STRATEGIC TASK FORCE AREA REPORT

Data Science

5.January.2018

ILLINOIS

Table of Contents

I. TASK FORCE CHARGE

Dear Colleagues:

The University of Illinois at Urbana-Champaign has seen tremendous change and progress over the past three years. The establishment of the Illinois Data Science Initiative (iDSI) as a campus-wide collective of people working to bring together data science resources has held various sessions across campus to discuss the emerging uses of big data and data analytics to advance research and teaching. To further build on this progress, to contribute to our next phase of strategic planning, and to create the best possible foundation for future success in this area, we are initiating the development of a comprehensive strategy for data science research, education, and engagement.

Based on your involvement with iDSI, you have been identified to contribute to this process by serving on the Data Science Strategy Task Force. Robert Brunner has agreed to lead this Task Force. Recently, a Task Force on Data Science Education was launched by Deans Jeff Brown, Andreas Cangellaris, Feng Sheng Hu, and Allen Renear. Please connect with this group as you begin to think strategically about the following questions we would like for you to explicitly address in this report.

- In what areas does Illinois have the greatest opportunity to positively influence state, national, and global data science through research, education, and engagement?
- What opportunities and challenges do we face in data science-related research and education over the next five to ten years?
- What specific actions would be most effective in realizing our potential in data science education, research, and engagement? How would we realize these actions given current resources? How do we coordinate the work across the various parts of the campus that engage in data science? Please distinguish short (24 months), medium (3-5 years), and long-term horizons (5-10 years) for the actions.
- Please reach out to the Task Force on Data Science Education as you consider how we can better leverage our research strengths to enhance the university's educational mission in data science-related fields.
- The area of data science presents many opportunities for community engagement. How best can we engage our local community through data science?

Your team might wish to consider additional questions to help guide the development of multiple strategies for data science at our institution. In producing the comprehensive strategy, we ask that you develop, at the outset of your efforts, a process for engaging key internal and external stakeholders and drawing on the diversity of disciplinary expertise on the campus.

We would like to receive your final report no later than January 5, 2018; and we would like to invite a representative of your group to a broader planning meeting on January 26, 2018 to present your findings. Members of the Office of the Provost stand ready to assist you in this work. Please contact Associate Provost for Academic Effectiveness Staci Provezis (sprovez2@illinois.edu) with question or request that emerge during your review.

We appreciate your willingness to serve on this task force.

Sincerely,

Robert J. Jones
Chancellor

John P. Wilkin
Interim Vice Chancellor for Academic Affairs and Provost

II. TASK FORCE MEMBERSHIP

Chair: Robert J. Brunner (School of Information Sciences; Accountancy, College of Business)

Vikram Adve (Computer Science, College of Engineering)

Brian Aldridge (Veterinary Clinical Medicine, College of Veterinary Medicine)

Gabrielle Allen (Education, College of Education)

Scott Althaus (Political Science and Communication, College of Liberal Arts & Sciences; Director of Cline Center for Advanced Social Research)

Matt Ando (Mathematics, College of Liberal Arts & Sciences)

Kathy Baylis (Agriculture and Consumer Economics, College of Agriculture, Consumer, and Environmental Sciences)

Roy Campbell (Computer Science, College of Engineering)

Peter Christensen (Agriculture and Consumer Economics, College of Agriculture, Consumer, and Environmental Sciences)

Jana Diesner (School of Information Sciences)

William Gropp (Computer Science, College of Engineering; Director of NCSA)

Matt Hudson (Crop Sciences, College of Agriculture, Consumer, and Environmental Sciences)

Heidi Imker (Library)

William H. Sanders (Electrical and Computer Engineering, College of Engineering)

Doug Simpson (Statistics, College of Liberal Arts & Sciences)

Jake Sosnoff (Department of Kinesiology and Community Health, College of Applied Health Science)

Ramanath Subramanyam (Business Administration, College of Business)

Venugopal Veeravalli (Electrical and Computer Engineering, College of Engineering)

Pramod Viswanath (Electrical and Computer Engineering, College of Engineering)

Ted Underwood (English, College of Liberal Arts & Sciences; School of Information Sciences)

III. EXECUTIVE SUMMARY

The University of Illinois has an international reputation and a rich history of innovation in computation, data, and information through its colleges, departments, institutes, laboratories, Research Park, and industrial partners. Our campus created the internet browser, spawned supercomputing data models and simulations of complex systems from the weather to the universe, forged data analytics and visualizations to study vision, social media, and the internet, and peered into the human body through MRI. Our graduates, many of them well versed in data science, are in high demand. However, the rewards of excellence can be lost through complacency. Recognizing data science as a key to discovery in the coming decades, other universities have moved quickly and deeply to both define Data Science as an academic discipline and to create new and high profile data science institutes.

In this report, we present what our task force sees as strategic opportunities for Illinois to surpass our competitors by leveraging existing campus strengths with a new organizational entity designed to innovate in data science: to create, coordinate, launch, and sustain interdisciplinary data science research projects, to train cadres of students and professionals, to build leadership to explore data science research opportunities, and to design programs and applications that address highly visible, significant, insightful problems with global societal impact. The return on investment identified by the task force is considerable: the proposed activity strengthens academic disciplines, creates interdisciplinary, cross-cutting, and fundamental research opportunities, supports a diverse but cohesive data science educational program within major colleges and departments, and builds focused outreach, government funding, Research Park and other industrial collaborations, partnerships, and initiatives.

The challenges around data (and, more broadly, data and information) science, analytics, and systems span all units across campus including agriculture, business, education, engineering, health, social sciences, humanities, the arts, and the natural and physical sciences. Realizing the opportunities around data science, data analytics, and data systems therefore requires synergistic coordination to effectively steward opportunities for new collaboration models, data sharing opportunities, and educational programs that engage all units across campus. As a focus, we have identified five grand challenge themes that have high global visibility, promise to fundamentally improve the human condition, and build on multiple areas of existing strength on our campus. Each of the grand challenges builds a team with significant methodological expertise, shared services, and data science knowledge, which is then integrated with the subject-matter expertise specific to the theme. The five themes act as a fabric within which we weave a world-class data science institute with a team of accomplished leaders. While these themes will be highly visible and broad, other research projects not related to these themes will also be encouraged.

Exploiting the existing silos of research expertise and the strengths of individual colleges and departments, we advocate a new institute focused on world-changing challenges to encompass the rapidly widening data science horizon on campus as well as national research and teaching leadership opportunities. The institute will consolidate and launch a novel, disruptive, and transformative cultural landscape of interdisciplinary data science, and will draw together an innovation community dedicated to solving significant practical problems while advancing the frontiers of research in ways that augment the campus' land grant mission with the imperative to extract knowledge from data. This institute will leverage our collective strengths in ways that will establish Illinois as a world leader in data science, promote and support data-intensive

research across our campus (both in the grand challenge themes and beyond), transform our approach to data science education at all levels and in every department, build meaningful collaborations with industrial partners who are also struggling under the data deluge, and purposefully engage with local, state, and national communities to help citizens everywhere better control their own destinies.

Thus, we propose a bold strategy to

1. nurture collaborative, world-changing, fundamental, and applied research into data science to address societal grand challenges by making connections between data generators, data analysts, and data consumers,
2. organize topical sessions to proactively identify and pursue emergent research themes that cross disciplinary and institutional boundaries,
3. unite new and existing campus-wide resources into a single coherent data- and resource-sharing framework to help break down real or perceived institutional barriers,
4. encourage every interested Illinois student to gain exposure to data science concepts and quality standards in a way that promotes basic data science literacy for all students enrolled on our campus,
5. serve as a central repository and innovation engine for undergraduate and graduate data science education curricula, materials and best practices,
6. apply data science to university operations in ways that improve student outcomes, enhance university services, detect risks more quickly, generate cost savings, and more equitably allocate resources,
7. engage broadly with industrial partners by collaborating with existing campus programs to bring innovations in data science to any interested company,
8. deliver data analytics training broadly across the state, country, and world, and in short, to
9. establish Illinois as a world-recognized leading innovator in data science, data analytics, and data systems.

