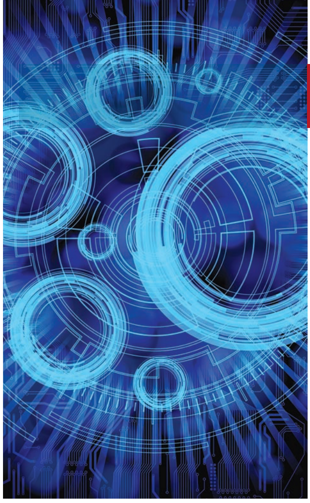

Editor: **Daniel Zeng,** Chinese Academy of Sciences & University of Arizona, zeng@email.arizona.edu

Next-Generation Team-Science Platform for Scientific Collaboration

Xiaolong Zheng, *Chinese Academy of Sciences*
Guanyan Ke, *National University of Defense Technology, China*
Daniel D. Zeng, *Chinese Academy of Sciences and University of Arizona*
Sudha Ram, *University of Arizona*
Hao Lu, *Chinese Academy of Sciences*

In the past two decades, many branches of science have shifted from individually oriented research toward team-based scientific collaboration.[1–3] Teams of researchers representing different disciplines are often brought together to better solve large-scale and often urgent problems of scientific, societal, and environmental relevance. In addition to combined subject matter expertise and the team's disciplinary composition, many contextual factors such as antecedent conditions, collaborative processes, and support technologies as well as a host of social factors such as team size and organizational complexity can directly influence outcomes in team-based research.[4,5]

From this perspective, emerging research on team science aims at better understanding the key contextual factors related to trans-disciplinary scientific collaboration processes and enhancing the outcomes of large-scale collaborative research programs. More specifically, team-science research combines problem-solving frameworks, specialized expertise, and research methods across disciplinary boundaries to help produce high-impact science.[6]

Team-science practice can range from a few participants working at the same site to a large number of researchers dispersed across multiple geographic and organizational venues. Additional aspects of a team-science project include technological resources and collaboration platforms that could influence team members' ability to communicate, integrate diverse datasets, and carry out joint research activities. Such platforms, often with the Internet as a technology enabler, have received much attention in the scientific community.[7–9] In the past two decades, we've witnessed a surge in investments in developing cyberinfrastructures for large-scale team scientific collaborations. Table 1 lists several examples.[10–13]

Despite the success of many of these cyberinfrastructure projects, a range of obstacles facing team science remain to be tackled. Virtual environments could never replace real-world social environments and setups to support the full gamut of collaborative activities.[5,14,15] Misunderstandings due to linguistic differences, disparities in management styles, and social conventions in different cultures can constrain the effectiveness of global teams.[5] Trust is also especially fragile and transient in virtual teams.[8]

In an effort to generate insights to inform team-science platform designs to better support scientific collaboration, we have been developing an experimental next-generation team-science-enabling platform (TSEP). With this platform, we have been collecting research intelligence related to the topics of scientific collaborations. Based on social computing theories[16–18] and empirical analyses, we model team-based research behaviors and conduct computational experiments evaluating different modes and setups of collaboration.

This article outlines the conceptual design framework of the TSEP. The key TSEP functions include research intelligence collection for scientific collaborations; empirical analysis of research intelligence for understanding the key factors underpinning team-science practice; computational experiments based on environmental, technological, and social-cognitive models to evaluate team-science

1541-1672/11/$26.00 © 2011 IEEE
Published by the IEEE Computer Society

designs and protocols; and training of researchers.

Research Intelligence Collection

Web 2.0 technology has been redefining scientific collaboration among participants in many research fields. Online knowledge repositories such as those supported through crowd sourcing, along with online libraries and scientific collaboration websites, provide us with the unprecedented abundance of research intelligence to understand contextual factors that influence scientific collaboration processes.

