

Version 6. 10/24/16

**Proposal for the Initiation of a Doctor of Philosophy of
Data Science and Engineering Program**

Submitted by

The University of Tennessee, Knoxville

Bredesen Center for Interdisciplinary Research and Graduate Education

A NEW PROGRAM LEADING TO THE DEGREE OF:

Doctor of Philosophy

Title of degree as on diploma

Data Science and Engineering

Title of major

CIP/THEC Code

DSE

Formal degree abbreviation

Doctor of Philosophy

Degree designation on student's transcript

August 2017

Proposed starting date

Table of Contents

Abstract.....	3
Estimated size of program	3
Program description	6
A. Mission.....	8
B. Curriculum	10
C. Admission standards	14
D. Faculty.....	17
E. Library resources	20
F. Administration/organization.....	21
G. Support resources.....	22
H. Facilities	23
I. Need and demand	24
J. Duplication	32
K. Cooperating entities	32
L. Diversity and access	32
M. Assessment/evaluation and accreditation	34
P. Cost/benefit	34
Financial estimate form.....	38
Appendix.....	39

Abstract

Institution: The University of Tennessee, Knoxville (UTK)

Division/Department: Bredesen Center for Interdisciplinary Research and Graduate Education
(Bredesen Center)

Program leading to the degree of:

Doctor of Philosophy in Data Science and Engineering (DSE)

This degree will be administered in the Bredesen Center but the degree will be awarded by the Graduate School of the University of Tennessee Knoxville.

Proposed startup date: August 2017

Number of anticipated students: 15 - 25 new doctoral students recruited and enrolled per year

Estimated headcount enrollment and graduates:

Year	New grad students	Head count DSE degree students	Graduates with DSE degree	Attrition
1	15	15	0	2
2	20	33	0	2
3	25	56	0	3
4	25	78	0	3
5	25	100	10	3
6	25	112	15	2
7	25	120	15	3

The plan is to build this Data Science and Engineering doctoral program to a steady state of 25 new graduate students per year, ramping up from an initial first-year class of 15. The elite nature of the program along with industry demand should lead to some attrition - perhaps three out of every 25 entering students will leave before finishing their degree. All of these students over the average of five years will be supported on the \$30K stipend. It is likely that there will be a few part-time students in addition. At the start of the sixth year of the Energy Science and Engineering doctoral program, there are at present seven part-time graduate student compared to 138 full-time students.

No new faculty lines will be established. Instead, Bredesen Center faculty will serve on a part-time unpaid basis and will be drawn from the ranks of existing UTK faculty and research staff at

Oak Ridge National Laboratory (ORNL). Many of the courses offered will be existing ones, therefore reducing the center and the program resources needed for instruction. A few new courses will be organized and offered, in which case some of the administrative funds for the center will be devoted to paying for course development and instruction.

New costs generated by the program:	Year 1	898,660
	Year 2	1,821,789
	Year 3	3,046,638
	Year 4	4,257,949
	Year 5	5,498,594
	Year 6	6,052,708
	Year 7	6,379,529

All of these costs will be covered by three sources of income. ORNL or UTK research groups will share the full cost of the graduate student once the student joins that group for dissertation research - stipend, tuition, and insurance, and then cover the full cost when the student has finished two years of course work. Secondly, UTK will provide a Research Incentive Fund proportional to the volume of grants and contracts that come to the university by virtue of Bredesen Center-related activity, e.g., ORNL support of DSE graduate students - the same model as for the ESE doctoral program within the Bredesen Center. Thirdly, Governor Haslam will ask the Tennessee Legislature to appropriate \$6M of *one-time* funds for support of the new DSE doctoral program.

Total credit hours required for degree: 72

Number of new courses proposed: 10

New courses proposed:

DSE 502 Registration for Use of Facilities (1-15). Required for the student not otherwise registered during any semester when the student uses university facilities and/or faculty time before degree is completed.

DSE 511 Introduction to Data Science and Computing I (3). Topics include: version control, scripting languages, relational and non-relational databases, proper use of data structures, introduction to data science work flows, introduction to project management, and applications.

DSE 512 Introduction to Data Science and Computing II (3). Topics include: platforms for scalable computing including Map Reduce, Hadoop, Spark, and HPC, setting up computing in cloud, and modern data science work flows.

DSE 537 Introduction to Data Analysis and Data Mining (3). Topics include: data visualization, data summaries, missing data, study design, communicating results, linear regression, ANOVA, decision trees, random forests, support vector machines, model diagnostics, cross validation, bootstrap, reproducible research skills. Hands on projects.

DSE 592 Internship (1-3). Individual project to further studies in data science research and allow students to engage in an established external entrepreneurship and/or policy environment.

DSE 593 Independent Study (1-3).

DSE 597 Special Topics (1-3).

DSE 599 Seminar (1).

DSE 600 Doctoral Research and Dissertation (3-15).

DSE 697 Special Topics (1-3).

The Core Curriculum requires that students must complete 21 hours in the following core courses (or substitute, approved in advance).

DSE 511

DSE 512

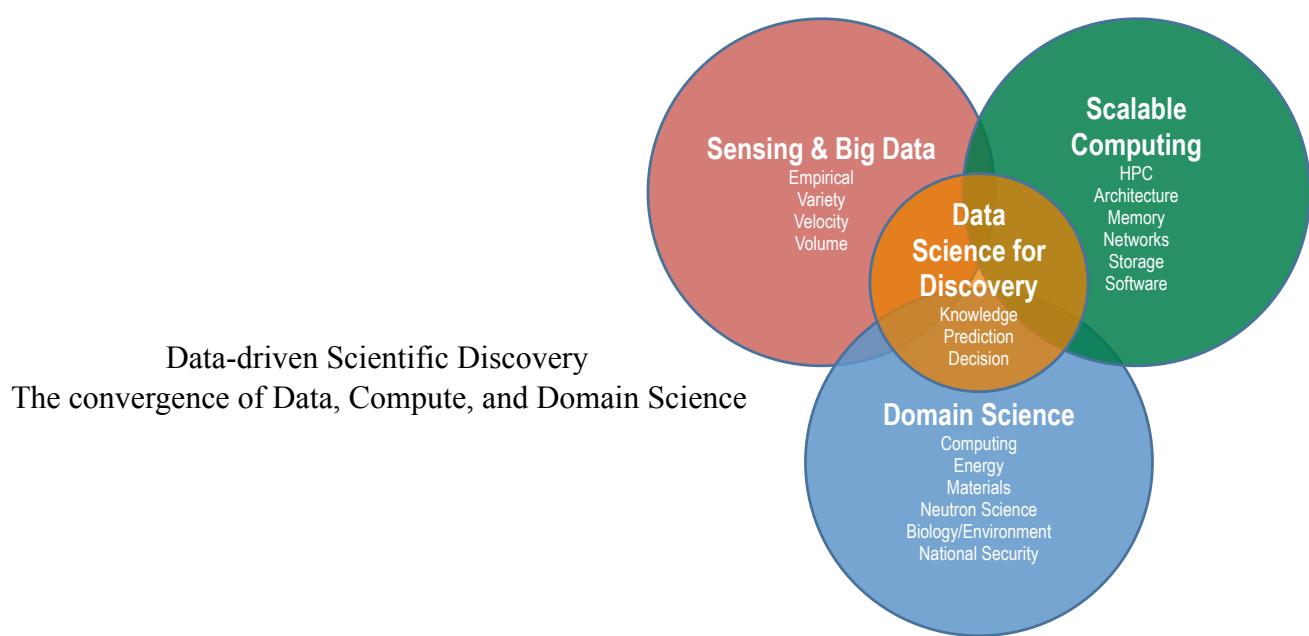
DSE 537

MATH 525 / STAT 563

MATH 526

MSE 510

BZAN 645 / BZAN 646 / ECE 571



Program Description

The University of Tennessee Knoxville (UTK) and Oak Ridge National Laboratory (ORNL) propose to create an interdisciplinary PhD program in the field of *Data Science and Engineering* (DSE). The program will recruit students from the world's leading undergraduate institutions and train them to apply data science and *big data* technology to solve critical problems in science, engineering, and society. Students will be mentored by top researchers at UT Knoxville, the UT Health Sciences Center, UT Chattanooga, and Oak Ridge National Laboratory. The program will directly impact many sectors of the state economy and drive UT to a leadership position in this important area. The proposed DSE doctoral program builds on the successful *Energy Science and Engineering* (ESE) PhD program initiated five years ago between UTK and ORNL. The Bredesen Center will be the academic home for the DSE degree, as with ESE.

A Critical Time. A massive structural transformation is overtaking the world economy based on powerful technological forces. The defining feature of this transformation is the tsunami of data-generating personal devices such as cell phones, tablets, laptops, and associated services. The power of this technological transformation, initially driven by the investments of government and major industries, has now reverberated back, opening the minds of scientists and technologists and spurring more rapid development of new and innovative tools that both consume and generate massive amounts of data. This digital resonance has given birth to the world we now live in, the world of *big data*. This phenomenon has penetrated every corner of society giving birth to giant consumer firms such as Amazon, Google and Uber that have truly revolutionized the worlds of retail, information and media, and transportation with even more significant changes such as robotic self-driving cars coming soon. Many more examples of the dramatic shift in society are readily obvious as older industries quickly disintegrate giving way to firms and that can quickly respond to data and information. Naturally, these changes are coupled with major changes in the scientific and technological research world. Researchers in life sciences, environmental sciences, and healthcare have developed powerful tools such as high-throughput sequencing machines, satellites, sensors, and imaging and genomic screening tools that are now on par with fields such as physics and engineering in their precision, power, and, necessarily, their ability to generate huge amounts of data. In a February 2016 *Wall Street Journal* article entitled “The Big-Data Future Has Arrived,” Michael Malone summed it up this way, “It is now possible to measure everything... and it’s beginning to show big results”.

All of the actors in this transformative play are convinced that the key to success in their field is the ability to combine these oceans of data from disparate sources into coherent information. Unfortunately, people have become the bottleneck restricting this flow of critical information and preventing these organizations, and society as a whole, from

reaping the full benefits of new technologies and research. Numerous articles in the popular press along with government and academic studies have brought attention to this need. For example, a recent Government Accountability Office *Data Analytics and Innovation forum summary* (Persons 2016) noted that participants “warned of an ongoing and potentially widening mismatch between the kinds of jobs that are or will be available and the skill levels of the U.S. labor force.”

It is not just academic and media accounts that point out the needs for these skills, as the market itself makes a powerful case. A 2014 survey by the KDnuggets website reported that the average salary for people with the job title “data scientist” was \$118,000 while those with supervisory roles were earning \$141,000 on average. Industry now unequivocally recognizes the value of the data scientist, a person with the skills to transport, transform, combine, and refine the mountains of data being generated into critical information and accurate decisions.

Foremost among the goals of the proposed program is to strengthen the competitive position of the Tennessee state economy. Data science is firmly established and growing in importance at leading firms such as HCA, FedEx, Eastman Chemical, Blue Cross and Blue Shield, and numerous mid-size firms and startups. Given the technical focus areas and entrepreneurial training, graduates of DSE will have both the technical and business skills to meet the employment needs of these key stakeholders. This program will also create a critical mass of talent in Knoxville with the potential to draw leading firms to the area to take advantage of these resources along with facilities such as the Cherokee Farm development. On the national level, firms such as IBM, Boeing, Lockheed Martin, and Amazon, with strong historical ties to the University, will also recognize the value of DSE graduates as they have with related programs such as the Haslam Colleges Business’s Analytics and Supply Chain degrees and numerous programs in the College of Engineering.

Students and young professionals are aware of the stakes and this had led to significant growth in introductory programs in data science. What has not been addressed in a significant way is the need for advanced interdisciplinary experts that can lead innovation and bring to bear a mastery of data science tools and techniques across critical disciplines in science, engineering, and the local and national economy.