IV. CHARGE QUESTIONS AND RESPONSES

INTRODUCTION

Data sciences (which we explain more precisely below) have begun to have a broad and fundamental impact on nearly every aspect of human endeavor. From targeted advertising to precision agriculture, large-scale sky surveys, financial risk control, personalized health care, autonomous vehicles, anomaly detection for national security, and better governance, the collection, analysis and decision-making associated with enormous volumes of data have had a profound impact on people and industry alike. And these changes are only scratching the surface: in every field, they reveal the deep potential impact of data sciences. Both companies and universities are now exploring all the ways in which their fields can benefit. However, these early advances in data sciences have not paid adequate attention to privacy and ethics issues that arise in the collection, analysis and use of data about the public at large. Nor have these early efforts addressed the challenge of making data science methods and their associated benefits

inclusive. At the same time, the need for workers—and researchers—with data science and analytics expertise has exploded far beyond the available workforce with the necessary skills: a simple search for job openings on LinkedIn that mention "Data Science" shows over 50,000 openings.

In this context, University research and education in data science has become a pressing need. Such efforts must span both the fundamental techniques of statistics, algorithms, data collection, and computing systems, as well as the application domains—spanning most of the disciplines on campus—in which data sciences will be used. While individual researchers and small teams have launched numerous data science efforts at the University of Illinois, there is still no coherent strategy that spans campus as a whole to ensure appropriate coordination, investments, and large-scale initiatives bringing together large teams of people to drive the most effective and impactful projects.

One of the biggest challenges in exploring ***data science*** is defining what, exactly, the term means. Our task force viewed this challenge as an opportunity, since the University of Illinois can lead in this space by defining data science in a manner that is inclusive of all. Formally, we choose a definition originally proposed by iSchool Dean Allen Renear:

> Data science is concerned with all aspects of the creation, management, and analysis of data, focusing in particular on the application of computational methods to digital data

This definition is not focused solely on analyzing data, as it addresses the full data lifecycle. Furthermore, this definition emphasizes the application of computational methods. This last aspect means that data science also involves ***how data are used*** and even ***why data are used***. From this definition, we see that data science encompasses areas known to be critical for the analysis of data including the computational, statistical, and domain disciplines, but also domains that guide the analysis and its impact including the social, philosophical, legal, and other more human-centric disciplines.

Broadening the scope to which the data science label is applied opens up new challenges. Yet, few of the immediate challenges facing data science-related research and education are new or confined to emerging fields. There has been an incredibly rapid growth in questions around data science, along with a corresponding demand for the relevant answers, so that both local and national capacity has been strained. Yet the specifics of what is and isn't ***data science*** remain broadly nebulous and lacking in clarity. As a consequence, many are aware of the buzz but fail to appreciate what it means to them individually. Conversely, people may appreciate how data science can benefit them directly, but have insufficient support—be it infrastructure, training, or collaboration—to integrate data science into their research or educational activities. Thus, a **first challenge is establishing cohesion** across our campus to align our many data science resources into a functional composite; and a **second challenge is developing data science in a manner that is inclusive** across all topical domains, levels of expertise, and modes of inquiry.

On our campus there is simultaneously too much and too little activity in data science, and we are not unique in facing this dilemma. Our campus currently offers an array of opportunities in the form of data science student groups, data science course work, as well as projects, programs, and even entire units (e.g., CSL and NCSA) that make extensive use of and share insights into data science principles. Yet because none of these cover the entire scope of data science, instead of having a cohesive and coordinated data science experience—with all of the accompanying

momentum, visibility, and credibility such an experience would afford our faculty, staff, and students—we have an organic patchwork of "find it if you can" opportunities that are not built nor supported to handle the variety of backgrounds or skill sets needed by the campus' diverse constituency. Not only is this clearly less than ideal from an individual's standpoint, but this deficit severely hampers the ability of our campus to attract and retain the talented students and faculty on which our world-class reputation and ability to meet our land-grant mission rely.

Many of our closest peers (e.g., Berkeley, Georgia Tech, Michigan, NYU, and Washington) have already made substantial commitments to data science, including the launch, support, and promotion of data science institutes. Thus, an unfortunate **third challenge is overcoming our campus' current data science deficit**. Individual members of our task force have experienced the very real consequences of Illinois' data science shortfall. For example, prospective Statistics students and faculty candidates routinely inquire about opportunities to affiliate with a Data Science Institute. Likewise, Engineering has experienced a lower than expected acceptance of offer letters and has even had advanced-stage negotiations, including a star faculty recruitment, thwarted at the last minute by the promise of well-funded data science opportunities at other universities. While these examples are anecdotal, it is increasingly clear that the University of Illinois' lack of investment in data science is becoming a strong, negative drag on multiple areas of campus, especially those that are vital to help us overcome this deficit.

These challenges also illuminate our opportunities. While our current data science activities may be uncoordinated and disparate, we are fortunate to have many resources already in place. Instead of starting from scratch, we have an opportunity to knit our collective data science activities together in new ways that will allow us to quickly recover ground lost to our peers and will set the stage for developing a distinctive reputation as a leading hub of fully-integrated data science pursuits. We have already glimpsed this opportunity through the actions of the Illinois Data Science Initiative (iDSI)[1], which held seven, highly successful topical summits in data science and organized the first campus-wide Data Science Day. Hosting eight major events for multiple domains is no small feat, and the energy, excitement, and engagement of campus stakeholders as well as the iDSI Steering Committee (SC) is a testament to the intensity of the interest and need at the University of Illinois. Of particular note is that the iDSI SC is an incredibly diverse group representing a full spectrum of disciplines on campus. That this group functioned so well and accomplished so much signals a highly promising—and highly unusual—level of productivity and inclusivity rooted in a shared vision for the synergistic potential that IDSI holds for our campus.

Thus, we have an opportunity to build on the demonstrated momentum and inclusivity of the iDSI to engage an even wider constituency as we work to create an ambitious, interdisciplinary future for data science at Illinois. We believe that the University of Illinois is unmatched in this potential. We have observed data science efforts at other universities that claim to be interdisciplinary, but on closer inspection their efforts are revealed to be energizing a narrow set of scholarly silos that have already been innovating in this space for some time. Our approach will be neither narrow nor shallow, and this combination of breadth and depth is our greatest opportunity to provide unrivaled opportunities for students and faculty in data science. This opportunity leverages a natural cultural advantage that our campus already possesses: our entrepreneurial spirit combined with our authentic ability to collaborate across lines of scholarly

[1] Not to be confused with IDSI, the acronym by which we refer to the proposed Illinois Data Science Institute.

difference, a known Illinois quality that is as critical to data science as to any other interdisciplinary undertaking.

AREAS FOR ILLINOIS GREATEST OPPORTUNITIES

We live in a remarkable period of human history. Humanity is facing multiple grand challenges due, in part, to explosive population growth and climate change within a world increasingly divided by economic, religious, and political lines. At the same time, there is an exponential growth in the quantity and quality of data generated or collected that offer the potential for fundamental improvements in the human condition. We see enormous opportunities for the University of Illinois to identify grand challenge themes that, if tackled responsibly and proactively, will create significant impacts locally, nationally, and globally. During the past year, the Illinois Data Science Initiative held seven topical summits and a data science day with eight topical sessions attended by hundreds of faculty, staff, and students from across campus that explored the impact of data science in different fields. From these events, the Data Science Strategy Task Force has identified five, interrelated grand challenge themes (defined in more detail later in this report) that we feel campus should support:

1. Improving Societal Resilience,
2. Living in a Connected World,
3. Personalizing Health,
4. Reducing World Hunger, and
5. Understanding Our Natural World.

Tackling these grand challenges will not be easy. Doing so will require a new approach, and not simply doing more of what we've already been doing: a "business as usual" approach that is leaving us behind the curve on integrating data science developments into existing campus strengths and is ill suited to sustain or support the revolutionary change envisioned in this proposal. To highlight this point, consider the "Personalizing Health" theme. In the near future, standard medical devices such as the stethoscope or otoscope will be replaced by new and more powerful diagnostic equipment. Individual health monitoring will become ubiquitous via wearable devices, local sensor networks, and even via in situ monitoring from the clothes we wear. Health and environmental data will be constantly collected, analyzed, and archived, in many cases autonomously, to provide personalized health guidance, which may involve recommendations to seek therapy, counseling, or specific medical attention.

