Executive Summary

Big Data: Big Challenges, Big Opportunities
The Reification of Consilience

Daniel Reed, Senior Vice President for Academic Affairs
University of Utah

- Each individual's Weltanschauung is shaped by the totality of their life experiences and it defines their perspectives, philosophy and understanding of the cultural, economic, and scientific milieu. An explosive growth of knowledge has had a negative effect on one's ability to see the integrative whole. The original seven liberal arts have given way to the speciation of disciplines. As a result, academics feel the deep loss to satisfy their needs for convergent conversation and reflection. Disciplinary, interdisciplinary, and transdisciplinary collaboration and shared insights are needed for a multitude of technological revolutions and socioeconomic disruptions. The emergence of big data and machine learning are potential opportunities to holistically reunify divergent domains.

- Three socioeconomic and technical developments have led to the explosive growth of data. First, interconnected mobile devices and the associated growth of social media have created volumes of consumer data of economic value. Second, new scientific instruments are changing the nature of academic research. With large scientific data readily available, hypothesis-driven experimentation is now being complemented by exploring what might the existing data reveal. Third, small sensors such as consumer health devices, environmental monitors and connected household objects are providing a rich source of data for understanding human behavior and interactions.

- The recent rise of machine learning depends on the confluence of rich sources of data, low-cost high-performance computing and deep learning. Deep learning's recent success depends on large volumes of training data and powerful computing systems. These tasks, once solely in the human cognitive domain, raise social, economic and ethical questions.

- The explosive growth of data has brought about many challenges such as issues with data retention. Security, privacy, and bias loom large in big data and machine learning discussions, particularly individuals' data. Big data and machine learning are accelerating, and we must recognize there are benefits as well as unexpected consequences. It will take an engaged and thoughtful debate to define both a social consensus and legal and ethical framework.

Growing Diversity in Data Science: Shared Lessons from Clinical Trials

Robert D. Simari, MD, Department of Cardiovascular Medicine
University of Kansas School of Medicine
University of Kansas Medical Center

- Data science must evolve with the social changes that are underway. Individuals who make decisions on how data sets are generated and analyzed make decisions based on their life experiences. In this paper, the importance and challenges of diversity in clinical trials is applied to the emerging field of data science.

- The lack of diversity in clinical trials can be attributable to factors such as lack of trust, and social and financial barriers of diverse populations. To overcome this, an area of important focus is the diversity of the investigative team, which may also be paramount in data science. Diversity in data science is important because it may limit bias in data sets and the analysis of the data. The likelihood that data science focuses on the social and medical issues affecting diverse groups is enhanced with a diverse investigative team. The data science workforce will suffer if diverse groups are not considered for advanced training in the field.

- There are similar challenges to developing diverse data science teams as there are to diverse clinical trial teams. Prepared doctoral graduates are needed to enter both fields and there remains a lack of diversity in the disciplines. Efforts need to be focused when students begin to develop aptitude for the STEM fields. The University of Kansas Medical Center has engaged with the Kansas City, Kansas community with a series of programs that attempt to meet the cornerstones of what such programs should include. These programs include a university run Head Start program, faculty involvement in multiple public K-12 programs, and faculty and staff advocacy that enabled sidewalks and grocery stores to be built in food deserts in the community.

Quantifying Biomedical Data Reuse in an Open Science Ecosystem

Lisa Federer, PhD, MLIS, Data Science and Open Science Librarian
Office of Strategic Initiatives, National Library of Medicine,
National Institutes of Health

- The amount of data that is available today has exploded due to advances in a range of research disciplines. Not only do we have more data today, it is freely available. More data sharing is the result of the adoption of policies requiring researchers to share their data. Though not all researchers share their data because they are required to, some share as a move toward open science. Part of the trend of open science includes open access publications, but it also encompasses digital objects across the research lifecycle including data and code.

