1. Introduction

This is a report summarizing a NSF sponsored workshop that was held in Chicago on June 26-27, 2017 on translational data science (TDS). The topics covered included:

1. What is TDS? What is a good definition of TDS? What is the relationship of TDS to the foundations of data science? Should it be a separate discipline? What are some TDS activities?

2. What are some examples of success stories and failures in TDS? What lessons can we learn from these? How do we interpret translation and consequently, what are viable and useful measures of impactful translation?

3. What are some research challenges, opportunities, and high priority areas in TDS? What would a research agenda or research program to further TDS look like?

4. What are some examples of models and methodologies for translation? What are some lessons for TDS methodologies and models that can be applied to a wide range of disciplines?

5. What is an education agenda for TDS? How should we teach TDS? How would a TDS educational agenda compare with data science as is currently being taught?

6. How will TDS evolve in the coming future? Will it be grow, carve a niche and sustain itself? How can it be deployed broadly in general beyond academia?

Broadly speaking, TDS is a new term for an emerging field that applies data science principles, techniques, and technologies to disciplines, including, but not restricted to, science, engineering, and biomedical research, especially when the result broadens our fundamental understanding of the discipline of data science or has a broader societal or human impact.

This report attempts to answer the above questions and provide a summary that could be used for further growth of TDS. The rest of the report is organized into the following sections:

- Section 2, background pertinent to basic, applied, translational research
- Section 3, definitions of TDS
- Section 4, examples of successful use of TDS
- Section 5, challenges for TDS endeavors and enterprises
In the final section, we provide a summary of the proceedings in Chicago over this past summer and point to the future.

2. What is Translational Data Science?

In this section, we first examine the notion of translation as exemplified in the phrase “translational data science” by discussing basic research and its consequences of use. A popular taxonomy, Pasteur’s Quadrant, is first discussed, which is then extended to include impact to the accompanying science and larger society. Next, translation in the context of medicine is introduced followed by likely deployment to more complex societal problems. Finally, we discuss how TDS can be viewed as the science of data science.

2.1 Basic versus applied: Pasteur’s Quadrant. It is standard to distinguish basic research from applied research. However, is there really a distinction between applied and basic research? It can also be helpful to think of graduations of pure and applied research along a dimension, with pure research at one end and applied research at the other end. In an impactful book first published in 1997, Donald Stokes introduced a second dimension for classifying research: whether research improves our fundamental understanding of a field [Stokes-1997]. With these two dimensions, research can be classified into three types: pure basic research, pure applied research, and use-inspired basic research (Figure 1). Use-inspired basic research, it can be argued, forms the basis and initial point for translation and TDS.

2.2 The impact of research. To understand TDS, it may be helpful to introduce a third dimension – the impact of the research outside its own field [Grossman-2017]. For example, within data science, there have been a number of projects, workshops, and conferences focused on data science for the social good, in which data science is applied to a problem, with the potential human or societal benefit of the solution an important measure of the success of the project. See [Catlett-2015], for example. Since at least 1997, NSF has used the broader impact of research as one of the evaluation criteria for awarding grants [NSF-2015].

NSF projects, in the aggregate, should contribute more broadly to achieving societal goals. These “Broader Impacts” may be accomplished through the research itself, through activities that are directly related to specific research projects, or through activities that are directly supported by, but are complementary to, the project [NSF-2015].
As another example, translational research within the biomedical community can be thought of as trying to reduce the time and effort required to move a basic research discovery into practice so that it has an impact on a patient or population of patients.

2.3 Translational research in the biomedical sciences. It has long been recognized that it is very challenging to take a basic discovery about the biological basis of a disease and use it to improve how physicians diagnose or treat the disease. Often the term “bench-to-bedside” is used to refer the process of translating medical or healthcare discoveries into medical practice affecting patients. One of the standard definitions of translational research is:

For many, translational research refers to the “bench-to-bedside” enterprise of harnessing knowledge from basic sciences to produce new drugs, devices, and treatment options for patients. For this area of research — the interface between basic science and clinical medicine — the end point is the production of a promising new treatment that can be used clinically or commercialized (“brought to market”) [Woolf-2008].

In 2011, the NIH set up a new institute focused on translational science called the National Center for Advancing Translational Sciences or NCATS. In the context of biological and medical research research, NCATS defines translational as “the process of turning observations in the laboratory, clinic, and community into interventions that improve the health of individuals and populations – from diagnostics and therapeutics to medical procedures and behavioral interventions” [NCATS-2017]. Similarly, “translational science” is “the field of investigation focused on understanding the scientific and operational principles underlying each step of the translational process” [NCATS-2017].

Translational research has received significant attention in the biomedical sciences because of the pressing need to enhance human health and well-being by applying findings from basic research to the production of promising new treatments, and ensuring that research knowledge reaches practice and is implemented correctly in practice. Thus, TDS bridges applied and basic research.

2.4 Translation beyond medicine. Is it possible to take the results of the laboratory and best practices of the clinic into society? Consider for instance, the opioid epidemic that is negatively impacting the Midwest. It takes more than 10 years for an idea rooted in logistics or therapy or device development to be effectively deployed among the afflicted populations, because there are roadblocks. Per the accepted definition of clinical translation [Woolf-2008], two specific roadblocks exist. The first obstacle often referred to as \( T1 \) prevents basic research findings from being tested in a clinical setting easily, while the second or \( T2 \) prevents proven interventions from becoming standard practice. What is missing from these definitions is the acknowledgement that the problems adversely affecting the human condition are wicked problems that defy systematic solutions like those pursued in engineering and science or even the clinic [Rittel-Horst-Webb-1973]. Planning problems are wicked because they are difficult or impossible to solve and are often incomplete, contradictory, and changing. Further, because of complex interdependencies, purported solutions of a wicked problem may reveal or create other problems.

Yet another take is offered by Boyer. To reiterate, defining translational data science in a larger context moves from the extraction of knowledge to its application. ‘Application’ is used in Boyer’s model of scholarship [Boyer-1990]. Boyer argues that the scholarship of discovery in higher education, and science more generally, primarily involves investigation and synthesis – and that it needs to be expanded to include the scholarship of application. He describes application as answering the questions, “How can knowledge be applied to consequential problems?” “How can
it be helpful to individuals?” and “Can social problems themselves define an agenda for scholarly investigation?”