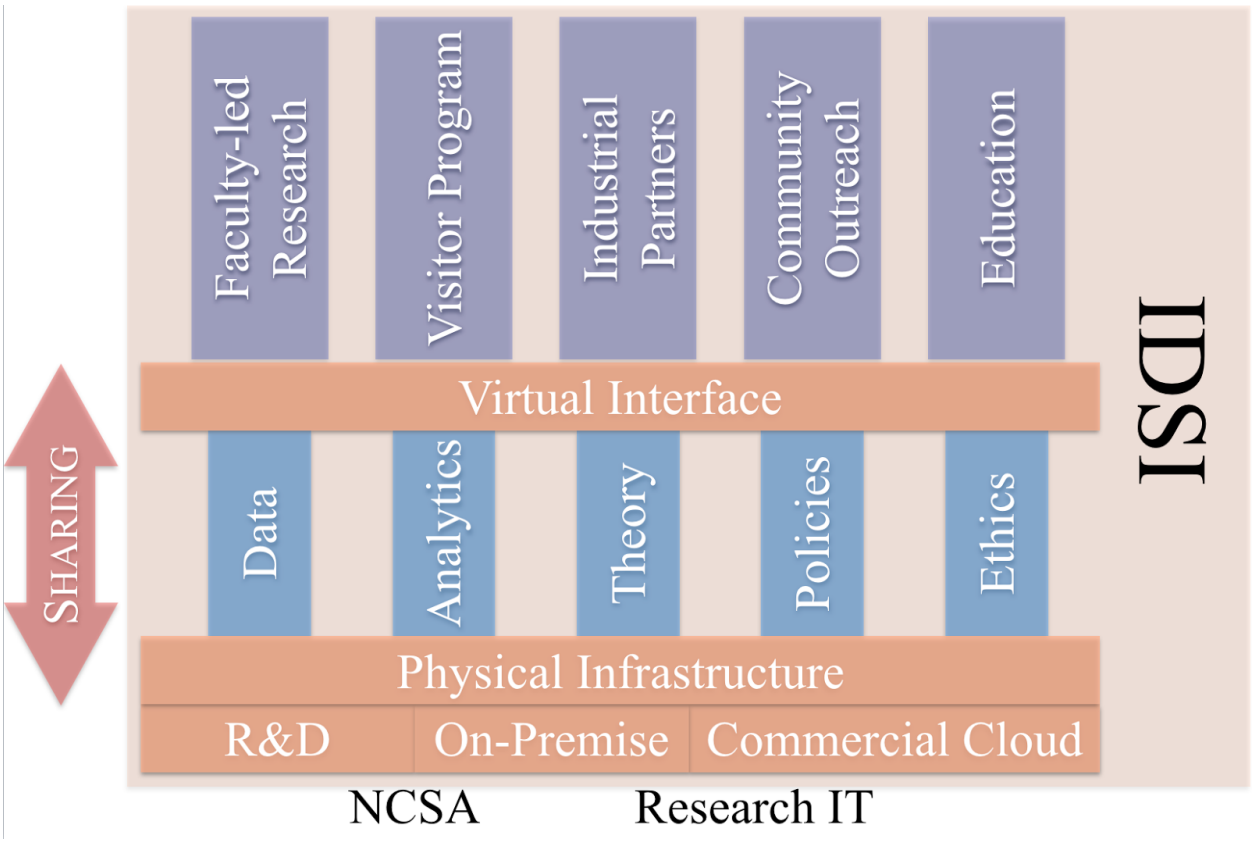
In this new system-based approach, physicians will increasingly focus less on data collection (e.g., using a thermometer or stethoscope) and more on data assimilation and patient advising, which should lead to improved patient outcomes for a larger population. But for this vision to become reality, vital data-related issues must first be solved, such as how data can be safely and securely collected, aggregated, managed, and preserved. New analytic approaches must be developed to gain insight from massive, heterogeneous data. Equally important, new policies must be developed to build robust analytics engines that can combine patient-level data with local environmental data for single patients while also aggregating the same data for population-level diagnostics and prediction. These same data collection and aggregation tasks raise major ethical concerns that must be openly addressed in order to ensure confidence in the eventual

solution. Finally, if we expect these solutions to remain viable as new and richer sources of data become available, we must be ready to pursue new theoretical investigations at the intersection of data, analytics, policy, and ethics, spanning in-home sensors to the hospitals themselves.

Yet these data-centric issues are not unique to the personalizing health theme; these same issues of data, analytics, ethics, policy, and theory pervade all grand challenges facing society today. While solutions to these data-related obstacles could be developed in a one-off fashion every time they are encountered, the greatest opportunity comes from working collectively on these data aspects and in leveraging a common data infrastructure. As a campus, we have considerable expertise in using and learning from data, including from institutes and centers such as ARI, Beckman, CSL, IGB, IHSI, ISEE, ITI, MNTL, and NCSA and from units such as CS, ECE, the iSchool, Law, the Library, Mathematics, Statistics, as well as those in the humanities, social sciences, and other disciplines. But we lack coordination and the synergistic benefits that come from a central, campus-wide home for data science.

These grand challenges are not only ambitious, globally important, and anchored in domain strengths on our campus, they also serve to unite existing campus strengths and resources across the data science spectrum. Rather than trying to *ab initio* define data science, these grand challenges provide focus on the most important aspects of data science. Thus, these focal points would serve to mutually advance both the domain science and data science as a discipline itself. From the preceding discussion, we can recognize that advancing these themes would involve computation, storage, and analysis, but also policy and regulations, ethical and societal concerns, as well as new theoretical works across the data science landscape. This big picture approach requires engineers to work with sociologists, computer scientists to work with humanists, and statisticians to work with lawyers, all in an environment where the contributions of each group are solicited and valued.

As a result, ***we propose that campus launch a data science institute***. This physical institute will focus on nurturing the people, experience, and skills required to unlock knowledge from data across different domains. Visually, this physical institute, known colloquially as the Illinois Data Science Institute, or IDSI, is outlined in the following figure. The IDSI will leverage physical infrastructure—including collaboration facilities supporting interdisciplinary research capable of video conferencing with collaborators in remote locations—that is provided by NCSA and Research IT, depending on the need and function. Over this infrastructure foundation are arrayed the five, currently defined IDSI pillars: Data, Analytics, Policies, Ethics, and Theory. Of these pillars, ***Data*** can arise from existing campus entities, such as the HathiTrust or the Cline Center; from faculty-led research, perhaps in partnership with the Research Data service; or from industrial partners. ***Analytics*** is primarily the application of statistical and machine learning to data. ***Policies*** addresses research into best practices that enable better use of data and enhance their impact on society. ***Ethics*** is broadly defined as assessing the impact of data science on society, including ethical issues raised by the collection, manipulation, and sharing of data but also extending to encompass public policies aimed at mitigating economic, cultural, and societal disruptions that the data science revolution is already starting to bring about. The final pillar, ***Theory***, spans these other areas and is centered on developing new algorithms, optimizations, systems, and approaches to problems driven either by projects within grand challenge themes, or arising from other faculty-led research or industrial partnerships.



All pillars involve both ***expertise***, which would be provided from highly trained academic staff, but also ***research***, where faculty and staff investigate solutions to new challenges as they arise. Spanning these pillars is a virtual interface to the world, through which faculty-led research, academic and industrial visitors, industrial partners, educational programs, and the broader community beyond our immediate campus environment can all interact with the IDSI. The faculty-led research projects that fall within the previously mentioned grand challenge themes would be competitively selected, provided seed funding, and expected to successfully obtain center-level funding, with sunset clauses to enable new groups to form and replace those who are unsuccessful. These groups will collectively help define the required physical infrastructure, analytics engines, policy development and enforcement, and ethical standards required not only for the IDSI to be successful, but that will influence data and information science across the nation and the world. To ensure these data science needs are met, we believe that campus must commit to hiring new faculty in these pillar areas to grow and sustain the IDSI. In addition, to ensure the institute remains nimble and relevant to impactful changes, we propose that the institute run a prize postdoctoral program where young researchers will develop independent programs in data science that broaden the capabilities of the institute while spotlighting new and potentially influential areas as they appear on the horizon.

