National Science Observatory

NSF Science and Technology Centers: Integrative Partnerships, IU pre-proposal

---

Katy Börner, Victor H. Yngve Professor, SLIS, Statistics, Informatics and Computing
Director of the Cyberinfrastructure for Network Science Center

March 24, 2011

Cognitive Science Program

School of Library and Information Science
Making Information Make a Difference

DLP Indiana University

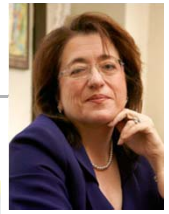
CYBERINFRASTRUCTURE for NETWORK SCIENCE CENTER
School of Library and Information Science, Indiana University

SCHOOL OF Informatics and Computing

COLLEGE OF ARTS & SCIENCES

CNetS Center for Complex Networks and Systems Research

IU Core Investigator Team








Bernice A. Pescosolido, Distinguished and Chancellor's Professor, Sociology
Alessandro Vespignani, Rudy Professor, Informatics and Computing, Statistics, Physics, Cognitive Science, Director of the Center for Complex Networks and Systems Research
Filippo Menczer, Associate Professor, Division Chair, Informatics and Computing, Cognitive Science, Physics, Associate Director of the Center for Complex Networks and Systems Research
J. Scott Long, Distinguished and Chancellor's Professor, Statistics and Sociology
Colin Allen, Professor of Cognitive Science and History & Philosophy of Science
Thomas F. Gieryn, Rudy Professor of Sociology and Adjunct Professor in the Department of History and Philosophy of Science, Vice Provost for Faculty and Academic Affairs
Olaf Sporns, Professor, Associate Department Chair, Psychological and Brain Sci.
William K. Barnett, Senior Manager, Research Technologies Life Sciences and Director, Advanced Information Technology Core.

Other IU Faculty: Sugimoto, Milojevic, Ying (SLIS), Bollen (SOIC), McDonald (DLP)

Motivation and Focus

1. Science and technology (S&T) is a multilevel system made up of
   - millions of individuals (scientists, researchers, policy makers, educators, and students)
   - in thousands of institutions (universities, government labs, corporate research centers, publishing houses, foundations and funding agencies)

   that constantly produce, evaluate, disseminate, and apply new scientific knowledge and technological artifacts.
2. As the complexity, scope, pace, and global scale of contemporary S&T increases, our ability to understand and utilize what exists, recognize and support what emerges, and preserve what is most valuable diminishes.
3. We propose a National Science Observatory that will provide theoretical foundations, integrative methodologies, and predictive outcome to the process and benefits of science.

The National Science Observatory STC will have four cores:

Research Core addresses basic foundational and theoretical questions regarding the creation, dissemination, and adoption of knowledge using a non-reductionist approach.

National Science Forecasting Core informs science and science policy decision making by drawing on foundational concepts, scientific discoveries, and novel tools that will allow the establishment of a science forecast infrastructure.

Technology and Computing Core builds on IU's extraordinary computing infrastructure and extensive prior work by IU scientists in the development of scholarly databases, tools and cyberinfrastructures, and community support platforms to create advanced methodologies for the study of S&T.

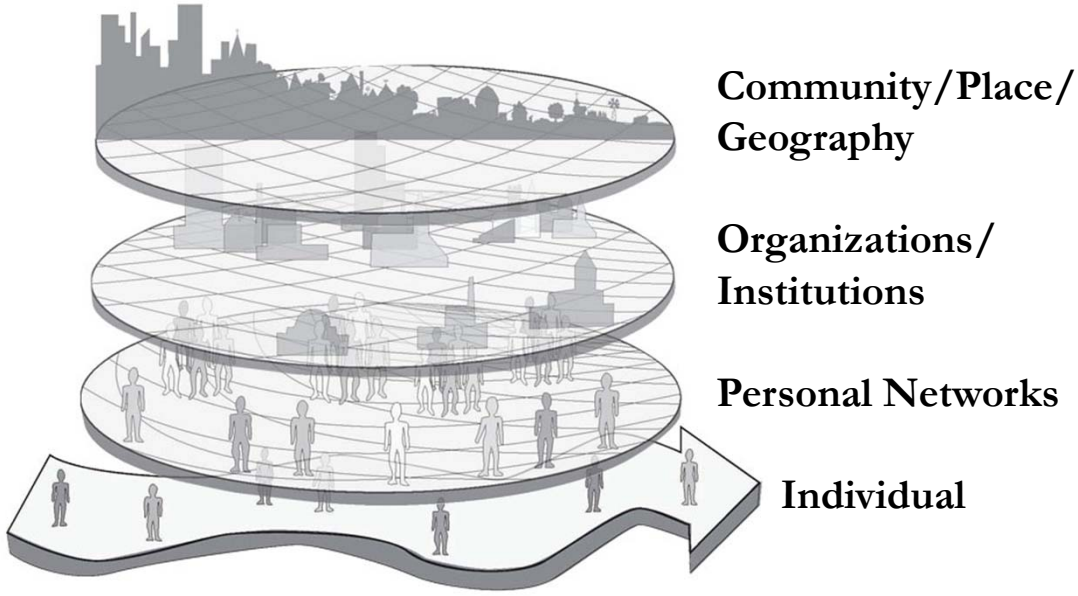
Education and Dissemination Core develops the first national, collaborative effort to develop state-of-the art materials that translate S&T theory, approaches, and predictions for educators, practitioners, policy makers, and the general public.