2.5 The science of data science. TDS is the systematic study of the processes involved in applying theoretical data science principles to problems of interest. As a distinct area under the larger umbrella of data science per se, TDS involves the interplay between technical areas and application domain areas, and is convergent in nature [Roco 2013].

In his summary “50 Years of Data Science,” David Donoho states that “despite the excitement around ‘data science,’ ‘big data’, and ‘analytics’, the ambiguity of these terms has led to poor communication between data scientists and those who seek their help.” [Donoho 2015]. We believe that this poor communication is, at least partially, due to the confusion between techniques and the application of techniques in practice – i.e., the foundational vs translational issue. Donoho proposes six divisions/activities of so-called “Greater Data Science,” the sixth of which is “Science about Data Science.” In that section he says, “Tukey proposed that a ‘science of data analysis’ exists and should be recognized as among the most complicated of all sciences.” He advocated “the study of what data analysts ‘in the wild’ are actually doing”, and reminded us that “the true effectiveness of a tool is related to the probability of deployment times the probability of effective results once deployed.” We believe that TDS includes this general category of science about data science.

3. Definitions of Translational Data Science

Using the discourse of the previous section, we offer four definitions that can be used to categorize translation. This list is no way complete; as TDS grows, it is only natural that both more nuanced and very comprehensive definitions are offered for consideration by the community. At the workshop, we discussed the definition of translational data science. Although we did not reach a consensus, there was broad support for two related definitions:

1. Translational data science is a field that applies data science principles, techniques, and technologies to problems in other disciplines, including, but not restricted to, science and engineering research, especially when the result broadens our fundamental understanding of the discipline of data science.

2. Translational data science is a field that applies data science principles, techniques, and technologies to challenging problems in other disciplines, especially those that hold the promise of having an important impact on human or societal welfare.

Note that the first definition is Pasteur's quadrant use-inspired basic research applied to the discipline of data science. With both definitions, data science is applied to problems in other disciplines (“applied data science”). With both definitions, there is also an impact. In the first definition, there is a broader impact, namely an improved understanding of the discipline of data science. In the second definition, there is also a broader impact, this time on human or societal welfare. Thus, a larger emphasis on impact leads us to the third definition.

3. For the purposes of this report, it may be helpful to think of TDS as a field that applies data science principles, techniques and technologies to problems in other disciplines with the hope of not just solving a particular problem but also of having a broader impact.

With this definition, TDS is the application of data science to problems with the goal of having a broader impact. Note that with this definition simply the application of data science to a problem would not be considered TDS unless there was the expectation of some type of broader impact.
Here “broader impact” is not used specifically in the NSF merit review sense [NSF-2015], but more generally as impacting our fundamental understanding of the discipline of data science, of having a human or societal benefit, or simply having an impact outside of the particular problem being addressed.

It is standard today to apply the principles, tools, and techniques of data science to a wide variety of problems in other disciplines. The importance of the discipline of translational research is not so much to ensure that research knowledge reaches practice, but rather to give careful thought to the structures and processes of translation and its implications, and to study and better understand these processes leading to a fourth definition.

4. There is value in seeking a process views of translational data science whereby methods of translation and its deployment are studied. Essentially, it is a Science about Data Science.

Research in TDS provides insights and understanding of issues and challenges related to applying data science theories in practice yielding its own methods, techniques, norms, best practices, and pedagogy. The methods of translation in data science are typically multi-stage and multi-step. Translational data science not only uses but also studies these methods; the principles behind them; and the transitions between stages, including the incorporation of feedback loops between the multiple stages. These stages can involve foundational data science, data wrangling, exploratory data analysis, modeling, scaling and optimization of data science solutions, and issues of ethics and policy. Note that this definition is beyond the abstraction of a workflow, which is purely technical. It is the whole “package” that includes technical, socio-technical, socio-economic, legal, and policy aspects of the comprehensive problem.

In light of Boyer’s concepts (see above) and the other definitions, a definition of TDS could be the practical use of knowledge derived in context of a specific societal problem through an iterative process of real-world application and stakeholder engagement. This definition focuses on societal versus scientific problems, and on practical use (e.g., prevent infant mortality) versus domain-specific expertise (e.g., public health). In this context, one could include issues of privacy, ownership, “liability” or implications of decisions, etc., as they relate to data science. It also becomes a team science rather than an individual or even a (small) group or laboratory endeavor.

To close this section, it should be noted that the above mentioned definitions will only grow and will capture other aspects of translation as TDS grows. Still, the four definitions can serve the purpose well.

4. Examples of Translational Data Science
The following describe representative successful application of TDS to a variety of problems, which were provided by some of the workshop attendees. The first two examples easily fit the second definition of TDS as listed in Section 3. The other three examples allow for the third kind of translation where the broader impact is achieved for very wicked problems. Many more examples undoubtedly exist. It will be the goal of subsequent workshops to divine more more meaningful examples that can be categorized by all prevailing definitions.

4.1 Data-driven exploration of multi-relational datasets for drug discovery
Large scale data analysis has become a crucial part of drug discovery. Biologists and chemists need to quickly explore and evaluate potentially effective yet safe compounds based on many datasets that are in relationship with each other. However, there is a lack of tools that support investigators
with utilizing these processes. To remedy this problem, researchers have developed ConTour, an interactive visual analytics technique that enables the exploration of these complex, multi-relational datasets. At its core, ConTour lists all items of each dataset in a column. Relationships between the columns are revealed through interaction, and selecting one or multiple items in one column highlights and re-sorts the items in other columns. Filters based on relationships enable drilling down into the large data space. To identify interesting items in the first place, ConTour employs advanced sorting strategies, including strategies based on connectivity strength and uniqueness, as well as sorting based on item attributes. ConTour also introduces interactive nesting of columns, a powerful method to show the related items of a child column for each item in the parent column. Within the columns, ConTour shows rich attribute data about the items as well as information about the connection strengths to other datasets. Finally, ConTour provides detail views that show items from multiple datasets and their associated data at the same time. Researchers have demonstrated the utility of the ConTour system via case studies conducted with a team of chemical biologists who investigated the effects of chemical compounds on cells and desired to understand the underlying mechanisms [1].

