomedical Image Informatics

This class provides an introduction to searching and retrieval in biomedical image analysis and imaging informatics. We will introduce some advanced biomedical image analysis, searching, and retrieval algorithm for fast and efficient image searching and retrieval. Prereq: BMI 734: Introduction to Biomedical Image Analysis.

CPH 535 Databases and SAS Programming

Students will learn how to construct and maintain databases with applications to public health. They will also learn how to program in SAS, the leading statistical analysis system. SAS skills include report writing, MACRO writing, and Programming using SAS Intranet. Lecture, two hours; laboratory, two hours per week. Prereq: STA 291 or equivalent.

CPH 580 Biostatistics I (core)

CPH 580 covers univariate statistical methods commonly encountered in public health studies. This includes descriptive statistics, hypothesis testing, paired and unpaired t tests, ANOVA, contingency tables, log rank test, regression and correlation. Prereq: MA 109 or higher. (Same as STA 580.)

CPH 630 Biostatistics II (core)

Students will learn statistical methods used in public health studies. This includes receiver operator curves, multiple regression logistic regression, confounding and stratification, the Mantel-Haenzel procedure, and the Cox proportional hazardous model. Lecture, two hours; laboratory, two hours per week. Prereq: STA 580 or CPH 580. (Same as STA 681.)

BST 655 Intro to Statistical Genetics

BST 655 presents an introduction to the statistical methodologies used today to investigate genetic susceptibility to complex diseases. The course focuses on linkage and association analysis with applications to real-world data. Commonly used (and freely available) software will be presented and used throughout. Because the field is constantly evolving, a focus of the material for this course will be recent statistical human genetics literature. Prereq: STA 580 or equivalent. (Same as STA 655.)

BST 763 Categorical Data Analysis

Multinomial and product-multinomial models; large-sample theory of estimation and testing, Pearson chi-square and modified chi-square statistics, Pearson-Fisher Theorem, Wald Statistics and generalized least squares technique; applications to problems of symmetry, association and hypotheses of no interaction in multi-dimensional contingency tables. Prereq: STA 603 and STA 606. (Same as STA 665.)

CPH 713 Pharmacoepidemiology

This course will provide an overview of the field of pharmacoepidemiology and its relationship to health care research. Various topics including methodology and analytical issues relevant to the conduct of pharmacoepidemiologic research will be covered. Time will also be spent reviewing existing papers in the field of pharmacoepidemiology. Prereq: CPH 605 and STA 580 or equivalent; may be concurrent. (Same as PPS 701.)

PPS 710 Techniques in Secondary Data Research.

A successful pharmaceutical outcomes and policy researcher must have the ability to independently assess the literature in order to identify a clinically relevant research question, design a study that will address the question and analyze and present the results appropriately to the scientific community. This course will provide an introduction to the conduct of pharmaceutical outcomes and policy research through in-depth didactic and practical instruction on the development, design, and presentation of relevant research study. The course will have two components. Approximately half of the course will involve didactic instruction on specific topics related to the conduct and execution of pharmaceutical outcomes and policy research and half will be a hands-on experience in which the learner develops their own research question and hypothesis, designs a study and begins to analyze an existing healthcare dataset in order to answer a relevant pharmaceutical outcomes and policy question and present the results. Prereq: CPH 580 or equivalent.

STA 524 Probability

Sample space, random variables, distribution functions, conditional probability and independence, expectation, combinatorial analysis, generating functions, convergence of random variables, characteristic functions, laws of large numbers, central limit theorem and its applications. Prereq: MA 213 and MA 322. (Same as OR 524.)

STA 525 Introductory Statistical Inference

Simple random sampling, statistics and their sampling distributions, sampling distributions for normal populations; concepts of loss and risk functions; Bayes and minimax inference procedures; point and interval estimation; hypothesis testing; introduction to nonparametric tests; regression and correlation. Prereq: STA 320 or STA 524 or consent of instructor. (Same as OR 525.)

MA 415G Combinatorics and Graph Theory

A basic course in the theory of counting and graph theory. Topics in enumerative combinatorics may include: generating functions, compositions, partitions, Fibonacci numbers, permutations, cycle structure of permutations, permutations statistics, Stirling numbers of the first and second kind, Bell numbers, inclusion-exclusion. Topics in graph theory may include: Eulerian and Hamiltonian cycles, matrix tree theorem, planar graphs and the 4-color theorem, chromatic polynomial, Hall's marriage theorem, stable marriage theorem, Ramsey theory, electrical networks. Prereq: MA 213 or MA 322. (Same as CS 415G.)

MA 416G Introduction to Optimization

The course is an introduction to modern operations research and includes discussion of modeling, linear programming, dynamic programming, integer programming, scheduling and inventory problems, and network algorithms. Prereq: MA 213 or equivalent. (Same as CS 416G.)

MA 522 Matrix Theory and Numerical Linear Algebra I

Review of basic linear algebra from a constructive and geometric point of view. Factorizations of Gauss, Cholesky and Gram-Schmidt. Determinants. Linear least squares problems. Rounding error analysis. Stable methods for updating matrix factorizations and for linear programming. Introduction to Hermitian eigenvalue problems and the singular value decomposition via the QR algorithm and the Lanczos process. Method of conjugate gradients. Prereq: MA 322. (Same as CS 522.)

DS 710 Research Seminar in Data Science (core)

An independent study course that aims to expose students to a broad range of research and professional topics in data science, and enhance students' ability to do research. Students read research papers and professional articles, attend colloquium/seminar talks on topics in data science as directed by the instructor, prepare written reports on topics and make in-class presentations.

