# STRATEGIC AND IMPLEMENTATION PLAN
## NOVEMBER 2016


## Center for Science of Information (CSoI)
## NSF Science and Technology Center - CCF-0939370

Lead Institution: Purdue University

Partnering Institutions: Bryn Mawr College
Howard University
Massachusetts Institute of Technology
Princeton University
Stanford University
Texas A&M University
University of California, Berkeley
University of California, San Diego
University of Hawaii
University of Illinois, Urbana-Champaign

# TABLE OF CONTENTS

# I. Mission

The Center for Science of Information (CSoI) was established on August 1, 2010 with Purdue University as the lead institution. The CSoI is comprised of ten partner institutions, each with unique and complementary strengths in research, education, and outreach. The CSoI is a research center sponsored by the National Science Foundation as a Science and Technology Center (STC Cooperative Agreement: CCF-0939370). The overarching vision of the CSoI is to develop *rigorous principles guiding the extraction, manipulation, and exchange of information*, integrating elements of *space, time, structure, semantics* and *context*. These principles are motivated by, and validated on applications drawn from various scientific, engineering, and socio-economic domains. The research and development mission of the Center is complimented by an education and outreach plan focused on training the next generation of students in this rapidly emerging discipline, significantly enhancing the diversity of students and researchers, and exposing them to novel concepts at the intersection of the science of information and its applications. This Strategic and Implementation Plan lays out a roadmap for CSoI to accomplish all of its stated goals.

## Science of Information STC Mission Statement:

*To advance science and technology through a new quantitative understanding of the representation, communication and processing of information in biological, physical, social and engineered systems.*

**Research Introduction**

Information is the basic commodity of our era – permeating every facet of our being. The foundations of modern data communication and storage, and its ancillary trillion dollar economic impact were laid in 1948 by Claude Shannon, who introduced a general mathematical theory of the inherent information content in data and its reliable communication in the presence of noise. While Shannon's Theory has profound impact, its application beyond storage and point-to-point communication (e.g., the Internet) poses fundamental challenges. The overarching vision of the CSoI is to develop a new science of information that incorporates common features generally associated with data/information that are not addressed by Shannon's Theory. It aims to develop a comprehensive science of information that includes definition of core theoretical principles, development of meters and methods based on these principles, and their application to important problems in diverse application domains. Beyond these core technical objectives, the Center offers:

- A venue for multi-disciplinary interactions and collaborations, while providing effective mechanisms for long-term, integrated scientific and technological research and education;
- A means for exploring effective ways to educate students and to train the next generation of researchers;
- Programs for broadening participation of underrepresented groups and to enhance exposure to deep and foundational problems in information sciences; and
- Mechanisms for timely transfer of advances in research to education and to the broader community.

CSoI brings together accomplished researchers from diverse disciplines (computer science, information theory, life sciences, chemistry, physics, statistics, economics, and social sciences) to gain a unique multidisciplinary perspective, and to develop solutions with significant broader impact.

Traditional formalisms associated with information do not adequately address several key aspects. By incorporating these elements into an integrated framework, the center fundamentally enhances the scope of these formalisms:

***Structure:*** Measures are needed for quantifying information embodied in geometric structures and networks (e.g., information in nanostructures, biochemical networks, social networks, networks of financial transactions). Often, these measures must account for associated context; and incorporate diverse (physical, social, economic) models of flow (dynamics).

***Time*:** Timely delivery of partial information often carries a higher premium than delayed delivery of complete information (e.g., real-time control systems, decision processes in financial markets). The notion of timeliness, however, is closely related to the semantics (is the signal critical), system state (is the system under stress), and the sender/ receiver.

***Space:*** In interacting systems, spatial localization often impacts information transfer (e.g., interference rates can be significantly reduced in wireless networks by localizing communication, cells perform vital functions by localizing chemical reactions to certain parts of the cell).

***Information and Control*:** In addition to delay-bandwidth tradeoffs, systems often allow modifications to underlying design patterns (e.g., network topology, power distribution, and routing). Information is exchanged in space and time for decision making, thus timeliness of information delivery along with reliability and complexity constitute basic objectives.

***Semantics:*** In many scientific contexts, one is interested in the role of signals, in the absence of precise knowledge of their semantics, (e.g., DNA sequences, spike trains between neurons, whale songs). These signals are known to convey information, but little more than that can be assumed a priority. Is there a general way to account for the meaning of signals in a given context?

***Dynamic Information*:** In a complex network, information is not just communicated, but also processed and generated (e.g., emergency response, stimuli). How can such considerations of dynamic sources be incorporated into an information-theoretic model?

***Learnable Information:*** Data-driven science has received considerable recent research attention. How much information can actually be extracted from a given data repository? Is there a general theory that provides natural model classes for (more structured) data? What is the cost of learning the model, and how does it compare to the cost of actually describing the data? Is there a scientific way to approach the problem of extracting relevant information?

***Resource Constraints:*** Our ability to communicate and manipulate information is often limited by available resources (e.g., computing devices, bandwidth of signaling channels). How much information can be extracted and processed with limited resources? This question relates to complexity and information, where different representations of the same data may vary dramatically when complexity is taken into account.

***Representation-invariant Information:*** How does one conclude whether two different data representations are information equivalent?

***Cooperation:*** Often subsystems may be in conflict (e.g., the problem of Byzantine generals, denial of service or selfish attacks in computer systems) or in collusion (e.g., price fixing, insider trades). How does cooperation/ collusion impact information?

The long-term goals associated with this challenging research agenda cannot be accomplished by any single group of researchers working in isolation. This is because most institutions (groups) lack a high level of expertise in all of the constitutive disciplines. The scope, scale, duration, and needed flexibility necessitate a multi-institution center embodied by CSoI.

## II: Integrated Research

Information provides the essential substrate and unifying theme for virtually all complex interacting systems. This is well-recognized in the context of various applications, more notably, in life sciences, modern communication, physical systems, social and financial systems, and economics, to mention a few. Understanding information flow, therefore, holds the key to comprehending and building more efficient systems. In view of this, the Center focuses its research around three application thrusts: **life sciences**, **communication**, and **knowledge extraction from massive datasets**.

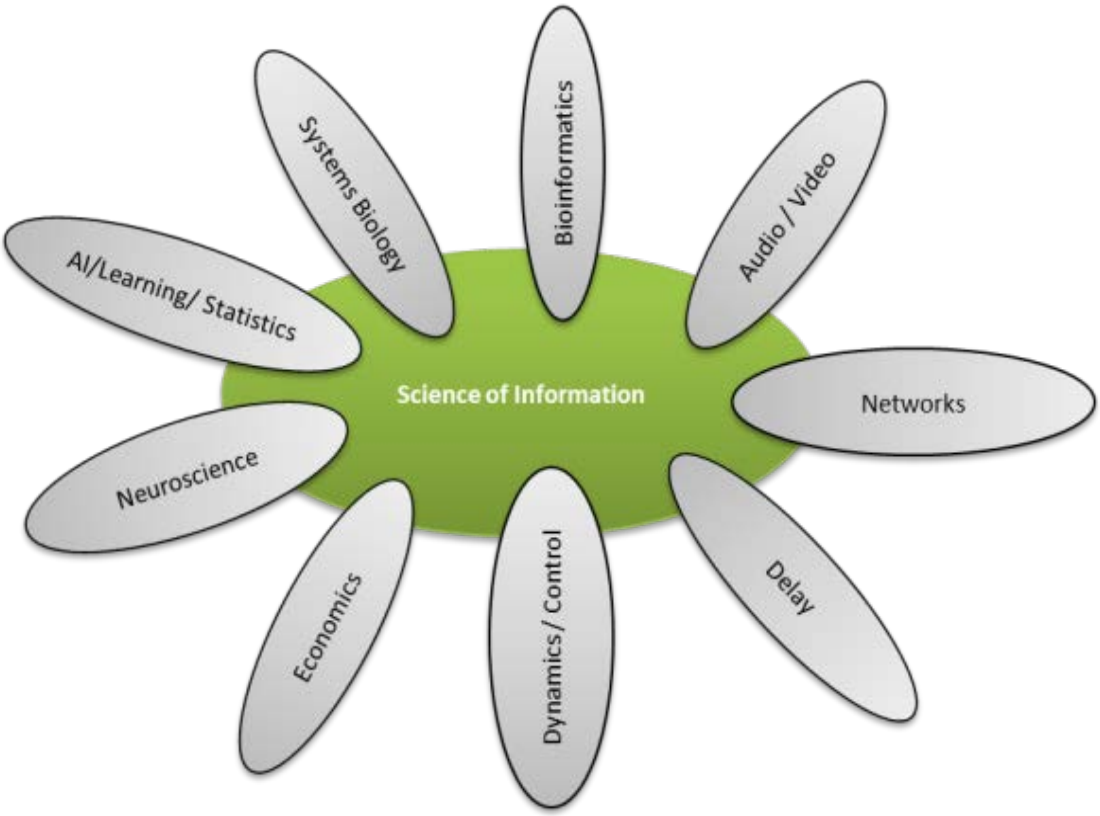



Figure 1: Intersection of Science of Information with other disciplines

Within these application thrusts, the Center investigates aspects of temporal, spatial, structural and contextual information. For example, in the study of DNA sequences and the associated proteins, while sequence-based methods for information extraction have been successful, there remains a need for appraising information in the structure of biomolecules (folds and misfolds in proteins, docking pairs), their interactions, functional annotation, and hierarchical (modular) organization. Paul Nurse, in the 2008 article in ``Nature" states ``*Focusing on information flow will help to understand better how cells and organisms work and ...the generation of spatial and temporal order*." In materials modeling, the information content of a crystalline structure is less than that of a polycrystalline material, which in turn is less than an amorphous material. How do we formally quantify this notion of information in structure, and what implication does this have for associated properties? These foundational questions motivate our study of structural, spatio-temporal and semantic facets of information in the presence of cooperation and dependence.