Figure 1 illustrates the four classes of research intelligence: collaborative readiness, collaborative process, collaborative outcomes, and effectiveness and impact. Research intelligence on collaborative readiness concentrates on social concerns and research issues, institutional support for collaboration, public and private investments in large-scale research initiatives, and participating members of a research team. The scientific collaborative process can be characterized along three dimensions: organizational, geographic, and analytic.[8] The organizational dimension includes not only intraorganizational partnerships but also interorganizational alliances. The geographic dimension ranges from local to regional to national or global collaboration. The analytic dimension incorporates micro to macro levels of intellectual analysis. Collaborative outcomes concern publications such as journal and conference papers, edited books, conference proceedings, patents, discoveries, and related applications. The effectiveness and impact of research outcomes can be evaluated from a temporal perspective, which ranges from early-stage collaborative synergy to mid-term scientific innovation and long-term societal impact.

Table 1. Sample cyberinfrastructures to enable team science.

Project	Objective
VIVO	Link profiles with grants or publications, and help scientists share research data
iPlant	Enable new conceptual advances in plant sciences through integrative, computational thinking
Liquidpub	Provide a range of tools to enable socially driven generation and dissemination of scientific knowledge
Biomedical Informatics Research Network	Provide a repository of neuroanatomical, clinical, genomic, and behavioral data and related analysis tools
Open Science Grid	Share resources, tools, and expertise for domain scientists, computer scientists, and technology specialists

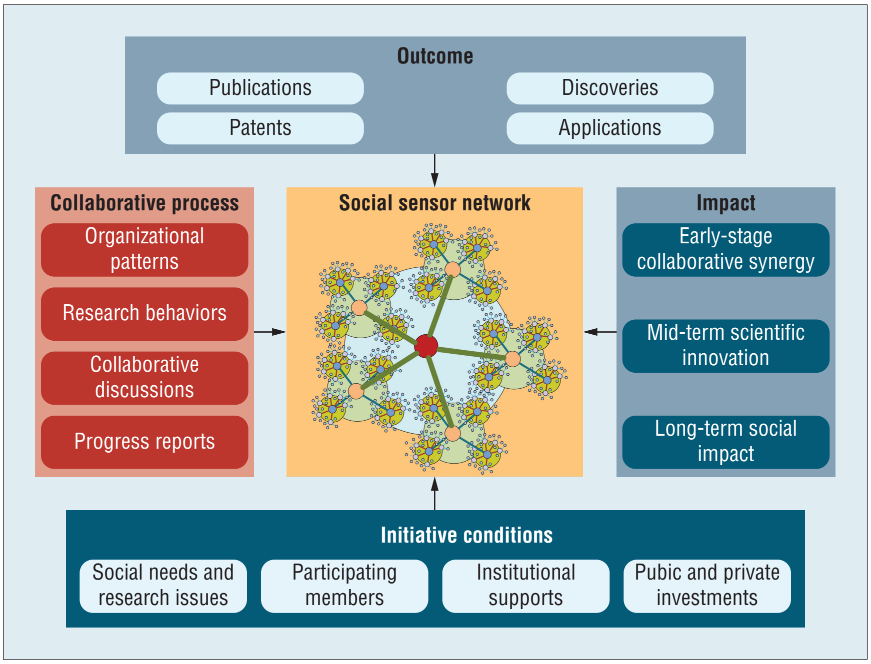

Figure 1. Research intelligence collection. The four classes of research intelligence are collaborative readiness, collaborative process, collaborative outcomes, and effectiveness and impact.

From an implementation standpoint, we have been applying application-specific knowledge engines (ASKEs) and social-sensor networks (SSN) to collect research intelligence over the Web. The basic idea of an ASKE is to use a knowledge-configuration file (KCF) to specify topics, keywords, search sequences, and schedules for query processing. Similar to physical sensor networks, a SSN based on an ASKE consists of many connected cyber-sensors assigned to topic-related websites. Each sensor is a topic-specific crawler. Depending on the topic, the SSN displayed in Figure 1 can be used to collect research intelligence at multiple levels.