4.2 Discovery of principles of nature from mathematical modeling of DNA microarray data

Recent advances in DNA microarray hybridization technology make it possible to record the molecular biological signals, e.g., mRNA expression levels and proteins’ DNA-binding occupancy levels, that guide the progression of cellular processes on genomic scales. The rapidly growing number of DNA microarray datasets holds the key to the discovery of previously unknown molecular biological principles.

Future predictive power, discovery, and control in biology and medicine come from the mathematical modeling of DNA microarray data, where the mathematical variables and operations represent biological reality. The variables and patterns uncovered in the data might correlate with activities of cellular elements, such as regulators or transcription factors, that drive the measured signals [A5]. Such techniques have recently been used successfully in cancer research to model patient-matched grades III and II, i.e., lower-grade astrocytoma (LGA), brain tumors and normal DNA copy-number profiles. For example, a genome-wide tumor-exclusive pattern of DNA copy-number alterations (CNAs) was revealed – it was encompassed in a previously uncovered glioblastoma multiforme (GBM; grade IV astrocytoma) where GBM-specific CNAs encode for genes associated with transformation and proliferation via growth and developmental signaling pathways in GBM relative to LGA [2].

4.3 Conservation

How many animals are in a population? How many survive until next year? How far do they travel? What habitats do they use? And with whom do they associate and why? These basic questions, critical for science and conservation, are among the hardest to answer because they require spatially and temporally dense observations of a large fraction of the animal population. For the first time, as the price of photography drops while quality and availability improve, the raw data needed to answer these questions can be gathered non-invasively, and at high volume and resolution over time, space, and individuals. Moreover, these data can come from traditional sources, such as camera traps, scientific projects, systematic surveys, as well as crowdsourced from citizen scientists, online social media, and tourists. There are several systems today that enable crowdsourcing and use of images for wildlife study and conservation (using computer vision, machine learning, and data analytics).

These systems include eBird.org (http://ebird.org/content/ebird/), leafsnap(http://leafsnap.com), eMammal (https://emammal.si.edu), and Wildbook (http://wildbook.org/doku.php). For example, crowdsourcing images of zebras and using those to perform sight-resight population size
estimates provided the most accurate census of the species (now used by the IUCN Red List as the official number for the species), gave the best evidence for evaluating existing conservation policies, and even led to a change in management policies for other species.

### 4.4 Urban studies

According to the United Nations Population Fund (UNFPA), the current state of urbanization can be summarized as, “The world is undergoing the largest wave of urban growth in history. More than half of the world’s population now lives in towns and cities, and by 2030 this number will swell to about a billion.” This unprecedented growth in urbanization, political conflicts and ensuing internal displacement of population, and increased poverty in the 20th century have resulted in the rapid increase of informal settlements. These unplanned, unauthorized, and/or unstructured homes, shantytowns, barrios, or slums, pose multiple challenges to the nations as these settlements are often located in the most hazardous regions and lack basic services. Though several World Bank and United Nation-sponsored studies stress the importance of poverty maps in designing better policies and interventions, mapping slums of the world is a daunting and challenging task. Recent TDS success in mapping and monitoring of urbanization included global settlement mapping and slum mapping [4][5].

### 4.5 Monitoring biomass

Monitoring biomass and its distribution over large geographic regions to identify changes are important tasks in many applications. With recent emphasis on biofuel development for reducing dependency on fossil fuel and reducing carbon emissions from energy production and consumption, the landscape of many countries is going to change dramatically in coming years. In the United States, continuous corn production is becoming a dominant cropping pattern as more and more soybean and wheat rotations are replaced by corn. It is also expected that more and more pasture lands will be converted to Switchgrass in the coming years. These changes are not limited to the United States alone. In developing countries like India, rural areas are facing increasing demand for energy. It is expected that energy crops like Jatropha curcas are going to be widely planted in Asian countries. A novel GEO-based biomass monitoring using Gaussian Process (GP) learning showcases the power of TDS [Chandola-Vatsavi-2011].

### 5. Research challenges of TDS

We see TDS as implying its own research agenda, distinct from that of core data science methods, and distinct from that of the domains served by data science. TDS research questions are those that emerge only when working directly across disciplinary boundaries, delivering data science technology into practice. These questions may be technical, social, or socio-technical, and will typically be defined and pursued with participation from domain experts and technical experts, and with input from policy and law scholars, social scientists, and ethicists. In defining and pursuing this research agenda, we encourage building lasting partnerships across disciplines, centered around an interest in TDS. The definitions offered in Section 3 strive to capture the essential nature of translation and also allow for also development of metrics and rubrics.

In this section, we describe some common problems and identify some common research challenges and technical themes. There are a variety of technical challenges associated with generating meaningful data products from multiple complementary data sources as well as challenges associated with the need to rapidly carry out computationally intensive machine learning methods.
5.1 Challenge 1: Responsible data science. As the tools and techniques of data science become democratized, the potential for harm is amplified: algorithmic changes in search engines can sway elections and incite violence; irreproducible results can influence global economic policy for years; models based on biased data can legitimize and amplify racist policies in the criminal justice system; algorithmic hiring practices can silently and scalably violate equal opportunity laws, exposing companies to lawsuits and reinforcing the feedback loops that lead to lack of diversity. These problems are only exposed when delivering data science techniques into practice, and are therefore a natural fit for TDS. Significant efforts to date have been limited to the data mining and machine learning communities, where research has been focused on analyzing the fairness, accountability, and transparency properties of specific algorithms in specific contexts. While these efforts are certainly worthwhile, their effectiveness is necessarily limited because they focus on the final step in the data analysis lifecycle, disregarding the problems that can be introduced during dataset selection, cleaning, pre-processing, integration, and sharing. This area is not, however, solely a technical issue, but also needs to involve other approaches, including ethics, law, and social research to identify problems and opportunities and to suggest solutions. Important aspects of these solutions, such as policy and institutional approaches, may fall outside data science, but will require input from and the support of data scientists. Societal impacts must be evaluated not only for individual methods or the immediate impacts of a particular project, but also with a view of the data science lifecycle, which induces system-wide effects and feedback loops.

5.2 Challenge 2: Data quality. The quality of data that is being shared, acquired, integrated and analyzed is important in all TDS domains. Building on existing work in the database and statistics communities, it is necessary to develop methods that quantify and improve data quality, addressing (1) uncertainty due to errors and noise, (2) bias, representativeness, coverage, and sufficiency, and (3) domain knowledge and information about the data collection methodology. To enable support for quantifying and improving the quality of data in multi-stage data science pipelines, techniques must be developed to annotate base data, and to propagate annotations appropriately through lifecycle stages. These methods should be developed with general applicability in mind, and should be complemented with a methodology for translation into specific domains.