Research Core

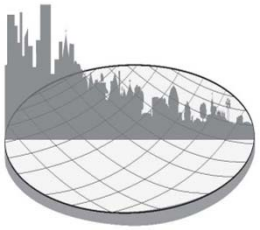
This core will lay out the theoretical foundation for

- the scientific understanding of system level, non-linear and non-local mechanisms governing the dynamics of knowledge,
- the diffusion of ideas, and
- the emergence, maturation, interaction, and dissolution of scientific and technological domains.

We will apply a systems science framework of vertical integration that aims to capture the structure and dynamics of the S&T system at different levels.

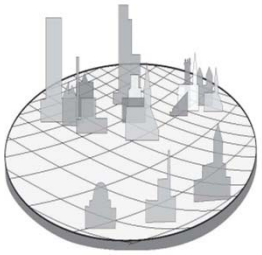


Prior Work by IU Investigators



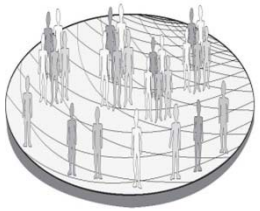
Communities

Börner – modeling the structural co-evolution of scientific networks and mapping the structure and dynamics of science • **Vespignani** – diffusion of scientific reputation • **Bollen** – measuring scientific activity from large-scale usage data: usage-based impact rankings and visualizing the flow of scientific activity • **Menczer** – crowdsourcing scholar annotations to compute universal impact measures and discipline maps



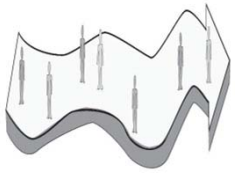
Organizations

Börner – institutional knowledge flows: sources and sinks of knowledge and expertise • **Gieryn** – culture of institutions, moments of science • **Allen** – modeling the hierarchical structure of concepts in dynamical reference works, using combination of data mining and machine reasoning about expert feedback on postulated relationships



Collaborations

Börner, Milojevic, Ding – team science level analyses and visualizations • **Pescosolido** – assimilation into scientific roles: influence of social networks on response to problem onset, diffusion of innovation • **Sugimoto** – Mentor-mentee relationships in doctoral education and academic genealogy • **Menczer** – relationship between citation networks and topicality/content; relationship between scholars based on crowdsourced annotations



Individuals

Long – structure of scientific careers • **Sugimoto** – contextualized, in-depth, multiple career line bibliometrics • **Sporns** – neurological basis of human decision making and innovation and human connectomics

Multi-Level Systems Perspective

The STC will unite and organize current and future transdisciplinary efforts in the theory, measurement, modeling, and communication of S&T using a

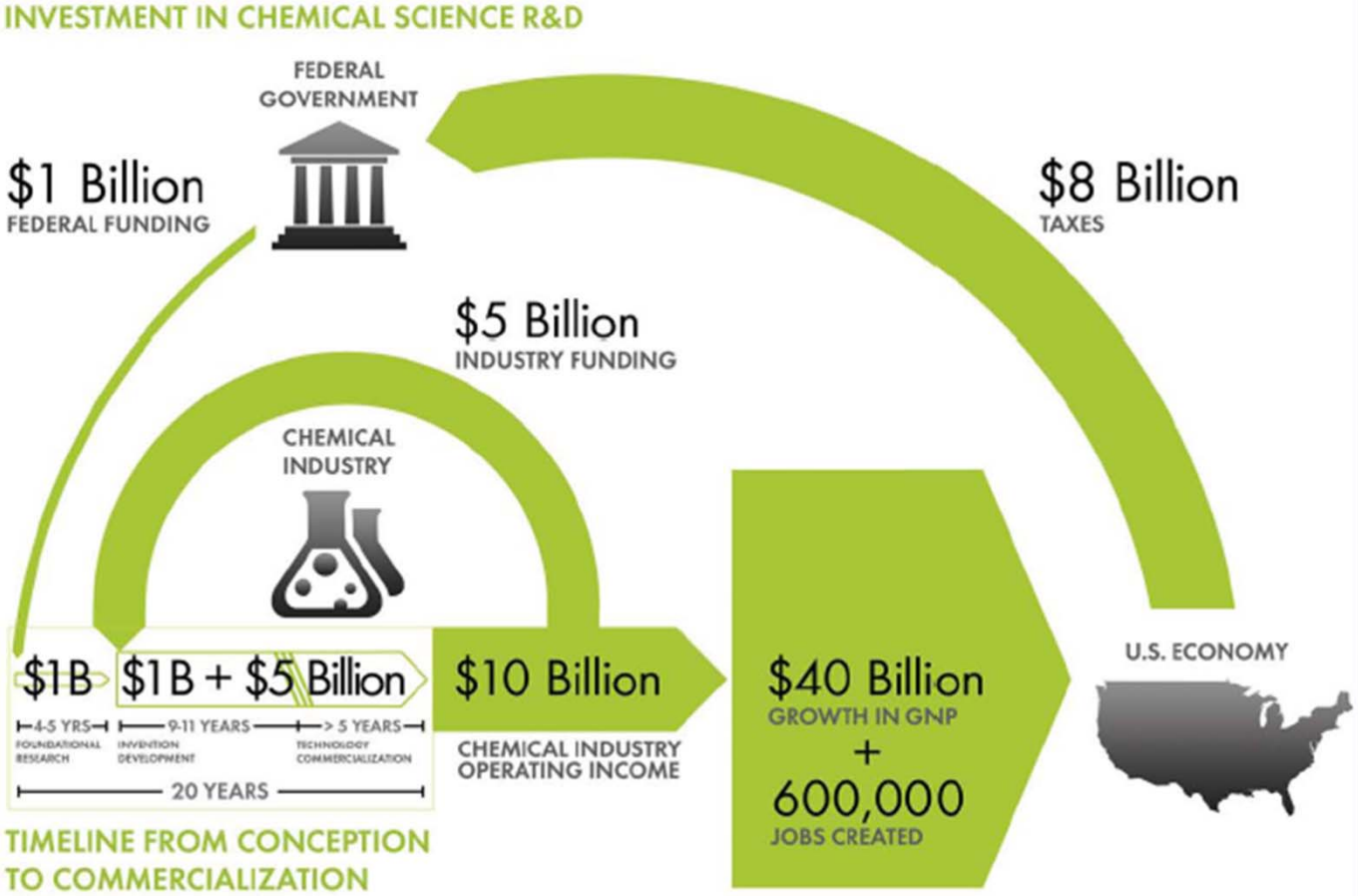
- Mixed data (survey data ... terabytes of streaming data) to capture S&T output (e.g., #papers, # citations, #reads, #clickstreams) but also outcomes (e.g., education, life quality)
- Mixed methods (qualitative to study semantics, scientific argumentation and narrative, meaning, culture, identity, philosophy … quantitative to understand the large scale structure and dynamics)
- Multi-level approach (micro … meso … macro)