The Bredesen Center at the University of Tennessee Knoxville embraces this challenge and proposes the creation of a joint PhD program in *data science and engineering* that combines the strengths of UTK, Oak Ridge National Laboratory, the UT Health Sciences Center, and UT Chattanooga to address key scientific and societal challenges in healthcare and life sciences, advanced manufacturing, materials science, environment and climate, national defense, urban systems, and transportation.

Model. To meet the critical need articulated above, UT Knoxville, Oak Ridge National Laboratory, the UT Health Sciences Center in Memphis and UT Chattanooga propose the creation of a new PhD degree in *data science and engineering* (DSE). The proposed DSE program is modeled after the successful UTK/ORNL interdisciplinary PhD in Energy Science and Engineering administered in the Bredesen Center for Interdisciplinary Research and Graduate Education. This program has grown remarkably in its first five years and currently includes 138 graduate students with 14 PhDs already awarded. The interdisciplinary nature of this program is ensured by a *Knowledge Breadth Curriculum* providing student experiences in entrepreneurship and/or policy relative to energy. One third of the students focus on entrepreneurship and intend to start new energy-related companies in Tennessee once PhD work is finished.

The Bredesen Center combines the educational resources of a comprehensive research university and the research capabilities of a major national laboratory to provide expanded opportunities for graduate students in energy-related sciences and engineering and, in the new proposed degree, in data science and engineering, fostering multidisciplinary research, large-scale problem-oriented research projects, and innovation and entrepreneurship.

A. Mission

As the preeminent research-based land-grant university in the state, UTK seeks to move forward the frontiers of human knowledge and enrich and elevate the citizens of the state of Tennessee, the nation, and the world. In parallel, the nearby Oak Ridge National Laboratory seeks to conduct basic and applied research that strengthens the nation's leadership in key areas of science; increases the availability of clean, abundant energy; protects the environment; and contributes to national security. The Bredesen Center's proposed PhD in *data science and engineering* seeks to join the missions of these two great institutions to create a rich and stable supply of cutting-edge experts that can help meet the needs of the State of Tennessee and the nation in vital data-centric disciplines.

It is important to detail the value of the bond that is being proposed for all stakeholders. At the state level, the proposed DSE program directly addresses the Tennessee Higher Education Commission Master Plan for 2015-2025, which calls for higher education to address the state's economic development workforce development and research needs and increase the efficiency of degree production. Even more directly, this program provides significant leverage toward meeting UT's Vol Vision 2020 goal by strengthening overall graduate education. It does this in obvious ways, by significantly increasing PhDs awarded and increasing federal and sponsored research expenditures at UT. Not so obviously, but more importantly, the program has the potential, by creating a leading program in a vital area, of increasing the visibility of UT nationally and attracting more of the best and brightest students from across the country and around the world. This increased visibility can also produce an important intended side-effect, increased diversity. This has already happened in the Bredesen Center's ESE program (the largest doctoral program at UTK), which composed of more than 1/3 women and minorities.

ORNL, one the largest and most economically critical organizations in Tennessee also stands to benefit significantly from the DSE doctoral program by expanding its mission to include the education of extremely high productivity workers, by expanding the skill set of its employees, by preparing their staff to tackle new high-priority programs for the Department of Energy and other federal agencies that are increasingly data oriented, and finally by increasing the scope and value of the supercomputing facilities and workforce based there.

In summary, by combining the resources of a comprehensive research university and a major national laboratory, the Bredesen Center provides expanded opportunities for graduate students in energy-related and data-related science and engineering, fostering scholarship and innovation, advancing multidisciplinary research, and accelerating development and deployment of new technologies. Finally, and importantly, the DSE program strengthens the overall partnership between both institutions by creating a foundation for broader and deeper collaboration in the future. The Bredesen Center is front and center in the partnership between UTK and ORNL.

Vision for the Center

Multidisciplinary: The Bredesen Center currently offers graduate students opportunities to engage in multidisciplinary research in energy-related sciences and engineering (ESE) and, in the future, data-related sciences and engineering (DSE), while preserving the rigor and depth of a PhD program.

Entrepreneurial: The Bredesen Center incorporates entrepreneurial experiences, including partnership opportunities in developing and implementing business plans to accelerate the deployment of new technologies in the broader economy.

Transformational: The Bredesen Center has become transformational in engaging graduate students in multidisciplinary projects, large-scale problem-oriented research programs, and science-to-applications research opportunities, enabling scientific breakthroughs and innovative solutions to energy-related challenges and, in the new program, big data-related opportunities.

Educational: The DSE program creates a core teaching competency and curriculum in applied computing and analytics that provides improved educational opportunities and access for departments across the UT campuses. In addition, the program provides continuing education for ORNL staff and serves to strengthen the educational contribution of the laboratory.

The proposed program will address one of the key missions of the university - broaden graduate education and generate more doctorates, a key aspect of the drive towards *Top 25* status among U.S. state-supported universities. The Bredesen Center will recruit students from the world's leading institutions and train them to apply data science and "big data" technology to solve critical problems in science, engineering, and society at

large. The program focuses on seven research areas: health and biological sciences, advanced manufacturing, materials science, environmental and climate science, transportation science, national security, and urban systems science. Since these areas are priorities for the State of Tennessee, the new DSE program will meet the state's growing need for data analytics specialists.

As Malone (WSJ 2016) suggests, the potential for “big data” analysis is limitless and much additional research is needed to learn how to extract information through the analysis of large disparate data sources. Examples of big data challenges that require data science include the study of electronic health records on a massive scale to identify more efficient and cost effective ways to treat patients, and the combination of data from cellphones, GPS, and traffic sensors, which could be used to improve traffic flow and emergency responsiveness. It is important to point out the value of data science can extend beyond science and engineering and be used to inform and enrich the arts and humanities by allowing people to communicate, create, and share in new ways.

The careers of many current students will be shaped in the future by the complex issues related to the generation of, use of, and extraction of new knowledge from large data sets. It behooves universities to prepare students for this increasingly dominant role of data science in so many future careers. Graduate degrees in the conventional fields of science, engineering, business, political science, and law will remain as important training for students to enter this field. However, future leaders in many data-driven fields will need training not only in a prime area of concentration but also in a broad spectrum of disciplines that are somehow related to data. It is time for the country to provide to students a broad training in the issues of data science in addition to a detailed ‘deep dive’ into one of the major fields related to data. This urgent need motivates the proposal for one of the first *interdisciplinary* PhDs in data science and engineering. The issues leading to the generation of most large data sets and the questions that they are used to address are extremely interdisciplinary in nature, thus the need to educate students accordingly in an interdisciplinary manner by coupling computer science, statistics, and analytical techniques to various domain sciences. From this coupled training come advances in *data-driven discovery*.

B. Curriculum

A minimum of 72 hours is required for the DSE doctoral program, and of this total a minimum of 36 hours of coursework is required beyond the BS degree. A typical academic load for a DSE student includes 39 hours of coursework and 61 hours of research credit, totaling 100 credit hours. Of the 36 hours of required coursework, the following courses (or their equivalent) must be completed at a minimum, including the Core Curriculum (21 credit hours), a Knowledge Breadth Curriculum (6), a Knowledge Specialization Curriculum (6), and Seminar Series (3), as summarized below.

All students are expected to take the graduate core curriculum in data science consisting of DSE 511-512, a two semester six-credit sequence emphasizing basic and parallel computing skills, and Math 525 - 526 or equivalently STAT 563 - 564, which are both two-semester six-credit sequences that introduce probability and principles of statistical inference. Students must also take MSE 510 (three credits) or an equivalent course on scientific computing and numerical problem solving, DSE 537 or STAT 577, a three-credit introduction to data analysis, statistics, and machine learning, and either BZAN 645 (Business Analytics), BZAN 646, or ECE 571, which are three-credit courses focusing on machine learning theory and algorithms.

Beyond these 21 hours of required core curriculum, students must also complete six hours from the knowledge breadth curriculum in areas such as policy, entrepreneurship, leadership, or management. IE 557 and ESE 520, 521, and 530 all qualify to meet this requirement. Students will also take DSE 599, a doctoral seminar presenting current research topics - they are required to register for at least of three semesters of this seminar. Knowledge specialization courses in areas such as health and biological sciences, advanced manufacturing, materials science, environmental and climate science, transportation science, national security, and urban systems science are encouraged depending upon area of specialization and can be used to meet remaining requirements for 36 hours of coursework. Additional courses, in mathematics, computer science, business analytics and statistics, or other technical areas are also acceptable.

A. *Core Curriculum (21 credit hours)*

The DSE curriculum will include an important *core* with the following goals for students:

- Understand and apply modern statistics and data analysis;
- Understand computing and coding skills in serial and parallel modes;
- Understand information storage and retrieval;
- Develop a comprehensive foundation in statistics, data mining, machine learning along with the ability to read current literature and implement software tools; and
- Have an understanding of other modeling and simulation approaches from applied mathematics.

Classes should be offered in a distance format and accessible to students at UTHSC and UTC, which will require some expenditure of operating funds. Courses should be offered MW or TTH to allow students to begin research work in the first year.

DSE 511 Introduction to Data Science and Computing I (3 credit hours). Topics include: version control, scripting languages, relational and non-relational databases, proper use of data structures, introduction to data science work flows, introduction to project management, and applications.

COSC 545 - Digital Archaeology which presents equivalent concepts - may be taken as a substitute for DSE 511.

DSE 512 Introduction to Data Science and Computing II (3 credit hours). Topics include: platforms for scalable computing including Map Reduce, Hadoop, Spark, and HPC, setting up computing in cloud, and modern data science work flows.

Math 525 Statistics I (3 credit hours); Topics include: probability spaces, random variables, distributions, conditional probability, maximum likelihood, the EM algorithm, statistical inference & testing, Monte Carlo simulation, important statistical models and estimation.

Math 526 Statistics II (3 credit hours); Topics include: decision theory, Bayes theorem, Bayes rules, priors, hierarchical modeling, Bayesian computations, importance sampling, MCMC simulation, model selection, Bayesian networks, important statistical applications.

MSE 510 Mathematical and Numerical Problem Solving Skills for Materials Scientists and Engineers (3 credit hours); Topics include: Formulation and solution of problems in materials science, including linear and nonlinear algebraic equations, ordinary and partial differential equations, and integral equations. Emphasize on use of modern computational tools.

DSE 537 Introduction to Data Analysis and Data Mining (3 credit hours). Topics include: data visualization, data summaries, missing data, study design, communicating results, linear regression, ANOVA, decision trees, random forests, support vector machines, model diagnostics, cross validation, bootstrap, reproducible research skills. Hands on projects.

BZAN 645/BZAN 646/ECE 571 (3 credit hours) Data Mining, Machine Learning, Pattern Recognition, Theory and Algorithms; Topics include: classification, regression, clustering, adaptive basis function models (trees), ensembles and boosting, kernels, neural networks/deep learning, mixture models, gaussian processes, state space models, undirected graphical models, Monte Carlo methods, graphical models.

B. Knowledge Breadth Curriculum (6 credit hours): select two courses from following areas:

- Political, social, legal, ethical and security issues related to data issues (3-4 courses, each 3 credit hours)
- Entrepreneurship, leadership, and management (3-4 courses, each 3 credit hours).

C. Knowledge Specialization Curriculum (6+ credit hours)

Choose courses from participating departments in a DSE area of specialization:

- Health and Biological Sciences
- Advanced Manufacturing
- Materials Science
- Environmental and Climate Science
- Transportation Science
- National Security
- Urban Systems Science

Currently available courses are listed in appendix A.

D. DSE 599 Seminar (3 credit hours; 1+1+1)

Topical seminars in the focus areas of the Bredeisen Center.

E. DSE 600 (minimum of 36 credit hours)

Doctoral research and dissertation.

Specialty areas and courses

Almost all courses needed for each of the seven specialty areas of the DSE degree exist in various departments. This section lists these courses by department, as shown in appendix A. As discussed above, these courses are needed for the minimum of six credit hours for the Knowledge Specialization Curriculum.