- With advances in technology, researchers have a wealth of public data available to them. The universe of publicly available research is vast with the National Library of Medicine (NLM) playing a significant role providing access to biomedical data. NLM is just part of the data sharing picture with the National Institutes of Health (NIH), institutional repositories and generalist repositories also contributing.

- Time, effort and funding has made research data publicly available, though what happens with the datasets remains unknown. Understanding data reuse can pave the way for rewarding researchers. Article citations are a measure to quantify scientific impacts, yet data and code are research outputs also meriting reward. Major research funders have formally recognized datasets as research products to demonstrate researchers' impact.

- Data are an important research output that merits a reward system to incentivize sharing. There are technological challenges that hinder rewarding data sharing. Though attempts have been made to standardize data citation, adherence remains low and there is still debate if data citations are the appropriate acknowledgement. Despite not having a reliable way to quantify data reuse, the future reward of data sharing is something to be considered. Not too far in the future, data reuse will feasibly be tracked and quantified and it is important to think about metrics to use for giving credit and rewards for the reuse of data.

Journal Programs and Cross-Disciplinary Research

Marianne Reed, Digital Initiatives Manager
University of Kansas Libraries

- In order for innovative cross-disciplinary research to find its audience, it must be easily discovered by scholars, professional practitioners, and the public. Journal publishing programs in libraries operate under the principle that investment in open access publishing of quality peer-reviewed research is the best way to make that research visible to a global audience and to shift control of publishing from commercial entities to the academy. Library publishers are therefore not constrained, as commercial publishers are, by the need to publish only research that will ensure a profit. This means that library publishing programs can provide a home for cross-disciplinary journals that break new ground and that may take time to find an audience.

- The lack of a profit imperative for library publishing programs also means that the platform for hosting journals is provided to journals at little or no cost, which makes library publishing very attractive to editors looking for a place to publish a new journal. Once the infrastructure is operational, the cost to add a new journal to the system is negligible because the costs of maintaining the technology are already covered. This lowers the financial barriers to starting new journals, allowing editors to focus on the task of finding and publishing excellent peer-reviewed research instead of fundraising.

- Journal platforms used by library publishers are designed so that journals published on those systems automatically follow best practices and standards, such as those outlined by the Open Archives Initiative Protocol for Metadata Harvesting (OAI-MPH) that make the content readily discoverable by internet search engines. These platforms also integrate the use of machine-readable licenses that clearly indicate how the content can be used. In addition to infrastructure that ensures visibility, library publishing programs benefit from existing library expertise in collaboration, technology, copyright, data management, scholarly publishing, information literacy, digital preservation, and the effective promotion of online research.

Convergence Research in the Age of Big Data: Team Science, Institutional Strategies, and Beyond

Daniel Sui, Vice Chancellor for Research and Innovation
Jim Coleman, Provost and Executive Vice Chancellor for Academic Affairs
University of Arkansas

- With the explosion of big data in the last ten years, discussions on interdisciplinary research now emphasize convergence research through a team science approach. The authors of this paper present how to facilitate convergence research in the age of big data by exploring the concept of convergence research, outlining key elements of a team science approach, and discussing institutional strategies, opportunities and challenges.

- More than ever we need interdisciplinary collaboration and team work to address the multiple challenges to society which cannot be resolved by any individual discipline. Big data and data science are emerging as the fourth paradigm, following the previous three paradigms in empirical, theoretical, and computational approaches to science. Now is the time for higher education to conduct convergence research through big data and team science to address grand societal challenges that reach beyond traditional boundaries.

- By emphasizing the need for convergence research, the authors are not abandoning the need for traditional-based research and individual-based inquiries. We need more cutting-edge discipline-based work to enhance our convergence efforts. All research must be conducted individually at some point, even in large team projects. Through the dialectal process of convergent/divergent, disciplinary/interdisciplinary, individual/team-based approach, our research enterprise has been propelled to a level of excellence to make the world a better place for all.