Our proposed approach to launching and operating a new data science institute has one additional, and very important, benefit. The data obtained or generated by the faculty-led research projects will come from both public and private sources. Thus, IDSI's data infrastructure must support both access modes, where data can be shared in three ways: publicly to a large group, privately to only those with explicit permission, or non-consumptively, where

analytics are run on data that researchers never see in order to preserve the anonymity of individual observations. We propose that the IDSI extend this infrastructure to operate a 'Data + Ideas Incubator' where data and project ideas from research efforts, industrial partners, local and state governments, and the Champaign-Urbana community can be shared. An analogy might be a statistical consulting service, but the approach we advocate is to openly solicit both data and ideas that can be used in classrooms, by student groups, or interested faculty and staff to advance research projects, improve university operations, or benefit local governments. This incubator would prove invaluable to data science education programs across campus, which often struggle to have fresh data and projects, while also providing direct engagement with local and regional communities.

OPPORTUNITIES & CHALLENGES

The grand challenge themes presented in response to the first charge question arose out of the summits organized by the data science initiative, which, in total, attracted over one thousand participants. Any research effort working to address one of these themes would itself be interdisciplinary, and we expect each theme to support multiple faculty-led research groups that will ignite large collaborative efforts to realize center-level grants. This same approach, where synergies can be leveraged between institutes and campus units, has been used with great success elsewhere at Illinois. For example, CompGen is a partnership between CSL and IGB (also involving ECE, BioE, CS, and Biology-related departments). We expect the Illinois Data Science Institute to function in a similar manner around data-science-centric opportunities. We now provide more detail on these proposed grand challenge themes.

1) **Improving Societal Resilience** seeks to use data/information science to address imminent and massive societal disruptions that can lead to societal instability, population displacement, and large-scale migration. These disruptions can arise from natural phenomena such as climate change and natural disasters, from human activity like cyber attacks and political upheaval, or from rapid technological innovations such as the growth of artificial intelligence and automation-driven unemployment. Example projects in this category include:

    a) Climate change is raising sea levels, changing weather patterns, and disrupting social systems built on an assumption of climate stability. Yet we know very little about how societies can effectively foster resilience in affected areas and populations. Given the complex and difficult-to-predict nature of climate change processes, data science solutions are required to not only identify and predict how climate change is affecting humans' ability to flourish, but also to develop and refine policies that can help governments minimize its disruptive societal effects.

    b) Automation, machine learning, and precision analytics not only have the potential to revolutionize many areas of society, but also to reproduce and enhance existing power inequities. For example, algorithms used by courts to predict criminal recidivism inadvertently perpetuate racial biases that are already present in legal institutions, and the economies to emerge from advances in robotics and artificial intelligence will create new challenges for those citizens who lack the social capital and educational credentials to ride the wave of a rapidly changing technological ecosystem. Data scientists will increasingly be called upon to identify these inequities and develop resilience strategies

that minimize the effects of societal disruption from these new technologies while maximizing the potential for all sectors of society to benefit from the opportunities that these innovations will increasingly afford.

2) **Living in a Connected World** is a challenge that accounts for the unprecedented growth of sensors and sensor networks, streaming data in general, and specifically the Internet of Things (IoT), as more data are collected from an increasingly diverse set of environments. This growth is already driving new challenges, including how to effectively harness these data for improving societal welfare, and how to securely manage these data streams to ensure that privacy is maintained. Specific projects might include:

   a) Combining learning management system (LMS), assessment, and lifestyle data to develop predictive models that encourage student behavior leading to positive learning outcomes across the educational landscape.

   b) In a world with self-driving vehicles, how do we evolve the transportation network and manage urban planning to optimize resources where individual vehicle ownership is no longer required or encouraged.

3) **Personalizing Health** aims to address the current reactive approach to healthcare, which is unsustainable due to rapidly increasing costs and population growth. We are at the cusp of a radical alteration in society's approach to health care. Instead of standard medical instruments for data collection, individuals will have wearable monitors continuously collecting physiological and environmental data, and analytics will be used—in many cases autonomously—to recommend whether patients should seek therapy, counseling, or medical attention. Physicians will increasingly focus on adding meaning to data assimilation, and tailoring this information to the specific patient, with the primary focus on improved patient outcomes. In this system approach, specific projects might include:

   a) Addressing important but unanswered questions about how these critical and sensitive data are collected, aggregated, managed, and preserved.

   b) Determining how to best present such data (or information) to both the physician and the patient in ways that maximize diagnostic accuracy, selection of appropriate and effective treatment options, and full consideration of the lifestyle implications entailed in particular medical choices.

   c) Addressing policy issues in order to build robust and reliable analytics engines that combine patient-level data with local environmental data that can be adapted for single-patient support and also in the aggregate for broader population- or community-level diagnostic support.

As these projects and others are explored, concerns about health disparities will need to be taken into account in this new, data-centric view of medical intervention. Algorithmic learning from these heterogeneous data is non-trivial and will involve entirely new approaches from what is understood today. The combination of all of these subareas of the Grand Challenge will likely involve new theoretical developments, ranging from ethics, sensors, algorithms, to data curation.

4) **Reducing World Hunger** is an ambitious effort to improve global food security by improving the production of staple crops in the developed world, improving growing approaches in underdeveloped areas, improving food and nutrient access for vulnerable

populations, enhancing stability of food systems, and improving the logistics, distribution, and supply of food goods and services around the world. This grand challenge theme is especially pressing in our era of increasing global population, decreasing availability of arable land, and growing forms of resource scarcity precipitated by our planet's rapidly changing climate. Specific projects might include:

a) Using novel data collection and analytics to model the global food system and the effect of climate shocks to predict food insecurity, increase yields, improve food access and societal resilience while reducing hunger and civil conflict.

b) Federating heterogeneous and multi-modal ground-, air-, and space-based data to develop short-, medium-, and long-range predictive models to identify location and causes of yield shortfalls, improve crop yields, enhance resource management and sustainability, and mitigate risks at the scale of individual farm fields.

c) Using models of plant growth and genetics to leapfrog conventional crops with grains, pulses, and vegetables designed for economies where agricultural shortages are serious problems.

d) Modeling the worldwide transportation network, which when combined with real-time information about commodity markets, weather events, and civil unrest, can optimize delivery of goods and raw materials around the globe.

5) **Understanding Our Natural World** is a thematic area that encompasses investigations from the very small to the entire cosmos. In the last century, our quest to understand the physical world has led to new fundamental theories such as quantum mechanics and general relativity, and inspired new innovations including the World Wide Web, Graphical Web Browsers, and GPS. In the next century, as we strive to comprehensively understand the natural world across its diverse spatial and temporal scales, we anticipate new theories and new innovations to emerge as we have the opportunity to incorporate experimental, observational, and simulation data across disciplines to increase our understanding of the natural world. Specific projects might include:

a) The materials genome initiative (MGI) seeks to identify and construct devices that can be built from new materials and that can be brought to market twice as quickly at half the cost when compared to traditional methods. Essential to the MGI is a major transformation in our approach to materials and manufacturing around data, which presents major challenges in the collection, archiving, and federation of digital data about materials from thousands of research groups around the world. Addressing these issues introduces additional requirements for the development of standard techniques for describing materials, for sharing these descriptions, and for enabling new insights to be extracted from these descriptions, including via automated analyses.

b) Climate and environmental studies strive to develop a predictive and comprehensive understanding of atmospheric, surface, and subsurface processes on the Earth that affect weather, oceans, air quality, continental watersheds, pollution, and other environmental factors critical for our planet's well-being. This involves integrating models and observations across many disciplines, including biogeochemical, genomic, ecological, geohydrological, aerospace, chemistry, physics, and coastal science.

c) Multi-messenger astronomy (MMA) federates observational data across the electromagnetic and now gravitational spectrum with data-driven simulations to identify, classify, and model millions of transient phenomena per night. Essential to MMA will be new tools and technologies to collect and analyze data from a worldwide network of telescopes including the LSST and DES surveys, both of which are archived and processed at NCSA, via a rapid communication and response network. Enabling this vision will require the delivery of data over optical networks, rapid image processing and analysis, comparison with observations from other instruments and survey archives, comparison with simulations, archiving these data and all derivative products, and linked associations to publications.