The tables below lay out a proposed timeline of courses to satisfy the curriculum requirements discussed above.

Year 1

FALL SEMESTER	SPRING SEMESTER
Required	Required
DSE 599 Seminar (1)	DSE 599 Seminar (1)
DSE 511 (3)	DSE 512 (3)
Math 525 (3)	Math 526 (3)
DSE 537 (3)	MSE 510 (3)

Year 2

FALL SEMESTER	SPRING SEMESTER
Required	Required
DSE 599 Seminar (1)	Knowledge Specialization Elective (3)
BZAN 645 or other (3)	Knowledge Specialization Elective (3)
Knowledge Breadth Elective (3)	Knowledge Breadth Elective (3)
Elective (3)	Elective (3)

Year 3+: Electives as needed. See appendix A for a detailed listing of the courses.

C. Academic Standards

Admissions

In order to be admitted to the PhD program in data science and engineering, student applicants must fulfill the general admission criteria for the Graduate School of the University of Tennessee Knoxville. In addition, the student must have a Bachelor of Science degree in either engineering or a scientific field (e.g., analytics, biology, chemistry, computational science, mathematics, physics, statistics, etc.), or the equivalent. Students with other undergraduate degrees may also be admitted on a case-by-case basis by the Bredesen Center Graduate Admissions Committee.

Dependent on the student's background, additional coursework may be required to satisfy co- and prerequisites.

Retention

A minimum of 72 hours is required beyond the bachelor's degree, exclusive of credit for a Master's degree, and completion of the core requirements, as outlined in Section B on Curriculum. Of this number, a minimum of 24 and up to 36 hours of 600 Doctoral Research and Dissertation and six hours of 600-level coursework at UTK will be required as specified in the UTK *Graduate Catalogue* under Academic Policies and Requirements for Graduate Students.

No later than one year after entering the program, each student must take a qualifying examination. A student must pass the qualifying examination to proceed in the PhD program.

No later than the first semester of the fourth year following entrance into the PhD program, each student must take and pass a comprehensive examination that includes presentation and approval of the proposed dissertation research. After passing the comprehensive exam, the student should submit the Admission to Candidacy Application to the Graduate School. Admission to candidacy indicates that the student has demonstrated the ability to do acceptable work in the area of study and has made satisfactory progress toward the degree. This action usually connotes that all prerequisites to admission have been completed and the program of study/research has been approved.

After completion of the dissertation, and prior to graduation, each student must pass a dissertation defense examination administered by the student's doctoral committee.

Tracking

Faculty complete performance assessments of the graduate student at each of the major milestones. The mentor provides additional feedback on performance of the graduate student on the PhD Qualifying Examination and the Comprehensive Examination through a standard instrument developed by the Bredesen Center. The DSE program director evaluates the curriculum of each student, verifies that all departmental and Graduate School requirements are met, and approves the course selection.

Milestones

Admission to Candidacy

Admission to candidacy indicates that the student has demonstrated ability to do acceptable graduate work and that satisfactory progress has been made toward the degree. This action usually connotes that all prerequisites to admission have been completed and a program of study has been approved.

A student may be admitted to candidacy for the doctoral degree after passing the *comprehensive examination* and maintaining at least a B average in all graduate coursework. Each student is responsible for filing the Admission to Candidacy form, which lists all graduate courses to be used for the degree, including courses taken at the University of Tennessee or at other institutions prior to admission to the doctoral program. The Admission to Candidacy form is signed by the doctoral committee and by the Director of Graduate Studies in the Bredesen Center.

Graduate Student Examinations

This section provides a description of the graduate student examination requirements for the PhD degree program. Three examinations are required as part of the doctoral program: qualifying examination, comprehensive examination, and defense of dissertation examination.

Qualifying Examination

The qualifying examination is developed, administered, and graded by the faculty (or designated subset of the faculty) of the PhD program under the coordination of the Bredesen Center Director. The Data Science and Engineering doctoral program requires students to be able to investigate and conduct research on a variety of problems. The qualifying examination tests the capabilities of a student through the preparation of a professional quality investigative research report and accompanying presentation that addresses one of several questions in data science and engineering. In case of failure, the candidate may appeal to retake the examination through the Bredesen Center Graduate Curriculum Committee within 30 days of notification of the result. If the appeal is granted, the student must retake the examination at the next offering. The result of the second examination is final. Completion of the qualifying exam enables students to begin working on dissertation research.

Comprehensive Examination

The Comprehensive Examination may be completed as early as the end of the second year following entrance into the PhD program and prior to admission to candidacy. Students should aim to complete the comprehensive exam by the end of the third year and must complete it no later than the first semester of the fourth year unless extenuating circumstances are involved. The timing is late enough in a student's academic program to permit most of his/her graduate course work to be covered on the examination, and early enough to permit modification of the student's program based on the results of the exam.

Two requirements must be satisfied before a student takes the Comprehensive Examination.

1. A written Dissertation Proposal, approved by the major professor, must be submitted to each member of the student's Doctoral Committee two weeks prior to the examination. Each member of the student's Doctoral committee must agree that the student is ready to take the Comprehensive Exam. The committee member will communicate to the major professor when they are satisfied that the student is ready to take the Comprehensive Exam. The Comprehensive Examination will consist of the student constructing and defending his or her dissertation research proposal to the committee in a format deemed acceptable by the student's Doctoral Committee. Typically, an oral defense is sufficient for this examination, although a written component may be administered at the discretion of the Doctoral Committee.
2. Once the Comprehensive Examination is passed, the student should file for and be admitted to candidacy. At the discretion of the Doctoral Committee, supplemental reexaminations for the Comprehensive Examination and/or proposed dissertation research may be required. In case of failure, the candidate may not apply for reexamination until the following semester. The result of the second examination is final.

Defense of Dissertation Examination

A doctoral candidate must pass an oral examination on the dissertation. The dissertation, in the form approved by the major professor, must be distributed to the committee at least two weeks before the examination. The examination must be scheduled through the Graduate School at least one week prior to the examination and must be conducted in university-approved facilities. The examination is announced publicly and is open to all students and faculty members. The defense of dissertation will be administered by all members of the doctoral committee after completion of the dissertation and all course requirements. This examination must be passed at least two weeks before the date of submission and acceptance of the dissertation by the Graduate School. The major professor must submit the results of the defense by the dissertation deadline.

Timeline

The academic level of the graduate student to be recruited is expected to be high, as a result of national recruiting and the unique nature of this program and the partnership between UT and ORNL in educating the student. Therefore, an aggressive timeline is assumed for most students that enroll, starting with the summer before graduate studies actually begin. It is assumed that many students will select an initial research mentor for the summer before coursework begins in the fall, and then continue working with this mentor (or perhaps another one) during the first year. This will guide the student to the intended choice of a dissertation research mentor in the first or second semester of the first year of graduate studies, resulting in a summer of research after the first year. This leads to the definition of a dissertation research topic and passing the *Comprehensive Exam* before the end of the fall semester of the fourth year.

D. Faculty

The Bredesen Center has no dedicated faculty lines. All faculty are appointed as part-time from the ranks of current UTK faculty and existing research staff at ORNL. All ORNL research staff members and UTK and UTIA faculty who fulfill the following criteria are eligible to apply for membership to the Bredesen Center faculty.

- Their appointment will substantially benefit the Bredesen Center and its mission.
- They have a strong record of research and leadership accomplishments in the Bredesen Center's mission areas.
- They are willing to commit the required resources (time, student support, expertise, etc.) to the DSE program or other Bredesen Center projects.

High professional standards will be applied in appointing Bredesen Center faculty. Membership on the Bredesen Center faculty is time limited but renewable. The initial appointment is made for five years and renewal appointments are made for five years.

Responsibilities of Bredesen Center faculty

- They should be actively engaged in Bredesen Center activities, which include mentoring, recruiting, teaching, course development, and committee service.
- They should commit to supervising and supporting at least one graduate student at any given time, ensuring timely completion of the PhD.
- They should provide descriptions of research opportunities, dissertation topics, and shorter research projects available in their groups for the benefit of graduate students searching for a research group and dissertation topic.

Bredesen Center faculty who are not fulfilling these requirements will in general not have their appointment renewed, and can, in severe cases, be terminated as Bredesen Center faculty prior to the end of their term.

Academic titles of Bredesen Center faculty

Bredesen Center faculty with ORNL as their home institution will hold one of the following three UTK titles of Joint Faculty: Joint Professor, Joint Associate Professor, or Joint Assistant Professor. Bredesen Center Faculty with UTK as their home institution will also hold an ORNL title (examples are Research Associate, Senior Research Associate, etc.). The initial title is determined at the time of the first appointment following the process for appointment of Bredesen Center faculty described below. Bredesen Center faculty can request promotions at the time of renewal. Promotion of a Bredesen Center faculty requires a vote by the Bredesen Center faculty, recommendation of the Bredesen Center Director, and approval of the Provost. In cases where a faculty member has an appointment within Bredesen Center and within another degree program, the Bredesen Center director will coordinate any change in title with the other degree program(s). The criteria for the use of the Joint Faculty titles within the Bredesen Center are given below.

Bredesen Center Joint Faculty Professors are expected to

1. hold the doctorate or other terminal degree of the discipline, or present equivalent training and experience appropriate to the particular appointment,
2. be accomplished teachers or mentors of graduate students,
3. have achieved and then maintain a nationally recognized record in disciplinary research, scholarship, and/or creative activity,
4. have achieved and then maintain a record of significant institutional, disciplinary, and/or professional service,
5. serve as mentors to junior colleagues,
6. have normally served as an associate professor for at least five years,
7. have shown beyond doubt that they work well with colleagues, staff, and students in performing their professional responsibilities.

Bredesen Center Joint Faculty Associate Professors are expected to

1. hold a doctorate or other terminal degree of the discipline, or to present equivalent training and experience as appropriate to the particular appointment,
2. be good teachers or mentors of graduate students.
3. have achieved and then maintain a recognized record in disciplinary research, scholarship, and/or creative activity,
4. have achieved and then maintain a record of institutional, disciplinary, and/or professional service,
5. have normally served as an assistant professor for at least five years,
6. have demonstrated that they work well with colleagues, staff, and students in performing their professional responsibilities.

Bredesen Center Joint Assistant Professors are expected to

1. hold a doctorate or other terminal degree of the discipline, or to present equivalent training and experience as appropriate to the particular appointment,
2. show promise as teachers or mentors of graduate students,
3. show promise of developing a program in disciplinary research, scholarship, and/or creative activity that is gaining external recognition,
4. have a developing record of institutional, disciplinary, and/or professional service,
5. show evidence that they work well with colleagues, staff, and students in performing their professional responsibilities.

Bredesen Center faculty appointment process

Requests for initial and renewal appointment as Bredesen Center faculty are submitted to the Bredesen Center Director.

- ORNL applicants who do not currently have a base appointment within an existing UTK degree-granting unit should submit their application through the ORNL division director, who will then forward the application to the Bredesen Center Director.

- Faculty applicants whose base faculty appointment is with an existing UTK degree granting unit should submit their application through the department head, who will then forward the application to the Bredesen Center Director.
- All applications will be reviewed by the Bredesen Center Faculty Credentials Committee. The Credentials Committee will provide a brief written recommendation concerning the decision of membership application and the proposed appointment level to the Director.
- The Credentials Committee is currently composed of senior Bredesen Center faculty who serve in the Energy Science and Engineering doctoral program. This membership will be increased by adding Bredesen Center faculty who also have worked in the data science arena.
- If a positive recommendation is made by the Credentials Committee, the application is brought to the Bredesen Center faculty for discussion and recommendation, which will require a simple majority of the votes with a quorum of at least half of the faculty required. The recommendation of the Credentials Committee and of the current faculty are considered by the Bredesen Center director in forming his/her recommendation, and all three, as well as the appointment level, are forwarded to the Provost for approval by the university.
- The appointment request is required to contain the following elements:
 - A current curriculum vita describing all the professional accomplishments of the applicant.
 - Full education history
 - Full employment history
 - Refereed publications
 - Invited and contributed talks
 - External research funding record
 - Teaching experience
 - Student supervision experience
 - Awards and recognition
 - A brief description (one page or less) of the reason(s) for the request and how the applicant fulfills the eligibility criteria.
 - For the initial appointment a letter of nomination from a current Bredesen Center faculty member or a unit leader at UTK or ORNL.
- The Bredesen Center Director will be responsible for an annual evaluation of all Bredesen Center faculty, shared with appropriate UTK department heads and ORNL division directors.