Science maps will be employed to interrelate different models, approaches, insights, and to communicate them across disciplinary boundaries to a general audience.

Although S&T data, studies, and insights are at the core of this proposal, we expect the results to transfer to other complex systems.

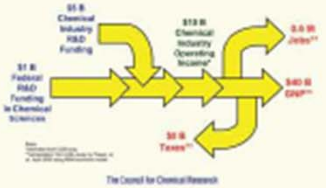
Chemical Research & Development
Powers the U.S. Innovation Engine

Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences



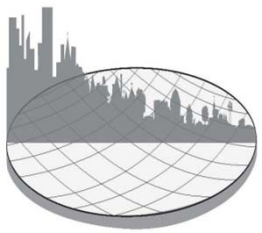
The Council for Chemical Research (CCR)

has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise and clear terms.



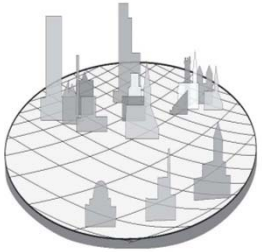
The design shows that an input of $1B in federal investment, leveraged by $5B industry investment, brings new technologies to market and results in $10B of operating income for the chemical industry, $40B growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of $8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles; the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase tax base that in turn is available for investment in basic research.

Council for Chemical Research - Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research - 2009



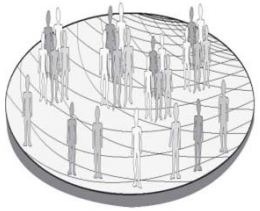
Communities

Modeling knowledge diffusion and production as economic system with global interactions (**Vespignani, Menczer**, **Börner**) • Patterns of S&T usage understood via download activity, science news, social media (**Bollen, Menczer**) • 'Return on Investment' patterns related to science funding (**Bollen, Börner**) • Emergence of technological areas as a cooperative phenomenon (**Vespignani, Menczer, Bollen**)



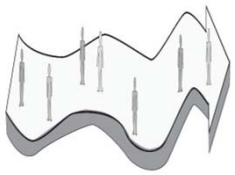
Organizations

Modeling the reputation of and knowledge flows among academia, industry and government institutions (**Börner, Ding, Gieryn**) • Studying the interplay of science and philosophy by tracking signatures of general terms and specialist terminology across different subspecialties, using Google books and news data (**Allen, McDonald**)