Empirical Analysis

Analyzing real-world team-science practice is key to understanding evolutions of the scientific community and their products—knowledge. In any discipline or transdisciplinary effort, researchers contribute various theories and techniques and generate diverse findings. Based on existing studies,[6,19] we propose a three-level analytic approach that can serve as

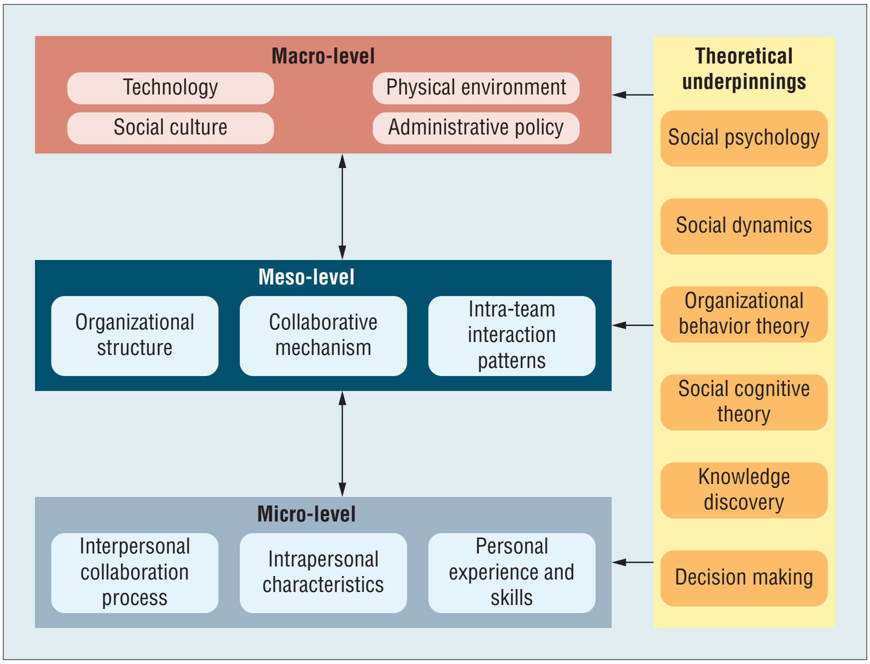

Figure 2. Empirical analysis based on research intelligence. A three-level analytic approach serves as a framework for organizing individual-, organization-, and team-level research efforts.

a framework for organizing the diverse forms of research by individual scientists, organizations, and teams (see Figure 2). Analysis at each level might use multiple approaches and provide different insights.

The micro-level analysis covers interpersonal collaboration process, intrapersonal characteristics, and personal experience and skills. The meso-level is concerned with organizational structures, collaborative mechanisms, and intra-team interaction patterns. Both micro- and meso-level analyses heavily rely on social-network-analysis technologies, where nodes represent participant members and link their relationships. We have developed tools to study and visualize the evolution of network structures and help summarize specific interaction patterns by examining communications.

The macro-level analysis provides insights about broad patterns of scientific collaboration in terms of effect and impact. Empirical analyses at the macro level also utilize social computing theories and might require large-scale computing infrastructures such as grid and cloud computing to obtain results.

To conduct meaningful analyses at these three levels, we need to incorporate social psychology, social dynamics, organizational behavior, and social cognition theories, in the specific context of scientific collaboration. Additionally, progress made in cultural computing, data mining, and visualization techniques can help identify interrelationships and patterns from empirical research intelligence.

Computational Experiment

Existing studies on team science have relied mostly on retrospective and prospective case-comparison studies. Experiments or quasi-experiments were rarely used.[2] Designers or evaluators of team-science platforms or projects, on the other hand, often need to capture and understand the collaboration process, and the changing environments and tasks, at a fast pace.