Faculty Recruitment

Members of the Data Science and Engineering task force have identified a long list of researchers who can contribute to the DSE research goals and demonstrate the viability of the funding model (see Appendix B). Besides these 61, there are many others that will contribute at the participating institutions. Applications for admission to the Bredesen Center faculty from some of these and other researchers will be accepted and considered by the Bredesen Center Credentials Committee. Among the many people that can and will contribute to the graduate

research directions of the DSE program in the Bredesen Center, there are UT and ORNL faculty who have national recognition in *data science* applied to various areas. Among these are anticipated faculty leaders (core data-science faculty) in our program: Vasileios Maroulas, Mark Dean, Audrus Mockus, Mike Langston from UTK, Bob Davis and Saunak Sen from UTHSC, and Arjun Shankar, Arwind Ramanathan, Thomas Proffen, Gina Tourassi, Edmon Begoli, Budhu Bhaduri, Dan Jacobson from ORNL.

Faculty Committees

Advisor/Major Professor

Each graduate student must have an advisor/major professor. This professor advises the student about course selection, supervises the student's research, and facilitates communication within the degree program and/or student's major department, to other departments, and with the Graduate School relative to requirements. A temporary advisor may be assigned to direct the entering student's work during the period in which the student is becoming acquainted with the institutions and determining the focus of research interests. Once the major professor is determined, the major professor and the student together select a doctoral committee. The student is expected to maintain close consultation with the major professor and other members of the doctoral committee with regard to progress in the program.

Doctoral Committee

The major professor directs the student's dissertation research and chairs the doctoral committee. The student and major professor identify a doctoral committee composed of at least four faculty members holding the rank of assistant professor or above, three of whom, including the chair, must be approved by the Graduate Council to direct doctoral research. At least one member must be outside the Bredesen Center faculty. Committee members should be chosen to ensure multidisciplinary breadth. The DSE Program Director has oversight responsibility to ensure the multidisciplinary nature of the committee. A doctoral student, in collaboration with the major professor, should begin to form the doctoral committee during the first year of study. Once formed, the doctoral committee, by request of the major professor, will meet annually, at the minimum, with the student to ensure timely progress toward the degree.

E. Library resources

No new library resources are needed. The library resources of UTK, ORNL, UTHSC, and UTC will serve the needs of this program. The library has extensive holdings in crucial areas such as science, engineering, mathematics, and the areas of *knowledge specialization* that are part of the DSE research areas. These current holdings will be sufficient for meeting the needs of the DSE program. In addition, the library has access to extensive on-line databases and electronic repositories of academic and research journals in these areas crucial to DSE.

F. Administration/Organization

In January 2010 the General Assembly of the State of Tennessee passed legislation authorizing The University of Tennessee to establish an academic unit of The University of Tennessee, Knoxville (UTK) for interdisciplinary research and graduate education in collaboration with Oak

Ridge National Laboratory (ORNL). This academic unit, known as the Bredesen Center for Interdisciplinary Research and Graduate Education, brings together extensive and complementary resources at UTK and ORNL to increase science, technology, engineering, and mathematics (STEM) academic and research activities of national significance focused on energy-related science and engineering. The Bredesen Center enhances collaborations between UTK and ORNL, increases the number of STEM graduate students, advances multi-disciplinary research in energy-related and by this proposal data-related science and engineering, and accelerates the translation of research results into beneficial technologies.

The Bredesen Center already offers one of the first interdisciplinary PhD programs in energy science and engineering. This five-year-old degree provides breadth while preserving the depth and rigor of a PhD program. With this proposal, a similar structure will be used for a PhD in data science and engineering. Topical areas have been chosen to align with unique ORNL capabilities and programs. Both the ESE and the DSE degrees include an emphasis on entrepreneurship and policy, including opportunities for interested students to develop and implement business plans.

Bredesen Center administration

The Bredesen Center is led by a Director (UTK employee) and Executive Director (ORNL employee) appointed jointly by the UTK Chancellor and ORNL Director. The Director is responsible for day-to-day operations, finances, personnel, appointment of Bredesen Center faculty committees, appointment of Bredesen Center faculty, performance appraisals of Bredesen Center faculty, recruiting and admissions, student life, and relationships with UTK departments and administration. The Executive Director assists the Director and is the primary interface with ORNL, including ORNL research programs, staff, management, operations and safety, security, and financial systems. The Director appoints a Credentials Committee, Curriculum Committee, and Graduate Coordinating Committee to assist in administering Bredesen Center programs. The director reports to the Chancellor. Lee Riedinger, Professor of Physics, was appointed to this position effective September 1, 2010. Ian Anderson of ORNL is the executive director.

DSE administration

The DSE program at UTK will be led by a director chosen from the UTK faculty, and Professor Russell Zaretzki from the Haslam College of Business is the initial choice. The DSE Program Director will be responsible for the operation of the Data Science and Engineering doctoral program. The DSE Program Director reports to the Bredesen Center Director.

Bredesen Center Board of Directors

A Board of Directors composed of senior officials at UTK and ORNL oversees the operation of the Bredesen Center. The current Board of Directors is shown below and is chaired by Dr. Davis (UTK) and Dr. Roberto (ORNL)

1	Thomas Zacharia	ORNL	Deputy Director
2	Alan Icenhour	ORNL	Associate Laboratory Director, Nuclear Science and Engineering
3	Jeff Nichols	ORNL	Associate Laboratory Director, Computing and Computational Sciences
4	Jim Roberto	ORNL	Associate Laboratory Director, Science and Technology Partnerships
5	Moe Khaleel	ORNL	Associate Laboratory Director, Energy and Environmental Sciences
6	Michelle Buchanan	ORNL	Associate Laboratory Director, Physical Sciences
7	Paul Langan	ORNL	Associate Laboratory Director, Neutron Sciences
8	Brent Park	ORNL	Associate Laboratory Director, Global Security
9	Taylor Eighmy	UTK	Vice Chancellor for Research
10	Wayne Davis	UTK	Dean of College of Engineering
11	Theresa Lee	UTK	Dean of College of Arts and Sciences
12	Dixie Thompson	UTK	Dean of the Graduate School
13	Steve Mangum	UTK	Dean of Haslam College of Business
14	Matt Murray	UTK	Director, Baker Center
15	Heidi Goodrich-Blair	UTK	Head of Microbiology
16	Bill Brown	UTIA	Research Dean
17	Stacey Patterson	UT	Associate Vice President for Research

DSE External Advisory Board

An external advisory board will provide independent advice and strengthen relationships with industry and other universities. The DSE Director and Assistant Directors will solicit nominations and present a list of potential members for Bredesen Center Director approval. The Advisory Board will meet in person or via teleconference at least once per year. At the initial meeting, members of the Advisory Board will select a Chair and Vice-Chair.

G. Support resources

Financial resources

The DSE doctoral program has been developed as a close partnership between UT and ORNL. Both institutions have committed resources to make this program a success, and in addition the Tennessee Legislature is considering \$6M of one-time funds for the Bredesen Center for the DSE program.

Physical resources

The university assigned the fourth floor of Greve Hall as the on-campus space for the Bredesen Center - its administration, students, and part-time faculty. ORNL and UTK have assigned space in the Joint Institute for Computational Sciences for the administrative functions of the Bredesen Center at ORNL.

Clerical resources

The organization chart shown in section F sketches the 3.0 FTE support staff that now help operate the Bredesen Center and the ESE doctoral program. It is anticipated that another 2.0

FTE staff would be needed for the addition of the DSE program to the Bredesen Center: 1.0 for business operations, 0.5 for development, 0.25 for graduate affairs, and 0.25 for media. Also, the Bredesen Center would pay for part of the salary of the DSE program director and some part of instruction of the DSE core courses. Also, 10% of the existing costs of operating the Bredesen Center would be borne by the DSE budget, as discussed in the section P of this proposal.

Advising resources

Each DSE student selects a committee as described in section D. This committee mentors and advises the student throughout each milestone. In addition, the DSE director and assistant directors help the graduate student in placement in a research group and in career development and employment opportunities. Bredesen Center graduate students will join interdisciplinary research teams at ORNL that will expose them to large-scale problem-oriented research and development, foster their ability to work across disciplinary boundaries, encourage them to approach research problems from new directions, and strengthen their ability to work in teams. Students will be encouraged to develop their research in the context of potential solutions to important national problems, and will be given the tools and support to follow paths consistent with their interests.

General resources

Students will have access to all campus facilities for recreation, health, computing, dining, and career placement.

H. Facilities

No new facilities are needed for this proposed doctoral program. The Bredesen Center occupies the fourth floor of Greve Hall, which houses at present the main administrative functions and staff and provides office and study and collaboration space for the graduate students in the energy science and engineering doctoral program. The plan is to include the graduate students from the data science and engineering PhD program in this space also. This will be a challenge and will result in converting meeting and collaboration space into office and study space for the expanded Bredesen Center graduate student body. Space for the Bredesen Center program is also available in the Joint Institute for Computational Sciences on the ORNL campus.

ORNL has available extensive and world-class user facilities to support graduate students in their research in areas of importance to the DSE program and research directions at the laboratory. There are eight national user facilities at ORNL, including Building Technologies Research Integration Center, Carbon Fiber Technology Facility, Center for Nanophase Materials Sciences, Center for Structural Molecular Biology, Manufacturing Demonstration Facility, National Transportation Research Center, Oak Ridge Leadership Computer Facility, and the Spallation Neutron Source. UTHSC has facilities, personnel, and equipment to support DSE graduate student research in healthcare areas.

I. Need and Demand

Importance

The term *big data* has become ubiquitous in industry and of growing importance to our country. The computerization of society through laptops, tablets, mobile phones, the web, satellites, sensors, and information technology has allowed humans to generate and store massive amounts of data across the entire range of experiences. Examples of this deluge of new data include giant databases of human web traffic, cell phone usage, health care services, economic and business transactions, academic and educational activities, and scientific studies involving the environment, life sciences, physical sciences, social sciences, and engineering. The term *big data* represents the potential, through its analysis, that these massive data have to improve the world.

Task Force

In July UTK Chancellor Jimmy Cheek and ORNL Director Thom Mason convened a Task Force to study opportunities in data science. The task force was chaired by Lee Riedinger (UTK) and Jeff Nichols (ORNL) and included top researchers from ORNL, UTK, UTHSC, and UTC. They proposed developing a PhD program based on seven key research areas that depend critically on the ability to analyze massive data: health and biological sciences, advanced manufacturing, materials science, environmental and climate science, transportation science, national security, and urban systems science. These areas of strength and investment are all current priorities for the State of Tennessee. The *Data Science and Engineering* doctorate will combine training in *computer science* relative to data bases and analytics algorithms coupled to the many big data challenges in specific *domain* areas of strength at our ORNL and UT institutions.