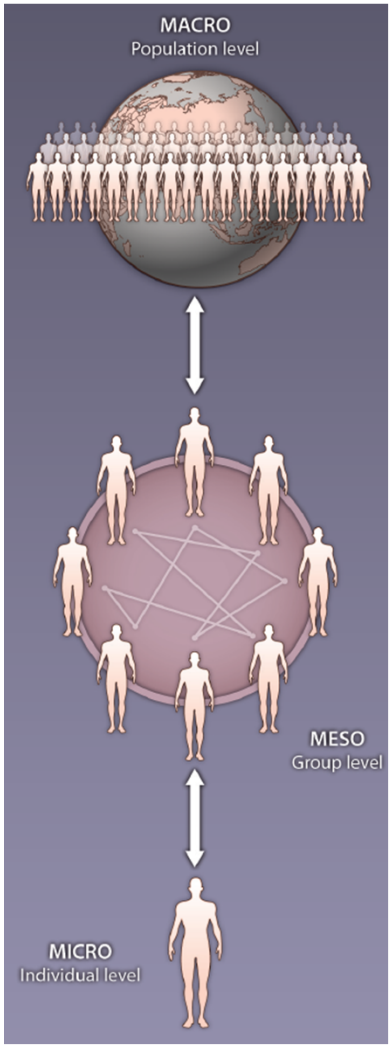
Collaborations

Science of Team Science: how to assemble highly productive teams (**Börner, Milojevic, Ding**) • Modeling the dynamics of collaboration using combined data on multiple types of collaboration (co-authors on publications, co-inventors on patents, co-PIs on grants) (**Börner, Milojevic**) • Assimilation into scientific roles: influence of social networks on response to problem onset, diffusion of innovation (**Pescosolido**)



Individuals

Multipartite network evolution in individual-knowledge-bibliometric space (**Vespignani, Börner**) • Career trajectories: identifications of 'academic superstars' and factors facilitating pathway to scientific success (**Long**, **Sugimoto**, **Sporns**)



TEAM SCIENCE

A Multi-Level Systems Perspective for the Science of Team Science

Katy Börner,[1*] Noshir Contractor,[2] Holly J. Falk-Krzesinski,[3] Stephen M. Fiore,[4] Kara L. Hall,[5] Joann Keyton,[6] Bonnie Spring,[7] Daniel Stokols,[8] William Trochim,[9] Brian Uzzi[10]

This Commentary describes recent research progress and professional developments in the study of scientific teamwork, an area of inquiry termed the "science of team science" (SciTS, pronounced "sahyts"). It proposes a systems perspective that incorporates a mixed-methods approach to SciTS that is commensurate with the conceptual, methodological, and translational complexities addressed within the SciTS field. The theoretically grounded and practically useful framework is intended to integrate existing and future lines of SciTS research to facilitate the field's evolution as it addresses key challenges spanning macro, meso, and micro levels of analysis.

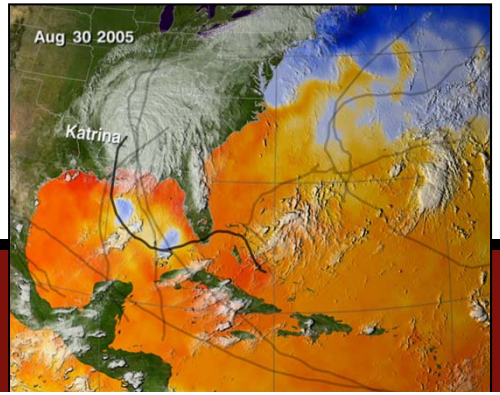
www.ScienceTranslationalMedicine.org 15 September 2010 Vol 2 Issue 49 49cm24

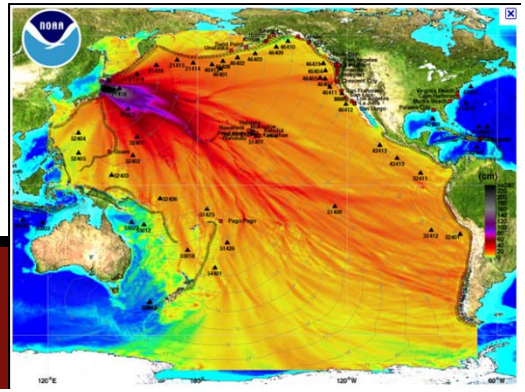
National Science Forecasting Core

We will design and deploy a data aggregation, analysis, and modeling infrastructure that extends VIVO and other semantic web efforts with existing scholarly databases to model the structure and dynamics of science in a comprehensive fashion. This will provide a national science forecast service via online interactive science maps.

A sample interface to seven paper, patent, and funding datasets relevant for the understanding of the research field of sustainability research can be found at http://mapsustain.cns.iu.edu. The ultimate science forecast would draw on these and other datasets such as paper download activity data, S&T relevant twitter messages, RSS job feeds, crowdsourced annotations, etc. as well as survey and qualitative data from ICPSR, OECD, NSF, etc. to render a comprehensive, near-real time view of past and future developments in S&T.

Results will be communicated for different temporal levels, data types, and levels of aggregation.