## Mission for Integrated Research:

*Create a shared intellectual space, integral to the Center's activities, providing a collaborative research environment that crosses disciplinary and institutional boundaries.*

Shannon laid the foundation of information theory, demonstrating that problems of information communication can be precisely modeled, formulated, and analyzed. He also provided basic mathematical tools for addressing these problems. Shannon's focus on what is fundamental, and his precise quantitative analyses, continue to motivate and inspire. In the current world, however, information is not merely communicated; it is also acquired, curated, suitably abstracted and represented, aggregated, analyzed, retrieved, inferred, secured, and used in various scientific, engineering, and socio-economic processes. A comprehensive Science of Information that fundamentally builds on Shannon's basic principles to address key challenges in transforming data to information to knowledge is critically needed (see Figure 2).

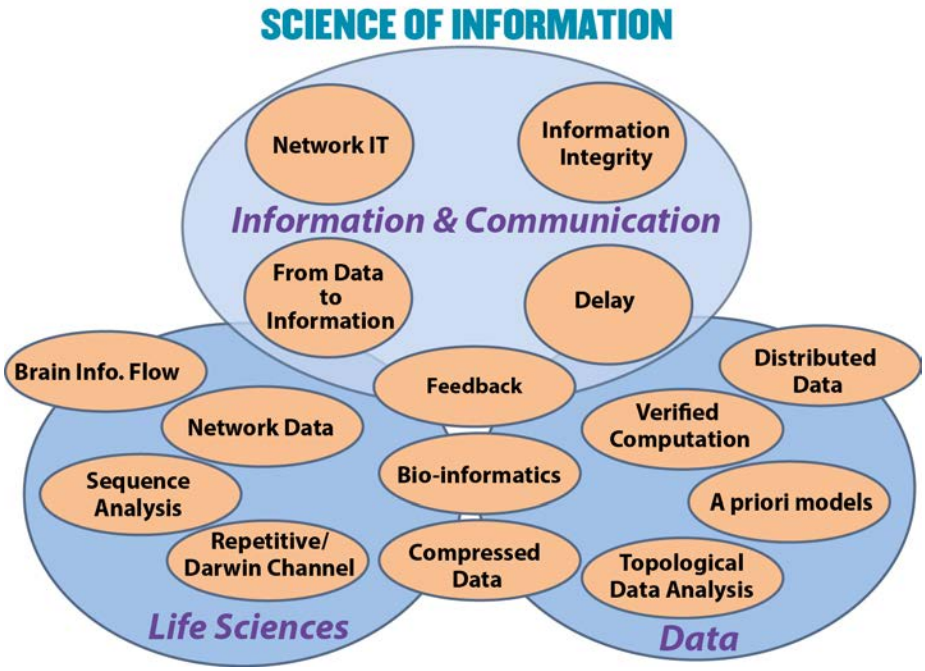



Figure 2: Overarching Science of Information

Our approach to developing the Science of Information uses a broad class of applications to precisely formulate problems. For instance, challenges in modern communications, biological processes, and socioeconomic systems motivate the need for modeling spatial information. A signal (in the form of a biomolecule) must arrive at the appropriate tissue at the corresponding receptor on the cell membrane at the right time. A control signal in a cyber-physical system must arrive at the corresponding actuator. These applications also motivate temporal information to meet strong real-time constraints of systems. In yet other applications it is essential to deal with challenges of incompleteness and dimensionality, along with considerations of privacy and obfuscation. Other challenges include semantics of information, learnable information, impact of cooperation, and constraints on resources. Following the principles of Shannon, the Center focuses on quantitative formulations of these problems with the goal of establishing fundamental limits and deriving methods with provable performance.

Our research plan follows this methodology—the emphasis on core principles aims to develop models and methods (e.g., structure, flow of information, temporal aspects, and value of information) that are relevant to a broad range of applications. These core principles provide bridging abstractions across applications in communication, data generation and data analysis, and life science. Information theoretic principles underlying these diverse applications have rich synergy, strongly motivating an integrated approach to these problems.

Creating a shared intellectual space that facilitates discussions, brainstorming, and problem formulation across disciplinary and institutional boundaries is essential to the Center's success. Creation of this space requires multiple components, including informing PIs about the disciplinary and interdisciplinary expertise and interests of other Center PIs, creating and disseminating pedagogical resources such as tutorial documents and seminars; creating a virtual collaborative workspace via the Center's website, and proactive planning and execution of workshops and brainstorming sessions. These sessions will be a core facilitator for formulating research problems for interdisciplinary teams to coalesce around. We envision two main mechanisms around which Center workshops will be formed: Regional and Opportunistic. We also plan annual meetings to report our progress and to create further opportunities for cross-disciplinary advancement.

Regional Workshops: The institutions associated with the Center coalesce in three main geographic regions: East Coast, Midwest, and West Coast. Each region, via the executive committee members and research thrust leaders at each institution, will plan 1-2 half or full-day workshops per year around specific interdisciplinary topics. PIs from outside the region will also be invited to attend. There will be a part of the Center website dedicated to these workshops, which can be used for both planning as well as reporting the outcome, in particular any problem formulations that were discussed during the meeting. Additional input on the workshop outcome will be solicited from all Center participants. PI meetings will also provide an opportunity for discussing these workshop outcomes. PIs will be expected to participate in at least one of these regional workshops per year. Workshops were held in 2013/2014 at UC Berkeley, MIT, Hawaii, Texas A&M, Purdue, San Diego, Howard and MIT.

Opportunistic Workshops: Many Center PIs attend the same international meetings, which is an opportunity to carve out some time during these meetings for brainstorming sessions or workshops associated with the Center. If the meeting takes place close to one of the Center institutions, then PIs from that Institution should also be invited to attend. The agenda for these brainstorming sessions or workshops should be set by the participants a few weeks in advance and shared with everyone via the Center website, where the workshop outcome and problem formulations that result should also be reported.

We propose a host of other activities such as exchange visits, joint student supervision, and a systematic dissemination of cross-disciplinary solution techniques. We believe that the latter, along with feedback, is at least as important as the tools and skill sets we may develop in the process of crossing barriers across disciplines. The dissemination process will be furthered by video records of presentations, a thesaurus and/or supplements (e.g., data sets) to foster model learning, a wiki space for questions and answers, and more standard publication and presentation methodologies. The process of building and disseminating such exemplars is meant to accelerate cross-disciplinary activity and thus enhance our integrative research through a learn-by-example approach.

While our detailed research plan for the next 5 years is described in the renewal proposal, below we highlight some aspects of our research objectives.

## A. LIFE SCIENCES THRUST

Applications in life sciences are primary research drivers for the Center for Science of Information. Given the ever-expanding repository of diverse data-sets, the complexity of underlying processes, and the importance of spatio-temporal context, life sciences applications serve as arguably the most interesting and challenging test-beds for models and methods. Broadly, challenges in life sciences targeted by the Center may be viewed in four categories.

1. Knowledge extraction from data

a. Integrating diverse datasets
b. Defining the granularity of data
c. Statistical methods with regularization
d. Biology-constrained methods
e. Information metrics
f. Dealing with context.

2. Dealing with noise in data
    a. Robustness of knowledge extraction to noise
    b. How to deal with missing data?

3. Classification of and modularity from data
    a. Specification and identification of modules (functional, spatial, temporal, etc.) from data
    b. Quantifying information content of modules
    c. Quantitative and qualitative comparison of modules.

4. Dealing with dynamical data
    a. How to deal with multivariate and high dimensional time series data?
    b. Understanding spatio-temporal information processing in systems
    c. Identifying suitable granularity and context for analyzing data.

To address these issues, we have several well-defined data sets that vary in time and context. The exemplar data sets that will be used include data on macrophage cells (http://www.lipidmaps.org), data from B-cells (http://www.signalinggateway.org) and data from cancer patients (NCI website). We address below a few problems that will be addressed in the first two years.

*Dynamical Data*: A key problem in life sciences is the development of models at different granularity from time series measurements of cellular constituents such as proteins, nucleic acids, and metabolites and phenotypes such as gene expression profiles, cellular proliferation, and cellular death. From cellular component measurements at different time instances following a stimulus, it would be desirable to build a biochemical pathway model. Such models may be correlative or causal and can contain myriad nodes and edges. If one were to consider modules that are varying with time, hypergraphs can be constructed based on a correlation metric or interaction data. These hypergraphs provide a glimpse of the dynamics of the system. However, it would desirable to convert these hypergraphs into necessary and sufficient models to quantitatively describe the cellular phenotypes. This is a major challenge for the Center.

*Many-to-many Network and Biochemical Pathways*: Shannon's methods deal with point to point interaction or communication. However all biological systems are many-point-to-many-point communications and there are no algorithms for understanding the information complexity of this system. We will develop methods to pose the following questions. What are minimal networks that will provide quantitative information on phenotypes? What is the sensitivity of different connections for a given phenotype? Entirely new methods need to be developed to address this problem.