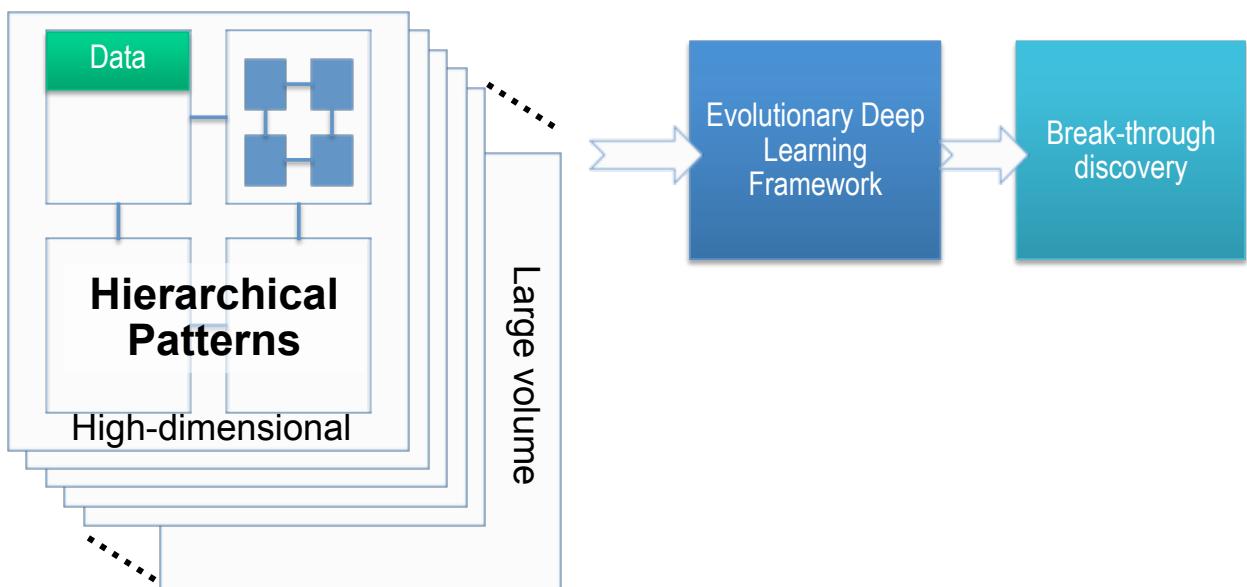
Research and education

Much research and development is needed to learn how to combine and analyze large disparate data sources to extract new information and knowledge. Data science addresses this challenge by bringing together established fields such as statistics, visualization, machine learning, and information retrieval. Electronic health records, when analyzed on a massive scale, should be able to identify more efficient and cost effective ways to treat patients. Cell phones, GPS technology, and traffic sensor data can combine to improve traffic flow and react to emergencies more quickly and effectively. Scientific resources like the Spallation Neutron Source at ORNL will become vastly more impactful by turning data into insights more quickly. On-line data repositories on the web will allow people to electronically access, share, and create music, arts, literature, and other cultural activities in new and innovative ways. The DSE doctoral program will meet Tennessee's growing need for data analytics specialists for science, healthcare, transportation, and a host of other industries.

Data science is a new field that requires knowledge of and expertise in a variety of areas. The chart below tabulates 12 key tools needed for dissertation research and eventual careers in data science. The DSE curriculum has been designed to provide graduate students with exposure and depth in these areas.

Signal & Image Processing	Mathematics	Data Mining	Machine Learning
Statistics	Computer Programming	Data Architectures & Workflows	Pattern Recognition
Visual Analytics	Uncertainty Analysis	Data-driven Modeling & Simulation	High-performance Computing

One cutting-edge tool of data science is *deep learning*, which identifies key patterns in huge data sets and leads to ‘break-through discovery’ in various aspects of science and engineering, e.g., healthcare, materials, cyber security, etc.



Competition

Some leading educational institutions (e.g., Cambridge, MIT, Duke, Stanford, Georgia Tech, and Ohio State) are beginning to invest heavily in data science. Recently the University of Michigan announced that it will invest \$100 million to support a new initiative aimed at working with large data sets to further research in areas like driverless cars, medicine, and climate change. *UT has the unique opportunity to leverage the vast expertise and resources at ORNL to initiate a leading data science and engineering program. With such resources, partnership, and DSE doctoral program, the University of Tennessee can very effectively compete with these universities.*

Research areas

The following seven areas of knowledge specialization are recommended as the initial portfolio of topics for dissertation research, due to the national need and the capabilities of the participating institutions:

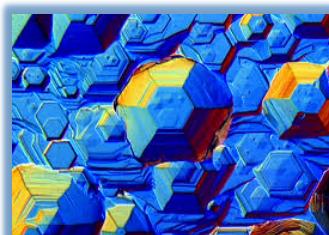
1. Health and Biological Sciences
2. Advanced Manufacturing
3. Materials Science
4. Environmental and Climate Science
5. Transportation Science
6. National Security
7. Urban Systems Science



Health & Biological Science



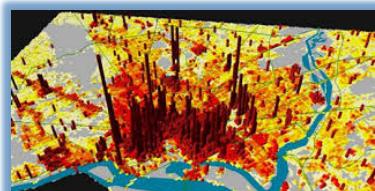
Environmental/Climate Science



Materials Science



Transportation Science



Urban Systems Science



National Security



Advanced Manufacturing

The following are descriptions of the research opportunities in each of these areas.

1. Health and Biological Sciences

Health sciences are undergoing a disruptive change due to the digital revolution of medical technologies, the shear volume and complexity of biomedical data generated on a daily basis, and the massive processing power available to analyze the data in a more holistic way than ever before. The data-driven transformation of health sciences in the 21st century requires a paradigm shift in the way we train the next generation of biomedical scientists and educators. Traditional academic programs are centered on distinct disciplines such as genetics, physiology, cell biology, biochemistry, biomedical engineering, or public health. To effectively tackle the complex multifaceted problems of health and healthcare, an integrative approach is essential to develop novel sensing technologies and to make the most of multi-scale multi-modal data in terms of understanding health processes and improving healthcare delivery systems.

The goal is to train graduate students to extract meaning from diverse, multi-scale data sets through the formation of multidisciplinary teams of faculty and students. These teams will tackle the problem of integrating such disparate data sets in order to derive actionable conclusions through concerted efforts in mathematical model building, computational models, statistical algorithms, and data mining. These teams will thereby decipher new broadly applicable rules and develop conceptual tools that will benefit the novel discipline of data science as a whole.

The Health Data Sciences discipline will have three research tracks: (i) personalized informatics, (ii) population health dynamics, and (iii) health economics and policy. The key differentiator among the three tracks is the unit for societal impact. Personalized informatics will target computational solutions for individual health benefit. Population health dynamics will target computational solutions for understanding and optimizing health processes and outcomes at the community level, while the health economics and policy track will target computational solutions for translational impact in cost-effective healthcare system management and policy.

Engineering also plays a key role in developing treatment delivery systems that respond and adapt to data at both the patient and organization level.

Healthcare is quickly adopting the tools and techniques of data science. For example, large databases of inter-institutional electronic medical records now exist and can be used to study risk factors leading to a variety of rare diseases and to understand the impact of treatment across key subpopulations.

2. Advanced Manufacturing

One of the keys to the economic security of the U.S. is the re-establishment of the U.S. as a world leader in manufacturing. ORNL has established a unique Manufacturing Demonstration Facility (MDF) that integrates state-of-the-art additive manufacturing capabilities, materials science, and sensor/control technology to enable the creation of unique parts that have been previously

impossible to manufacture via traditional casting or milling approaches. A critical step is to be able to certify that an additively manufactured part will achieve all of the mechanical properties necessary to perform at the required level in the final system.

A key example is the development of critical, flight-enabling, additively manufactured parts for aircraft. This simply cannot be done today due to inconsistencies in the manufacturing process, which lead to unpredictable mechanical performance of the part. Additive manufacturing systems monitor thousands of system variables (temperatures, pressures, flow rates, optical/thermal images, spot sizes, etc.) during the printing process, generating huge data sets as a by-product. However, although critical information is contained in this data, currently it is not being used effectively to optimize part quality. New sensor technology R&D is needed to effectively characterize the process. Also, manufactured parts are using x-ray CT scanning systems, mechanical endurance testing systems, etc., to measure voids, cracks, and mechanical performance characteristics. These evaluations generate gigabyte and terabyte data sets per part. Big data science/analytics (e.g., deep learning) should be used to analyze both system censor and post-manufacturing diagnostics data to determine what sets of system parameters lead to the highest quality manufactured parts. These data can be used to both inform existing and create new data-driven models of the manufacturing process. Furthermore, these data science methods can play an important role in achieving the 5-10 year goal of part certification to expand the uses of additively manufactured components in key industries such as aerospace.

3. Materials Science

At its core, materials science and engineering seeks to establish the relationships among structure (electronic to macroscopic)/composition (local to bulk) properties, performance, synthesis, and processing. In recent times, increasingly large amounts of data from ever-improving characterization and computational techniques and methodologies (nano to macro) portend astounding progress in developing unprecedented depth and breadth of these relationships and, consequently, the ability to direct and control matter and to design molecules, materials, and interfaces for specific functionalities. However, to achieve this potential, advances and efficiencies in data acquisition, manipulation, and analytics are needed. High-throughput computational and experimental screening of structures and properties provide new opportunities for discovery and development, but are increasingly dependent on efficient ways of sorting and analyzing the data. State-of-the-art capabilities in materials characterization, including electron microscopy, scanning and atom probes, and neutron and x-ray scattering, produce enormous amounts of data. Such streams however, require effective analytics to correctly interpret structure, composition, and properties, particularly when dynamical processes are being investigated and rapid data analysis, interpretation, and simulation are needed to inform and guide experimentation or for sensing and control. Given the increasing amount of computationally generated, experimental, and engineering data, data mining and knowledge management approaches are critically needed to identify core relationships. Insights and methodologies that expose the coupling of data representing different length and time scales can verify outputs from theory, modeling, and simulation, and provide critical information about the design and engineering of materials and material processing. In all of these aspects, frontier

research in analytics and data science is essential to maximizing the potential of materials science and engineering and increasing its impact on energy, environment, global security, health, and the economy.

4. Environmental and Climate Science

The analysis of large-scale simulations that predict the future climate requires strength in a wide variety of data science techniques ranging from data storage, analysis, uncertainty quantification, visualization, and feedback into models. At the same time, data gathered from radar systems and terrestrial sensors need to be integrated in a well-defined work flow that helps refine models and design better experiments. A recognized need by the Biological and Environmental Research office in the Department of Energy increasingly relies on a principled data processing methodology and infrastructure. This need extends across academia and throughout the private-sector climate change assessments spanning terrestrial, ocean, and atmospheric models. ORNL has a defined leadership role in hosting major environmental data (in the Atmospheric Radiation Measurement Data Archive (ARM) and Carbon Dioxide Information Analysis Center (CDIAC)), and a principal role in executing leadership-class climate simulations. Data science will play a crucial role in combining these world class data sources and producing critical scientific findings. In turn, ORNL and the UT/ORNL Joint Institute for Computational Sciences play an important leadership role in large-scale scientific computing that will contribute to the quality of the DSE program.

5. Transportation Science

Effective and efficient transportation systems are critical to maintaining and enhancing the quality of life of individuals and to enabling the growth and prosperity of businesses. Timely access to accurate and reliable transportation network data is of paramount significance for decision making related to the planning, design, operations, maintenance, and management of transportation networks. Recent advances in technology and communication systems have enhanced the quantity and quality of both spatial and time ordered attributes of the transportation networks and their users. Other data that need to be incorporated into strategic management of transportation systems include demographic, economic, environmental, and socio-behavioral characteristics. These data elements are obtained from a myriad of sources. They have wide-ranging scales of resolution in the spatial and temporal dimensions. These seemingly disparate datasets need to be integrated to support decision making. However, this area is relatively nascent and it offers fertile opportunities for research, development, and deployment in the transportation systems context.