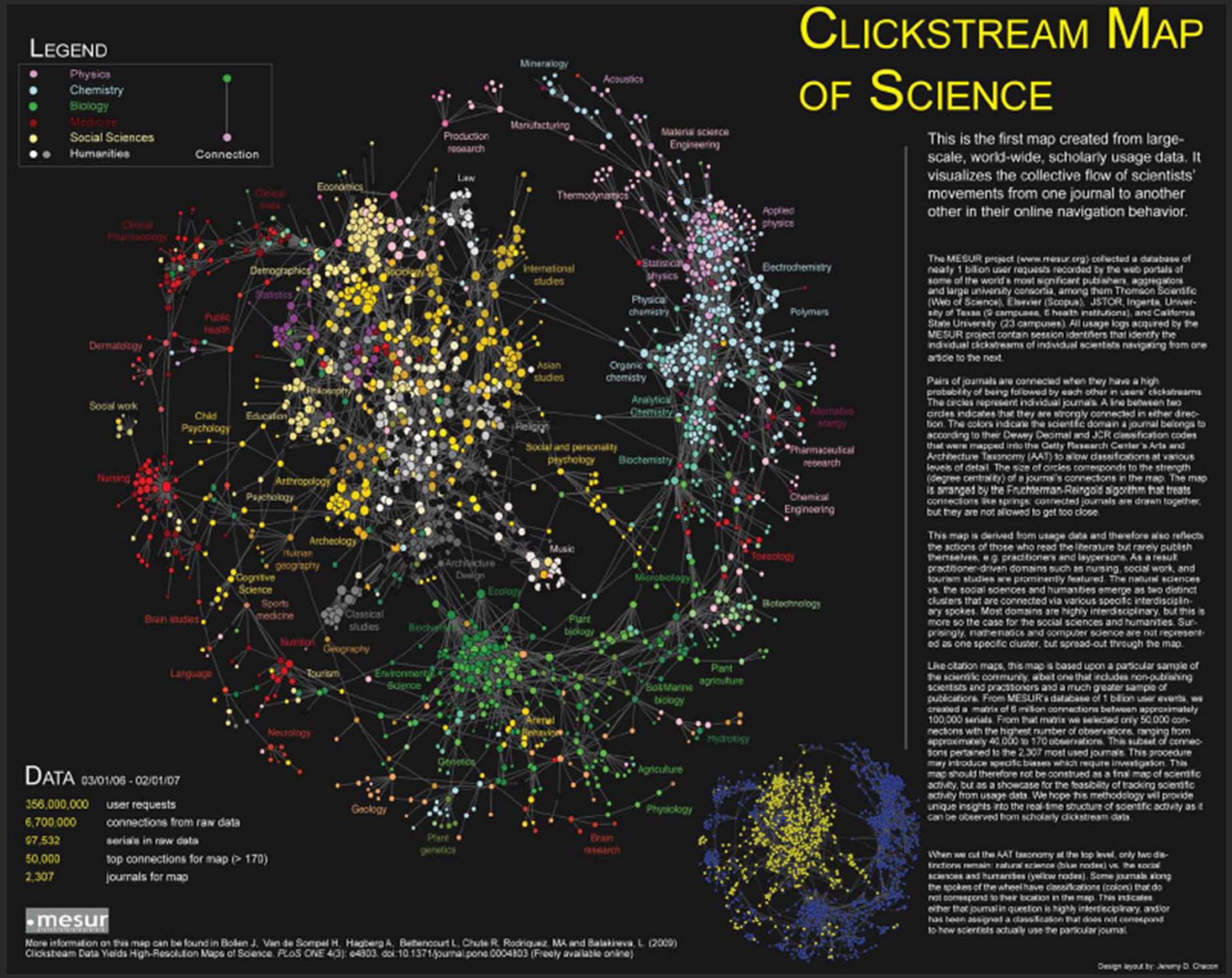
Technology and Computing Core *(Lead by Barnett)*

This core will build on and extend the extensive prior work by IU scientists in the development of high quality, high coverage, unique **scholarly databases** but also **tools and cyberinfrastructures, and community support platforms** for the study of S&T.

Datasets:

- Scholarly Database with 25 million paper, patent, grant records (http://sdb.cns.iu.edu) (**Börner**)
- Mesur with 2 billion user request download data (http://mesur.org) (**Bollen**)
- VIVO Researcher Network (http://vivoweb.org) (**Börner, Ding, McDonald, Barnett**)
- The DIRECT Network (http://direct2experts.org) of faculty expertise from 29 CTSAs (**Barnett**)
- MPACT scholarly family trees data (http://www.ibiblio.org/mpact) (**Sugimoto**)
- Data from commercial and governmental providers, national surveys, and in-vivo experiments will be used to validate and optimize the multi-level models.

The study of outcomes as opposed to output (papers, citations) requires cracking the new data collection problem. Large-scale, high quality, high coverage datasets need to be collected and analyzed in near real-time. Quantitative and qualitative research (results) have to be combined.