*Modularity in Networks*: We will develop algorithms for deciphering modularity in systems. Amongst the interaction networks, biologists have painstaking identified cliques that have relevance for chosen phenotypes. However, there are few methods that can a priori predict modules in networks. One quest of this center will be to identify modules from complex pathways.

*Genome Encoding and Evolution*: A large fraction of the Human Genome codes for gene expression control during the life time of an organism. Driven by exciting new technologies, the field of Genomics is now beginning to decipher the language of gene control. This process holds many challenges related to Information Theory. At a pragmatic level, it requires the integration of large amounts of heterogeneous, noisy and missing data, which nonetheless describe the action of robust networks. There are also fascinating questions of classification and identification of the different functional components of the regulatory networks. Also, by comparing the genomes of different individuals and different species we

stand to learn about modes of information transmission through the generations. In many ways the genome is the ultimate information repository, and using Information Theory to better understand it is a major challenge.

## B. COMMUNICATION THRUST

The communication thrust is a direct extension of Shannon point-to-point communication to a networked environment. Over the next two years the Center focuses on integrating elements of delay in information theory, information in networks, new measures in information, and the so called non-asymptotic limits. We elaborate on these goals below:

*Delay in Information Theory:* Quantifying the performance limits for coding systems complying with various restrictions on their operations, including causality, delay constraints, memory and complexity constraints, and the availability of feedback and of side information. We aim to understand the performance attainable under various combinations of such restrictions. Do the fundamental limits here give rise to measures of uncertainty and dependence other than entropy and mutual information? Among others, the Center aims to develop a better understanding of:

- Quantifying the temporal value of information
- Information theory for finite block lengths
- Tradeoffs between delay, distortion, and reliability in feedback systems.

Such understanding will fundamentally influence the design of practical systems.

*Information in networks*: Studies at the Center are aimed at understanding communication between agents, subject to a variety of different assumptions on communication. Agents may also interact with the physical world. The configuration may be dynamic: agents may move, and their communication capabilities and external interactions may change over time. In such scenarios, one needs to understand communication, computation, and coordination among the agents. How does information flow through a distributed network? What is the type and amount of information necessary to solve distributed tasks in dynamic networks? These investigations will lead to new classes of problems relating to scaling laws and the microscopic-macroscopic transition.

*Information and computation*: The distributed and increasingly interconnected nature of information leads to new challenges in the study of information networks and interactive computation. Most information today does not lie in any single location in the network, is often replicated and sometimes inconsistent. And the spatial location of information is itself important to establishing its context and utility. These novel aspects lead to a host of new problems. We list some of these problems which we plan to attack in the next two years:

- Quantifying fundamental limits of in-network computation, and the computing capacity of networks for different functions
- Complexity of distributed computation in wireless and wired networks
- Information theoretic study of aggregation for scalable query processing in distributed databases
- Fundamental limits of interactive communication.
- Applications and implications (e.g., for collaborative computation).

*New measures and notions of Information*: In order to solve aforementioned problems, the Center aims to develop new measures of information that quantify among others:

- *Soft-information* (beliefs) in rate distortion theory
- *Semantics* in information: framework, probabilistic modeling.
- Modern communication networks, such as the Internet and its overlaid networks (e.g. social networks) pose several new and important kinds of questions relating to what kind of ``information'' is carried over them. Examples include notions of *structural information* such as the information carried in the routes used by the flows

(which it is often of interest to hide) and the information carried in patterns of behavior (which it is often of interest to extract, e.g. for targeted advertising).

*Interface with Life Sciences Thrust*: The Center engenders a tight collaboration between the life sciences and communication thrusts. On the communication side, the Center addresses the following problems relevant to Life Sciences:

- Furthering our information theoretic understanding of deletion, substitution, and insertion channels.
- Information theoretic models for evolution
- Models for stimuli
- Communication models for intra-neuron signaling.
- Models to predict the behavior of various systems, ranging from intra-cellular signaling, to tissues, individuals, colonies, and ecosystems.

**C. KNOWLEDGE THRUST**

The Center targets two broad and fundamental areas of knowledge management, motivated by three transformative application domains.

*Information Science for Collaborative Computing and Inference*: In many applications, high-value data is distributed among parties that share some common goals and have some individual goals. There are important questions involving what data to share and who to share it with to accomplish desired tasks. These issues are particularly important in the face of limited resources such as time, power, and bandwidth, and other considerations such as privacy and security. We will explore fundamental problems in distributed inference and collaborative computing, and particularly the role of information in these tasks.

Often parties may be reluctant to share information, even though all would gain from collaboratively computing using everyone's data. The reluctance to share can be quite rational if the drawbacks of revealing one's private, proprietary information, and the loss of control against its further dissemination and misuse, can outweigh the benefits gained from sharing private information. Quantification of the information gained and the private information leaked, would enable rational cost-benefit analysis by potential collaboration participants. In the absence of this, risk aversion dominates, and many potential "win-win" collaborations may not take place. One major challenge in this endeavor is the *impact of time* – the time-value of information versus the time to compute it (e.g., a data disclosure may be harmless if computing the confidential information from the disclosed data takes long enough). A second major challenge is *mitigation* – perturbing the disclosed data to protect private and confidential information, without damaging its usefulness for the purpose of collaborative computing and inference. A third challenge is *quantifying* the mitigation afforded by secure multiparty computation protocols, which makes possible "computing with data without knowing it" yet must inherently leak the information that can be inferred from knowing one's own inputs and the computed outputs.

In addition to computing and inference, another fundamental challenge we will explore are methods to *summarize* complex or *high dimensional* datasets, for example nonlinear dimensionality reduction and various techniques for making complex datasets easy to interpret (data visualization). This is particularly important in many of the applications that will be investigated (e.g., biology, economics, social networks, environmental modeling).

*Semantic, Goal-Oriented, and Communication*: One of the goals of the Center is to propose a modern theory which integrates *computing and communication* right from the start. Such a theory would attempt to formalize the "problems" that devices attempt to solve by communicating, i.e., the goals of communication. By then focusing on these goals, we hope that efficiency and reliability measures can be proposed that allow various solutions to be analyzed rigorously and compared quantitatively.

Such a theory is already in the nascent stages in the works of co-PI Sudan, postdoc Juba, and their collaborators. Explicit tasks for the coming years would include articulating the goals in some of the common communication tasks; proposing

concrete measures by which they should be evaluated; and some explicit analyses of some protocols under these measures.

*Economics and Information Theory*: Much of modern dynamic economic theory formulates models by examining how continuously optimizing agents will interact in markets. This has been important in allowing consistent treatment of economic behavior, but the models postulate continuous optimization, implying very rapid responses to policy changes and to market signals, whereas actual behavior is more sluggish. Approaches to address this (e.g., by postulating "adjustment costs") have an ad hoc flavor and are not grounded in direct microeconomic observations. The existing "rational expectations" theories with continuous optimization imply infinite mutual information, in Shannon's sense, between the stochastic process for market signals and the stochastic process of a person's action. At least qualitatively, recognizing that this rate of information flow must be finite explains a broad array of observed facts about economic behavior that has in the past been explained with ad hoc postulates of inertia or adjustment costs. Our work will attempt to integrate a formal information-theoretic approach into dynamic economic theory. This seems to be a promising avenue for both explaining observations and improving the formulation of economic policy.

*Learning and Inference in Networks*: In order to model decision-making and behavior in networks, it is important to be able to efficiently estimate joint distributions over possible *network structures* and accurately assess the significance of *discovered patterns*. For example, one network mining task is to estimate the joint distribution of node attributes (e.g., the political views of users in Facebook) conditioned on the network structure, modeling dependencies among neighboring nodes (e.g., similar political views among friends). The resulting distribution is useful to jointly predict the unknown features of nodes in a network, exploiting dependencies among nodes to improve predictions. While there are some recently developed methods for this problem, little is known about the theoretical foundations of these methods or of the underlying estimation problem. Another fundamental problem is to estimate probability distributions over the *graph structures* themselves. Accurate estimation can improve understanding of the underlying network generation process and is a necessary precursor for anomaly detection in network activity graphs (e.g., intrusion and fraud detection). Current methods result in estimated models that fail to capture the natural variability of real world social network domains. These and other foundational problems in social networks will be pursued.