Making Mountains out of Molehills: Challenges for Implementation of Cross-Disciplinary Research in the Big Data Era

Daniel Andresen, Director, Institute for Computational Research in Engineering and Science. Professor, Department of Computer Science
Eugene Vasserman, Department of Computer Science
Kansas State University

- In this paper, the authors present a "Researcher's Hierarchy of Needs", based loosely on Maslow's "Hierarchy of Needs" in the context of interdisciplinary research in a "big data" era. As in Maslow's model of needs, those needs at higher levels can only be expressed if lower levels needs are met. In the research environment, these levels are shared vision, social capital/relationships, domain expertise, technical expertise, and data and software. In this researcher's model of needs, two researchers who have a shared vision but lack the data for their research will be unsuccessful. Considering the hierarchy model for researchers, they suggest that researchers and institutions recognize that interdisciplinary research is both difficult and rewarding.

- There are several overarching issues when considering interdisciplinary research centered around big data. Interdisciplinary research is extraordinarily challenging in universities where the environments are typically siloed by departments in which individuals within the unit communicate at a much higher level. Cybersecurity and privacy present more challenges as data sharing occurs between research groups. So, too, research environments are lacking in the support needed for today's interdisciplinary research. Those institutions who can best support researchers will have a strong competitive advantage. The molehills need to be converted to mountains at every level of the hierarchy: infrastructure should be well resourced, professionals trained in big data across disciplines, and institutions should be committed to a planned and systemic infrastructure.

Training for Cross-Disciplinary Research and Science as a Team Sport

Jennifer L. Clarke, PhD, Professor, Food Science and Technology, Statistics
Bob Wilhelm, Ph.D., Vice Chancellor for Research and Economic Development
University of Nebraska-Lincoln

- Members of the University of Nebraska, a land grant highly research active university, recognize the increasing significance of data and computing across disciplines. With faculty, postdoctoral scholars, and students working together in a cross-disciplinary environment and leveraging advances in data and computing, we can further institutions and make discoveries that benefit humankind.

- This way of thinking—scientific research as a team sport for communal benefit—represents challenges to faculty such as lack of knowledge and resources to plan for use of data beyond their own projects. Another challenge includes the faculty time and effort to prepare data or code to meet the guidelines for proper data sharing. A more sustainable model for cross-disciplinary data services and management is needed, as it is difficult for researchers to secure all the financial support needed. Proper documentation and sharing of code are a requirement for some publications, and institutions are challenged with how to support researchers with meeting this requirement. The University of Nebraska-Lincoln is addressing this campus need with the training of individuals with advanced training in data science who manage multiple projects across disciplines. These application specialists will be tasked with facilitating transdisciplinary research through data knowledge and advanced cyberinfrastructure. As repositories of institutional memory, they will enable the use and reuse of data and code from university projects.

- Research has evolved into a team sport with members from various disciplines working toward a shared goal, enabled by advances in data science and cyberinfrastructure. Institutions of higher education must enable convergent research through avenues of support such as: reproducibility of data, University Libraries, application specialists, and strategic investments in transdisciplinary teams.

Protecting the Value of Interdisciplinary Collaborations in the Development of a New Budget Model

Carl Lejuez, Interim Provost and Executive Vice Chancellor,
University of Kansas

- If you want to know an administrator's priorities, you need look no further than the budget. In its efforts to build a more stable and fiscally healthy institution, the University of Kansas developed a new budget model. A Responsibility-Centered Management (RCM) model, or a hybrid of it, has been adopted by many U.S. higher education institutions. This model offers a decentralized budget with a percentage of revenue controlled by the unit that generated the revenue. KU refers to its hybrid RCM, where less than 100% of the funds are returned to the unit, as a Priorities Centered Management (PCM) model. Our budget is aligned with our priorities including research, student success, career development, outreach, and diversity, equity and inclusion.