A Clickstream Map of Science – Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva - 2008

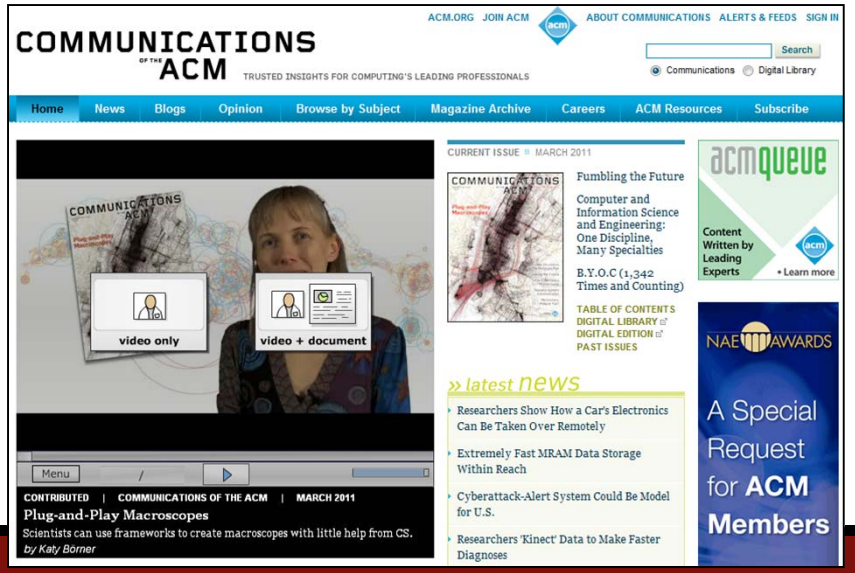
Technology and Computing Core *cont.*

IU **Tools, Service & Cyberinfrastructures**

- Network Workbench Tool (http://nwb.cns.iu.edu) (**Börner, Vespignani, Wasserman**)
- Science of Science Tool (http://sci2.cns.iu.edu) actively used by NSF, NIH, and USDA (**Börner**)
- Epidemiology Tool and Marketplace (**Vespignani, Sherman, Börner**)
- Scholarometer crowdsourcing tool (http://scholarometer.indiana.edu) (**Menczer**)
- Usage-based journal and article ranking tools and visualization (www.mesur.org) (**Bollen**)

IU Hardware

- Big Red with 40.7 PetaFLOPS
- Quarry with 8.9 PetaFLOPS
- 1,000 TB of storage
- 5.7 PB of mass storage
- Research Database Cluster with 50 TB of dedicated storage, all aligned with HIPAA
- TeraGrid (http://teragrid.org)
- Virtual Reality Theaters, display walls
- Founding member of HUBZero Consortium



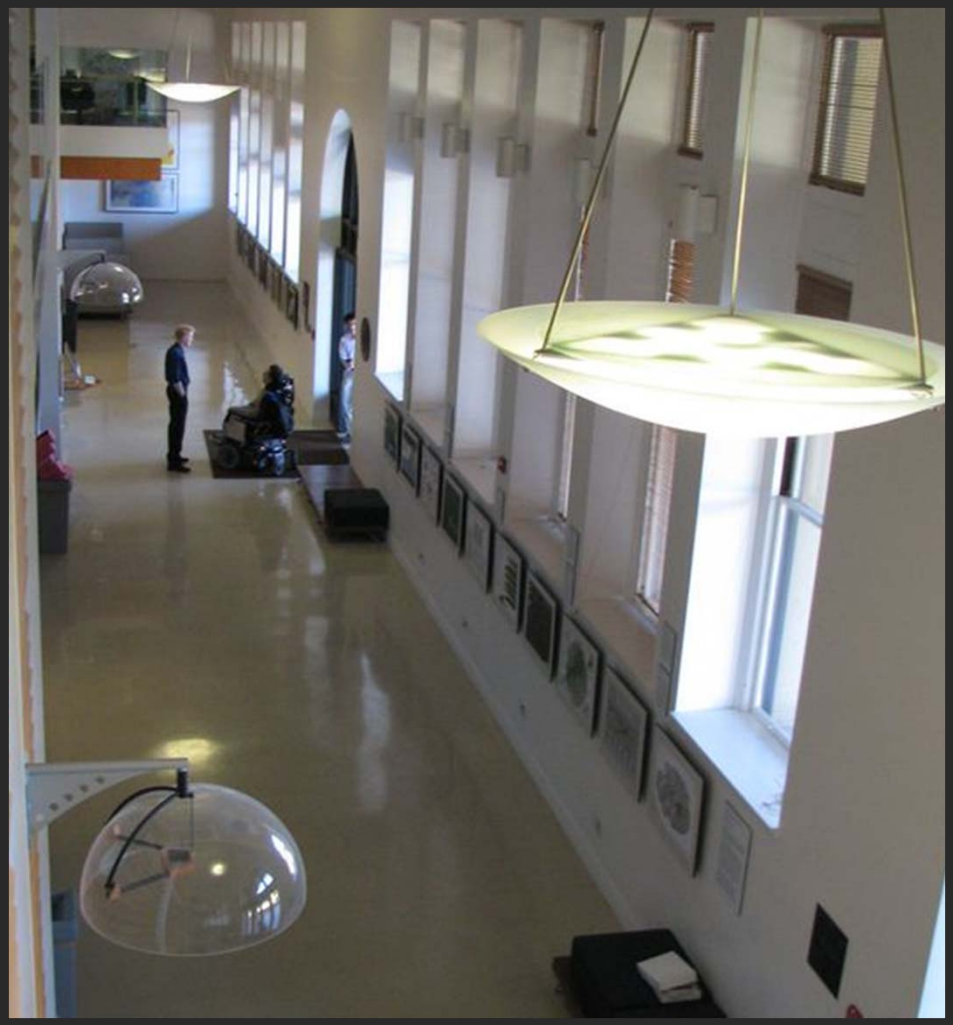
Education and Dissemination Core *(Lead by Pescosolido)*

Partners: James Moody and the Network Analysis Center at Duke University and David Lazer and the Program on Networked Governance at Harvard University

This core will develop the first national, collaborative effort to develop state-of-the art materials that translate S&T theory, approaches, and predictions for educators, practitioners, policy makers, and the general public.

