# Science of Science Management

## Knowledge Discovery and Management Topics

October 2-3, 2008

NIH Campus

Bethesda, MD

# SCIENCE MANAGEMENT USING ORGANIZATIONAL KNOWLEDGE: The Wisdom of Crowds*

Science of Science Management Conference
National Institutes of Health

Mary Kane, President
Concept Systems, Incorporated

© 2008 Concept Systems, Inc.

* *Jim Surowiecki, 2004*

# *Focus on Knowledge for Management*

- "Knowledge management concepts and systems are both an opportunity for science managers to improve planning and management, and a tool for science research that requires increased expertise in the management process."
- We need *prospective* thinking on what is required for science advancement, as well as retrospective analysis research data
- Independent contribution and consideration of knowledge by individuals in groups can yield higher quality decisions than those of selected experts only: the *wisdom of crowds*

*Meeting Highlights: Field Specific Conference Calls. Provided by SOSM Conference Staff*


© 2008 Concept Systems, Inc.

# *Knowledge Development to Support Research Programs Management*

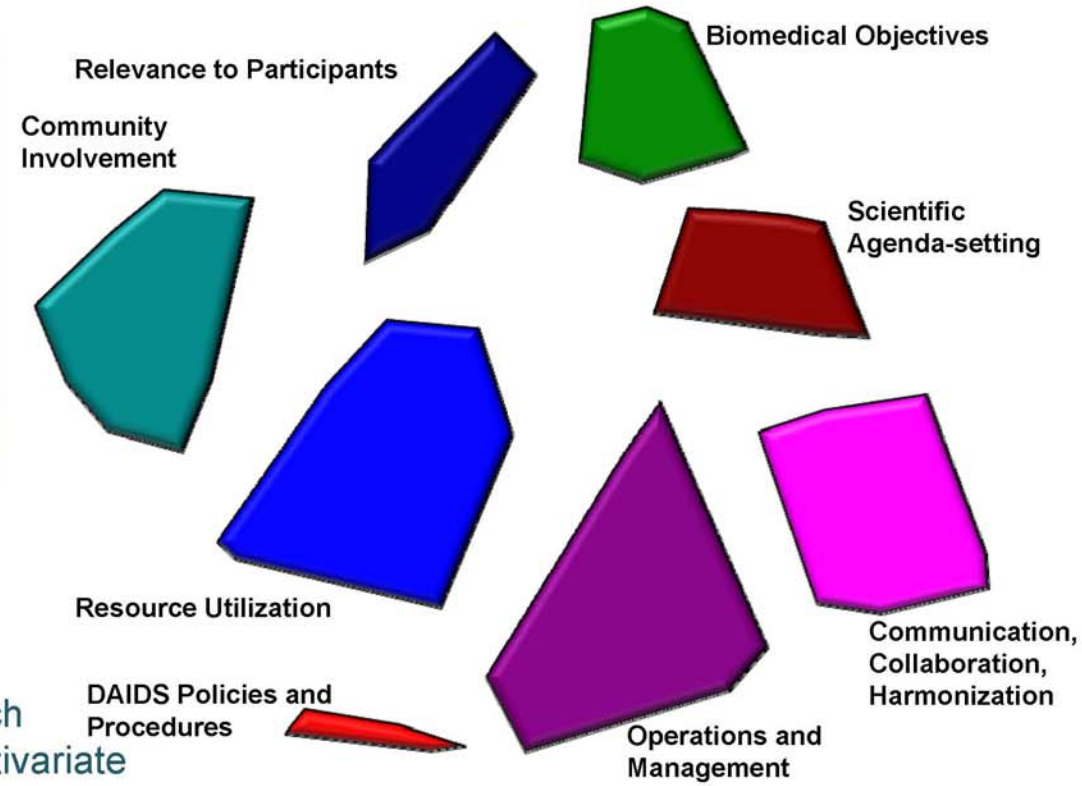
- Our emphasis: the *management* of science initiatives and the organizations that support them, to:
  - Establish commonly understood, purposeful research agenda
  - Accumulate organize and apply knowledge from those who know
  - Identify and reduce gaps in knowledge to improve use and results
  - Enable focused resources and assessment of results based on scientific priorities
- *Collaborative* construction of emergent concepts to
  - Tackle thorny issues (from defining what we mean by science management to systematic support of discoveries)
  - Get agreement on research and operational priorities
  - Align operations to support framework priorities

© 2008 Concept Systems, Inc.