These five grand challenge themes emerged from and were championed by **just the first wave of engaged faculty** who were eager to apply data science to their areas of interest and who vocalized this interest through the summits and sessions organized by the iDSI. But they are no means complete, nor indicative of the entire range of opportunities arising from the broader University of Illinois research community. The Specific Action section below further outlines our proposed process of engaging key internal and external stakeholders to enable even wider engagement across campus and the region. To make the most of these opportunities, our ultimate intention is for the University of Illinois to build up unparalleled strengths in data science such that any project that can benefit from the application of data science does benefit, whether that be through IDSI directly or through coordination with other affiliated units such as NCSA or Research IT.

SPECIFIC ACTIONS RELATED TO EDUCATION, RESEARSCH, AND ENGAGMENT

To organize our data science efforts in ways that not only mirror our peers, but surpass them, we must complete several near-, intermediate-, and long-term actions. Broadly, our short term (less than two years) actions must focus on fostering a coherent on-campus data science community, creating a robust framework for sustained long-term excellence through the creation of a formal data science institute, and initiating steps to realize our intermediate- and longer-term goals. Our intermediate goals (two to five years) focus on an ambitious growth plan that puts Illinois on a path to being recognized as an undisputed world leader in several key aspects of data science (e.g., the proposed pillars of the Illinois Data Science Institute) through research excellence, educational innovation, and broad societal impact. Our long-term goals (greater than five years) build on and sustain our reputation for excellence while identifying new disciplines in which to integrate data science more broadly.

Short-Term Actions in the First Two Years

To establish data science as a first-class academic discipline at the University of Illinois, we must coordinate and expand on our current patchwork of research capacity, educational efforts, and physical infrastructure. To aid in this effort, we first propose a distinguished lecture series focusing on both general data science itself as well as he application of data science to societal grand challenges. This lecture series would generate visibility on- and off-campus and encourage those researchers and communities not currently engaged in data science to see both the ***why*** and the ***how*** in relation to data science and its application in a wide variety of domains. To supplement the lecture series, we will compile curated examples of data analytics and applications, including code and data, to facilitate adoption of these novel techniques. We also

recommend the creation of a campus-wide data science minor, which would be flexible enough to meet the needs of students both north and south of Green Street.

We also will need to attract and recruit research staff who are capable of facilitating world-class scholarship on the application of data science to our proposed grand challenge themes from within the IDSI pillars. Hiring these personnel and hosting training programs required to bring individuals up to speed on data science methods and concepts would further reduce the technical barriers faced by the broader campus community when trying to break into data science research. As a result, we would effectively 'grow the data science pie' by creating basic but widespread data science literacy across the entire disciplinary spectrum, which would generate additional support and grant opportunities for these proposed efforts. To aid in this process, we also recommend the creation of seed funding to support faculty-led research projects in the grand challenge themes that would be selected via a competitive process and with appropriate sunset clauses. This approach would not only increase the visibility of data science activities on our campus, but also further expand the data science constituency at the University of Illinois.

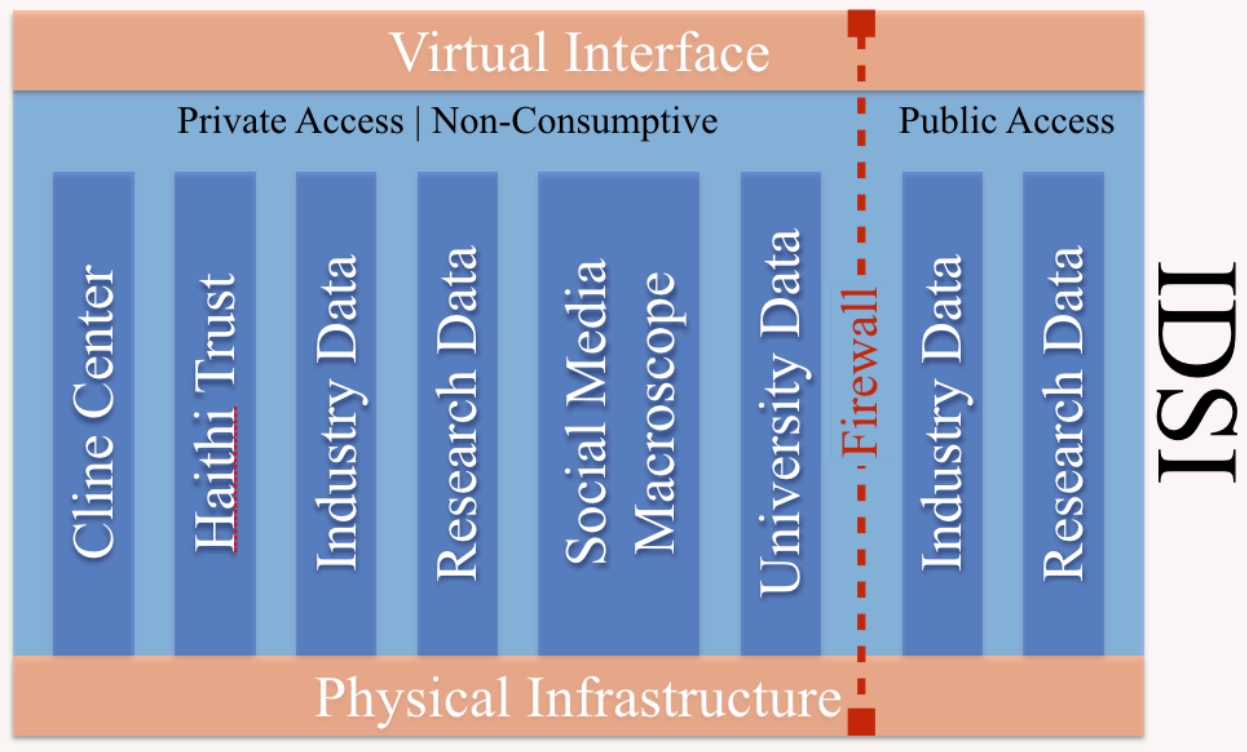
We strongly recommend the creation of a formal data science institute that will serve initially as a central coordinating body for the enormous range of (currently uncoordinated) data science activities around campus. This approach will ensure that, in the intermediate and long term, the University of Illinois will be seen as one of the leading data science institutions in the world. In this process, we already have examples from which to learn, as one of the major strengths of Illinois is its portfolio of highly multidisciplinary institutes that enable broad new collaborations, center-level funding, and provide high external visibility and credibility. Long-lived institutes such as the Beckman Institute, the Coordinated Science Laboratory (CSL), the Information Trust Institute (ITI), the National Center for Supercomputing Applications (NCSA), and the Institute for Genomic Biology (IGB) are respected around the world in their various fields.

These institutes serve as powerful catalysts for large-scale, cross-campus efforts. For example, IGB and CSL have catalyzed multiple, large centers including a Big Data to Knowledge (BD2K) center funded by NIH and, together with the Institute for Sustainability, Energy, and the Environment (ISEE), the $115M DOE funded CABBI center on biofuels. The Coordinated Science Laboratory (CSL) has launched numerous interdisciplinary centers over its long and distinguished history, most recently the $25M ARL-funded center for a next generation Internet of Battlefield Things (IoBT). The Information Trust Institute has received more than $100M of funding from DOE, DHS, and NSF to support research in building trustworthy, secure, and resilient energy delivery systems. Because of the breadth and potential impact of data sciences, a new data science institute has the potential to produce a similar, long-term impact across campus.

When launched, this new data science institute will require a formal organizational structure along with an appropriate formal budget, including campus and college-level contributions, and a five-year plan to become a self-sustaining entity (for example, through a combination of government, industry, and foundation funding; ICR; and non-credit professional training opportunities provided to industrial partners and the broader community). Any campus and college contributions should be strategically planned so that the participating units can expect a sufficient return on investment through new research initiatives and revenue-generating instructional activities. A formal institute would also help sustain the momentum built up over the past year by the Illinois Data Science Initiative that includes multiple successful summits and the campus-wide data science day. This momentum is visible in the numerous new interactions

among researchers from different domains across campus and the engagements with University leadership around data science and its application at multiple levels. Another important means to leverage this momentum is to explicitly include the data science institute in the new capital campaign, which can include goals to fund postdoctoral scholars, visitors, research infrastructure, research projects, or a dedicated physical space.