SnetS (Science of Network Science) Portal will use the IU developed HUBzero technology to develop a model national curriculum at undergraduate and graduate levels for Network Science, provide cutting edge teaching materials, develop the platform for streaming live video and Webinars.

NSF-supported education portals exist for other sciences, e.g., http://www.biosciednet.org/portal/, http://www.merlot.org/merlot/index.htm) but not yet for the transdisciplinary science of networks.



Results will also be disseminated via the international Mapping Science Exhibit, here on display at Wallenberg Hall, Stanford University in 2010 (http://scimaps.org)

National Collaborators

The proposed STC will build on and expand long-term, close collaborations with leading national and international experts from academia, industry, and government:

Academia:
Mark E. J. Newman, Gorge Alter (ICPSR), University of Michigan
Noshir Contractor, Luis A. N. Amaral, Brian Uzzi, Northwestern University
James Moody, Jonathon Cummings, Duke University
Lee Giles, Caroline Wagner, Pennsylvania State University
Albert-László Barabási, David Lazer, Northeastern University
Woody Powell, Jeffrey Herr, Stanford University

Industry:
Google, Yahoo, Elsevier, Mendeley, Rand, ESRI, SRI International

Government:
NSF, NIH, USDA, IEDC, NIST, OSTP, Private Foundations

International Collaborators

Yves Gingras, Canada Research chair at Université du Québec à Montréal (UQAM) and Scientific Director of the Observatoire des sciences et des technologies, **Canada**

Abel Packer, BIREME/OPS/OMS lead of Lattes National Researcher Infrastructure, **Brazil**

Yuko Harayama, Directorate for Science, Technology and Industry, OECD, David Chavalarias, CNRS/Ecole Polytechnique, Co-director of the Complex Systems Institute of Paris, Bruno Latour, Professor and Vice-President for Research at Sciences Po Paris, **France**

Prof. Stefan Hornbostel, IFQ Institut für Forschungsinformation und Qualitätssicherung, Bonn; Andreas Trepte, Max-Planck-Society, Munich, Andreas Hotho, Universität Würzburg, and Gerd Stumme, Universität Kassel, **Germany**

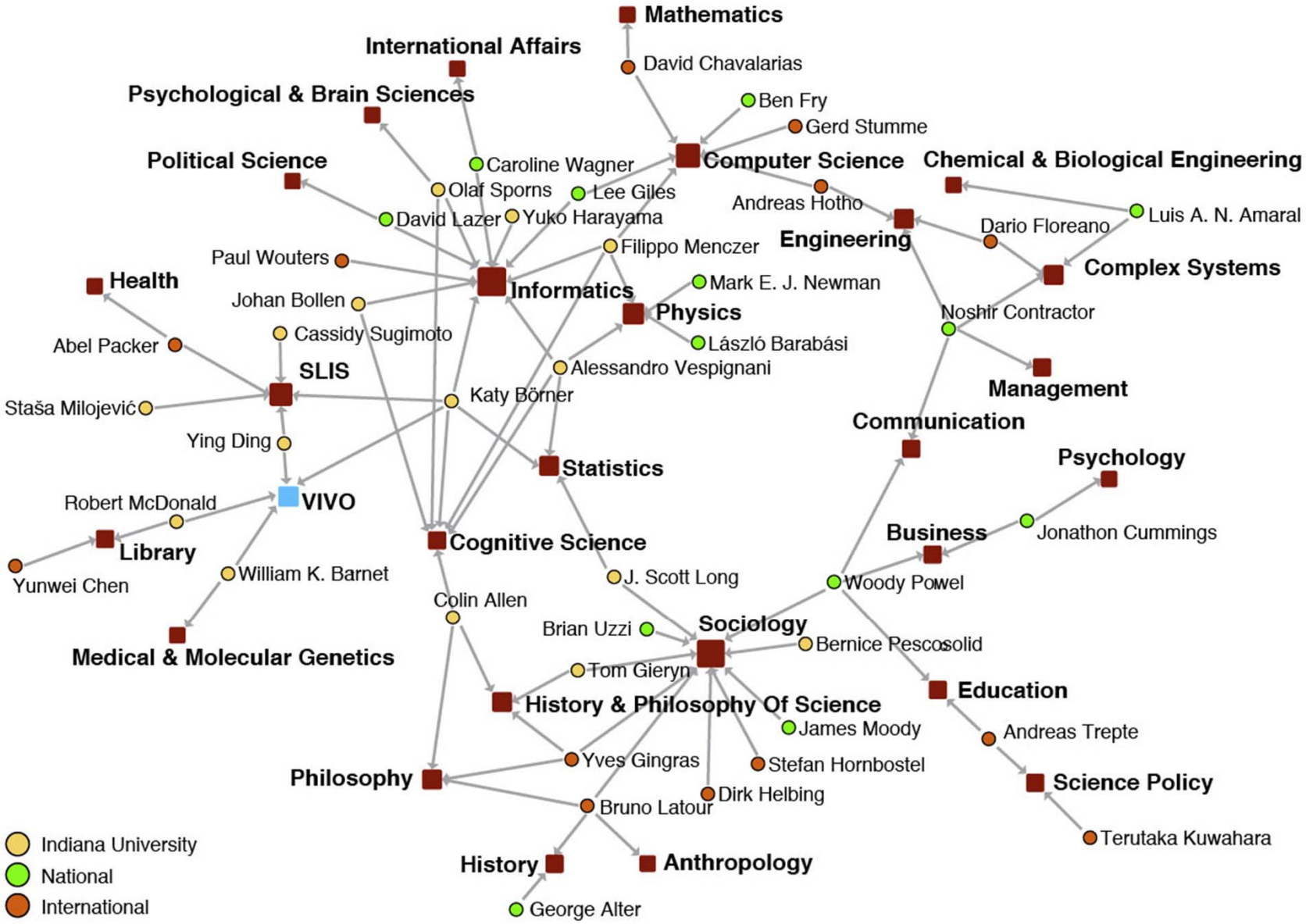
Dirk Helbing , ETH, Dario Floreano, EPFL, **Switzerland**