5.3 Challenge 3: Data profiling, triage, and cost planning. Early in a project’s lifecycle, and common across all application domains, is the need to “triage” datasets by their basic characteristics – volume, quality, freshness, variety, and more. These characteristics drive technology decisions and hiring decisions, and therefore strongly influence cost planning. As computing moves from the desktop to the cloud, cost planning becomes an important skill for any researcher in data-intensive science. The solutions and skills that are effective at 1 GB can be useless at 1 TB. A transportation engineering study may require real, rigorously clean data to produce relevant results, while synthetic or dirty data may be sufficient to evaluate a generic machine learning method. Some studies need fresh data, maybe even real time feeds, while others can work with a historical snapshot. A study involving the integration of 100 different small data sources can easily be 100 times as expensive as a study involving one large data source; heterogeneity of sources correlates with lines of code and number of special cases, and therefore costs. Tools and approaches for automatically or semi-automatically estimating project costs based on the size, complexity, and freshness of the data, along with estimates of required effort, would be valuable. A general set of best practices around data triage and cost planning with respect to scale, quality, freshness, and heterogeneity would have broad applicability in TDS.

5.4 Challenge 4: Data and model commons. It is important to make data, models, ontologies, and “entity” definitions available to the domain scientists. A commons should accommodate all types of data that are appropriate in the domain, and all methods to handle the data. For instance, in the legal domain, text as a data form will be predominant. In radiology, images will be key. Other data
forms of increasing prominence are audio (e.g., in social sciences and psychiatric counselling), video (e.g., in law enforcement), and non-tabular data structures (e.g., graph networks in biomedicine, transportation and social science).

The model commons will, in turn, inform what models work well for the (sub)domain, which in itself will be worthy of propagation. Data and models must be annotated appropriately to be discoverable. An important aspect of research under this challenge is to enable a data scientist to identify a model that was developed in another domain (or one that is “generic” – domain-agnostic), and to instantiate it in a specific application domain – translate or transfer the model. Support for model translation must incorporate methods to validate applicability of models to new data and in new settings.

This challenge also includes development and evaluation of domain-agnostic predictive analytics, of hypothesis generation and testing tools and methodologies, and of methods for validation and maintenance of the models during operational deployment.

5.5 Challenge 5: Data integration. Science has become increasingly integrative. Linking data from multiple data sources — spatial, sensor, financial, clinical, genomic — has become the norm rather than the exception, but the methods and technology to support data integration have not kept pace. Although data integration is a perennial research topic in the database field, there have been relatively few practical tools, and the focus has traditionally been on relational data only. Beyond tools to facilitate data integration, a better understanding of issues of quality, bias, and uncertainty introduced by data integration processes is needed. For example, decisions made about mappings between attributes or ontologies may involve tradeoffs between completeness and bias – records that fail to satisfy a mapping may be skewed. Visualization tools to debug and inspect the results of integrated data, exposing data quality issues, sources of bias, noise and uncertainty that would otherwise be masked by the integration process, are also motivated.

5.6 Challenge 6: Data-intensive social science. The further development of data-intensive social science is needed to address the social issues that arise in TDS. This can be addressed in multiple ways. The teaching of TDS would benefit from systematic analysis of the determinants of successful outcomes, a social science problem that could benefit from data-intensive techniques. The success of TDS research groups, even groups focused purely on the physical sciences, may be enhanced by systematic analysis of the determinants of the success of such groups. This situation is again a social science problem that could benefit from data-intensive approaches. Finally, many problems TDS groups seek to address are socio-technical and solutions require a strong grasp of social science approaches. For example, medical outcomes are substantially affected by patient behavior and choices. Addressing crime, education, poverty, and many other issues involves social science and may benefit from data intensive approaches.

5.7 Challenge 7: Interactive, convenient, scalable, and high-performance computing. Interactivity can be crucial for agile hypothesis generation and testing. In order to facilitate at-scale interactive data analytics, methods are needed to transparently scale computationally intensive data analytic methods, such as convolutional neural networks, to large processing platforms. This challenge encompasses both development of cloud and HPC tools, and methods for optimizing algorithm scalability, efficiency and performance. This challenge can also include mapping of machine learning/deep learning algorithms to novel computer architectures. Beyond existing research agendas in scalable computing, the TDS angle would emphasize direct interactivity by domain experts with limited computational expertise. Evaluation of these methods and tools may involve a combination of performance experiments, user studies, and design critiques.
The above listed challenges are those more commonly observed. Using viable definitions of translations, the various challenges can be met with useful and functional models. The following section describes a variety of models for translation.

6. Models and Methodologies for Translation

Translation of ideas into “tangible products” or “viable concepts” will occur beyond the mere generation of ideas. “Ideation” is celebrated in academia. While the ideas generated in the corridors of academia are often invigorating, they often do not lead to innovation or useful products. In essence, several of the ideas are not deemed worthy of translation. Models that include impact seem to be most relevant (Section 3). Successful examples of translation also point to impact. The earlier section (Section 5) lists the many challenges faced by practitioners of TDS. It should be noted that solutions proposed to address the challenges indeed can be assembled into a viable model or methodology of translation. These models can be based on meaningful definitions of translation, which in turn give rise to more definitions.

What is still is required is a more meta and summary view of translation that transcends elementary methodologies. A likely and successful working model of translation will include under its purview metrics and rubrics of impact and innovation and the identification of challenges and roadblocks for translation. It is also plausible that working models of translation make a case for the availability and value of copious data through agile and robust methodologies. Some viable models of translation are described below with some sharing common threads while others emphasize specific aspects given the contextual bias of the authors. The summary below does not attempt to weave them all together but presents commentaries obtained from position papers and presentations of various participants in a stream-of-consciousness style.

6.1 Models of medicine. To be considered translational, a proposed solution, a prototype, or a collection of ideas should lead to a complete and well-defined solution that result in a positive impact on the human condition. This view is often repeated throughout the document. Successful examples abound in medicine and certain definitions of translation have gained ample currency. A natural question that arises is “could the lessons learnt in medicine be applied to other disciplines and contexts?” An oft-stated hypothesis is whether translational data science as a widely-accepted discipline will thrive if the “best practices” from clinical science are adopted in disciplines across-the-board. For instance, genetics-driven bioinformatics workflows that utilize many well-known results from computer science and statistics have shown that it is quite possible to process complex data leading to meaningful and useful interpretation of results. Could this be adopted as a model?