- KU had a $20 million budget reduction in Fiscal Year 2019. A series of townhall meetings were held to educate the Lawrence campus community on how the new model will align resource allocations with strategic priorities. A working group, in consultation with campus leadership, developed and shared guiding principles for the new budget. The overall PCM model was enhanced with meetings with the provost's direct reports, town hall presentations, and meeting with faculty, staff and students.

- The new budget allocation model will take effect in Fiscal Year 2021. The first structural feature of the model is the creation of three broad categories in which budgetary resources can be allocated: 1) foundational priorities, 2) institutional strategic priorities and 3) units allocations, including academic and support units. The funding for academic units will be based on performance in a set of priority areas. Additional budget strategies include subsidies outside of unit allocation and implementation of guardrails to reduce the impact of potential budget fluctuations to units. The PCM Model provides support for interdisciplinary collaborations at a time when there is great awareness of the benefits of interdisciplinary initiatives.

Cross-Disciplinary Research: From Nuclear Physics to Cosmic Ray Detection and Medical Applications

Christophe Royon, Foundation Distinguished Professor
Tommaso Isidori
Nicola Minafra
Department of Physics and Astronomy
University of Kansas

- The Large Hadron Collider (LHC) at Cern, Switzerland is the highest energetic collider in the world. The collider provides a better understanding of proton structure and reproduces conditions as close to possible to the Big Bang, where new particles might be produced General purpose detectors including ATLAS and CMS are large detectors built to identify all kinds of particles produced after the interactions. Recently, strange events were observed at the LHC where protons are found to be intact after interacting, though they lost part of their energy. This means that it is possible to detect intact protons after the interaction with detectors.

- At the University of Kansas (KU), fast silicon detectors together with their readout electronics have been developed to achieve this goal. At KU, multipurpose electronics boards were designed to measure precisely the time when particles cross the detector. A test-stand was built to test the full chain from the detector to the read-out electronics. The amplifier that was designed at KU can be used for a full range of detectors and applications. The performance of the amplifier designed at KU is better than commercial ones and the cost is much lower.

- Three possible applications using Ultra Fast Silicon detectors and electronics that were developed at KU are discussed. The first is a project in collaboration with NASA to measure within one single detector the nature and the energy of cosmic ray particles originating from the sun. This will eventually help with the precise measurement of radiation between Earth and Mars, needed to send astronauts to Mars. The second application will measure radiation in cancer treatment with millimeter squared precision and, if successful, will allow a more optimized dose during treatments. An additional medical application deals with PET imaging. The third application is a better understanding of catalysis in chemistry. This could have implications for the way medicine is absorbed and improve the interface of human cells and medicine. It will also improve the methods to desalinize sea water.

Complexities of Conducting Cross-Disciplinary Biomedical Research

Jennifer Larsen, MD, Vice Chancellor for Research
W. Scott Campbell, PhD, Senior Director of Research and IT
University of Nebraska Medical Center

- Solving complex health related problems requires large teams with a broad range of skills. There are many "complexities" that must be considered when building an effective team for cross disciplinary biomedical research. An effective team must define the rules of engagement, which takes time and effort. Teams should form an environment where all members are understood which includes not only the vocabulary in conversations, but also using a common format for capturing and storing data. An effective multidisciplinary team includes team members with terminology expertise and the ability to translate between disciplines.

- Another complexity in cross disciplinary biomedical research is data transfer and storage. More teams are working with large research files which need to be stored, and moving the data is time consuming. Data sharing can create new risks if those researchers who are sharing data are not as knowledgeable about privacy and security issues. Protected health information, protected individual information, as well as other sensitive data might require special controls for the access of data, as well as the ability to audit who has accessed data.

- There are special considerations with global sites, teams or focus. There are increasing and changing rules and regulations on moving data, samples, equipment or team members between countries. Lastly, the public needs to be part of the communication before and after data is shared, to understand the value and make use of the of the results that are found.