# *"Coordinated Clinic Trials Research Networks Will Be Successful If…"*

**Relative Importance, All Participants**

Biomedical Objectives
Relevance to Participants
Scientific Agenda Setting
Community Involvement
Resource Utilization
DAIDS Policies and Procedures
Collaboration, Communication, Harmonization
Operations and Management



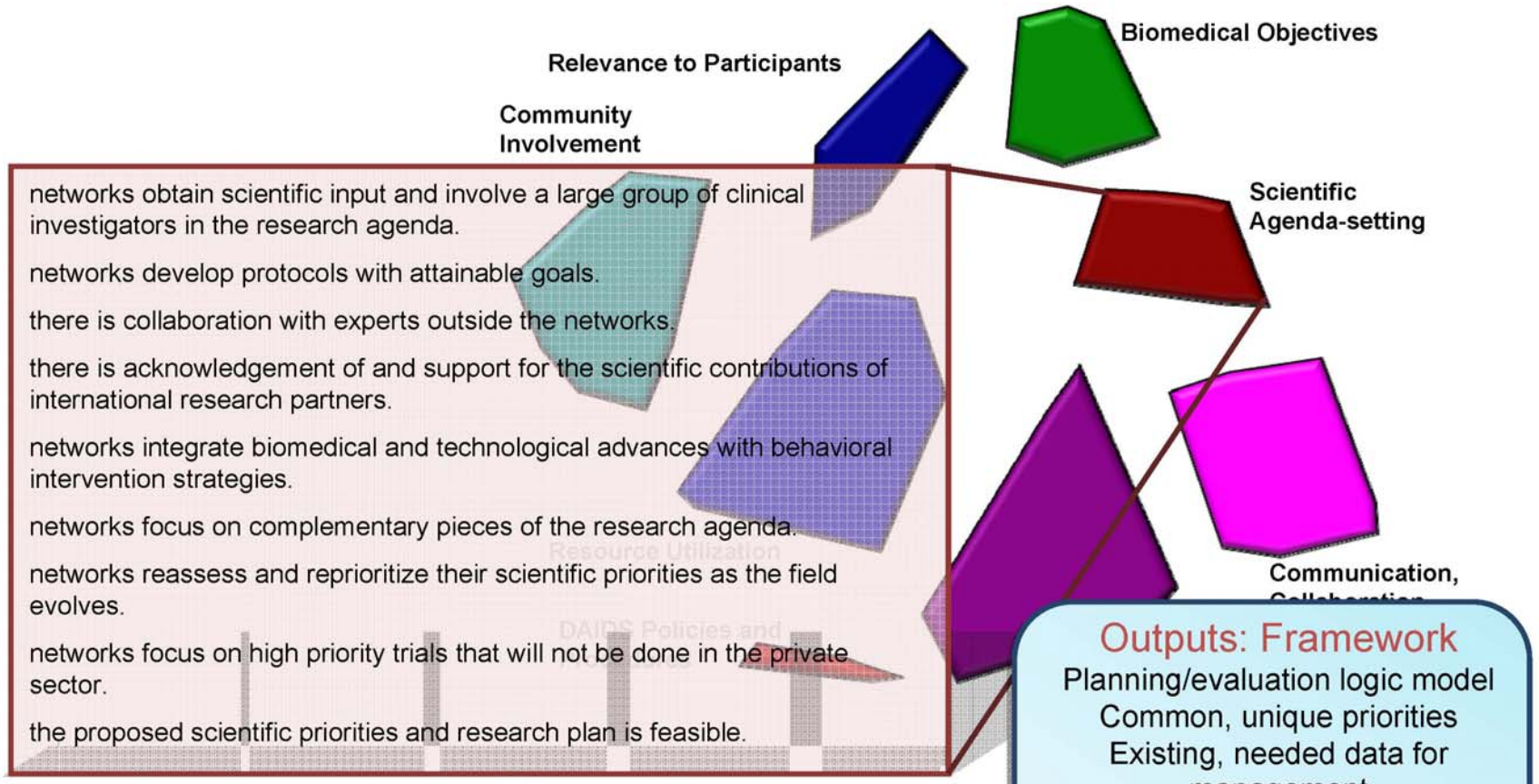

## Concept Mapping

Mixed methods approach
Group process and multivariate statistical applications
Systematic, structured and broadly applicable