6.2 Impactful models of medicine. Sharad Irani of the University of California, San Francisco contends that this could be done by creating Learning Health Systems (LHS). He postulates that LHS is a scalable way to bring the power of translational data sciences to medical practice and health maintenance, serving all segments of society. An LHS is thus a formal way of demonstrating a path for translation. It is suggested that medicine’s TDS agenda be focused on clinical data, including new forms that are emerging. He proposes a list of first and foremost research priorities, centered on clinical data, which in turn will enable a sequence of other priorities. Entity recognition, data quality, and natural language processing are especially fundamental to solve. Irani carefully lists a series of challenges and solutions thereof to address the various challenges. This should also reduce the “wastage” of historical clinical data that today goes untapped for research. Lastly, it is proposed that as an early step in their training, researchers should be educated to discern under which data range (enumerated below) their needs fall, to get a faster start to their research.
6.3 **Sum of many parts.** Joel Saltz of Stony Brook University certainly thinks so and defines the role of a translational data scientist using the clinical context. He considers TDS to be a multidisciplinary generalization of biomedical informatics. Saltz provides a concrete example from cancer biology in making his case, which entails unearthing the meaning of data within a whole slide image (WSI). The translational data scientist embraces the entire problem from understanding of a glass slide, to how the pathologist thinks about it, to how a cancer diagnosis is made. This information, when integrated with the patient’s genetics and the tumor’s genomics is critical to solving the question at-hand – how will a given patient respond to therapy, and what can we learn from the WSI about cancer detection, therapy, and outcomes? This comprehensive approach is essentially the model that Saltz espouses when he essentially posits that TDS is more than the sum of the parts by taking data across multiple areas of research and driving advances in science by bridging the chasm between data and knowledge.

6.4 **Data science environments.** Juliana Freire of NYU emphasizes the need to have data science environments (DSEs) at universities that facilitate translation and serve as a model. Such environments have been created at institutions such as NYU, UC Berkeley, and U. of Washington (http://msdse.org). The goal is to build infrastructure that produces new data science methods, which in turn accelerate data-intensive science in other domains. The DSEs are organized around working groups, which include space for collaboration, careers, software and tools, open science and reproducibility, and education. A key component of the DSEs is “people.” These institutions have hired a number of postdoctoral fellows and research scientists who conduct interdisciplinary research and are tasked with translating methods into tools that can be deployed by scientists to solve ‘real’ problems. Industrial and national labs offer other models for translation and it will be important to document best practices of successful efforts; however, Freire contends that the essence of translation remains.

6.5 **Collaborative team science.** Rama Vasudevan of the Oak Ridge National Laboratory provides a perspective derived from his experience as a researcher in the material sciences. A working model is proposed by meeting the prevailing challenges of unfamiliarity with statistics and computer algorithms and language barriers towards the widespread adoption of TDS in the physical sciences. Like Freire and Saltz, Vasudevan’s models of translation include individuals in multiple disciplines and also bridging organizations that serve to connect applied mathematicians and computer scientists with physicists, chemists, and biologists such as at the Institute for Functional Imaging of Materials (IFIM) at Oak Ridge National Laboratory. These umbrella organizations provide the necessary framework for establishing collaborations, incentivizing cross-discipline research, and greatly reducing barriers, not only due to physical proximity, but by enabling seminars, workshops, and tutorials in the respective areas. Vasudevan's model of translation also predicates the semi-automated analysis of data leading to a high degree of automation by dramatically reducing the time spent on routine analysis and maximizing attention on areas where new discoveries are likely to be made.

6.6 **Combining data production and use.** Julia Lane also of NYU proposes a two-pronged approach for data management and analysis that includes both technical and human aspects and serves as a medium and infrastructure for translation. She argues that an “if you build it they will come” approach is not sufficient. It is necessary to ensure that the interests of a disparate group of data providers are addressed. On the technical side, she contends that one can learn much from the successes and failures associated with the development of new types of research infrastructures. An integral part of this infrastructure should be a systematic approach to data stewardship in conjunction with data providers. Data providers will be looking for a way to control access, and they will need evidence that their data are both securely housed and being used for the purposes for which they were provided. But as well as providers, the technical infrastructure must be designed
to produce value for users. Researchers must be able to discover what data are available and how they can be used. Data sets could be searched and found on the basis of their use, rather than by the way they were produced. Lane proposes an institutionalized solution in the form of an initiative that would combine innovative program ideas, professionalised data centers, and a common set of interoperability standards.

6.7 Multidisciplinary structures. Chris Johnson of the University of Utah draws a parallel with computational sciences. In his opinion, there exists synergy between computational science and data science that he sees as two sides of the same coin, and posits that there is a significant overlap in expertise required and the methodologies used – computational science and data science are both rooted in solid foundations of mathematics and statistics, computer science, and domain knowledge. Indeed, many computational science competencies translate directly to the analysis of massive data sets at scale with high-end computing infrastructure. This view is reflected in NSF’s Computational and Data-Enabled Science and Engineering (CDESE) program, which includes all divisions within MPS, ENG, ACI, and CISE – a truly interdisciplinary program. As complementary interdisciplinary endeavors, both data science and computational science suffer from the entrapments created by disciplinary boundaries. To provide rigorous, multifaceted educational preparation for the growing ranks of computational scientists and data scientists needed to optimally advance scientific discovery and technological development in the years to come, universities will need to implement new multidisciplinary structures. In essence he makes a case for interdisciplinary teams much like Freire and Vasudevan.

6.8 Agile data collection. Raghu Machiraju of The Ohio State University contends that across-the-board translation therefore requires an approach where collecting data is the starting point and the mainstay. A first requirement of meaningful translation is the generation of data with a purpose that is often derived from the context. Thus, it will be beneficial to create ecosystems that reward purpose-driven data generation and collection from experiments and observations. A proposed model of translation is therefore an adoption of a simple baseline procedure that will be iteratively refined by collecting more and more data to drive the construction of methods and systems at every stage. In addition to developers and domain experts, either human or automated data gatherers should spur data collection at every stage of the method development. Additionally, the scope and extent of data gathering should also scale and change with the methods. In essence, agile data collection and generation using automated and semi-automated processes are required. Further, when combined with agile development methods, this synergy will create translational methods that have widespread impact to several stakeholders and not just a few.