Paul Wouters, Leiden University and Director of CWTS, **Holland**

Yunwei Chen, National Science Library, Chinese Academy of Science, **China**

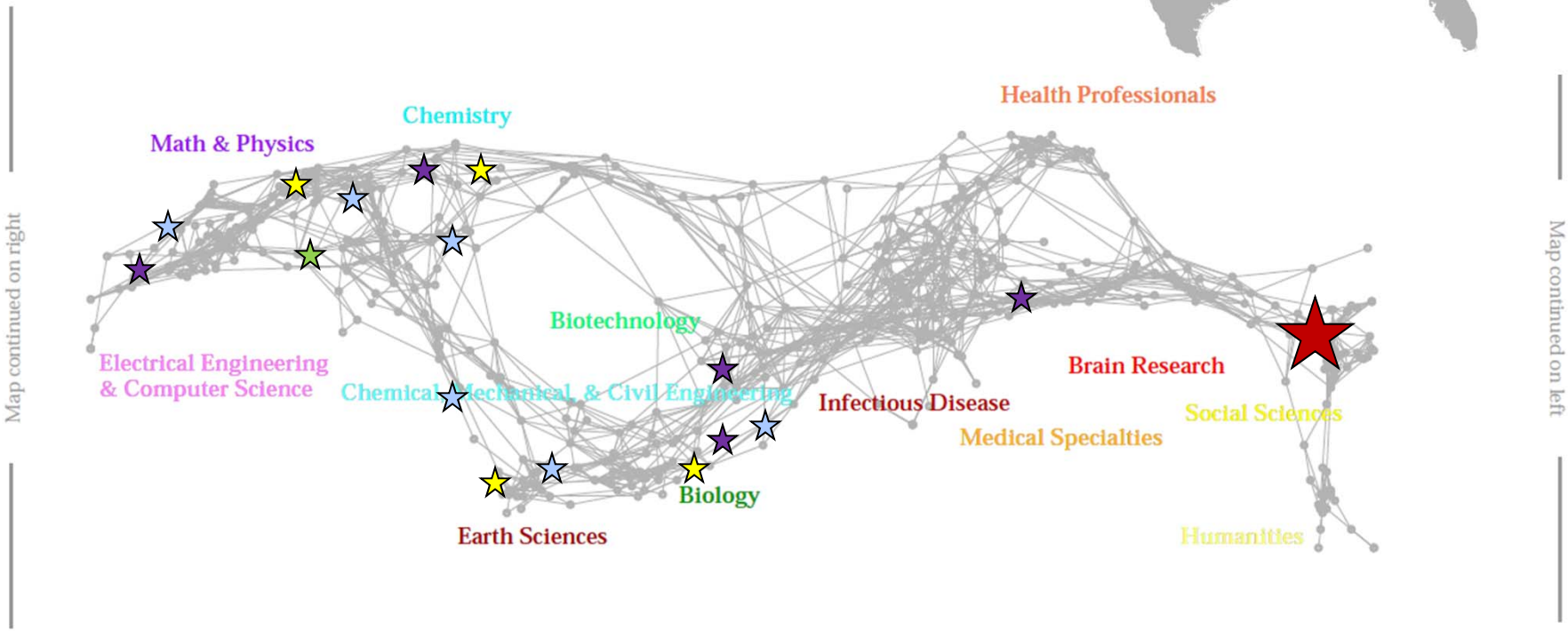
Terutaka Kuwahara, Director General, National Institute of S&T Policy, **Japan**

Collaborators and Fields of Science



Position in Geospatial Space and in the Landscape of Science

STCs awarded in 2002 | 2005 | 2006 | 2010



Science & Technology Forecasts @ Times Square in 2016