*Kane and Trochim (2007) Concept Mapping for Planning and Evaluation.*


© 2008 Concept Systems, Inc.

# *"Coordinated Clinic Trials Research Networks Will Be Successful If…"*




networks obtain scientific input and involve a large group of clinical investigators in the research agenda.

networks develop protocols with attainable goals.

there is collaboration with experts outside the networks.

there is acknowledgement of and support for the scientific contributions of international research partners.

networks integrate biomedical and technological advances with behavioral intervention strategies.

networks focus on complementary pieces of the research agenda.

networks reassess and reprioritize their scientific priorities as the field evolves.

networks focus on high priority trials that will not be done in the private sector.

the proposed scientific priorities and research plan is feasible.

**Outputs: Framework**

Planning/evaluation logic model
Common, unique priorities
Existing, needed data for management
Supporting cross-network planning
Complementary scientific agenda development


© 2008 Concept Systems, Inc.

# *Utility*

- Questions like: *What will or should a system to manage science contain, support, achieve?*
- "Wisdom of crowds:" each here is expert; as a group we are new at this
- Link knowledge and positional priorities to create "common sense"
- Use agreed-upon knowledge framework rooted in scientific objectives and methods
- Result: greater than the sum of its parts

© 2008 Concept Systems, Inc.

# Science of Science Management

## Knowledge Discovery and Management Topics

October 2-3, 2008

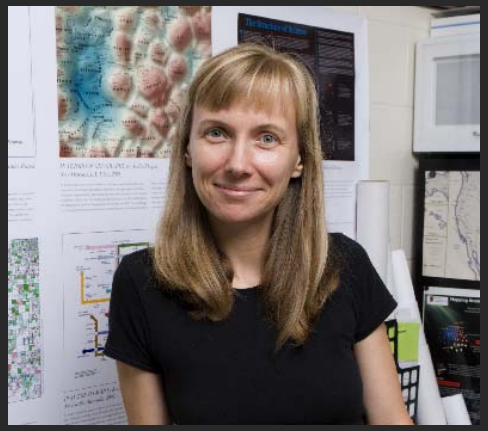
NIH Campus

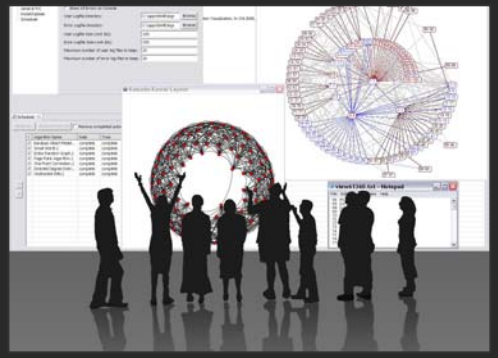
Bethesda, MD

# Computational Scientometrics

*Studying Science by Scientific Means*

**Dr. Katy Börner**
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
*katy@indiana.edu*





*Science of Science Management Meeting*
*NIH Office of Portfolio Analysis and Strategic Initiatives (OPASI)*
*National Institutes of Health, Bethesda, MD*
*October 2 & 3, 2008*

"Features that distinguish science from pseudoscience are repeatability, economy, menuration, heuristics, and consilience."
*E. O. Wilson in Consilience: The Unity of Knowledge (1998)*

# Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

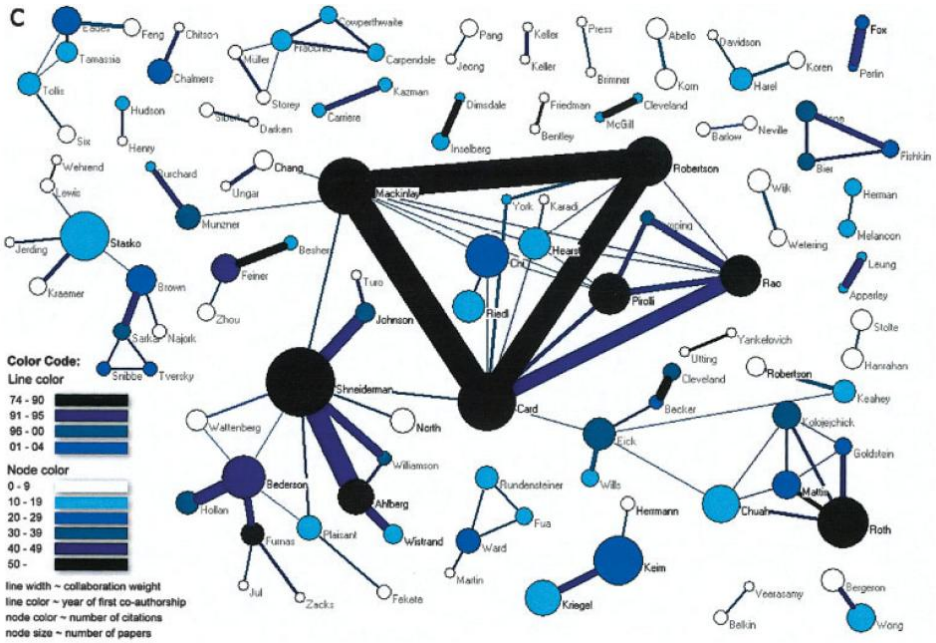
*Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.*

**Research question:**

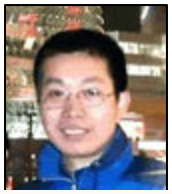
- Is science driven by prolific single experts or by high-impact co-authorship teams?

**Contributions:**

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.



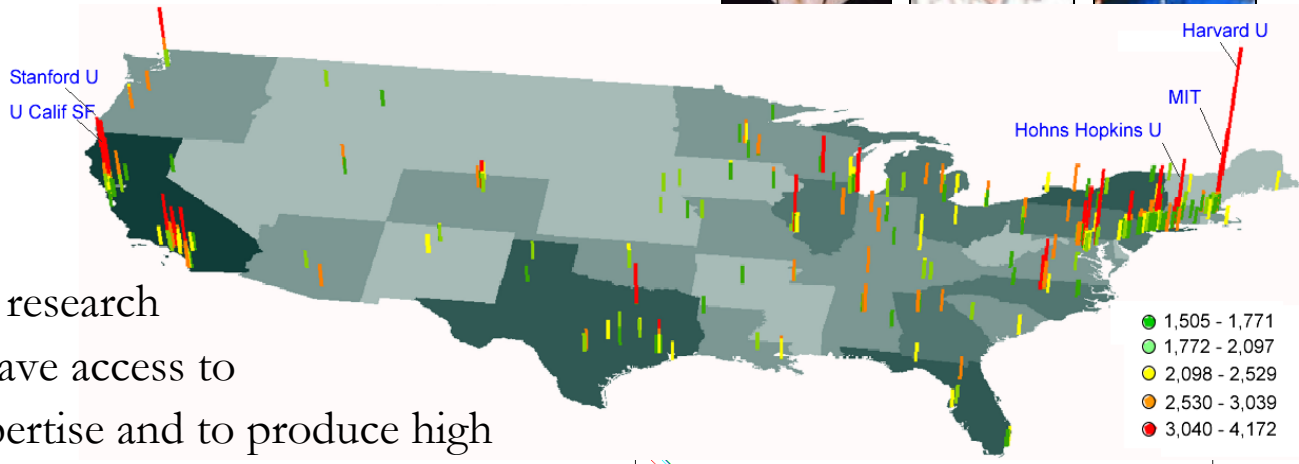
# Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions



*Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.*

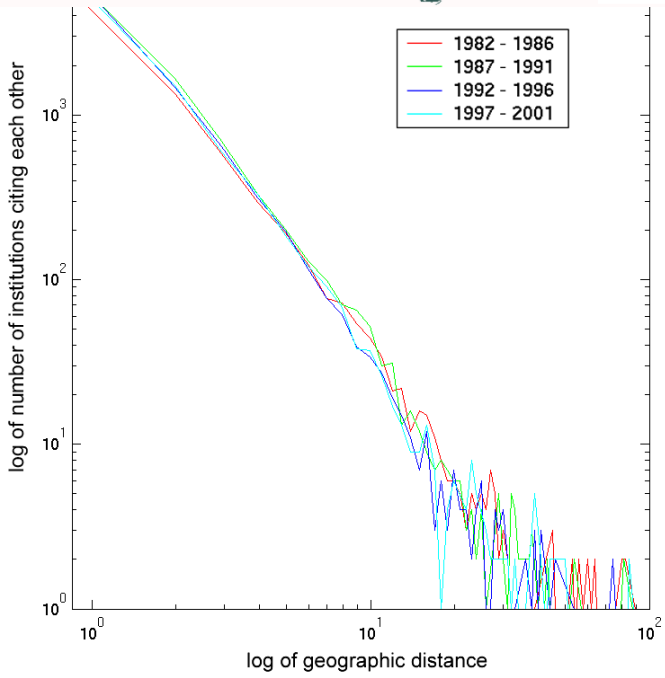
**Research questions:**

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?




**Contributions:**

- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.



# Mapping Topic Bursts

*Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.*

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

**Insight gained:**

Most bursts occur before words experience wide-spread usage.

RNA
CLONING
receptors
gene expression regulation
molecular weight
messenger/genetics
89' kinetics
cultured
cells
inhibition
polymerase chain reaction
growth
antigen
96' transfection
activation
restriction mapping
93' sequence homology
EXPRESSION
protein binding
models
92' & 94' comparative study
viral
97' induction
protein
signal transduction
CELL LINE
nucleic acid hybridization
90' antibodies
89' GENES
2000' tumor cells
87' structural
88' inbred strains
89' plasmids
blotting
rats
MOLECULAR SEQUENCE DATA
99' mutagenesis
nucleic acid
binding
apoptosis
mutation
growth-factor
95' base sequence
amino acid
amino acid sequence
in vitro
receptor
sequence
gene expression

**Color Code**

82 - 85
86 - 89
90 - 93
94 - 97
98 - 01

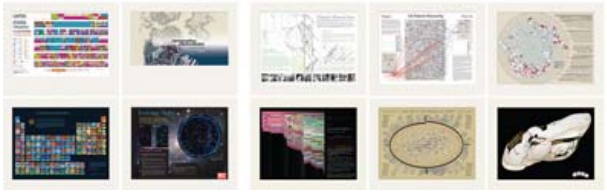
circle size ~ burst weight
cicle color ~ burst onset
ring color ~ year of max word count
years of 2nd and 3rd burst are given in color

# Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org/

**The Power of Maps (2005)**

**The Power of Reference Systems (2006)**



**The Power of Forecasts (2007)**

**Science Maps for Economic Decision Makers (2008)**

**Science Maps for Science Policy Makers (2009)**
**Science Maps for Scholars (2010)**
**Science Maps as Visual Interfaces to Digital Libraries (2011)**
**Science Maps for Kids (2012)**
**Science Forecasts (2013)**

**How to Lie with Science Maps (2014)**

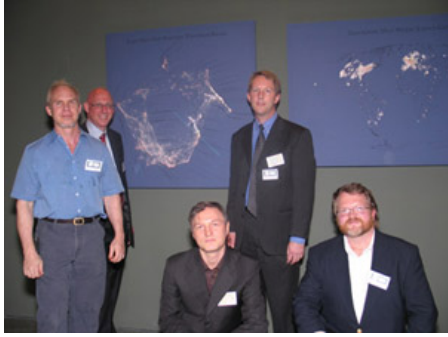
Exhibit has been shown in 49 venues on four continents. Also at
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA.
- National Research Council in Ottawa, Canada, April 3-Aug. 29, 2008.
- Chinese Academy of Sciences, China, May 17-Nov. 15, 2008.