The Transportation Science focus area in the Data Science and Engineering program focuses on developing and applying data science (i.e., knowledge, tools, and products) to enhance societal goals of safe and efficient transportation systems. It addresses challenges related to data integration, synthesis, modeling, simulation, and visualization to help individuals and transportation system managers make better-informed transportation and logistics decisions and engineer more robust and cost effective networks. High impact opportunities in Transportation Science include connected and autonomous (self driving) vehicles, transportation safety,

sustainability, freight and logistics, transportation planning, and transportation operations. These opportunities will leverage ongoing research and state-of-the-art facilities at the UT/ORNL National Transportation Research Center (NTRC). There also are significant synergistic opportunities to collaborate on cross-disciplinary efforts that would include many of the other application areas envisioned for the DSE PhD program.

6. National Security

The heavy investment in and reliance upon cyber infrastructure in the U.S. and other nations around the world are both a vulnerability for national security attacks as well as a potential platform to solve some of the most complex national security problems. Cyber security is an extremely challenging issue for the U.S. and the world, and data science is at the heart of many technology solutions that can address the growing cyber security threat. Furthermore, there has been an exponential increase in the number of sensors and other devices connected to the cyber infrastructure, generating data that are relevant to national security challenges. The Department of Defense, Department of Homeland Security, and the intelligence community are collecting orders of magnitude more data than can be intelligently analyzed by humans and computers, leading to a big data bottleneck across national security agencies. The data being collected are multi-modal and streaming 24/7 in many cases. There is a tremendous need for data scientists that can tackle the unique big data challenges encountered in the national security space. Engineering data collection, management, and analytic platforms to process multi-modal, streaming data sources and then provide real-time knowledge to decision makers are essential. In addition, the DSE program focuses on extending and improving the sources of data for national security challenges. To that end, trained scientists are needed to research and develop next-generation smart sensors that more effectively measure specific signals of interest and also have on-board capability to intelligently convert data to knowledge at the point of measurement. The proposed DSE program focuses on training the next generation of data scientists who have (1) the domain knowledge to understand the complexities of the national security landscape and (2) the expertise to research and develop smart sensors, methods, algorithms, and data analytic platforms to tackle national security challenges.

7. Urban Systems Science

In the rapidly urbanizing world, the unprecedented rate of population growth is mirrored by increasing demand for energy, food, water, and other natural resources, which also creates detrimental impacts on environment and human security. The *Internet of Things*, which is expected to include 50 billion devices by 2026, provides the ability to observe and measure human impacts through direct instrumentation of the man-made environment from the individual to the building and finally to the planet scale. This explosion of data, including citizen sensors, provides a unique opportunity to manage and increase efficiencies of existing infrastructures as well as design and engineer a more sustainable future.

Urban informatics requires characterization of the interactions between the human dynamics and critical infrastructures. A recent ORNL initiative aims at understanding

infrastructure domains, such as transportation, water, and public health, coupled with developing novel algorithms and computational strategies utilizing machine learning and data mining. Development of novel technical capabilities is needed to enable innovative pervasive sensing combined with scalable data collection, integration, analysis, and knowledge dissemination techniques for cross-domain applications. Capabilities to observe, measure, analyze, and model urban dynamics will allow unique data-driven understanding of complex societal systems that are governed by both physical and behavioral sciences. This will lead to transformational pathways guided by novel science and technology shaping a future society where our energy, environment, and mobility goals are attained simultaneously.

Role in economic development

The DSE degree program presents profound opportunities for economic development in the State of Tennessee. The seven initial research areas of focus summarized above all have great importance, not only to the country, but even more to the state.

1. Health and Biological Sciences - the medical industry based in Memphis and Nashville is deeply involved in issues of managing health data and using such data to better manage all aspects of disease and treatment and to help control costs throughout the system. Research on how to analyze and mine health-care data in new ways can have enormous pay-off for the State of Tennessee. UTHSC and ORNL can play a central role in this.
2. Advanced Manufacturing - ORNL and UTK have new programs, new hires, and vast new opportunity in additive manufacturing. This effort was bolstered by a huge federal grant recently awarded to these two institutions. Better use of data will be key in the growth and success of additive manufacturing to a wide industrial sector by helping to improve reliability and reduce costs just as earlier generations of manufacturers benefited from data based reliability and quality management programs such as Six-Sigma.
3. Materials Science - ORNL and UTK are leaders in a number of areas of materials science, in part via the Institute for Functional Imaging of Materials in the Center for Nanophase Materials Science. Harnessing big data, deep data, and smart data from state-of-the-art imaging might accelerate the design and realization of advanced functional materials and lead to new companies making devices based on these ‘smart’ materials.
4. Environmental and Climate Science - the country’s largest climate modeling calculations are now performed on the Titan supercomputer at ORNL. Huge data sets across a range of geographical regions and from a host of monitoring platforms on climate conditions present a big challenge for understanding climate warming of the past 100 years and projecting that forward in time. The State of Tennessee will be very dependent on future temperature, precipitation, and violent storm increases as the climate continues to warm and thus it is important to understand and model these quickly changing trends in addition to gaining understanding of other potential impacts.
5. Transportation Science - UTK and ORNL have well established transportation research programs which benefits the state directly. Because of its central location between eastern and midwest population centers, Tennessee is home to a large transportation and logistics industry that includes both ground and air business. The use of data to reduce

accidents and increase transportation system capacity is a key to both industry function and individual quality of life. Making use of the large data sets collected by road monitors and sensors relative to transportation is a vital challenge that data science can address. The SimCenter at UTC performs computations in this arena and the creation of new data scientists with expertise in this arena can clearly benefit the state.

6. National Security - in no other area is the handling, analysis, and understanding of big data trends more critical than in national security. Securing critical data and protecting public institutions and facilities is also of critical importance to the state. ORNL is a leader in these programs.
7. Urban Systems Science - modern urban environments are rich in both sensor enabled technologies and opportunities for growth in systems that improve quality of life and decrease costs of living. For example, renewable energy resources will only operate effectively if demand for electricity is managed and synchronized to occur when the devices are generating power. The key to converting sensor information into high impact control systems that improve urban the quality of life is through data analytics.

The State of Tennessee has projected that the largest growth in jobs in the next 8 – 10 years will be in the fields of computing, mathematics, engineering, and healthcare (jobs4TN.gov). STEM graduates are essential to the economic health of the state and region. The graduates of this program will be employed by leading firms in Tennessee and across the nation. They will contribute to research teams in industry, government, retail, healthcare, energy, manufacturing, aeronautics, and transportation positions. The core skills of the graduates will enable them to become leaders in these industries. Many of these graduates will make their own way as innovators and entrepreneurs, creating employment opportunities for others as they build new visions and applications of data science.

J. No unnecessary duplication

There are no comparable programs at the University of Tennessee or in the state of Tennessee. There are a few such data science and engineering programs at other universities, and funds have been allocated for other universities to initiate such a degree program. The opportunities for research, funding, graduate student interest, and jobs after the PhD are great, and what is proposed will be unique.

K. Cooperating Entities

Faculty/research leaders and administrative leaders at four major institutions in Tennessee have pledged their support in the development of the data science and engineering doctoral program: ORNL, UTK, UTHSC, and UTC. In addition local industry (e.g., Cisco in Knoxville) is very supportive of this new doctoral program since it will generate experts that they will try to hire.

L. Diversity and Access

Graduate students will be nationally recruited for this new doctoral program in a large annual campaign led by Oak Ridge National Laboratory. Each year recruiters from ORNL visit 20 top universities across the U.S., to attend job fairs and advertise the ESE doctoral program to top

undergraduate students. This on-campus recruitment will be extended to promote the DSE doctoral program. Diversity will be a strong consideration in this recruitment process. Part of this ORNL/UTK emphasis is recruiting from high quality minority institutions, especially Historically Black Colleges and Universities.

There is a rich diversity in the current 138 doctoral students in the energy science and engineering PhD program. One third are female. Most of the Bredesen Center graduate students are placed with mentors for dissertation research, including 89 supported and guided by an ORNL faculty mentor, 23 by a UT mentor, and 20 by Governor's Chairs (joint between UTK and ORNL). A total of 17 other countries have sent 36 current students to the ESE program, including India (9), China (6), South Africa (4), Nigeria (3), South Korea (2), and one each from Argentina, Bangladesh, Colombia, Ecuador, France, Hong Kong, Kenya, Mexico, Pakistan, Puerto Rico, Togo, and Uganda.

The recruiting and placement of graduate students into the DSE program follows the general model developed for the ESE degree program, which has been highly successful. Each year applications to the Graduate School and to the degree program will be accepted through January, then faculty of the Bredesen Center will help decide which applicants should be invited for interviews at ORNL and UT at the end of February. Rigorous on-site interviewing will help determine which students are suited for an interdisciplinary doctoral program and which students should be offered a fellowship and pursued heavily. Fellowships with stipends competitive with similar high-valued degree programs elsewhere in the country and compatible with those of the ESE program will be offered one week after the on-site interviews to a subset of the applicants.

The placement of DSE graduate students into different research tracks begins with the interview process. Bredesen Center faculty associated with the seven DSE research tracks will participate in the interviews, to help decide the best and brightest and also to spot students that they would like to recruit to their research group. It is important to recruit students into various DSE research areas, hopefully matched up via the interview process with specific faculty. Faculty from all four participating institutions need to be involved in the interview process. Offers of fellowships need to be made to achieve a good distribution across research areas and institutions. For example, finding students interested in the healthcare aspects of the DSE degree program is important, in part to increase research partnerships in data science with UTHSC.

As a second step in the admissions process, graduate students are encouraged to start their engagement in the DSE program in the summer before admission. As with the ESE doctoral program, summer internships for incoming graduate students (the summer before the August start) are an important way to aid in rapid placement of students in the appropriate research area and group. ORNL has hosted summer internships for incoming ESE graduate students, and over half of the ad hoc placements for these appointments end up becoming the home for dissertation research. The task force envisions a similar process for incoming DSE students, at ORNL and at each of the UT campuses. Funding each of these initial summer programs will be a challenge, but the outcome would be quick involvement of students in a research program.

The third step in the placement of students in a research program is a fall semester seminar (DSE 599) hosting talks by faculty in different research areas. The goal of the recruitment and placement process is that all first-year graduate students will join a research effort by the end of the first semester. This has functioned well in the ESE program and helps to ensure an early focus on research and prompt progress toward a PhD.

Faculty from UTHSC and UTC will be able to direct dissertation research in the DSE program, once they are approved to be members of the Bredesen Center faculty and certified by the Graduate School.

M. Assessment/Evaluation and Accreditation

This degree will not be subject to accreditation, as there is no accreditation process or body for such an interdisciplinary program.

The Bredesen Center Board of Directors is appointed by the UTK Chancellor and ORNL Director to oversee the development and operation of the Bredesen Center. The Board has balanced representation from UT and ORNL. The Bredesen Center Director and Executive Director serve ex-officio on the Board. A DSE External Advisory Committee will provide advice to the Bredesen Center Director and strengthen relationships with industry and universities.

The Board of Directors and the External Advisory Committee will perform regular evaluation of the DSE PhD program. The Board of Directors will evaluate performance relative to established goals and an in-depth university evaluation of the center and the graduate program will be performed every five years.

Specific key performance indicators for the DSE program are similar to those already used in the energy science and engineering doctoral program within the Bredesen Center. These include journal publications, national fellowship awards, entrepreneurial successes (e.g., startup of new companies and receipt of venture capital funds), published contributions to policy discussions relative to energy (and data science for the DSE program), and placement in high profile positions upon graduation.

P. Cost/benefit

The costs of this proposed program are included in the table at the end of this document. Costs have been escalated at 2% per year on average. The one-time allocation of \$6M from the Governor and the Tennessee Legislature has been requested and is shown in Revenues.

Financial estimate form

The anticipated expenditures related to the DSE program are shown in the Financial Estimate Form below. The Bredesen Center was established in 2010 to manage the new Energy Science and Engineering (ESE) PhD program, and the financial model for the proposed Data Science and Engineering (DSE) PhD program is built in the same way. Infrastructure that has been built for

the ESE program will be shared with DSE. Experience in the operation of the ESE financial model for five years gives confidence that a similar model for DSE will work.