To address this challenge, we propose applying computational experimentation ideas[20–22] that are well-suited to model complicated behavioral, organizational, and social issues and evaluate system design alternatives in a holistic manner. Team-science researchers can construct various computational models based on empirical findings and conduct agent-based experiments to study the impact of various team-science organizational principles or platform choices. This approach is a viable modeling framework to address the open, dynamic, complex, and unpredictable nature of team-science practice. First, we construct an artificial society-based model of the team-science ecosystem using agent-based modeling technology, with the empirical findings as the backdrop. The related research issues include social reasoning of agents' beliefs, motivations, goals, emotions, intentions, and trustworthiness. Then, we design scenario-based computational experiments, with formal specifications on the tasks, task assignments, teaming structure, and platform choices, among others. After the experiments are specified formally, the team-science artificial society will be activated with agents interacting with each other over time. Through observations of the emergent behavior, and system- and individual-level performance, we validate models and evaluate design alternatives.

Training

Training researchers to improve their scientific collaboration skills involves many challenges, including learning domain knowledge, navigating within and between disciplines, acquiring communication skills, and appreciating cultural differences in different disciplines, among others.[23] In addition, the distinctions between single-discipline, interdisciplinary, and transdisciplinary forms of scientific collaboration are often directly relevant to the development of criteria for evaluating the success of team science. These challenges and distinctions have been largely neglected.[2,24,25]

The TSEP is designed to help deal with some of these challenges. Through the TSEP, we can empirically determine the main determinants of successful scientific collaboration training, such as individual trainer characteristics, key institutional setups, and training structures and processes. Based on both qualitative and quantitative methods, various effective training models can be constructed under different conditions. The TSEP is also expected to provide assessment methods and measures to monitor ongoing scientific collaboration processes and to evaluate outcomes.

Most pressing problems of significant impact—such as pollution, social and national security, economic crisis, public health, and ecological balance—are complex and call for transdisciplinary scientific efforts in team settings. Team science aims to better understand such team efforts and help design more effective approaches to enable more effective and efficient scientific collaboration. The proposed TSEP system provides research intelligence collection capabilities and a suite of analysis and modeling tools to help understand key factors that influence the effectiveness of large-scale team-science initiatives and evaluate different team-science design alternatives. These capabilities and tools are expected to enhance the research platform of the team-science research community.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China grants 71103180, 71025001, 91124014, 60875049, 60921061, 90924302, 70890084, and 91024030 and by Chinese Academy of Sciences grants 2F09N06, 2F07C01, and 2F11N06.