ORDER MAPS

# Illuminated Diagram Display

*W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007)*

*Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007.*



**Large-scale, high resolution prints illuminated via projector or screen.**

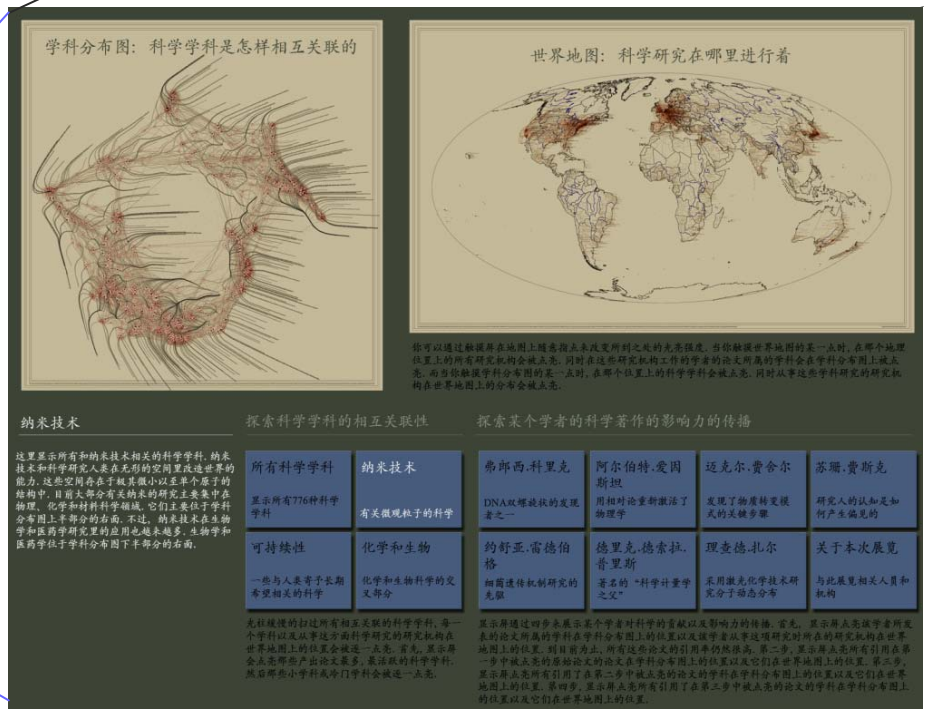
## Questions:

- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

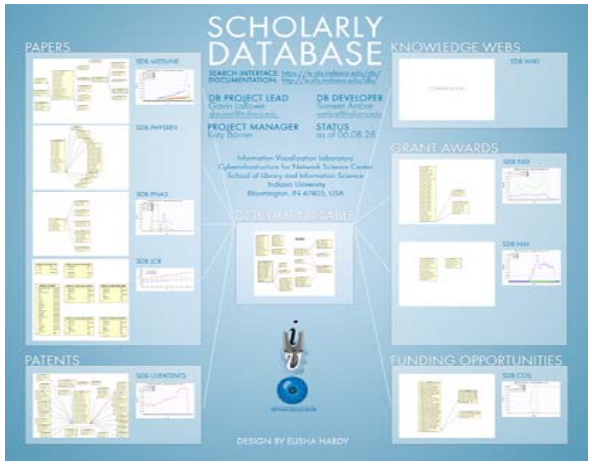
**Interactive touch panel.**

## Contributions:

- Interactive, high resolution interface to access and make sense of data about scholarly activity.



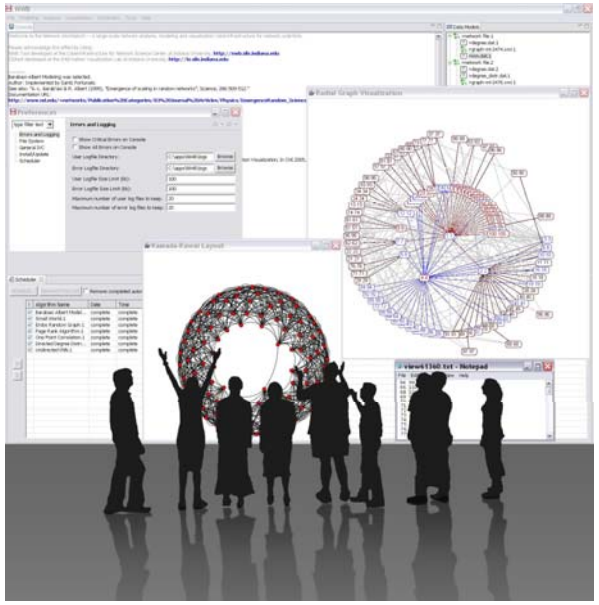
# Cyberinfrastructures for a Science of Science