As in the ESE program, the costs of the new DSE doctoral program are covered by three sources of ‘income.’ ORNL or UT research groups pay part of the cost of the graduate student once the student joins that group for dissertation research - stipend, tuition, and insurance. When the student is finished with coursework and spends full time on dissertation research, then the group covers the full cost of the DSE student. Secondly, UTK provides a Research Incentive Fund (F&A recovery) proportional to the volume of grants and contracts that come to the university by virtue of Bredesen Center-related activity, e.g., ORNL support of DSE graduate students. Thirdly, the top programmatic priority item from UT for FY18 is a requested appropriation from the Tennessee Legislature for \$6M of one-time funds in support of the proposed DSE doctoral program; the Governor is supportive of this request. When this one-time fund is depleted (after year 5), the university will provide part of the earned tuition to replace the state allocation.

The budget form below shows total expenditures of \$1,005,250 in year 1 of the program, building to \$6,501,368 in year seven. By far the biggest component of this is the support of the graduate students in the DSE program: stipend, tuition, insurance - \$6,087,144 of the \$6,501,368 in year 7.

The DSE financial model has the following assumptions and initial conditions built into it.

- **Student body size.** The goal is to ramp up to 25 new doctoral students per year, starting with 15 in year 1, 20 in year 2, and 25 thereafter. This has been the profile for ESE also.
- **UT/ORNL distribution of DSE graduate students.** The goal is that 70% of the DSE graduate students are supported in their dissertation research by ORNL groups, 30% supported at UT research groups. More than the 70% will work on dissertation research at ORNL, but supported by UT grants to joint faculty such as Governor’s Chairs. Cost sharing of their stipends and tuition (between the Bredesen Center and the research group) takes place in both scenarios, but Research Incentive Funds come to the Bredesen Center for support of the program when a graduate student works in an ORNL group and that mentor writes a subcontract to send part or all of the student support package to UTK, paying also the off-campus overhead. The budget sheet includes as revenue \$511,344 of such Research Incentive Funds coming to the Bredesen Center in year 6 (‘institutional reallocations’).
- **Bredesen Center support of DSE graduate students.** The model for ESE will be carried over to DSE. The Bredesen Center covers on average 3/4 of the graduate student package in the first two semesters of graduate school, assuming that the student joins a research group during the first semester and the group is able to provide part of the funding in the second semester. The Bredesen Center part of the package is half in the second year and zero thereafter, since it is expected that the graduate student is finished with coursework in year 3 and can devote more time to dissertation research, with that group assuming the financial responsibility.
- **Sharing of Bredesen Center infrastructure.** The budget form reflects the sharing of Bredesen Center infrastructure between the ESE and the DSE program. That is, the DSE

budget covers 30% of the salaries and benefits of the Bredesen Center director and the 3.0 FTE staff that operate the Bredesen Center. Also, the space allocated to the Bredesen Center (the fourth floor of Greve Hall) is sufficient to house the DSE in addition to ESE students. The space available at ORNL in the Joint Institute for Computational Sciences will also suffice.

- **New staff.** The number of graduate students and the workload of the Bredesen Center will double with the addition of the DSE program. One new clerical staff will be hired in year 1 and one more in year 3. One of these will be devoted to business operations and the other to graduate students affairs: recruiting, following course work and graduate school requirements, etc. Also part of the existing salary of the DSE program director (a UTK College of Business faculty member) will be paid.
- **Instruction.** The financial worksheet includes \$25,000 for ‘faculty,’ which is needed to pay for part of an existing faculty member’s salary to teach a few of the core courses in the DSE curriculum. Since the start of the Bredesen Center, the director has taught the ESE core course (a two-semester graduate energy course), which has incurred no extra expenditure of funds. However, DSE core courses will be taught by various faculty, thus the need for partial salary support.

The THEC financial projections worksheet below includes expenditures and revenues. By definition, the ‘revenues’ include all revenues to the university. In year 6, the tuition line and fees line is \$1.85M that is paid to the university for the tuition of the 125 graduate students in the program. This total is paid in part by the Bredesen Center, ORNL research mentors, and UT research mentors. The ‘federal grants’ line is the ORNL support of graduate student stipends and health insurance for those that are working on dissertation research at the laboratory. The ‘institutional reallocation’ line is the Research Incentive Fund (F&A recovery) proportional to the volume of grants and contracts that come to the university by virtue of ORNL support of DSE graduate students and the subcontracts sent to the university. This is a revenue to the Bredesen Center.

The ‘other’ line on the revenue part of the worksheet below contains several other pieces of revenue to the university, as described in the table below using year 6 as an example.

Year 6 finances in thousands		
Calculation of "Other"		
UT PI support of student stipends and insurance	\$853.5	
BC administrative, staff, operating	\$406.0	\$406.0
BC share of student stipends and insurance	\$856.5	\$856.5
Subtract F&A recovery	\$511.3	\$511.3
	Sum = "Other"	\$1,604.7
BC share of tuition		\$404.8
Subtract State support		\$30.0
Needed from central administration for Bredesen Center		\$1,126.0

The ‘other’ revenue includes \$853.5K of stipend and insurance coverage of DSE graduate students (30% of the total) that work on campus for UT mentors (the tuition for this 30% is included in the ‘tuition’ line). It needs to also include revenue to cover the Bredesen Center share of graduate student stipends and insurance (\$856.5K), and Bredesen Center costs for administrative leadership, staff salaries, and operating for the DSE program (\$406K). But, part of this needed revenue for Bredesen Center expenses is covered by the F&A recovery for year 6, \$511.3K. The total of ‘other’ revenue is thus \$1,604.7K in year 6.

The last column in this table shows the items of Bredesen Center spending for year 6. This includes administrative, staff, and operating (\$406K), the Bredesen Center share of the stipends and insurance for the student body (\$856.5K), and the Bredesen Center share of tuition for students (\$404.8K). These expenses for the Bredesen Center are offset by the F&A recovery (\$511.3K) and the last part of the \$6M state allocation (\$30K). The remaining \$1,126K of Bredesen Center support would come from the UTK central administration, part of the \$1,850K in tuition that comes to the university for the 125 graduate students. In years 1 through 5, the state \$6M allocation covers the Bredesen Center expenses (along with the F&A recovery), but the last part of this principal is spent in year 6 (\$30K). In year 7, the required support from part of the earned tuition would be \$1,185K.

A similar scenario has evolved for the ESE program. A one-time allocation of \$6.2M from the State of Tennessee started the program. Cost sharing and Research Incentive Fund return has been done in the same way; \$540K of F&A recovery funds in FY2016 came to the Bredesen Center. The state one-time allocation was spent to zero after five years and in year six a total of \$1.2M was transferred to the Bredesen Center from tuition earnings. For DSE the state allocation would last through year 5 also. The funding model and expected outcomes are very similar.

It is anticipated that grants and contracts will provide additional support to the Bredesen Center in coming years. In the ESE program Bredesen Center faculty have occasionally made grant proposals through the Bredesen Center, and a few have been funded. This does not provide direct financial support of the Bredesen Center, but a small amount of Research Incentive Funds come to the center and help in covering the costs of operation.

Tennessee Higher Education Commission
Attachment A: THEC Financial Projections
The University of Tennessee Knoxville
Data Science Engineering (DSE) Doctoral Program

DSE PROPOSAL

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
I. Expenditures							
A. One-time Expenditures							
New/Renovated Space							
Equipment	\$1,000			\$1,000			
Library							
Consultants	\$2,000						
Travel							
Other							
Sub-Total One-time	\$3,000			\$1,000			
B. Recurring Expenditures							
Personnel							
Administration FTE							
Salary	\$134,259	\$136,944	\$139,683	\$142,477	\$145,326	\$148,233	\$151,197
Benefits	\$45,627	\$46,539	\$47,470	\$48,419	\$49,388	\$50,375	\$51,383
Sub-Total Administration	\$179,886	\$183,483	\$187,153	\$190,896	\$194,714	\$198,608	\$202,580
Faculty							
Salary	\$25,000	\$25,500	\$26,010	\$26,530	\$27,061	\$27,602	\$28,154
Benefits	\$0	\$0	\$0	\$0	\$0	\$0	\$0
Sub-Total Faculty	\$25,000	\$25,500	\$26,010	\$26,530	\$27,061	\$27,602	\$28,154
Support Staff FTE							
Salary	\$77,884	\$79,442	\$114,844	\$117,141	\$119,484	\$121,873	\$124,311
Benefits	\$28,574	\$29,145	\$40,886	\$41,704	\$42,538	\$43,389	\$44,257
Sub-Total Support Staff	\$106,458	\$108,587	\$155,731	\$158,845	\$162,022	\$165,262	\$168,568
Graduate Assistants							
Salary	\$450,000	\$1,050,000	\$1,800,000	\$2,550,000	\$3,300,000	\$3,600,000	\$3,750,000
Benefits	\$28,120	\$70,205	\$128,777	\$195,204	\$270,300	\$315,514	\$351,666
Tuition and Fees* (See Below)	\$199,537	\$479,553	\$846,754	\$1,235,555	\$1,646,923	\$1,850,542	\$1,985,478
Sub-Total Graduate Assistants	\$677,656	\$1,599,759	\$2,775,531	\$3,980,759	\$5,217,222	\$5,766,056	\$6,087,144
Operating							
Travel	\$5,000	\$5,100	\$5,202	\$5,306	\$5,412	\$5,520	\$5,631
Printing	\$1,000	\$1,020	\$1,040	\$1,061	\$1,082	\$1,104	\$1,126
Equipment	\$0	\$0	\$0	\$0	\$0	\$0	\$0
Other	\$7,250	\$7,395	\$7,543	\$7,694	\$7,848	\$8,005	\$8,165
Sub-Total Operating	\$13,250	\$13,515	\$13,785	\$14,061	\$14,342	\$14,629	\$14,922
Total Recurring	\$1,002,250	\$1,930,844	\$3,158,209	\$4,371,092	\$5,615,361	\$6,172,158	\$6,501,368
TOTAL EXPENDITURES (A + B)	\$1,005,250	\$1,930,844	\$3,158,209	\$4,372,092	\$5,615,361	\$6,172,158	\$6,501,368

*If tuition and fees for Graduate Assistants are included, please provide the following information.

Base Tuition and Fees Rate	\$13,302	\$13,702	\$14,113	\$14,536	\$14,972	\$15,421	\$15,884
Number of Graduate Assistants	15	35	60	85	110	120	125

Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
--------	--------	--------	--------	--------	--------	--------

II. Revenue

Tuition and Fees ¹	\$199,537	\$479,553	\$846,754	\$1,235,555	\$1,646,923	\$1,850,542	\$1,985,478
Institutional Reallocations ²	\$0	\$38,070	\$105,579	\$224,883	\$361,862	\$511,344	\$573,439
Federal Grants ³	\$146,424	\$406,074	\$864,936	\$1,391,778	\$1,966,708	\$2,205,535	\$2,267,198
Private Grants or Gifts ⁴	\$0	\$0	\$0	\$0	\$0	\$0	\$0
Other ⁵	\$659,289	\$1,007,146	\$1,340,940	\$1,519,875	\$1,639,869	\$1,604,736	\$1,675,253
BALANCED BUDGET LINE	\$1,005,250	\$1,930,844	\$3,158,209	\$4,372,092	\$5,615,361	\$6,172,158	\$6,501,368

Notes:

(1) In what year is tuition and fee revenue expected to be generated and explain any differential fees. Tuition and fees include maintenance fees, out-of-state tuition, and any applicable earmarked fees for the program.

Tuition is generated in all years; no differential fees

(2) Please identify the source(s) of the institutional reallocations, and grant matching requirements if applicable.

F&A recovery - 100% retained by Bredesen Center.

(3) Please provide the source(s) of the Federal Grant including the granting department and CFDA(Catalog of Federal Domestic Assistance)

UT-Battelle (DOE) contracts

(4) Please provide the name of the organization(s) or individual(s) providing grant(s) or gift(s).

n/a

(5) Please provide information regarding other sources of the funding.

