SIMPLEX

Performers selected for negotiations. Please note that this is not a guarantee that a contract will be awarded.

Prime organization *PI* *PI Email* Teammate, organization	**Proposal title**	**TA1 Efforts**	**TA2 Efforts**
American University *Michael Robinson* *michaelr@american.edu* Chris Capraro, SRC, Inc Cliff Josyln, PNNL	Conglomeration of Heterogeneous Content using Local Topology (CHCLT)	Develop a unified framework for global representation and modeling of hybrid information systems, integrating both quantitative and qualitative information, using the mathematical theory of sheaves. Project aims to provide a mathematical basis for all data fusion problems requiring local to global inference, independent of their source, nature, or data type.	N/A
Georgetown University *J.C. Smart* *smart@georgetown.edu* V.S. Subrahmanian, SentiMetrix, Inc.	AvesTerra: A Global Hypergraph Approach to Reducing Complexity for Accelerated Multi-Disciplinary Scientific Discovery	Theory and architecture of a distributed hypergraph-based approach to large-scale knowledge representation and analysis	N/A
American Museum of Natural History *Ward Wheeler* *wheeler@amnh.org* Daniel Janies, University of North Carolina at Charlotte	Integrating Linguistic, Ethnographic, and Genetic Information of Human Populations: Databases and Tools	N/A	The project proposes to use algorithmic complexity theory to analyze a diversity of data sets (cultural, linguistic, genomic) bearing on human populations and groups. The historical origins as well as social and biological relationships among groups will be summarized on GIS informed graphs of kinship.

Prime organization *PI* *PI Email* Teammate, organization	**Proposal title**	**TA1 Efforts**	**TA2 Efforts**
The Trustees of Columbia University in the City of New York *Liam Paninski* *lmp2107@columbia.edu*	Deciphering the cortex: Circuit inference from large-scale brain activity data	N/A	Inferring connectivity of biological neuronal networks from optical calcium imaging video data and electrophysiological data.
Massachusetts Institute of Technology *Wojciech Matusik* *wojciech@mit.edu*	Multi-Scale Representation and Translation for Complex, Heterogeneous Materials	N/A	We provide a multi-scale, data-driven material model that represents physical properties of complex heterogeneous materials that can be manufactured using new fabrication methods. We also provide a translation mechanism between this model and corresponding material structures. This allows us to compute and express a multi-dimensional gamut of material properties.
QuesTek Innovations LLC *Greg Olson* *golson@questek.com* Alok Choudhary, Northwestern University Mercouri Kanatzidis, Northwestern University Chris Wolverton, Northwestern University Ji-Cheng Zhao, Ohio State University Taylor Sparks, University of Utah	Data-driven Discovery of Novel Thermoelectric Materials	N/A	We will construct a database of thermoelectric materials properties, including data of various modalities, time-scales, length-scales, and dimensions. Machine learning tools will be developed to identify key physics and functional relationships from this dataset. The discovery and experimental verification of novel, high-efficiency thermoelectric materials is a practical goal of this project.

Prime organization *PI* *PI Email* Teammate, organization	**Proposal title**	**TA1 Efforts**	**TA2 Efforts**
Harvard University *Vahid Tarokh* *vahid@seas.harvard.edu*	Behavioral Approach to Simplified Learning from Space-Time Data	We will determine time-varying relationships between space-time data series, and will use those for detection, inference and prediction of the underlying phenomena. We will apply our methods to dynamic frequency allocation and scheduling of autonomous transmitters and receivers, and to financial data (prices of stocks, etc.)	N/A
Johns Hopkins University *Joshua T. Vogelstein* *jovo@jhu.edu*	From RAGs to Riches: Utilizing Richly Attributed Graphs to Reason from Heterogeneous Data	We will develop technology to convert multi-terabyte brain datasets into richly attributed graphs (RAGs), as well as theory and methods for performing statistical inference on such objects	We will develop the tools to analyze said billion vertex, trillion edge RAGs, including graph traversal and matrix operations.
Stanford University *Jure Leskovec* *jure@stanford.edu* Stephen Boyd, Stanford University Olivier Lichtarge, Baylor College of Medicine David Gleich, Purdue University	MINER: Multimodal Networks-A General Representational Language Applied to Bio-Medical Hypothesis Generation and Validation	In TA1, we will develop a mathematical framework, a working software system, and inference algorithms for multimodal networks, which are complex networks with a number of modes or layers. Nodes in these networks link inside the layer as well as to nodes at other layers.	In TA2, we will apply the methods and the tools from Tasks 1 and 2 to a scientific domain of drug discovery. We will build a single, large multimodal network, integrating heterogeneous biological knowledge. We will then harness the network and use knowledge discovery methods for generation of biological hypotheses with the ultimate goal of discovering new molecular interactions, gene functions, disease-causing genes and pathways, in cancer and in other diseases. These biological hypotheses will be validated by laboratory experiments.

Prime organization *PI* *PI Email* Teammate, organization	**Proposal title**	**TA1 Efforts**	**TA2 Efforts**
Stanford University *Chris Re* Gill Bejerano, Stanford University	The Stanford Resident (Reason-Syndicate-Identify) Project: Towards Knowledge-drivenThe Stanford Resident (Reason-Syndicate-Identify) Project: Towards Knowledge-driven Medical Genomics	Extend our DeepDive system (deepdive.stanford.edu) to handle representation of rich data structures and knowledge structures from Genomics. We plan to also extend the system to features involving signal processing and vision.	Construct an automated framework to build and update the world's best knowledgebase for linking human genomic mutations with their phenotypic and disease consequences. This "Genomic DeepDive" knowledgebase will be crucial to our ability to solve the imminent flood of patients genomes, which is just starting to take off.
University of California, Los Angeles *Song-Chun Zhu* *sczhu@stat.ucla.edu*	Learning Homogeneous Knowledge Representation from Heterogeneous Data for Quantitative and Qualitative Reasoning in Autonomy	Study a domain agnostic knowledge representation from to heterogeneous data: images, video, language, sound and force sensors etc.	Focus on the use case of robot autonomy and develop a unified and common knowledge representation for multi-tasks: Visual perception, Commonsense reasoning, Situated dialogue with humans, and Robot demonstrations
Vanderbilt University *Sandeep Neema* *sandeep.neema@vanderbilt.edu*	N/A	Development of modeling languages and model-based representation and transformation tools for knowledge representation, building on Vanderbilt University's Model Integrated Computing Toolsuite (GME/WebGME). Development of model transformation tools for datafication and tools for exploration of high-dimensional combinatorial/parametric search spaces.	N/A
Sotera Defense *Kathleen Lossau* *Kathleen.Lossau@soteradefense.com* Justin Gawrilow, Sotera Defense	Unified Datafication and Discovery Toolset (U-DADT)	Leverage and extend a suite of opensource tools that operate over the data to generate temporal analysis, graph clusters, co-occurrence, pairwise correlations, Bayesian networks, and other analytics. Provide a flexible open source data architecture for all TA1 and TA2 performers and provide support for all SIMPLEX performers to foster collaboration and integration of data, representations and tools.	