References

1. S. Wuchty et al., “The Increasing Dominance of Teams in Production of Knowledge,” *Science*, vol. 316, no. 5827, 2007, pp. 1036–1039.
2. D. Stokols et al., “The Science of Team Science: Overview of the Field and Introduction to the Supplement,” *Am. J. Preventive Medicine*, vol. 35, no. 2, supplement 1, 2008, pp. S77–S89.
3. B.F. Jones et al., “Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science,” *Science*, vol. 322, no. 5905, 2008, pp. 1259–1262.
4. D. Stokols, “Toward a Science of Transdisciplinary Action Research,” *Am. J. Community Psychology*, vol. 38, no. 1/2, 2006, pp. 63–77.
5. G.M. Olson and J.S. Olson, “Distance Matters,” *Human-Computer Interaction*, vol. 15, no. 2/3, 2000, pp. 139–178.
6. K. Börner et al., “A Multi-Level Systems Perspective for the Science of Team Science,” *Science Translational Medicine*, vol. 2, no. 49, 2010, p. 49cm24.
7. P.A. Schad et al., “Building a Biomedical Cyberinfrastructure for Collaborative Research,” *Am. J. Preventive Medicine*, vol. 40, no. 5, supplement 2, 2011, pp. S144–S150.
8. D. Stokols et al., “The Ecology of Team Science: Understanding Contextual Influences on Transdisciplinary Collaboration,” *Am. J. Preventive Medicine*, vol. 35, no. 2, supplement 1, 2008, pp. S96–S115.
9. W. Chismar et al., “Health Cyberinfrastructure for Collaborative Use-Inspired Research and Practice,” *Am. J. Preventive Medicine*, vol. 40, no. 5, supplement 2, 2011, pp. S108–S114.
10. V. Gewin, “Collaboration: Social Networking Seeks Critical Mass,” *Nature*, vol. 468, no. 7326, 2010, pp. 993–994.
11. L.D. Stein, “Towards a Cyberinfrastructure for the Biological Sciences: Progress, Visions and Challenges,” *Nature Reviews Genetics*, vol. 9, no. 9, 2008, pp. 678–688.
12. V. Gewin, “Networking in VIVO,” *Nature*, vol. 462, no. 7269, 2009, p. 123.
13. F.-Y. Wang, “SciTS: Is it the Lighthouses of Scientific and Technical Cooperation of the 21st Century?” *Science & Technology Rev.*, vol. 29, no. 12, 2011, p. 81.
14. B.W. Hesse, “Of Mice and Mentors: Developing Cyber-Infrastructure to Support Transdisciplinary Scientific Collaboration,” *Am. J. Preventive Medicine*, vol. 35, no. 2, 2008, pp. S235–S239.
15. B.W. Hesse et al., “Temporal Aspects of Computer-Mediated Communication,” *Computers in Human Behavior*, vol. 4, no. 2, 1988, pp. 147–165.
16. F.-Y. Wang, “Toward a Paradigm Shift in Social Computing: The ACP Approach,” *IEEE Intelligent Systems*, vol. 22, no. 5, 2007, pp. 65–67.
17. F.-Y. Wang et al., “Social Computing: From Social Informatics to Social Intelligence,” *IEEE Intelligent Systems*, vol. 22, no. 2, 2007, pp. 79–83.
18. X. Zheng et al., “Analyzing Open-Source Software Systems as Complex Networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 24, 2008, pp. 6190–6200.
19. L.C. Mâsse et al., “Measuring Collaboration and Transdisciplinary Integration in Team Science,” *Am. J. Preventive Medicine*, vol. 35, no. 2, supplement 1, 2008, pp. S151–S160.
20. C. Xie et al., “Computational Experiments for Science Education,” *Science*, vol. 332, no. 6037, 2011, pp. 1516–1517.
21. F.-Y. Wang, “Computational Experiments for Behavior Analysis and Decision Evaluation in Complex Systems,” *J. Systems Simulation*, vol. 16, no. 5, 2004, pp. 893–897.
22. Y. Gil et al., “Wings: Intelligent Workflow-Based Design of Computational Experiments,” *IEEE Intelligent Systems*, vol. 26, no. 1, 2011, pp. 62–72.
23. J.M. Nash, “Transdisciplinary Training: Key Components and Prerequisites

for Success," *Am. J. Preventive Medicine*, vol. 35, no. 2, supplement 1, 2008, pp. S133–S140.

24. D. Rhoten and A. Parker, "Risks and Rewards of an Interdisciplinary Research Path," *Science*, vol. 306, no. 5704, 2004, p. 2046.
25. J.M. Nash et al., "Training the Transdisciplinary Scientist: A General Framework Applied to Tobacco Use Behavior," *Nicotine & Tobacco Research*, vol. 5, supplement 1, 2003, pp. S41–S53.

Xiaolong Zheng is an assistant professor in the State Key Laboratory of Management and Control of Complex Systems at the Chinese Academy of Sciences' Institute of Automation. Contact him at xiaolong.zheng@ia.ac.cn.

Guanyan Ke is a project manager for science and technology in the College of Mechanical Engineering and Automation at the National University of Defense Technology, China. Contact her at keguanyan@gmail.com.

Daniel Dajun Zeng is a professor in the University of Arizona's Management Information Systems Department and a research professor in the State Key Lab of Intelligent Control and Management of Complex Systems at the Chinese Academy of Sciences' Institute of Automation. Contact him at zeng@email.arizona.edu.

Sudha Ram is the McClelland Professor in the Department of Management Information Systems at the University of Arizona. Contact her at ram@eller.arizona.edu.

Hao Lu is a research engineer in the State Key Laboratory of Management and Control of Complex Systems at the Chinese Academy of Sciences' Institute of Automation. Contact him at hao.lu@ia.ac.cn.