DS 711 Master's Project in Data Science (core)

Design and implementation of a large data science project under the supervision of a member of a faculty member.

1. Assessment – Program-level objectives and SLOs

The program-level outcomes

Students graduating from the program are expected to have professional skills that will allows them to

1. Obtain employment and advance in careers appropriate to an advanced science/technical degree;
2. Be leaders in the industrial sector, research and development, or entrepreneurship and business, or be pursuing further graduate study;
3. Use their science, technical, and professional skills to make a positive impact on society and the world; and,
4. Engage in continued professional development and life-long learning.

These program-level outcomes will be assessed primarily based on data gathered from regular alumni surveys. This will include job placement data and self-assessed alumni satisfaction with the professional skills acquired in the program in support of objectives 1-4. The survey will be administered and analyzed by the graduate committee every three years.

Together with the alumni survey data, every three years the graduate committee will also review secondary measures of the overall quality of the program: the currency of core courses in their support of objectives 1-4; the relevance of final projects to objectives 1-4; time-to-graduation; and full-time and part-time enrollment numbers, GPA.

Student learning objectives (SLOs)

SLOs are assessed in each of the core courses by artifacts (parts of exams, homeworks or projects) designed specifically to address them individually; through the written report submitted at the end of the semesters 1, 2 and 3 for the fulfillment of the completion requirements for DS 710 (supporting a rubric-based direct measure), and through the written and oral components of the Master's project (supporting a rubric-based direct measures).

The following table maps *required* courses, and course and project reports to SLOs:

	SLO1	SLO2	SLO3	SLO4	SLO5
CS626	I/D	I/D	I/D		
CPH 580	I/D	I/D	I/D		
CPH 630	I/D	I/D	I/D		
DS 710 reports	I/D/M	I/D/M	I/D/M	I/D/M	I/D/M
DS 711/MS project report	D/M	D/M	D/M	D/M	D/M

I – introducing concepts
D – developing concepts

M – mastering concepts

For assessing DS710 reports we will use a rubric developed to assess the level of mastery of the SLOs (attached). These reports will summarize talks attended, papers read, and topics discussed. The expectation is that at the end of the semester 1, the DS 710 reports will show that students are familiar with basic concepts of data science, at the end of semester 2 that the students developed and expanded this understanding, and at the end of semester 3 that they mastered them.

SLOs will also be assessed based on the written and oral components of the Master's project (DS 711 MS in Data Science course). They will be assessed according to the same rubric as that for assessing DS 710 reports.

The SLOs are assessed annually in three-year cycles as shown in the table below.

	SLO1	SLO2	SLO3	SLO4	SLO5
Year 1	X	X	X		
Year 2				X	X
Year 3	X	X	X	X	X

The graduate committee for the program will meet annually to review the SLO assessment data, and assess the degree to which students master them..

All SLOs are assessed directly based on evaluated artifacts collected in the core courses; the DS 710 reports, and on the final MS examination (DS 711 MS in Data Science courses, a written report and an oral presentation). A rubric with separate sub-elements for all SLOs will be completed for each student submitting DS710 report by the DGS or the designee. Rubrics for students taking their MS exam will be completed by the committee members at the time of the final MS examination. The expected target performance is good or better (at least 3.0 on the numeric scale).

SLO1, SLO2 and SLO3 are assessed every three years (year one of the cycle). The DGS collects grade statistics for the appropriate courses; and assessments of DS710 reports and the MS exams. The graduate committee evaluates the data and, if needed, proposes an action plan.

SLO4 and SLO5 are assessed every three years (year 2 of the cycle). The DGS collects grade statistics for the appropriate courses; and assessments of DS710 reports and the MS exams. The graduate committee evaluates the data and, if deficiencies are identified, proposes a corrective action plan.

2. SLO assessment rubric – MS degree program

Outcome	Excellent (4)	Good (3)	Acceptable (2)	Poor (1)
Identify, analyze and solve technical problems related to data science and big data analytics	Excellent literature review; precise formal problem statement; comprehensive comparison of approaches; appropriate approach selected, selection informed by the comparison	Good literature review; formal problem statement; some comparison of approaches; selected approach appropriate for the task, selection has some justification in the comparison	Some literature review; correct but only informal problem statement; selected approach appropriate for the task	Poor or no literature review; no, incorrect, or ambiguous problem statement; No or incorrect approach selected
Collect, organize, store, process, visualize and analyze large data sets	All components demonstrated; tools required for each component appropriately chosen, well understood and used correctly	All components demonstrated; tools required for each component appropriately chosen, and used correctly	Most components demonstrated, tools required for each component appropriately chosen and used correctly	None or only few components demonstrated; tools poorly selected, used incorrectly, with little understanding of their functionality
Assemble computational pipelines to support data science from off-the-shelf or self-developed tools	Mastery of the process and the tools demonstrated	Good understanding of the process and of the tools used demonstrated	Ability to build working solutions demonstrated; no efficiency considerations and gaps in understanding the tools used	Little or no understanding of the process; little or no understanding of tools used
Communicate technical concepts effectively, both orally and in writing	Well organized and well executed written report; clear, well timed and well delivered oral presentation	Organized, mostly clear written report, and mostly clear and well delivered presentation	Written report and oral presentation convey key points but are partially disorganized, unclear or ambiguous	Disorganized and unclear written report and oral presentation
Work as data scientist aware of and according to policy, privacy, security and ethical considerations	Demonstrated knowledge of security, privacy and ethical aspects of data science	Aware of the issues, some evidence that they informed the project	Aware of the issues but no evidence they informed the project	Little or no awareness or sensitivity to the security, privacy and ethical issues