Scholarly Database of 18 million scholarly records
https://sdb.slis.indiana.edu

James S. McDonnell Foundation

Information Visualization Cyberinfrastructure
http://iv.slis.indiana.edu



Network Workbench Tool and Community Wiki
*NEW* Scientometrics plugins
http://nwb.slis.indiana.edu

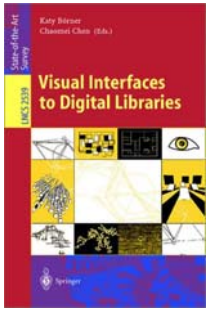
Epidemics Cyberinfrastructure
http://epic.slis.indiana.edu/

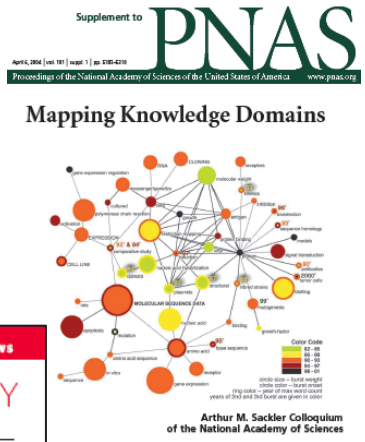
# Relevant References

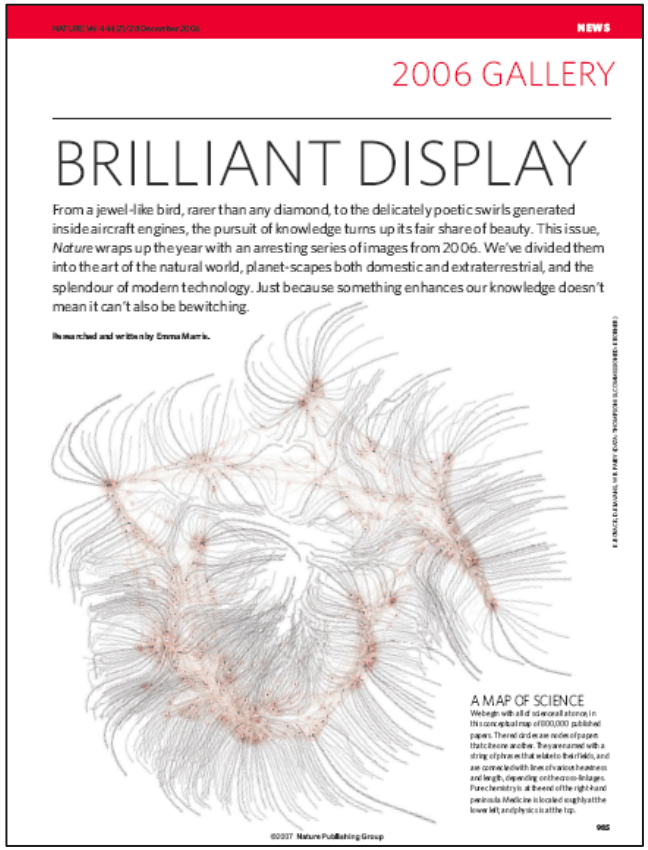

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255.
http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).
http://www.pnas.org/content/vol101/suppl_1/

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.
http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf