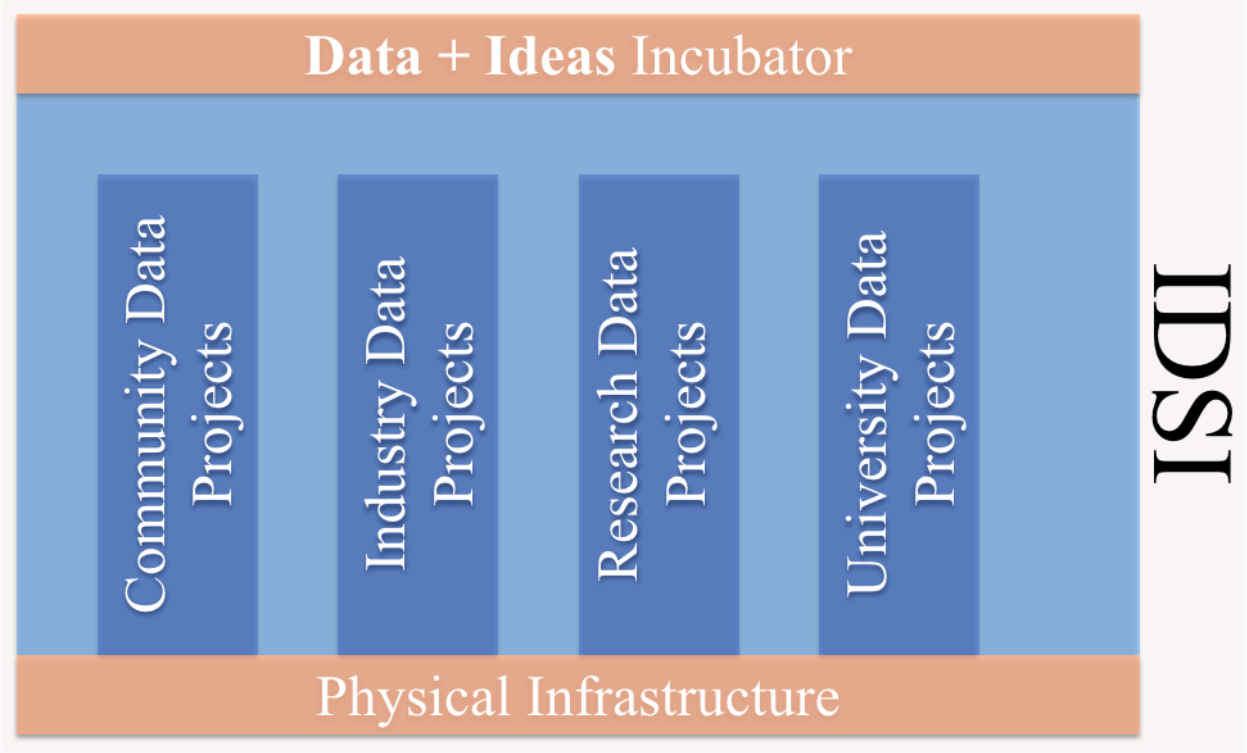
Another venue by which this new institute can benefit campus is to broaden the participation among researchers from around campus (and beyond) in using data collected or generated from campus research. Presently, the library provides the Research Data Service that aids faculty in storing research data, especially as required by funding agreements. At the other end of the spectrum are large data projects, such as the HathiTrust or the Cline Center, that manage their own data archives and services. While this approach is functional, it fails to provide the maximal benefit to both the data providers and those who might wish to leverage these data projects. The proposed data science institute will, as shown in the following figure, be able to provide private and non-consumptive access to data sets as well as public access to data sets as desired. In this manner, campus research will immediately gain a broader impact as well as a higher visibility since both the data and associated analyses can be easily shared. This same infrastructure will benefit both researchers and industry partners who face many of the same data concerns.



To encourage the rapid growth and impact of this new institute we have three specific, related recommendations. First, we recommend establishing a visiting scholar/fellow program dedicated to interdisciplinary data science. The primary benefit of this program will be to increase the number of data science researchers on campus with minimal capital expenditure. However, we also will gain national and international visibility for our data science efforts through the advertisement, recruitment, and hosting of these world-class visitors. Second, we recommend establishing a prize postdoctoral fellowship program in data science. This competitive program

for independent, young researchers will serve to energize our data science activities while also ensuring a steady supply of fresh ideas. This program can also be used as a means to recruit and retain world-class junior faculty in data science, by getting them on campus before they are widely known and building a data science home for them that is second to none. Finally, we recommend new faculty hires with an emphasis on interdisciplinary-bridge builders who would provide further cohesion to the wide array of data science activities across campus. We envision this program will be supported by various training grants, and we emphasize the need for new faculty hires to establish our data science institute as a leader in defining policy locally, nationally, and globally.

One additional benefit of our proposed approach is that the same infrastructure that supports research data needs and the needs of industrial partners can also be used to transform data science education. The main challenge in teaching data science is getting real-world data along with projects that can make use of these same data. Furthermore, to keep courses fresh, there is a continuous need for new data and projects. We call our solution the 'Data + Ideas Incubator', as shown in the following figure. In practice, we will solicit data and project ideas from research efforts on campus, from university administrators and their staff, from industrial partners, from local and regional communities, and beyond, in a standard manner to enable the broadest possible uses of the data and to extract the maximal impact of any resulting analyses. For example, some project ideas might be sufficient for an introductory data science course, while others might be suitable for a senior thesis or capstone project. In fact, the same data might be used under different project ideas at these and other educational levels. As this concept grows, we will plan to demonstrate to other universities how they can launch their own 'Data + Ideas Incubator' to spur additional real-world data science educational opportunities.



To successfully launch a new data science institute, our campus must provide sufficient administration and staffing to ensure the new institute meets the previously stated goals. First,

the data science institute will need to recruit an Executive Director with demonstrated data science research expertise and a track record of building and leading significant research programs. Second, the institute will need sufficient staff support to operate the proposed programs and run workshops and summits on the grand challenge themes or new data science topics, especially those suitable for large-scale, center-level funding, and to support faculty teams developing large-scale proposals. The institute will also need to fund senior faculty at a part-time level to lead center-level funding proposals, to lead the new grand challenge themes, including the competitive process for obtaining seed funding for new theme-based projects, and to liaison with the broader community around these themes. Finally, the new institute will need dedicated staff to build an industrial partner program, to communicate our success and goals effectively via traditional mechanisms and social media, to engage with interested stakeholders from around the state (e.g., via the extensions program), and to manage the 'Data + Ideas Incubator' program.

Intermediate-Term Actions in Years Two through Five

With the creation of a new data science institute and all that it entails, the issuance of seed funding to new projects in the identified grand challenge themes, and the creation of a new campus-wide data science minor, we will have rapidly altered our data science footprint on the national and global stage. Our main actions for the intermediate-term are to help transition these new research and education programs into more sustained and visible entities. First, all seed-funding for new research efforts within grand challenge themes will have explicit sunset clauses. Any group receiving seed funding will be expected, as part of the application process, to identify and pursue follow-on funding ideally at the center level (greater than $10M). Any group that is unsuccessful within the first two to three years (or other agreed upon time limit) will be disbanded. New competitions for seed funding will be held frequently (e.g., every two years) to enable new projects and ideas to be developed and submitted for new external funding. To continue building momentum in these areas as well as grow the data science institute, we expect to continue to hire new faculty to support both our core data science areas as well as these interdisciplinary grand challenge themes.