6.9 Measures of data productivity. In the same vein, David Maier of Portland State University introduces the concept of data productivity. He defines data productivity as the ratio of the benefits obtained from data versus the investment in collecting or generating (and managing) data. Benefits can take the form of expanded scientific insight, better reliability, improved management decisions, or more effective policies. Resources include both money and people’s time. The working tenet is that any data science technology or methodology that does not increase data productivity is a poor candidate for translation to practice because it represents a negative return on investment. Strategies are provided for improving productivity that also decrease the cost of data collection, generation, and management. Further, guidelines are provided for collecting data more wisely and getting more value out of data.

6.10 True partnerships for synthesis. Rachael Croson of Michigan State University asserts that successful translational research requires true partnerships. Per Croson, one of the most common errors in translational work is Maslow’s Hammer (1966), which submits that “when you have a hammer every problem looks like a nail.” Successful translational work recognizes the existence
of nails, screws, and even paper clips. True partnerships enable each participant to understand the needs and constraints of the other parties, ensuring that the problem is formulated so as to be solvable, and that the solution is customized to fit the problem. She further states that successful translational research requires synthesis. Sometimes this synthesis can come from a collection of tolerant and open-minded experts from different arenas. Typically however, bringing individuals from different domains together is not sufficient to achieve synthesis. Projects either need structured group processes through which synergies are achieved, or participants whose roles are explicitly recognized as that of synthesizer.

6.11 Process-oriented characterization. Dan Green of the US Navy and the University of California San Diego provides the most expansive model and introduces the notion of the value of data, its transformation, and ensuing information. In fact, this model can be traced to the process-like definition of translation (Section 3). He argues that there is a need for a broader collaboration enabling us to translate data into action, to translate action into progress, and to translate progress into improvement of the human condition. The notion of TDS, therefore, becomes an applied science of Big Ideas, Big Challenges, and Big Results, not just Big Data. At the core of TDS is the ability to dynamically derive “expected value” from information diffusion. Attempting to predict future value rather than simply define outcomes helps orient our collective effort. It supports strategic planning, may lead to more ethical decisions, and helps create courses of action that become the “root cause” of beneficial future events.

In this framework, data and analytical products are treated as digital assets. Digital asset valuation might be considered along three information generating events or primitives: sensing, conceiving, and responding. If we conceive of these primitives in terms of inputs and outputs to nodes of a notional value stream, this results in nine event-pairs (sense-sense, conceive-respond, respond-sense, etc.) representing a generic input-output model. Orchestrating these transforms and translations into workflows create value chains. The expected value of information, in this context, becomes a function of the qualities associated with the workflow (i.e., time, cycle time, precision, accuracy, completeness) at each stage of the value-chain lifecycle. The integration of these value adding transforms reflects the measure of expected utility to the decision maker and can be mapped to the expected value of the data. Also, increasingly, information value propositions are influenced by sense and response patterns of behavior executed by non-human actors that dynamically intersect and influence human behavior and perceptions sometimes on a global scale.

In closing, TDS is an analytical discipline and suite of powerful new tools that provide us the opportunity to reflect on the nature of knowledge creation. Greens’ commentary provides some general themes such as digital assets, value functions, and action cycles to explore opportunities, articulate threats, define contributions, and encourage all to rethink the thinkable as we move forward with developing various communities of practice.

7. Educational Approaches to Translational Data Science

Education in TDS differs from a core data science curriculum in key ways. One important distinction is the focus on applying data science within a larger process. Data science is tightly focused on methods and approaches, whereas TDS education needs to develop and communicate how data can be applied broadly. This is not just tool-based learning of applications, but a consideration of how data science questions and applications become integrated and influence the broader research questions and task objectives we encounter.

Developing the capacity to undertake TDS requires both education in knowledge engineering, and integration of knowledge engineering principles into a curriculum. As a community, we need to
identify the shared components that allow knowledge to be transferred and the core language that allows fluent and effective communication among stakeholders. A common foundation or framework that integrates assessment of what is a reasonable result and an unreasonable result within an application. Foundational ways of how we think about and learn from data, upon which we can layer a method of analysis and visualization. Need to understand all aspects and components to the data lifecycle, which include: ethical concerns, stewardship, domain environments (what is an appropriate framework for one domain versus another), communication principles between domain experts and data science, sustainability, and feedback into the data science community. These curriculum components need to be integrated or re-visited or repeated across various courses. For instance, how might we re-engineer data science performance functions to account for a broader social utility function of bias or ethics?

Feedback into curriculum can operate through the integration of case studies and internships. Integration of case-studies of successes and failures, but also applications of various ways that TDS operates across the arts and humanities, the social sciences, manufacturing, and industry. These case-studies will help illustrate the role of different stakeholders in the translational process as well as how the nature of how protocols, language, and knowledge diffusion operate. Internships demonstrate a context in which TDS operates in practice or allows TDS to inform and change how data science is applied within domains as well as to curriculum.

Introduction of data reasoning or thinking with data as well as algorithmic thinking early into primary and secondary education. The differences between predictive inference and causal inference. How we develop a bridge from data findings to theoretical concepts or practical concepts. There is a critical need for team-based learning perspective and performance, an integration of various types of expertise within the TDS process. The data scientist should be exposed to the messiness of the data creation that may be under the domain expert’s role. We do not teach people to be translational data scientists, but rather, how different stakeholders can work together on a TDS process. That includes instruction on identifying roles, standard protocols, and creation of language that enable successful translation. Figure 2 captures various elements of this discussion.

To summarize, as it is currently being taught, data science advances the research agendas and resides in departments of mathematics, statistics, and computer science – and occasionally business. It favors algorithmic, computational, and statistical concepts and thinking. Translational data science needs to favor pragmatism, considering the value of the concepts and derivative algorithms and theories based on the likely success of their practical use. The education agenda for TDS needs to
Fit the work environment of its users – so teaching of translational data science demands specialization and tracks aligned with societal problems

Find ways to engage stakeholders – both by having them participate in curriculum development and by designing programs that can be delivered beyond traditional classroom settings.