DSE State Appropriation is showing in Other row, and is mostly spent out in first five years.

Appendix

A. Courses

Existing courses available for DSE graduate students

Course Code	Department
AE	Aerospace Engineering
BZAN	Business Analytics
CBE	Chemical and Biomolecular Engineering
CE	Civil Engineering
COSC	Computer Science
ECE	Electrical and Computer Engineering
ECON	Economics
EEB	Ecology and Evolutionary Biology
ENVE	Environmental Engineering
EPP	Entomology and Plant Pathology
GEOG	Geography
IE	Industrial Engineering
INSC	Information Sciences
LFSC	Life Sciences
MATH	Mathematics
ME	Mechanical Engineering
NE	Nuclear Engineering
POLS	Political Science
PUBH	Public Health
SOCI	Sociology
STAT	Statistics

Existing and Proposed Electives:

General Electives in Computer Science and Programming

COSC 462 – Parallel Programming

COSC 565 – Survey of Programming Languages

COSC 581 – Algorithms

ECE 575 – High Performance Computer Modeling and Visualization

Statistics, and Machine Learning, and Information Science

Regression Modeling: Theory and Application (Proposed)

Matrix Algebraic theory of linear models and application of regression models using R language

Advanced Regression Modeling (Proposed)

Regression Models for non-linear and categorical outcomes. Mixed model extensions for panel data. Non-parametric regression. Introduction to survival analysis and lifetime data.

Existing Courses

BZAN 552 – Multivariate and Data Mining Techniques
BZAN 553 – Design of Experiments
BZAN 642 – Advanced Stochastic Analysis
BZAN 645 – Advanced Topics in Data Mining
BZAN 646 – Modern Multivariate Techniques
BZAN 648 – Advanced Topics in Design of Experiments and Linear Models
BZAN 649 – Observational Studies and Causal Models
COSC 425 – Machine Learning, Multivariate Statistics
COSC 526 – Introduction to Data Mining
COSC 528 – Introduction to Machine Learning
COSC 557 – Visualization
ECE 471 – Introduction to Pattern Recognition
ECE 571 – Pattern Recognition
ECE 697 – Network Theory
MATH 527 – Stochastic Modeling
MATH/COSC 571 – Numerical Analysis
MATH/COSC 572 – Numerical Analysis II
MATH 577 - Optimization
INSC 585 - Information Technologies
INSC 587 - Mining the Web
INSC 592 - Big Data Analytics
STAT 567 – Lifetime Data Analysis
STAT 575 – Time Series Analysis and Forecasting
STAT 578 – Categorical Data Analysis
STAT 579 – Multivariate Analysis

Health Care/Life Sciences

Analysis of observational data, Missing values, and causal inference (Proposed)
Epidemiological designs and analysis (Proposed)
Statistical genetics (Proposed)
Statistical analysis of high throughput data (Proposed)
PUBH 530 - Biostatistics
PUBH 537 – Fundamentals of Program Evaluation
PUBH 540 – Principles of Epidemiology
PUBH 542 – Advanced Epidemiological Methods
PUBH 637 – Applications in Program Evaluation
ECON 582 – Econometrics 2
ECON 682 – Panel Data Econometrics
EPP622 - Bioinformatics Applications
LFSC 520 (GST I) - Advanced Genetics/Genomics
LFSC 507 - Bioinformatics/Computational Biology Lab I
LFSC 521 (GST II) - Analytical Technologies

LFSC 517 - Bioinformatics/Computational Biology Lab II
MATH/EEB 581 - Mathematical Ecology I
MATH/EEB 582 - Mathematical Ecology II
MATH/EEB 585 - Mathematical Evolutionary Theory
CBE 672 – Computational Bioinformatics

Cyber Security
ECE 554 – Computer Security and Forensics

Image Analysis and Processing
ECE 572 – Digital Image Processing

Materials, Environment, Climate Engineering
COSC 471 – Numerical Analysis, Linear Algebra
COSC 472 – Numerical Algebra
MATH/COSC 574 – Finite Element Analysis
MATH 513-514 – Principles of Fluid Mechanics I and II
MATH 515-516 - Analytics and Applied Math
MATH 523-524 – Probability I and II
MATH 535-536 – PDE
MATH 578 – Numerical Methods for PDE
MATH 673-674 – Advanced Topics in Numerical Partial Differential Equations
ME 517/518 – Finite Element Analysis, CFD
ME 529 – Linear Algebra
ME 541-542 – Fluid Mechanics
AE577 - Neural and Fuzzy Approaches in Engineering
NE 579 – Empirical Models for modeling and diagnostics
NE 582 - Monte Carlo Analysis
NE 671 – Advanced Topics in Artificial Intelligence
ENVE 561 – Climate and Environmental Informatics
ENVE 562 – Three dimensional climate modeling

Optimization, Transportation, Supply Chain
BZAN 610 – Probability and Stochastic Processes
BZAN 620 – Prescriptive Analytics
BZAN 630 – Decision and Operations Analytics
BZAN 640 – Advanced Prescriptive Analytics
BZAN 641 – Advanced Supply Chain Analytics
ECE 611 - Convex Optimization
ECE 612 – Discrete Optimization
ECE 616 – Nonlinear Programming
ECE 619 – Application of Constrained Optimization
IE 522 – Optimization Methods

IE 526 – Advanced Systems Modeling and Simulation
IE 552 – Advanced Linear Programming
IE 556 – Data Mining in Manufacturing
IE 602 – Nonlinear Optimization
IE 604 – Network Planning and Scheduling
IE 607 - Stochastic Processes
IE 608 - Advanced Optimization via Simulation
IE 609 - Stochastic Programming
IE 610 - Heuristics in Optimization
IE 611 - Integer Programming
CE 651 – Analysis techniques for transportation I.
CE 652 – Analysis techniques for transportation II.
GEOG 517 – Geographic Information Management and Processing

Social Science, Policy, Program Evaluation

ECON 581 – Econometrics 1
ECON 582 – Econometrics 2
ECON 681 – Time Series Econometrics
ECON 682 – Panel Data Econometrics
POLS 515 – Maximum likelihood Methods in Political Science
POLS 516 – Dynamic models in political science(time Series)
POLS 518 – Bayesian Modeling in political science
SOCI 631 – Advanced Quantitative Methods
SOCI 633 – Survey Design and Analysis

B. Faculty - potential Bredesen Center faculty for the DSE program

Possible faculty	Organization	Unit	Discipline/Focus Area
Bob Cook	NASA	DAAC	Mentoring a student and DAAC data stores and the long history of work they have done
Inatwat Nookaew	ORNL	BSD	Large-Scale Bioinformatics – focus on Human Genomics
William Heller	ORNL	BSMD	Modeling of small angle scattering data of soft and bio molecules, MD simulations
Jack Fellows	ORNL	CCSI	Climate change science
Timmy Ramirez-Cuesta	ORNL	CEMD	DFT allied to neutron vibrational spectroscopy, database of neutron spectra, method development
Sergei Kalinin	ORNL	CNMS	Oxide physics and scanning probe techniques, uses of big data for functional imaging
Chad Steed	ORNL	CSED	Big Data Mining and Visualization
Devin White	ORNL	CSED	Urban Remote sensing, Photogrammetry, Geocomputation
Gina Tournassi	ORNL	CSED	Medical Informatics/Imaging Data Sciences
Robert Stewart	ORNL	CSED	Geographic Data Sciences, Exploratory Spatiotemporal Analytics
Arjun Shankar	ORNL	CSED	Computational sciences and health care
Arvind Ramanathan	ORNL	CSED	Big Data Mining in Biology
Chris Symons	ORNL	CSED	Computational Data Analytics, Machine Learning/Model Selection
Shaun Gleason	ORNL	CSED	Cyber security/large data bases
Stacy Prowell	ORNL	CSED	Cyber security
Michael Leuze	ORNL	CSMD	Computational Science – Computational Genomics at Scale
Kody Law	ORNL	CSMD	Data Assimilation and Scientific Computing
Rangan Sukumar	ORNL	Ctr. Comp Sciences	Big Data Analytics in Healthcare Delivery
Sreenivas Rangan Sukumar	ORNL	Ctr. Comp Sciences	Machine Learning and Big Data Computing in Health Sciences
Garrett Granroth	ORNL	Data Analysis/Visual	Advanced data analysis and modeling related to neutron scattering, instrument simulations
Mathieu Doucet	ORNL	Data Analysis/Visual	Automation of data reduction tasks, web monitoring, reflectometry and related data analysis
Thomas Proffen	ORNL	Data Analysis/Visual	Modeling related to diffuse scattering, feedback loop/steering of neutron scattering
Shih-Chieh Kao	ORNL	ESD	Environmental data analysis
Lonnie Love	ORNL	ETSD	Advanced manufacturing
Jamie Morris	ORNL	MSTD	Computational materials science with most experience in metals and carbons
Ryan Dehoff	ORNL	MSTD	Alloy science and metals processing with current emphasis on additive manufacturing
Weiju Ren	ORNL	MSTD	Compilation, assessment, manipulation, and accessibility of nuclear materials database
Mike Frame	USGS/ORNL	Information Science	Data science at USGS, in Oak Ridge; adjunct professor in the School of Information Sciences
Joseph Kizza	UTC	Computer Science	Computer science and cyber security
Mina Sartipi	UTC	Computer Science	Computer science
Reinhold Mann	UTC	Engineering	Computational fluid dynamics
Rob Williams	UTHSC	Genetics, Genomics	Computational science and genomics
Bob Davis	UTHSC	Pediatrics	Database epidemiology
Saunak Sen	UTHSC	Preventive Medicine	Database epidemiology
Teresa Waters	UTHSC	Preventive Medicine	Data science and preventive medicine
Ham Bozdogan	UTK	Bus Analytics/Stat	Business Analytics and Statistics/Mathematics, High dimensional statistics
Russell Zaretzki	UTK	Bus Analytics/Stat	Business analytics
Asad Khattak	UTK	ECE	Transportation Science
Chris Cox	UTK	ECE	Civil and environmental engineering
Joshua Fu	UTK	ECE	Environmental and Climate Science
Lee Han	UTK	ECE	Transportation Science
Nick Wierschem	UTK	ECE	Urban Infrastructure Systems (structural health monitoring)
Michael Langston	UTK	EECS	Big Data Analytics, Graph Theoretical Algorithms, Life Science Applications
Mark Dean	UTK	EECS	Computer engineering
Mike Berry	UTK	EECS	EECS/Mathematics, Machine Learning
Nicholas Nagle	UTK	Geography	Spatial demography, Geostatistics
Shih-Lung Shaw	UTK	Geography	Geographic information science (GIS), spatio-temporal analysis
John Kobza	UTK	IE	Industrial engineering
Suzie Allard	UTK	Information Science	Scientific data curation to support team science.
Carolyn Hank	UTK	Information Science	Data management
Wade Bishop	UTK	Information Science	Data management specialty area provenance & GIS
Jan Rosinski	UTK	Mathematics	Mathematics, Probability and Statistics
Judy Day	UTK	Mathematics	Mathematics/EECS, Immunology
Ohannes Karakashian	UTK	Mathematics	Applied Mathematics Modeling and Numerical Simulations
Steve Wise	UTK	Mathematics	Mathematics, Scientific Computing and Numerical Simulations
Vasileios Maroulas	UTK	Mathematics	Analytics and Statistics, Stochastic Optimization, Computational Bayesian statistics
Colleen Jonsson	UTK	Microbiology	Mathematical environment science
David Keffer	UTK	MSE	Materials science
Daniela Corbetta	UTK	Psychology	Psychology, Computational Neurosciences
Ralph Lydic	UTK	Psychology	Neuroscience
Stephanie Bohon	UTK	Sociology	Social and applied demography, immigration policy analysis, ethnic economies.