| Integrative Research Goal Statement | | |
|---|---|---|
| *Create a shared intellectual space, integral to the Center's activities, providing a collaborative research environment that crosses disciplinary and institutional boundaries.* | | |
| **Objectives & Metrics** | | |
| Formulate research problems for interdisciplinary team to investigate. (ongoing) | | |
| Refine existing and identify new grand challenge problems and share throughout Center (ongoing) | | |
| Investigator exchange visits for immersive activity (ongoing) | | |
| | | |
| Expand existing collaborations and initiate new collaborations through joint supervision, student exchange, joint publication, and presentations. | | |
| Development of two pedagogical resources (e.g. books, survey, papers, lecture series) at the interface of applications and theory. Special Issue of IEEE Journal by spring 2015 and also "Foundations and Trends" | | |
| Development proposals for sustained external funding within five years. | | |
| **Creation of a shared intellectual space for the development of the research priorities.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Record and post all talks, seminars, and presentations to website. | Ongoing | M. Atwell, B. Ladd, B. Brown |
| Share grand challenge problems on website. | January 2015 | Research Thrust Leaders |
| **Bridge the language barrier that divides different research disciplines.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Develop a web space for immersive tools and applications - commonly used in the different disciplines. | Ongoing | Set-up: M. Atwell<br>Population: ALL |
| Create and maintain and integrate web space for expert question and answer interaction. | Ongoing | M. Atwell, B. Ladd, B. Brown, K. Andronicos |
| **Extend beyond our comfort zone in order to take risks and experiment with new collaborations and possibilities.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Maintain website of research interests and expertise of all Center participants in order to identify potential collaborations. | 2010 - ongoing | M. Atwell, B. Brown |
| Make available datasets for the development of theoretical methods. | 2011 and Ongoing | S. Subramaniam, Research Thrust Leaders |
| Pair up senior investigators with junior faculty to provide mentoring. | Ongoing | Executive Committee and Research Thrust Leaders |
| Initiate new collaborations and expand existing collaborations through joint supervision, student exchange, joint publications, and/or presentations. | 2012 - ongoing | Executive Committee and Research Thrust Leaders |
| Formulate research problems for interdisciplinary teams to coalesce around. | 2011 - ongoing | Research Thrust Leaders |
| **Secure additional funding/collaborations for interdisciplinary activities in Science of Information.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Invite participants from group/agencies, labs, industry, etc. to collaborate and/or initiate new research. | 2011 - ongoing | W. Szpankowski, Executive Committee, B. Brown |
| Develop joint proposals for external funding. | 2012 and ongoing | Executive Committee, Research Thrust Leaders, coordinator |

## III: Education & Diversity

**Vision:** Our long term vision is to educate the next generation of scientists in various STEM disciplines with the core fundamentals, methods, algorithmic thinking, and understanding of applications in the Science of Information. Moreover, CSoI will be a leader in increasing diversity among students, scientists and the STEM workforce for disciplines related to the study of the science of information.

## Education and Diversity Mission Statement:

*Integrate cutting-edge, multidisciplinary research and education efforts across the Center to advance the training and diversity of the work force*

**Education and Diversity Management Structure:** The management structure for the planning, coordination, and execution of all education and diversity programs is comprised of the following: Director of Education (Brent T. Ladd, Purdue), Director of Diversity (Kelly Andronicos, Purdue) and the following Center faculty: Associate Director for Education & Diversity (Deepak Kumar, Bryn Mawr), Assistant Director for Education (Mark D. Ward, Purdue), and Assistant Director for Diversity (Todd Coleman, UIUC).

### ***Education Strategic Plan***

#### *Education Mission Statement:*

Advance interdisciplinary understanding, knowledge, skills of analysis, and application of science of information techniques and tools to problems across multiple domains at both undergraduate and graduate levels.

#### *Education Goals:*

The education program has developed an ***Information Frontiers Initiative*** that successfully supports and integrates the research mission of the Center with a long term vision of: (1) developing the next generation of scientists who will continue to strengthen the community of science of information and solve grand challenge information problems, and (2) a set of courses, modules, and teaching resources available to all that provides information literacy at foundational and advanced levels.

A cornerstone of the Center is the effort to increase our understanding of the process from data to information to knowledge. This effort guides the Center's research objectives. Correspondingly, the Center's education mission and goals integrate with this effort; the courses our faculty teach, the online modules developed, education activities offered, and the research training and mentorship of students all provide a strong bridge between research and education. **Two primary goals support our *Information Frontiers Initiative* and educational mission:**

***Build a community of practice in the science of information.*** The Center's education program has created activities where students can enhance their knowledge, learn new methods and tools, interact with peers and faculty, and collaborate in teams leading to interdisciplinary experiences. We develop activities that that increase their professional and technical capacity adding significant value to students' departmental experience.

***Develop courses and modules supporting the emerging field.*** We support and coordinate new courses and modules, both classroom-based and online offerings guided by the Center's theme of data to information to knowledge. The purpose is to encourage science of information literacy and skills across undergraduate and graduate levels at a diverse set of institutions. Central to this effort are faculty members translating research into the classroom by enhancing their courses with new topics that result from Center research. These courses and modules provide training for our core students, students outside the typical domains, and those participating in the Center's diversity program.

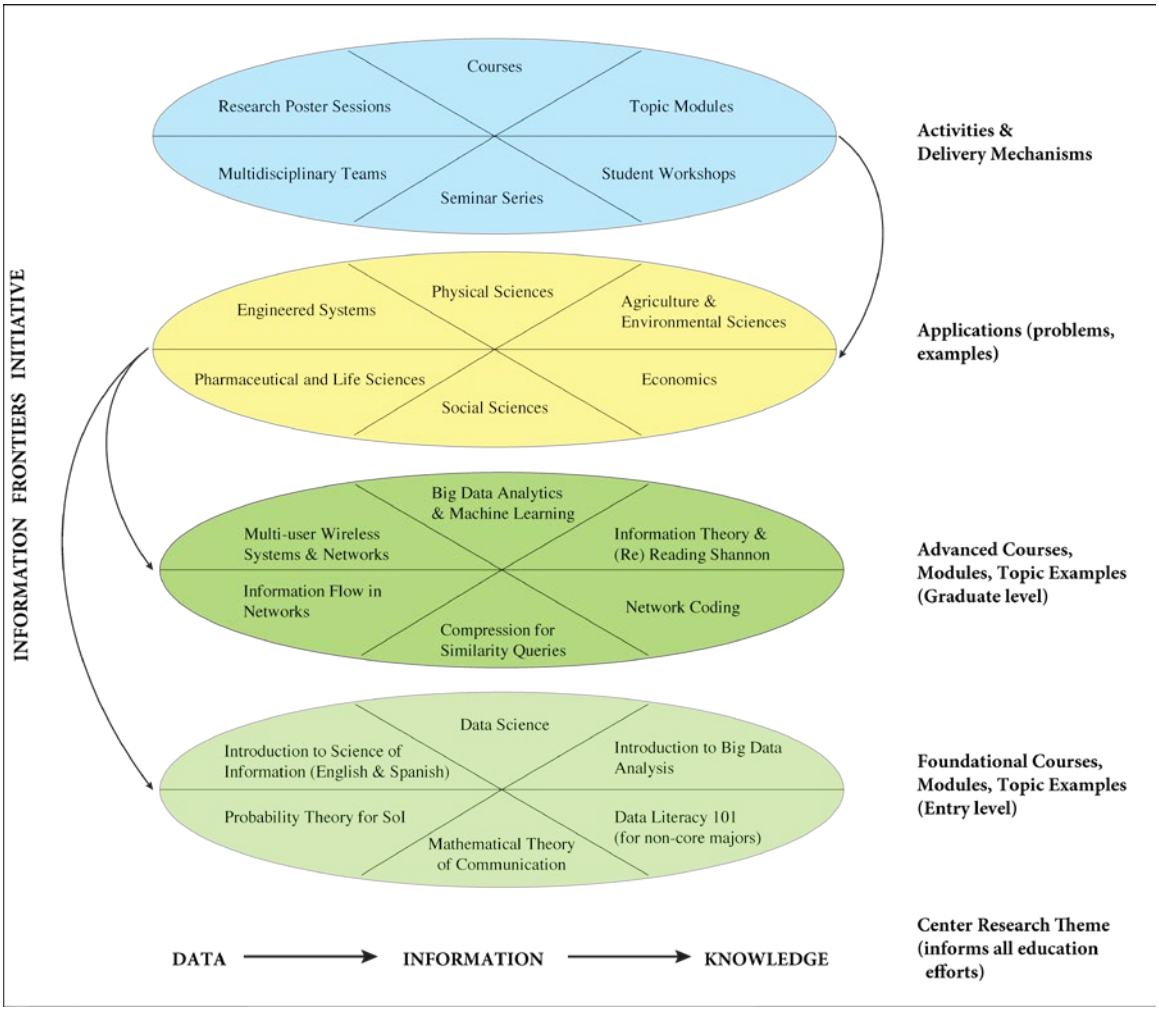



***Figure 1. Information Frontiers initiative.*** *Foundational and advanced topics are informed by Center research; information problems in many domains influence these topics, while Center activities lead to knowledge exchange, new skills, tools, and methods strengthening the field, and then applied to other domains by students.*

**Education Indicators and Metrics**

A logic model for education has been used to chart overarching metrics against expected quantitative outcomes. Additional specific evaluations of student workshops, and an extensive evaluation of new courses are undertaken in period 6 annual report. Qualitative findings are presented as student feedback as part of activity and yearlong student assessments and surveys.