Finally, it should be noted that curricula for data science are also being discussed in various forums. The recent report for undergraduate curriculum [PNAS-2017] describes the needs mostly from an academic perspective, while another discusses needs in general from a demand-supply market perspective [BHEF-PwC-2016] while clearly distinguishing the needs in data science from related disciplines of computer science, mathematics, and statistics. Such close scrutiny will also be required to establish curricula in TDS; forums like these can only start the conversation and even feed into the ongoing discussion over data science.

8. Deployment of TDS
In this section, a roadmap for the evolution of TDS is sketched. While not exhaustive, the points of discussion capture some of the prevailing thoughts of the participants of the Chicago workshop.

8.1. Growing a TDS community
Advances in method, analysis, processing, and algorithm development come from many disciplines; these advances are increasingly being applied in an equally diverse panoply of settings. Data science itself must thus be understood as existing within an ecology of institutions. Managing, describing, and sharing the large and diverse data sets that are characteristic of modern data science is best described as a cross-sectoral or cross-institutional, rather than individual, activity. Novel data science draws data together in order to establish plausible causality or demonstrate correlation, but it also provides consilience, which relies on the heterogeneity of data to produce new knowledge across otherwise unrelated sets of data [Wilson-1998]. An excellent example of data-laden consilience can be seen in the historical study of climate change, where disparate and heterogeneous data from ice core samples, sediment, historical documents, tree rings, and even paintings are brought together to understand the near- and long-term history of the global climate [Lamb-2002].

Translational data science promises to add more layers to the traditional data science stack, as knowledge moves beyond the walls of industrial and academic laboratories to be tested in the laboratory of society [Mitchell-2005]. This also means further refinements in how data-intensive science informs a broad swath of academic activity. TDS not only engages disparate, heterogeneous, cross-sectoral groups of stakeholders to work towards particular outcomes, it also potentially develops new tools, methods, and techniques for these stakeholders to better understand themselves and their processes. TDS as a tool is not limited to the more visible outcomes education, training, or even democratization efforts — [Lewis-2013; Hartzog-Selinger-2013; Espinoza-2014] — it provides tools for self-understanding at various scales (cf. [Choe-2014]) and can potentially open the imaginative space of new research in a great many domains.

8.2 Aspects of translation
In this section, we outline several possibilities for what TDS has the potential for becoming. Our practice-focused approach is driven by the fact that when we speak of the need ‘to translate’, we are using a verb – an action word. Above all, TDS implies action. A successful model for TDS also implies that its work will produce a change, be it in decision-making processes within an organization; the creation of new public policies; or a broader shift in our understandings about ourselves and the world in which we live. These can come about in a number of ways – for example
through data-intensive research, or in teaching, outreach and service activities oriented towards computationally-intensive methods.

8.2.1 Translation across domains and sectors
One mode of translation occurs across multiple scientific domains. Within the academy, disciplines learn from each other through the sharing and exchange of tools, techniques, and data. The techniques of large scale data analysis promise to illuminate fundamental social processes, motivated by existing theories about how people and organizations work. In addition, translation across scientific disciplines allows us to see outcomes even absent theorizing around fundamental processes. For example, data mining has allowed medical researchers to see where it would be useful to repurpose drugs, without necessarily understanding the mechanisms of action driving the drugs’ effectiveness; in turn, some of these very drugs can also be used to explore mechanisms driving tumor activity in the oncology domain [Vignola-Gagné-2017]. The importance of mechanisms and understanding processes varies across domains, influencing the kind of collaborations required to do data science well.

A second mode of translation involves translating knowledge from academia into practice. For example, the NSF Big Data Regional Innovation Hubs (BDHubs) initiative and the Data Science for Social Good (DSSG) programs translate knowledge outward from what are typically siloed academic disciplines and institutions, deploying the knowledge products of data science in nonprofits, private firms, and at multiple levels of government as a means for guiding more evidence-based decision making. The Grand Challenge-centered mission of BDHubs, and the demonstration ‘Spoke’ projects with which it is linked conduct research that aims to bridge data about health disparities and extreme weather events. Both can prove useful to industry, for example, in assessing supply chain needs, or to government in assessing the extent of drug use problems, or vulnerability to extreme weather events, and to nonprofits, which offer casework assistance or mobilize to make political claims.

A third mode of TDS involves importing both data and innovative data analyses conducted in industry, NGO, or governmental settings into the academy. Governments at all levels are currently assessing which data can be made publicly available; how to make it available; how to protect its most sensitive elements; and how it can be made useful. Investments in making data available, such as the Police Foundation’s Policing Data Initiative, seek to build partnerships with academics who define questions in conjunction with communities of interest in public policy.

Finally, nonprofits, industry, and government may share data or data science with each other, including databases, tools, analyses, and representations, without the involvement of academia. For example, the private sector is focused on generating revenue for shareholders while nonprofits must undertake work for the public good. When sharing data sets and analyses techniques across the private-nonprofit sector boundary, perhaps the carbon footprint or sustainability of land-based activities can be addressed while being mindful of the private sectors’ need to be profitable. In sum, TDS can be the translation of data sets, tools, methods and implicit knowledge from one academic domain to another; from the academy to other sectors (NGOs, industry, policy); from those external sectors back into the academy; and among the external sectors themselves.

8.2.2 Translation across the data lifecycle
In addition to cross-sector TDS, translation can occur across the data lifecycle, and these have different forms and effects. For example, the “data wrangling” process — cleaning and preparing data for use — is crucial to translation. The data not collected are also an occasion for translation, as the need to account for omitted variables will differ for different purposes. Once a data set is in
place, different ontologies can be used to represent it. Visualization experts develop representations of a pattern, model, and result. Organizations agree to use multiple data sets to address specific types of problems, like systems questions. This data lifecycle view implies a wide range of knowledge and skills for people working and collaborating in data analytics. It also implies that we need a wide range of cognate areas’ expertise to be effective.

Successful translation thus requires we address communication and interpretation challenges with relevant communities. People interpret information differently, conclusions are not obvious, analyses can yield no helpful results, and organizations quite reasonably discount conclusions that do not comport with their substantive experience in the field. Because data science works with data that include audio, text and video, core elements of TDS include understanding the aesthetics of information, communication, and the political and social environment in which the analysis is being provided.