We also propose to create a new data science major and an associated ***X + DS*** major. Graduates from these programs would be in high demand, and students in these programs would also be able to help programs that are traditionally outside the data science framework to enhance their curricula in general and specific courses in particular to be more data-friendly. We also propose to create a graduate certificate in data science that will enable graduate students who decide to leave academia to have an easier entry into alternative career pathways. There exist a myriad of funding opportunities for these types of programs, both private and public, and having an institute to focus and coordinate these efforts will significantly increase our likelihood of competing successfully for these resources. We also propose to identify and collaborate with faculty and staff who have expertise in data science topics or who are teaching data science courses on campus to create non-degree professional training opportunities. Finally, we expect the new institute will have a broad and robust industrial program that complements existing campus- and system-level efforts such as the Discovery Partner Institute. That requires our 'Data + Ideas Incubator' to pair courses and students in need of projects with interested parties from across the research enterprise, including industrial partners, university administration and staff, and the broader community.

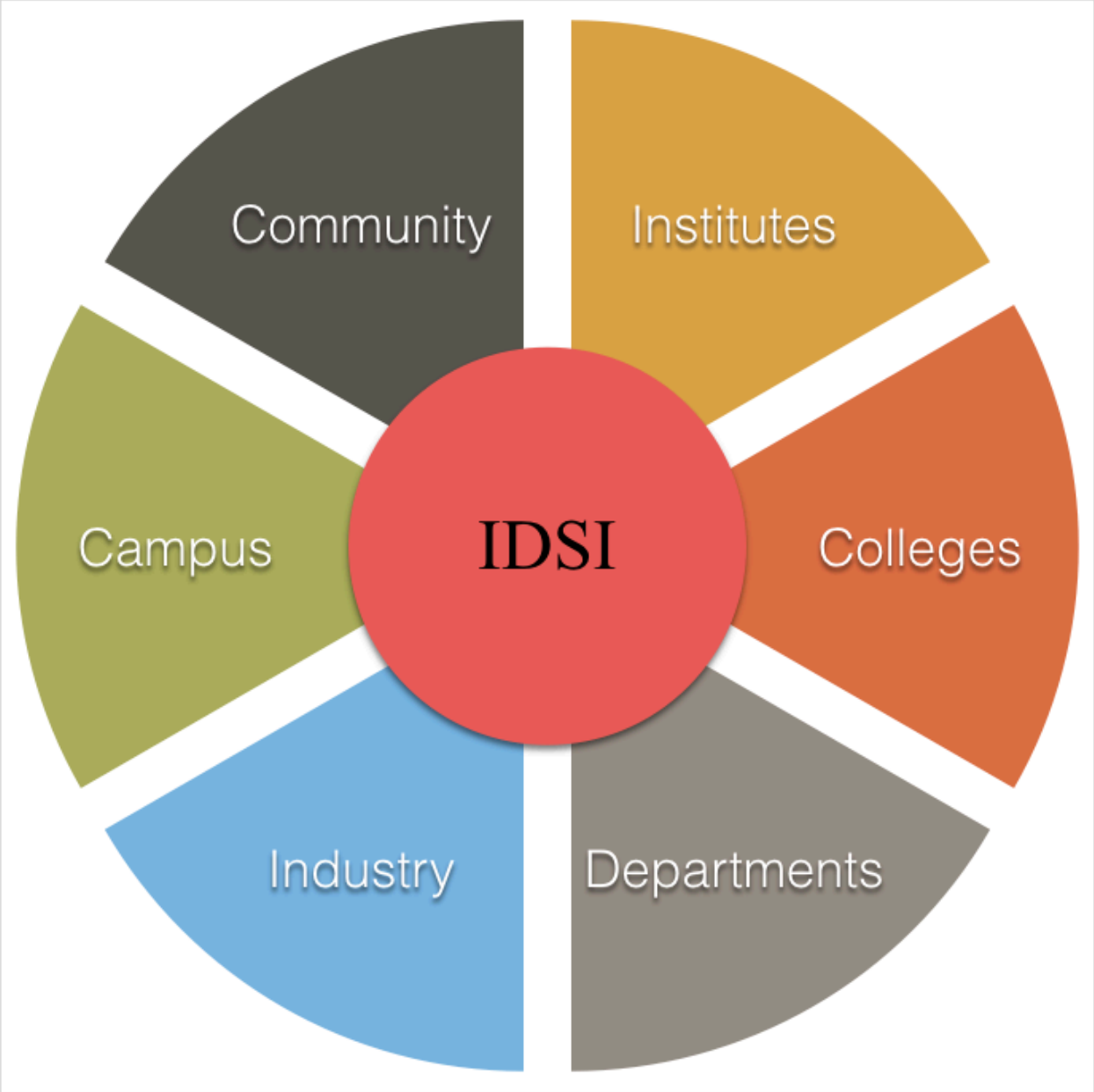
Long-Term Actions After Year Five

By this period we expect the data science landscape at Illinois will have changed dramatically. Beyond continuing with the previously recommended data science institute programs (i.e., research projects within grand challenge themes, visitor program, postdoctoral program, new faculty hires, and distinguished lecture series), we see two long-term actions that will be vital for the continued success of our proposed data science institute. First, the grand challenge themes will need to be revisited, perhaps at five-year intervals, to ensure that they remain national and global priorities and that they still align with the research interests of sufficient faculty and staff on campus. Second, by this point we expect that the data science institute will have grown considerably, requiring an expansion of previously available dedicated physical space. If we achieve the goals laid out for IDSI in the first five years, then continued momentum might require that IDSI have its own building. At whatever scale, the continued success of the Illinois Data Science Institute will depend on identifying and appropriately developing the physical infrastructure required to support a mature data science presence that is fully integrated with all corners of campus at all levels of instruction. For this we expect a major gift and will work aggressively with Advancement to ensure philanthropic funds are obtained.

WAYS TO LEVERAGE RESEARCH STRENGTHS

The challenges that we have identified in this report are significant; we have not yet capitalized on our emerging potential to be a national and global leader in the vitally important area of integrating the data science revolution into higher education's research, education, and engagement missions. Our proposed solution to overcoming these challenges is ambitious. However, we are not starting from scratch in this endeavor. Since our proposed approach leverages our existing, and unmatched strengths in computation, information science, and many relevant domains, we are confident that the opportunities that we have also identified in this report can be quickly realized. In the following figure, we highlight how we envision the Illinois Data Science Institute as a central costive hub linking the relevant parts of our university together to create an uniquely Illinois approach to data science research, engagement, and education.

Key to this concept is a partnership with existing campus institutes. For example, while NCSA is rightly recognized for its high performance computing excellence, it also leads numerous data-centric projects including the NSF-funded Midwest Big Data hub and the National Data Service, which could also partner with a nascent data science institute on our campus. In addition, NCSA has partnered with multiple large projects to provide public and private data processing and archiving solutions. For example, just in Astronomy this includes the archives for both the Dark Energy Survey and the Large Synoptic Survey Telescope project. The real-world knowledge and experience gained from these efforts provides our campus with competitive advantages over many of our peers. As one example, NCSA has developed a HIPPA-compliant non-consumptive data portal in collaboration with the Mayo Clinic. By partnering with NCSA staff, IDSI will be able to leverage this knowledge and physical infrastructure to provide this same capability more broadly to new research projects, industrial collaborations, and educational programs (e.g., our 'Data + Ideas Incubator').



IDSI will also be able to directly benefit from existing data-intensive research projects across campus to help bootstrap the new interdisciplinary research efforts we have proposed within our grand challenge themes. For example, the CompGen project has successfully demonstrated how to make groundbreaking, interdisciplinary advances by collaboratively leveraging computational and domain expertise. With IDSI, we will build on the lessons learned from this and other similar efforts to build even stronger links between our campus institutes and even more academic units. Likewise, IDSI will not need to start a new industrial partner program. Instead, we expect IDSI to both provide direct value to existing industrial programs, especially around new expertise IDSI affiliated researchers will develop in analytics, policies, ethics, and theory, while bringing even more companies to the table because of the many engagement opportunities around data science including professional training and our 'Data + Ideas Incubator.'

Finally, we have an urgent, unmet need around data science education, especially at the undergraduate level. Researchers across campus have already initiated efforts to enhance data science education in STEM fields and beyond. The breadth and intensity of these efforts is striking, involving multiple faculty, departments, and colleges. Collectively they demonstrate the strong commitment of our data science researchers to data science education. The growth in enrollment in data science courses has been phenomenal, and the emergence of massive amounts of data across so many fields of inquiry means that this growth will continue for the foreseeable future, so that data science ideas will permeate much of our teaching.