| Activity | Expected Outcomes | Measures | Targets | Timeline |
|---|---|---|---|---|
| **Goal 1: Build a community of practice in the science of information** | | | | |
| Students participate and collaborate in Center activities | Students from a broad range of universities participate in activities | Number of students and universities participating in education activities | At least 300 students and 25 universities participate each period | August 2011 and ongoing. |
| | Center workshops are developed and offered | Number of workshops, schools, SoI days offered | Offer 2-3 student workshops and coordinate with IT society on 1 school each period. | July 2012 and ongoing |

| | | | | |
|---|---|---|---|---|
| | Student teams form around interdisciplinary problems, and produce solutions | Number of student interdisciplinary teams, number of posters, presentations, papers | Add 1-2 new collaborative teams each period | September 2012 and ongoing |
| | A growing percentage of students collaborate with faculty and students beyond their major professor | Percentage of Center students collaborating with Center members beyond their major professor. | 50% of Center students report collaborating with other students, postdocs and faculty in the Center on posters, presentations, papers, and research. | August 2010 and ongoing |
| Students present their work to peers and faculty | Research by students shared in the Center and presented among broader audiences outside the Center | Number of posters and presentations | Students present at least 100 posters and presentations each period. | August 2011 and ongoing |
| | Quality and impact of posters and presentations in the broader community is recognized | Number of awards/recognitions for student research | At least 5 students receive significant awards for their research (best paper, best poster) | August 2011 and ongoing |
| Students publish their work | Students publish their research in journals and conference proceedings | Number of Center students and postdocs publishing<br>Number of conference and journal publications | At least 50 Center students publish their work each period. At least 100 publications at conferences and in journals each period | August 2011 and ongoing |
| Alumni attain industry and academic positions in the U.S. | Alumni continue working in science of information areas in both academia and industry | Number of alumni of CSoI continuing in academia, industry, nonprofits in the U.S. | 95% of our graduates find employment in academia, industry, non-profit. | August 2011 and ongoing |
| **Goal 2: Develop courses and modules supporting the emerging field** | | | | |
| Use<br>Center web portals for delivery of content | Center web site and learning hub (soihub.org/learninghub. soihub.org) are used worldwide by students and faculty to access content | Views and downloads of content<br>Number of U.S. states and countries where educational content accessed | Each period at least 100,000 page views from users and multimedia. All 50 U.S. states, and at least 100 countries represented. | August 2011 and ongoing |
| Develop and offer courses & modules online and classroom-based | New courses/modules, and established courses with new content are developed to reach both undergraduate and graduate students | Number of courses and modules developed<br>Faculty involved<br>Universities involved<br>Students enrolled across all courses and modules<br>Video tutorials watched online | 40 total courses with at least 1,000 students enrolled:<br>20 new courses developed<br>20 existing courses with new material<br>30 faculty involved<br>20 universities involved<br>10,000 views of video tutorials online | August 2011 and ongoing |

***Diversity Strategic Plan***

***Diversity Mission Statement:***

To increase the participation of women and underrepresented minorities in the Science of Information by providing opportunities for research, mentoring, and networking in order to build a more diverse, dynamic workforce.

***Diversity Goals:***

Our diversity program is strongly integrated with the Center's research and education agendas. It builds on the strength of the Center infrastructure, to realize our vision of building a more diverse, dynamic workforce in science of information. The Center's goals are to increase participation of underrepresented groups in this emerging discipline, by:

- Providing funded research and mentoring experiences with the field's top faculty at the most prestigious schools, and by
- Opening networking channels through educational programming, bringing students and faculty together in ways that improve the field's culture and climate.

It is our belief that involving the widest possible breadth of backgrounds, experiences and ideas enriches science and contributes to increasing the long-term involvement of students in the Science of Information. We see our diversity initiatives as a web of activities, connecting our educational and research areas, and supporting the advancement of students from undergraduate experiences, through graduate studies, and into professional development. We also focus on additional recruitment and support, to further enlarge the pool of potential post-docs and faculty who are interested in research opportunities with the Center.

| Actions/ Programs | Expected Outcomes | Measures | Targets | Timeline |
|---|---|---|---|---|
| **GOAL 1: Provide funded research and mentoring experiences** | | | | |
| **Expand Undergraduate Research/Mentoring Opportunities** | Increased numbers of diverse undergraduates participating in the Channels Scholars Program and REU Program | Number of URM scholars participating in the program | Increase the number of URM scholars to one for each institution. | August 2012 and ongoing |
| | Increased numbers of institutions represented. | Number of partner/non-partner institutions with participating scholar | Increase the number of partner institutions sponsoring an undergraduate scholar from five to all partners (11). | Summer 2014 and ongoing |
| | Increased numbers of undergraduates nationwide with an awareness of the Center and the science of information | Number of applicants to the program | Increase the number of applicants from non-partner institutions from six to 12. | Spring 2014 and ongoing |
| **Precision recruitment of women graduate** | Increased numbers of women graduate students recruited and mentored by faculty at partner | Number of women graduate students recruited and mentored by | Increase the number of women graduate students from 13% Center-wide to closer to the national average of 22%, with special emphasis and | Spring 2016 and ongoing |

| Actions/ Programs | Expected Outcomes | Measures | Targets | Timeline |
|---|---|---|---|---|
| **students** | institutions. | faculty at underperforming partner schools. | focus on partner schools who are underperforming in this area. | |
| **Precision recruitment of U.S. citizens/Permanent Resident graduate students** | Increased numbers of U.S. Citizen/Permanent Resident graduate students recruited and mentored by faculty at partner institutions. | Number of U.S. Citizen/Permanent Resident graduate students recruited and mentored by faculty at underperforming partner schools. | Increase the number of U.S. Citizen/Permanent Resident graduate students from 24% Center-wide to closer to the national average of 43%, with special emphasis and focus on partner schools who are underperforming in this area. | Spring 2016 and ongoing |
| **Precision recruitment of African American/Black/Mixed Race graduate students** | Increased numbers of African American/Black/Mixed Race graduate students recruited and mentored by faculty at partner institutions. | Number of African American/Black/Mixed Race graduate students recruited and mentored by faculty at underperforming partner schools. | Increase the number of African American/Black/Mixed Race graduate students from 2% Center-wide to closer to the national average of 3-5%, with special emphasis and focus on partner schools who are underperforming in this area. | Spring 2016 and ongoing |
| **Incentivized leveraged funding to diversify Center-wide post-doctoral short list** | More diverse applicants on short list. | Number of diverse applicants on short list. | Increased numbers of U.S Citizens/Permanent residents and diverse candidates hired as Center-wide post-docs. | Spring 2015 and ongoing |
| | | | | |
| | | | | |
| **Goal 2: Open networking channels through educational programming** | | | | |
| | | | | |
| **Promotion of research and researcher through Virtual Brown Bags** | Enhanced retention of graduate students through promotion of research and researcher | Number of women and URM's invited to give presentations | 1. One or more URM as speaker<br>2. One Brown Bag devoted to STEM diversity | Fall 2014 and ongoing |

| Actions/ Programs | Expected Outcomes | Measures | Targets | Timeline |
|---|---|---|---|---|
| **Embedded diversity programming at major CSoI educational events** | Students learn about the importance of diversifying STEM and the challenges involved | Number of participants | One dedicated session at the annual Summer School and other prominent, Center-sponsored events. | Summer 2014 and ongoing |
| | | | | |

## Knowledge Transfer Mission Statement:

*Develop effective mechanisms for interactions between the Center and external stake-holders to support the exchange of knowledge, data, and the application of new technologies*

**Broader and Transformative Impact.**

The Center has significant, broad, and transformative impact beyond the foundations of the Science of Information. The cross-disciplinary research effort has led to fundamentally new explorations in areas such as life sciences, economics, communication systems, and social sciences, among others. In addition it has fostered an active and thriving community of students and scholars within academia and industry, and broadened its geographic scope to national and international levels. Novel research results and software artifacts developed at the Center have made tremendous scientific impact. The Center collaborates with industry through a number of relatonships; in life sciences (Amgen and Pfizer), communications (Bell Labs and HP), and large-scale data handling (Google and Technicolor). To enhance its scientific impact, the Center has developed a Science of Information portal and conducted several Science of Information Days that reach out to the broader community.

The Center's Knowledge Transfer mission includes effective mechanisms for interactions in the Center and for interactions between the Center and external stakeholders, to support the exchange of knowledge, data, and the application of new technologies. To achieve this mission, the Center continually works to engage with associated stakeholders. Knowledge outflow takes the form of disclosures to the public domain (including publications, presentations, software tools, and instructional materials), and technical collaborations.

The Center builds on a significant set of accomplishments and programs, as well as best practices and lessons learned during the current period. These programs include efforts aimed at transitioning technologies to commercial entities (intellectual property), outreach to industry, to the broader community, and enhancing international visibility. Accomplishments during the current period include patents and licensed technologies, significant industry funded research (over $1M in the most recent year), large number of personnel exchanges, significant outreach to the community, through frequent meetings and workshops, a comprehensive web portal (accessed by 11,054 unique new visitors over the most recent twelve month period), and a number of international partnerships aimed both at research collaborations and educational efforts.

The Center is now a strong partnership of 11 institutions. Since the Center has a very broad geographic distribution, it is well poised to reach out to industry in different geographic regions, with relatively little overhead, while maintaining close personnel proximity. Leveraging this opportunity, we present a comprehensive plan for continued knowledge transfer efforts, whose primary goals are: (i) to publicize Center efforts and scientific developments to industry; (ii) to match specific industry needs with suitable Center research groups; (iii) to provide a single point of contact for personnel exchanges between industry and the Center researchers; (iv) to organize periodic meetings (e.g., Science of Information Days) at all of the project sites to reach out to the broader community; (v) to establish international collaborations and

manage associated opportunities; (vi) to maintain a vibrant Industry Advisory Board and solicit feedback on related efforts; (vii) to reach out to the scientific community through publications and exhibits; and (viii) to maintain a community-wide web-presence for disseminating all Center artifacts.

### Context/ Current State

Each of the partner institutions and affiliated PIs has significant industry contact. The challenge is to provide an overall structure where the integrated whole is greater than the sum of individual parts. Specifically, the center must provide an attractive and compelling proposition to external partners for active engagement.

We initiate contact with a number of companies/labs/institutions with a view to identifying specific structure, scale, and scope of our partners program. Our Science of Information Hub, showcases current projects and partnerships.