Thus, TDS will require skills beyond those required in traditional data science. Conceiving of a particular data problem as translational implies that the techniques, tools, methods, and knowledge must be co-produced as part of a feedback loop involving mutual adjustment, negotiation, and conversation. Translation is not a ‘whole-cloth’ dumping of a data set and its accompanying analytic techniques onto an industry, but rather an adjustment of the data itself (as well as the methods by which it is processed and analyzed) to apply to a space that may not be part of its original conception. In sum, the lifecycle of translational data science is focused around its applications – it incorporates problem formulation, real-world impacts, and a selection of the best methodology, resources, and outputs to move from problem to solution.

9. Summary

In this report, working definitions, various challenges, and models of translation are defined. An educational curriculum is proposed that allows for a pervasive teaching of translational data science or TDS. Finally, a comprehensive roadmap for mechanisms of translation with broad applicability is provided.

TDS thus offers a path for data science to enter the mainstream through the practical methods and tools that its practitioners should create. The TDS community has the potential to bring data science to the mainstream with methods and tools that are accessible and comprehensible. In sum, TDS can provide pivotal insights into the science of data science itself (much like the Science of Science Policy program at the NSF).

What is amiss is more focus on TDS in sectors beyond the corridors of academia. Some of these aspects were dwelled upon but not examined in detail. Examples of translation from industry and nonprofit enterprises will be especially valuable. This is the purported goal of the Second Workshop of Translational Data Analytics to be held on the campus of University of California, Berkeley on November 13-14, 2017.

Finally, this report is to be discussed at the same workshop. The report will receive additional scrutiny and it is likely to result in recommendations for the third workshop to be held at New York University (NYU) and beyond.

9. Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant No. 1742814. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

10. References


[Qiello-2016] K. A. Qiello and O. Alter, *Platform-Independent Genome-Wide Pattern of DNA Copy-Number Alterations Predicting Astrocytoma Survival and Response to Treatment Revealed by the GSVD Formulated as a Comparative Spectral Decomposition*, Public Library of Science One 11, article e0164546 (October 2016).


[Vatsavai-2013] R. Vatsavai. *Gaussian Multiple instance learning approach for mapping the slums of the world using very high resolution imagery*. 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL August 11-14, 2013, pages 1491-1426


[Wilson-1998] TBD.
Appendix

Workshop Participants

<table>
<thead>
<tr>
<th>First</th>
<th>Last</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephanie</td>
<td>August</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>Chaitan</td>
<td>Baru</td>
<td>University of California San Diego</td>
</tr>
<tr>
<td>Tanya</td>
<td>Berger-Wolf</td>
<td>University of Illinois at Chicago</td>
</tr>
<tr>
<td>Alan</td>
<td>Blatecky</td>
<td>RTI International</td>
</tr>
<tr>
<td>Hillary</td>
<td>Child</td>
<td>The University of Chicago</td>
</tr>
<tr>
<td>Elaine</td>
<td>Collier</td>
<td>National Center for Advancing Translational Sciences</td>
</tr>
<tr>
<td>Melissa H</td>
<td>Cragin</td>
<td>University of Illinois at Urbana–Champaign</td>
</tr>
<tr>
<td>Rachel</td>
<td>Croson</td>
<td>Michigan State University</td>
</tr>
<tr>
<td>Michael</td>
<td>Franklin</td>
<td>The University of Chicago</td>
</tr>
<tr>
<td>Juliana</td>
<td>Freire</td>
<td>New York University</td>
</tr>
<tr>
<td>Matt</td>
<td>Gee</td>
<td>The University of Chicago</td>
</tr>
<tr>
<td>Dan</td>
<td>Green</td>
<td>US Navy / University of California San Diego</td>
</tr>
<tr>
<td>Robert</td>
<td>Grossman</td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Andrew</td>
<td>Hoffman</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Nick</td>
<td>Horton</td>
<td>Amherst College</td>
</tr>
<tr>
<td>Bill</td>
<td>Howe</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Sharat</td>
<td>Israni</td>
<td>University of California San Francisco</td>
</tr>
<tr>
<td>Chris</td>
<td>Johnson</td>
<td>University of Utah</td>
</tr>
<tr>
<td>Julia</td>
<td>Lane</td>
<td>New York University</td>
</tr>
<tr>
<td>Christopher</td>
<td>Lucas</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Raghu</td>
<td>Machiraju</td>
<td>Translational Data Analytics Institute</td>
</tr>
<tr>
<td>David</td>
<td>Maier</td>
<td>Portland State University</td>
</tr>
<tr>
<td>Tim</td>
<td>McKay</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>David</td>
<td>Mongeau</td>
<td>Translational Data Analytics Institute</td>
</tr>
<tr>
<td>Peter</td>
<td>Muhlberger</td>
<td>National Science Foundation</td>
</tr>
<tr>
<td>Fernando</td>
<td>Perez</td>
<td>The University of California Berkeley</td>
</tr>
<tr>
<td>Renata</td>
<td>Rawlings-Goss</td>
<td>Georgia Institute of Technology</td>
</tr>
<tr>
<td>Name</td>
<td>Last Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Joel</td>
<td>Saltz</td>
<td>Stony Brook University</td>
</tr>
<tr>
<td>Greg</td>
<td>Shannon</td>
<td>Carnegie Mellon University</td>
</tr>
<tr>
<td>Scott</td>
<td>Shearer</td>
<td>The Ohio State University</td>
</tr>
<tr>
<td>Steve</td>
<td>Slota</td>
<td>University of California Irvine</td>
</tr>
<tr>
<td>Cory</td>
<td>Smidt</td>
<td>Michigan State University</td>
</tr>
<tr>
<td>Susan</td>
<td>Sterett</td>
<td>Virginia Polytechnic Institute and State University</td>
</tr>
<tr>
<td>Julia</td>
<td>Stoyanovich</td>
<td>Drexel University</td>
</tr>
<tr>
<td>Ian</td>
<td>Thomas</td>
<td>RTI International</td>
</tr>
<tr>
<td>Rama</td>
<td>Vasudevan</td>
<td>Oak Ridge National Laboratory</td>
</tr>
<tr>
<td>Raju</td>
<td>Vatsavai</td>
<td>North Carolina State University</td>
</tr>
<tr>
<td>Elena</td>
<td>Zheleva</td>
<td>University of Illinois at Chicago</td>
</tr>
</tbody>
</table>