As a result, we expect that IDSI will significantly enhance our educational and outreach efforts in several ways:

- It will create a community of teacher-scholars who can share ideas about data science education.
- It will create a focal point for collecting and disseminating information about our educational and outreach work in data science.
- In particular, it will create a natural point of contact for those seeking expertise about data science on campus, in the local community, and across the world.

Right now, incorporating data science ideas and projects into the classroom is challenging. It is typically done by those faculty who have already made a substantial commitment to data science in their research. Yet, there are tremendous opportunities to incorporate data science ideas into courses across the university, and to help advance the use and understanding of data science beyond the academy. Whether they are data science researchers or not, faculty who want to incorporate data science into their teaching and research need somewhere to physically go to seek advice, materials, and assistance. Thus, while the teaching of data science will belong to the colleges and departments, IDSI will help coordinate these offerings as well as support the development and implementation of innovative data science courses, curricula, and programs that are not currently being offered at Illinois.

Furthermore, the expertise and personnel that IDSI will nurture to support faculty and students in data science education and research can also be deployed to support data science education and outreach in the community, locally and statewide. Outreach efforts in data science education—for example, helping K-12 teachers to incorporate data projects into their teaching—are of great potential importance and impact. Having experience in learning from data will create important pipelines for students in the coming years that will eventually deliver to campus a community of learners who already possess basic skills and conceptual fluency in data science methods, in much the same way that years of incorporating computers into standard classroom experiences from kindergarten through high school has now developed a pool of undergraduates who are ready to capitalize on the more advanced opportunities that the data science revolution can provide if our campus is able to effectively harness their energy and ambition throughout the university curriculum. Thus, by developing a strong commitment to outreach, IDSI will create new ways for the university to benefit the local community and the people of Illinois.

WAYS TO ENGAGE COMMUNITY

One key goal for IDSI will be to promote the application of data science to a wide range of important, social and societal issues. We propose that this process start at the local level, and, as we demonstrate success, expand to the regional, state, and national levels. We believe that this goal will be best achieved by fostering an 'outward-looking' institutional culture that ensures the resources and expertise of the Institute are broadly accessible to members of local and regional communities. Achieving this result will require developing strategic partnerships with both individuals (e.g., local businesses) and private or public organizations (e.g., foundations, city or county agencies) that can benefit from actionable intelligence informed by data science approaches to improve decisions and to optimize operational efficiency.

Our 'Data + Ideas Incubator' is perfectly suited to form the foundation for these relationships. If an 'off-the-shelf' solution is not readily available for a given problem, the incubator can be used to share the data and a description of its associated problem broadly within the University of Illinois community. In this manner, problems can be tackled in multiple venues ranging from classroom settings to appropriate registered student organizations (e.g., hackathons), knowledgeable IDSI affiliate faculty and staff, or even interested community groups (e.g., members of the data science user group run out of the Research Park). We should emphasize that this is not simply a hopeful appeal to the altruistic nature of our campus community: demonstrating practical experience in data science requires producing visible results. Solving real-world problems that positively impact our local, state, and nation benefits the local community just as much as it supports the learning goals of the budding data scientists on our campus.

Solving problems for others has immediate benefits for interested stakeholders, but longer-term impact will come from building and sustaining a training pipeline to enlarge the pool of data science talent in ways that generate waves of solutions across multiple problem cycles and a wide range of beneficiaries. While we do that naturally through undergraduate and graduate training, we also see a need to develop initiatives that both arouse interest and provide training to people in our community and state that have limited exposure, experience, or competency in the field of data science. This point becomes even more important when viewed through the lens of our grand challenge themes: as data and technology become more pervasive, citizens will require continuing education to keep pace. Programs that equip and inspire community members can leverage existing curricula developed by IDSI affiliated researchers and staff, or be developed directly by the institute as necessary. Our campus has multiple examples of these types of programs, for example, the Osher Lifelong Learning Institute (OLLI) at the University of Illinois and the 'Genomics for Judges' program developed by IGB.

However, a 'build-it-and-they-will-come' approach will be an ineffective way to engage broadly with our community. Instead, we feel strongly that the first community to benefit from our engagement efforts should be the University itself. In many ways the University campus, with all of its varied activities and operations, is an ideal exemplar of a multidimensional social community, with a wide range of data science competencies, and a great diversity of practical and operational needs. Thus, it makes considerable sense that IDSI should reach out to the wider campus community to develop relationships that can be used as demonstration cases illustrating how data collection, analysis, and interpretation can improve operations, support better decisions, and allocate scarce resources more efficiently.

This potential was displayed through the widespread interest, engagement, and enthusiasm of attendees at the iDSI-sponsored University Data Summit held in August 2017. Over 200 individuals actively participated, representing almost every college, institute, administrative and service unit, and almost every sphere of University activity and operations from student services to teaching, research, facilities, finances, advancement, administration, finances, athletics, and beyond. The University Data Summit was also supported, and attended, by numerous senior administrative leaders, many of whom provided the foresight, encouragement, and philosophical framework for the event.

In some ways, this summit acted as a surrogate for the IDSI by successfully bringing people, data and technology together. Fostering interaction between data providers, data analysts, and data consumers from across the campus community showed what could be achieved if our vision for a campus data science institute becomes reality. A series of more than twenty lightning talks, from individuals representing a wide range of operational units (e.g. registrar, housing, athletics, provost's office, data security, and library operations), described successes, needs, and challenges related to the use, access, analysis, and interpretation of data at the University. Interestingly, many of these presentations highlighted existing campus expertise, tools, and data sources, which are readily available for use by other members of the University, but that are often unrecognized and underused. The summit also included facilitated breakout groups, which allowed individuals to participate in an envisioning process to imagine, and devise, schemes by which our campus could become a 'data smart' institution.

During this event, several strong themes for additional exploration emerged that engaged and motivated a large, interdisciplinary group of competent individuals to address them. These themes included (1) a data driven student performance dashboard to inform and advise the learner, and to prompt institutional intervention; (2) multidimensional data collection and analysis to enhance student and faculty recruitment, engagement, wellness, retention and success; (3) the need for a comprehensive, central, catalogued, searchable repository of University data; (4) an initiative to generate solutions to problems related to data governance and confidentiality, and to improve access and dissemination of campus, college and departmental level data; (5) a predictive algorithm for maximizing the efficiency of classroom and teaching resource allocation; and (6) a process of data-informed optimization of research revenues and expenditure. Each of these project ideas could form the basis for developing specific, interdisciplinary task forces that would be assigned the task of outlining a project roadmap. Our proposed data science institute would provide a natural home for these interested individuals to gather so that they might identify the processes and resources needed to realize these initiatives on our campus.

The success of this summit and the many other data science focused summits organized by the iDSI over the past nine months powerfully illustrates the potential benefits that IDSI could have for our local community. These experiences demonstrate how a centralized collaborative could support stakeholders in forming new partnerships and in generating new ideas for gathering, analyzing, and applying data to make better plans, decisions, and outcomes for the good of all partners. Clearly, the benefits of such a collaborative should not be restricted to our University alone. Many local and regional agencies, small businesses, educational institutions, and nonprofit entities collect data but lack suitable expertise, capacity, or ability to analyze and interpret the information effectively. At the same time, many governmental and commercial institutions are interested in using data to drive policy decisions, guide resource allocation, maximize efficiency,

and ensure fiscal responsibility. To identify and recruit these communities we propose to leverage university infrastructure, such as the extension offices, through which groups can present their needs, goals, and interests. In this manner, IDSI can either identify and recommend 'off-the-shelf' solutions, identify related efforts so that new challenges can properly leverage existing or on-going work, or load novel challenges into the 'Data + Ideas Incubator.' While no list, especially at this stage, would be complete, we can provide several examples of potential community partners.

- Local city and government (e.g., public services, energy utilization, economic evaluations, risk analysis, and tourism analyses)
- Regional health care organizations (e.g., Carle)
- Local law enforcement agencies (e.g., public safety, resource allocation, and crime prevention)
- Regional education establishments (e.g., Parkland College or K-12 school districts)
- Local businesses (e.g., market surveys, economic analysis, and risk-benefit constructs)
- State NGOs.