### Action Plans/Activities

A comprehensive set of activities associated with our knowledge transfer goals is presented in the Table below. These activities can be broadly classified as: (i) developing a framework for the knowledge transfer program. (ii) staff various aspects of the knowledge transfer program, including day-to-day management, online presence, execution of action items, (iii) identify and leverage institutional support for various aspects of the program at various partner institutions, (iv) recruit partners by developing and publicizing a compelling value proposition for external stake-holders, and (v) constitute procedures for periodically evaluating and tuning the program.

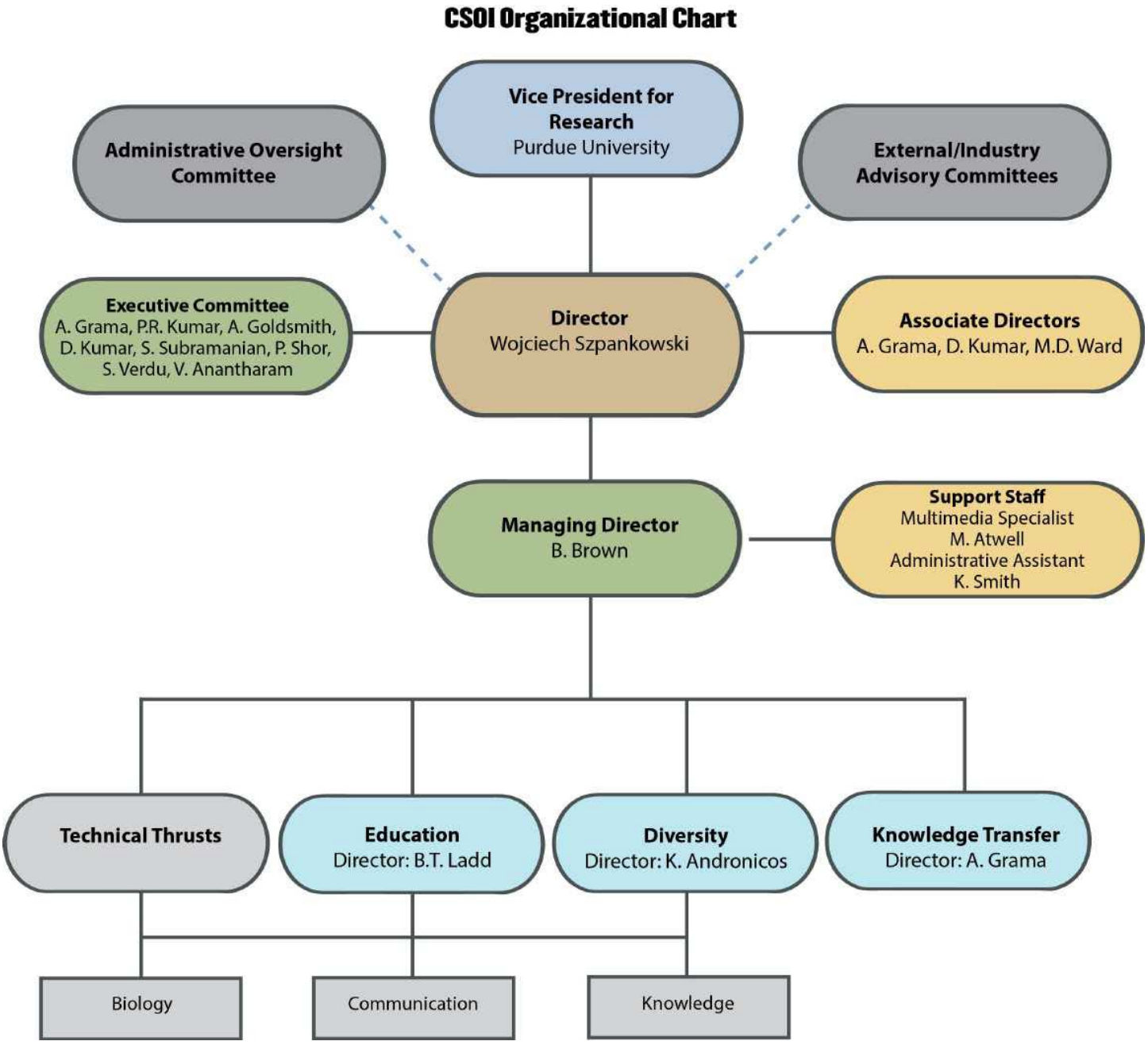
### Target End-State

We identify our goals with respect to each group in our set of external stake-holders. Detailed goals are provided in the table below. We specifically identify growth in terms of number of partners. We establish qualitative and quantitative targets for our online portal, and for our outreach efforts. We identify precise mechanisms for evaluation of the knowledge transfer program. Annual reviews examine benchmarks, identify shortcomings and fixes, and recommend course of actions. Detailed reports will also be made available at annual reviews, for formal evaluation of the program.

| Knowledge Transfer Goal Statement | | |
|---|---|---|
| *Develop effective mechanisms for interactions between the Center and external stakeholders to support the exchange of knowledge, data, and the application of new technologies.* | | |
| **Objectives and Metrics** | | |
| Increase the number of private/industrial affiliates to 8 by December 2016. | | |
| Increase the number of industrially funded faculty projects to 15 by December 2016 | | |
| Hold at least 3 Science of Information Forums on partner campuses annually | | |
| Increase the number of complete online modules/courses available each year on the learning hub. | | |
| Increase number of visits to Science of Information hub every year. | | |
| Increase the number of external links to the Science of Information hub by 50% each year. | | |
| **Initiate an Active Partners Program** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Develop a compelling value proposition for potential partners and mechanisms for publicizing collaborative opportunities | March 2011 - ongoing | A. Grama |
| Make targeted visits to selected potential partners to solicit partnerships/collaborations | Ongoing | A. Grama |
| Meet annually with the Industrial Advisory Board from among the partners and initiate periodic meetings | August 2011 | W. Szpankowski/A. Grama |
| **Human power for development, quality control, and publicity for Science of Information hub.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Maintain the Science of Information Hub. | Ongoing | M. Atwell |
| Institute periodic review procedures for content on the SoI hub, and mechanisms for publicizing them. | Annually | M. Atwell, All Staff |
| Institute mechanisms for cooperation and contribution from partner sites, as well as broader community to the SoI hub. | December 2016 | M. Atwell, B. Brown |
| **Institutional support and multi-institutional cooperation for course development.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Identify leads for course content distribution/use | 2015 | B.T. Ladd, M.D. Ward, D. Kumar |
| Outline the resources and institutional support needed for the creation of course content. | Ongoing | B.T. Ladd, Education Committee |
| Constitute mechanisms for integrating courses into undergraduate and graduate curricula across all partner institutions | Ongoing | B.T. Ladd, Education Committee |
| **Run Science of Information Workshops for STC and external participants.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Convene a workshop development team; including representatives of each research thrust area, which will identify workshop opportunities and themes. | Ongoing | W. Szpankowski, B. Brown, T. Weissman, S. Subramaniam, M. Sudan, A. Grama, Exec. Comm. |
| Create mechanisms for organizing and publicizing the workshops to diverse scientific communities (beyond traditional Information Theory). | Ongoing | W. Szpankowski, B. Brown, All staff |

## V: Leadership and Management

Integrating diverse, geographically separated teams of researchers into an effective common unit is a major challenge—one that needs considerable commitment and forethought. We adopt a management plan that leverages extensive existing resources to integrate project sites and teams to articulate a shared vision of Science of Information as a key enabling discipline. To accomplish this, we have implemented an organizational structure and procedures aimed at achieving research, education, diversity, knowledge transfer, and outreach goals. This strategy has enabled us to transition the Center from individual partners to an integrated team that stretches across disciplines and institutions.



The following high level goals guide the Center in its efforts to articulate and work towards a shared vision: (i) effectively coordinate efforts across project sites, and leverage research efforts in a timely manner; (ii) engage the broader scientific community to evolve our shared vision of Science of Information; (iii) utilize an industrial outreach plan focused on dissemination of our results to industry and enabling them to influence the Center's research and development efforts; (iv) leverage interdisciplinary opportunities and the Center's geographic scope in engaging underrepresented groups in all aspects of research and education.

**Center Leadership**
The Center is led by Purdue University and consists of the following partners (with lead PI listed below): Purdue University - Lead Institution (Wojciech Szpankowski), Bryn Mawr College (Deepak Kumar), Howard University (Mugizi Rwebangira), MIT (Peter Shor), Princeton University (Sergio Verdu), Stanford University (Andrea Goldsmith), Texas A&M University (P.R. Kumar), University of California, Berkeley (Bin Yu), University of California, San Diego (Shankar Subramaniam), University of Hawaii, Narayana Prasad Santhanam, University of Illinois, Urbana-Champaign (Olgica Milenkovic). Through past engagement with the Center, we have identified the University of Hawaii as an excellent potential partner, offering unique collaborative opportunities in all phases of our mission of research, education, and diversity. As part of the next project period, we propose to add the University of Hawaii (Narayana Prasad Santhanam) as a partner. Key individuals and their area of responsibility are: Director - W. Szpankowski, Purdue; Associate Directors - A. Grama and M. Ward, Purdue; D. Kumar, Bryn Mawr; Executive Committee Chair - P.R. Kumar, Texas A&M; Managing Director - Bob Brown; Education Director - Brent Ladd; Diversity Director - Kelly Andronicos; Knowledge Transfer - A. Grama and A. Goldsmith; Life Science Research Thrust Leaders - S. Subramaniam and A. Grama; Knowledge Management Thrust Leaders - J. Neville and S. Kulkarni; Communications Thrust Leaders - T. Weissman and D. Tse.

The leadership team is eventually responsible for accomplishing all project goals in a timely manner, and to communicate developments to various agencies.

**Primary Structure of the Management Plan**.
As originally planned, the Center involves over approximately 40 faculty across project sites. Several other affiliated faculty members are also involved in the Center from other institutions. Each partner site has research and development responsibilities, along with well-defined roles in education and diversity initiatives. The Center engages, on average, over 60 graduate students and post-doctoral participants each year.

**Management Structure, Processes, and Responsibilities**
The Center's management structure is shown in Figure 2. The structure consists of the following key elements: the leadership team, the advisory committees, and the administrative team. The managing director leads the Center's administrative efforts and reports directly to the Center Director. He is responsible for the day-to-day operation of the Center. He is assisted by an administrative staff member and business office staff. The education and diversity directors work closely with the Center Director, Managing Director, and faculty at the various project sites. The Center employs a multimedia specialist, who oversees all web related activity and development, and assists with online interactions. These interactions are facilitated by the SoI Hub (http://soihub.org).

**Center Administration.**
The Executive Committee meets once a month and oversees all major decisions regarding management, finances, and goals of the Center. During these meetings, Center business and progress are constantly discussed and evaluated. The evaluation process includes an independent evaluation subcommittee consisting of the Executive Committee chair and Center research thrust leaders. The committee evaluates progress of all PI's, and makes recommendations to the Director, who, with counsel of the Executive Committee, implements the recommendations.

**External Advisory Boards.**
The leadership team is guided by an External Advisory Board that consists of researchers from academia and industry. (J. Gibson, Chair; S. Bruck; A. Ephremides; J. Garcia Frias; M. Guzdial; C. Heegard; A. Karlin; M. Luby; N. Shroff; E.

Soljanin). The advisory board annually reviews project progress and recommends necessary changes and new directions. It also provides a channel for communicating developments of the Center to outside constituents.

At the behest of the renewal site visit report, the Center is now looking strategically at how it will best utilize its External Advisory committee to enhance the impact and longevity of the Center. In particular it is discussing how this group can assist to:

- Forge and/or strengthen communications between industry and the Center
- Help identify Center-related research ripe for tech transfer, and facilitate such transfer
- Help establish connections between the Center and other research organizations with overlapping interests.
- Help identify new funding mechanisms for Center-related research to complement existing funding and maintain aspects of the program after NSF funding ends.

The Director and executive committee have already begun discussing this issue and will continue to work closely with the NSF Program Director to strategically address this issue.

**Institutional Oversight.** On the recommendation of the 2012 site vist team, the Center added the Administrative Oversight Committee, chaired by Purdue's Vice President for Research. It includes administrative representatives from each partner institution. The purpose of the group is to provide executive oversight to the Center's management team and executive committee. The committee accomplishes this through a review of the Center's annual report in consultation with the Center Director and Managing Director. This is done once a year in between the completion of the annual report and the site visit.

**Technical Organization and Integrative Activities.**
The Center is organized around the technical thrusts, and Center faculty align with one or more of these thrusts as appropriate. Each thrust has two thrust leaders, who work to involve all of the research partners in effectively addressing their specific research challenges, as well as broadening participation from the scientific community.

The Center has several programs aimed at integration of partners into all Center activities: *Center Fellows*. A program that has been extremely successful in integrating disciplines and institutions across the Center is our Center Fellows program. Each year the Center solicits faculty proposals for funding for a post-doctoral researcher to work on collaborative projects across institutions. These proposals are evaluated based on their intellectual merit, as well as their integrative aspects. Center Fellows are also expected to provide leadership to our education and outreach activities. *Graduate Student Research Teams*. This program gives graduate students and post-doctoral researchers from partner institutions the opportunity to form interdisciplinary and cross institutional research teams that are funded by the Center. These teams also foster faculty collaborations by bridging research groups. *Seed Funding for Research.* The Center uses savings from yearly operations for collaborative research projects proposed by Center and/or affiliated faculty. In our second five years, this program will be institutionalized. The Center will provide guaranteed funding to a fewer number of faculty at each institution, and use remaining funds to build a "collaborative seed fund." A subcommittee will review faculty proposals and
will rate them according to their relevance to the Center's research agenda, interdisciplinary nature, and cross institutional participation. The committee will recommend to the Center Director the top proposals for funding.

**Center Succession Plan**. With guaranteed support from Purdue University and ongoing support from industry, the Center is well-positioned to thrive well beyond the next five year funding period. Purdue has committed to recurring funding that will support the Center beyond the next five-year period. Starting July 1st, 2014 CSoI will be designated as one of the research centers in Purdue University Discovery Park, a cluster of interdisciplinary centers, each with a large-scale research mission. In addition, the Center plans to put in place an Industrial Affiliates program with multiple tiers of membership and commensurate financial commitment. It is anticipated that 5–10 prime members and 20–30 regular members will annually contribute research support to the Center. Similar models are currently used to support the Center for Integrated Systems and Computer Forum at Stanford and the Berkeley Wireless Research Center.

We hold the following high level goals for our management plan:

- effectively coordinate efforts across project sites, and leverage research efforts in a timely manner;
- engage the broader scientific community to evolve our shared vision of information sciences;
- involve the applications community to build up novel methods, formalisms, and products of the research and development plan;
- develop an industrial outreach plan focused on timely dissemination of results to industry and enabling them to influence research and development efforts;
- develop a coherent educational and diversity plan across various project sites focused on research involvement at all levels;
- develop a novel curriculum across project sites, integrating theory and applications of information science;
- leverage tremendous interdisciplinary opportunities in engaging underrepresented groups in all aspects of research and education.

| Leadership and Management Goal Statement | | |
|---|---|---|
| *Accomplish the Center's mission through:*<br>• *inspirational leadership.*<br>• *inclusive and transparent decision-making.*<br>• *catalyzing new research opportunities.*<br>• *facilitating collaborative efforts.* | | |
| **Objectives & Metrics** | | |
| Executive committee will hold formal activities at least once per year to allow all Center participants to have a voice in the decision making of the Center. | | |
| **The project must be interdisciplinary** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Executive committee conducts yearly evaluations of each investigator's contribution to the STC. | Annually | W. Szpankowski, PR Kumar |
| **Instill a sense of pride in membership in the Center.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| The Center Director and Managing Director will conduct regular visits to each of the partner institutions and assist with organizing SoI Workshops | Annually | W. Szpankowski, B. Brown |
| **A harmonious leadership and management team.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Clearly articulate expectations for each participating investigator. | Ongoing | W. Szpankowski, Exec Comm |
| Make STC website the key resource for visibility of our accomplishments. | December 2010 - ongoing | M. Atwell, All Staff |
| Timely press releases highlighting research accomplishments for the general public. | Yearly | Staff, W. Szpankowski. Exec Comm |
| Create a reputation for excellence in research and innovation. | Anytime | ALL |

# VI: Ethics

**Background**

The America COMPETES Act requires that any academic institution seeking NSF funding for science and engineering research or education provide a plan for providing appropriate training and oversight in the responsible and ethical conduct of research to undergraduate students, graduate students, and postdoctoral researchers participating in the project. NSF's Implementation Plan to address the America COMPETES Act provisions requires all academic institutions to have a plan to provide the appropriate training and oversight for proposals submitted on or after January 4, 2010. Although the Science of Information STC proposal did not fall under the requirements, we intend to leverage the efforts of all the partner academic institutions to comply with the NSF Implementation Plan of the America COMPETES Act.

The Science of Information STC Ethics programs will be aimed at two groups of constituents:

- Students – undergraduate, graduate and postdoctoral researchers
- Faculty and Staff

**Ethics Mission:**

*Implement a multidisciplinary and multi-institutional program to inform and guide all members of the Center on the ethical and responsible conduct of scientific research*

**Ethics Program for SOI STC Students**

Each undergraduate student, graduate student and postdoctoral researcher being funded by the STC (with either sponsor or cost sharing funds) will satisfy the requirements of their home institution's responsible conduct of research (RCR) plan, including all on-line training and/or discussion-based classes. The SOI STC Co-PI or faculty lead for each institution (or their designee) will notify the SOI STC Ethics Program Coordinator (STC Managing Director) when each of their students completes RCR training and certification so a central record can be maintained for the STC.

**Ethics Program for SOI STC Faculty and Staff**

Each faculty and staff member being funded by the STC (with either sponsor or cost sharing funds) will complete ethics training by one of the following:

- Satisfy the requirements of their home institution's plan for faculty and staff training and certification.
- Complete the appropriate on-line training in RCR (via CITI online courses or other resources)
- Other approaches as agreed by the center director and the partnering institutions.

The SOI STC Co-PI or faculty lead for each institution (or their designee) will notify the SOI STC Ethics Program Coordinator (STC Managing Director) when each of their faculty and staff working in the STC completes RCR training and certification so a central record can be maintained for the STC.

Information detailing the RCR plans and programs at each SOI STC institution can be found in Appendix A.

| **Ethics Goal Statement** | | |
|---|---|---|
| *Implement a multidisciplinary and multi-institutional program to inform and guide all members of the Center on the ethical and responsible conduct of scientific research.* | | |
| **Objectives and metrics** | | |
| Design, create, and conduct a program that reaches all personnel working on the STC and that addresses the following critical areas: plagiarism, notebook documentation, authorship, social surveys, testing on animals, human subjects in experiments, and additional topics to be determined during the creation of the program and/or as a result of the program assessment over time. | | |
| Success in establishing clear guidelines for the ethical and responsible conduct of all work associated with the center.<br>Success in educating center members about these guidelines**.**<br>Success in all members of the center following these guidelines, which will result in the highest level of integrity and ethical conduct for work done within the center**.** | | |
| **Action** | **Completion Date** | **Point of Contact** |
| Establish ethics program that spans all partner institutions. | March 2011 - Ongoing | STC Manager and STC Ethics Program Coordinator |
| Ensure timely participation of all center members. | Students: follow institutional requirements<br>Faculty & Staff: complete training | STC Ethics Program Coordinator<br>Ethics representatives for each institution |
| Ongoing assessment and evolution of the program is needed to ensure it is meeting its objectives. | Ongoing | STC Ethics Program Coordinator<br>Ethics representatives for each institution |

## Appendix

| RCR Programs in place | Current Status | Institutional Point of Contact | Web Info |
|---|---|---|---|
| **Bryn Mawr** | All Students | Nona C. Smith, Director, Office of Sponsored Research | http://www.brynmawr.edu/grants/ |
| **Howard** | Grad students only | Chontrese Doswell Hayes, Assistant Dean, Graduate School | http://www.gs.howard.edu/omrs/rcr.htm |
| **MIT** | All Students | Michelle Christy, Director, Office of Sponsored Programs | http://osp.mit.edu/compliance/responsible-conduct-of-research-rcr/rcr-implementation-plan |
| **Princeton** | All Students | David N. Redman, Associate Dean, Academic Affairs | http://gradschool.princeton.edu/about/docs/academics/RCR_Plan_2010_Final.pdf |
| **Purdue** | All Students | Peter Dunn, Office of the Vice President for Research | http://www.purdue.edu/research/vpr/rschadmin/rcr/ |
| **Stanford** | All Students | Ann George, Administrative Services Manager, Vice Provost for Graduate Education | http://ora.stanford.edu/ora/osr/proposal_development/fastlane/default.asp |
| **UC Berkeley** | All Students | Patrick Schlesinger, Assistant Vice Chancellor, Research Administration and Compliance | http://rac.berkeley.edu/rcr.html |
| **UC San Diego** | All Students | George Tynan, Associate Vice Chancellor for Research | http://www.sandiego.edu/sponsored-programs/documents/USDProtocolforNSFandNIHRequirementforTraininginResponsibleConductofResearch.pdf |
| **UIUC** | All Students | Ravi K. Iyer, Interim Vice Chancellor for Research | http://research.illinois.edu/rcr/ |

## SOI STC Ethics Goal

| Programs in place at SOI-STC institutions | UNDERGRADS | GRADS | POST-DOCS | CERTIFICATION | INSTITUTIONAL POINT OF CONTACT |
|---|---|---|---|---|---|
| **Purdue** | CITI -- subset; encouraged to attend seminars/workshops | CITI; discussion-based RCR education | CITI; discussion-based RCR education | Doc by grad program, postdoc mentor | Peter Dunn, Office of the Vice President for Research |
| **UC Berkeley** | UC San Diego handout | CITI | CITI | PI cert within 60 days | Patrick Schlesinger, Assistant Vice Chancellor, Research Administration and Compliance |
| **Stanford** | CITI or Stanford course | CITI or Stanford course | CITI or Stanford course | PI must notify students | Ann George, Administrative Services Manager, Vice Provost for Graduate Education |
| **UC San Diego** | On-line RCR certification | On-line RCR certification | On-line RCR certification | PI must provide certificates before getting money | George Tynan, Associate Vice Chancellor for Research |
| **Bryn Mawr** | CITI or college's preferred program or through classroom work | CITI or college's preferred program or through classroom work | CITI or college's preferred program or through classroom work | Doc by RCR training and certification program | Nona C. Smith, Director, Office of Sponsored Research |
| **MIT** | CITI | CITI | CITI | CITI upload to MIT system for tracking and reporting | Michelle Christy, Director, Office of Sponsored Programs |
| **Princeton** | Princeton-based courses (with CITI modules for partial support); discussion required | Princeton-based courses (with CITI modules for partial support); discussion required | Must have RCR training prior to or when starting employment (Princeton or CITI) | Academic departments required to track all students and postdocs | David N. Redman, Associate Dean, Academic Affairs |
| **UIUC** | College's preferred program as stated by the UIUC Office of the Vice Chancellor for Research (OVCR) | College's preferred program as stated by the UIUC Office of the Vice Chancellor for Research (OVCR) | College's preferred program as stated by the UIUC Office of the Vice Chancellor for Research (OVCR) | All students and postdocs tracked as stated by the UIUC Office of the Vice Chancellor for Research (OVCR) | Ravi K. Iyer, Interim Vice Chancellor for Research |
| **Howard** | Not Applicable/Grads Only | Must complete RCR workshop to advance to candidacy | Must complete RCR workshop to advance to candidacy | Certificate sent to the candidate's acedemic department | Chontrese Doswell Hayes, Assistant Dean, Graduate School |

# Intellectual Property Agreement
# For the Center for Science of Information
# (A National Science Foundation funded Science and Technology Center)

THIS Agreement made this12th day of August, 2010, by and between PURDUE UNIVERSITY and its Academic Partners (as listed below). This Agreement supersedes any Consortium Agreements by and between Purdue and its Academic Partners for the Center for Science of Information.

WITNESSETH:

THAT, WHEREAS, Purdue University ("Lead Institution") has been awarded a Science and Technology Center by the National Science Foundation (NSF). The Center for Science of Information shall include the following Academic Partners: Bryn Mawr College, Howard University, Princeton University, Massachusetts Institute of Technology, Stanford University, University of California - Berkeley, University of California - San Diego, and The Board of Trustees of the University of Illinois and;

WHEREAS, NSF requires that the Academic Partners and Lead Institution execute an Intellectual Property Agreement prior to the release of funds to the Center.

NOW, THEREFORE, The Parties agree as follows:

1. "Center Intellectual Property" means the legal rights relating to inventions (including Subject Inventions as defined in 37 CFR 401), patent applications, patents, copyrights, trademarks, mask works, trade secrets, and any other legally protectable information, including software, first made or generated during the performance of this Agreement.

2. The rights of the Parties to Subject Inventions made by their employees in the performance of this Center shall be as set forth in the Patent rights clause of 37 CFR401.14 (Bayh-Dole Act). The NSF may obtain title to any subject invention not elected by a Party as set forth in the Patent rights clause.

3. Unless otherwise agreed in writing, Center Intellectual Property shall be owned by the Party whose employees make or generate the Project Intellectual Property. Jointly made or generated Project Intellectual Property shall be jointly owned by the Parties unless otherwise agreed in writing.

4. In the event Joint Intellectual Property is created, the involved Parties shall work together in good faith to negotiate an Interinstitutional Agreement to address the management and distribution of said Joint Intellectual Property.

5. Should it be necessary for any Party to disclose confidential information to another Party, then:

   A. Each Party may designate a Primary Contact with whom disclosing parties shall communicate regarding Confidential Information. Prior to the disclosure of confidential information, the disclosing Party shall confirm with the receiving Party's Primary Contact that the receiving Party is willing to accept the confidential information and to handle it in accordance with this Agreement.

   B. The receiving Party and its staff agree to use reasonable efforts to prevent disclosures of any such information furnished by the disclosing Party which is labeled in writing as confidential at the time of delivery or if oral, visual, or other non-written manner of disclosure of otherwise undisclosed confidential information is made, such information shall be entitled to protection only if identified as confidential upon initial disclosure, and if a written summary of all such disclosures, appropriately stamped or marked, is delivered to the receiving Party addressed as noted hereafter in this Agreement within thirty (30) days of such disclosure.

C. In no event is information Confidential Information if it (a) was in the receiving Party's possession before receipt from the disclosing Party; (b) is or becomes a matter of public knowledge through no fault of the receiving Party; (c) is received by the receiving Party, without restriction as to further disclosure, from a third Party having an apparent bona fide right to disclose the information to the receiving Party; or (d) is independently developed by the receiving Party without use of the disclosing Party's Confidential Information. For purposes of this Agreement, the receiving Party's students are not third parties vis-a-vis the receiving Party.

D. Nothing in this Agreement shall be construed to prevent the receiving Party from disclosing Confidential Information as required by law, or pursuant to an order of a court or other governmental authority of competent jurisdiction, as long as the receiving Party promptly notifies the disclosing Party of its obligation to disclose and provides reasonable cooperation to the disclosing Party in any efforts to contest or limit the scope of such order or subpoena.

E. Following termination of this Agreement, the receiving Party shall, at the direction of the disclosing Party, either destroy or return to the disclosing Party all documents, materials, and other tangible manifestations of the disclosing Party's Confidential Information and shall destroy any electronic or digital manifestations of the disclosing Party's Confidential Information, except that the receiving Party may retain one copy of the Confidential Information solely for the purpose of monitoring its obligations under this Agreement.

6. Funding for the Center activities shall be obligated to the Academic Partners through subcontracts issues in accordance with the Lead Institution's Cooperative Agreement from the NSF.

7. This Agreement may be executed in any number of counterparts. Each executed counterpart shall be deemed to be an original. All executed counterparts taken together shall constitute one agreement. Facsimile signatures shall constitute original signatures for all purposes.

IN WITNESS WHEREOF, the parties hereto have executed this Agreement as of the day and year first above written.