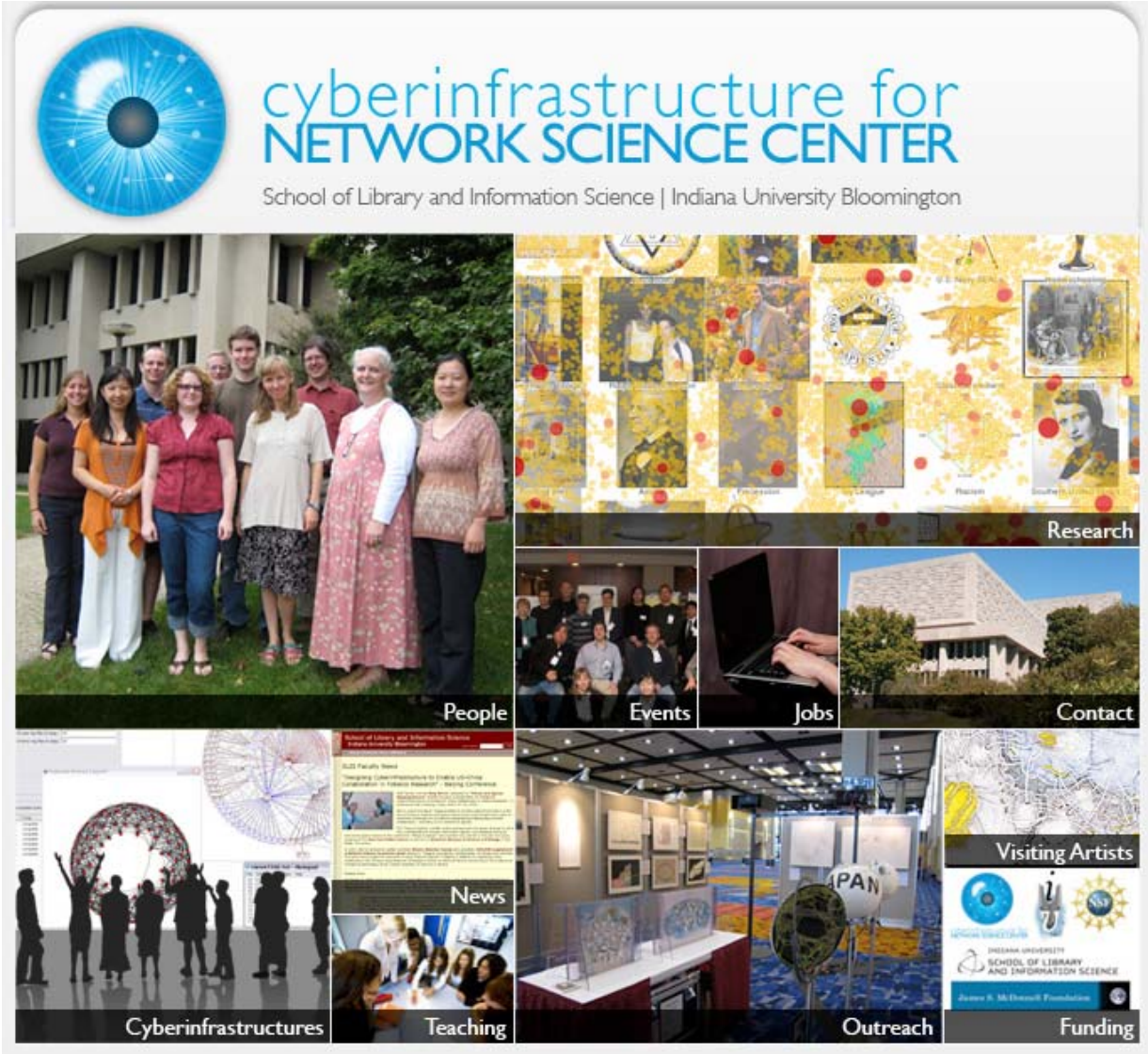



## Science of Science Studies Challenges

- (Different) user groups and their needs and priorities have to be identified.
- Major terms, e.g., 'impact' or 'interdisciplinary', need to be defined and operationalized.
- (Standard) datasets have to be federated and made available so that science of science studies can be replicated.
- A common science of science cyberinfrastructure is desirable.
- There is a need for well documented case studies and evaluation.
- Major results, good practices, and new datasets/tools have to be communicated widely.

## Note that

- Science of science studies can augment but not replace human judgment.
- Incomplete, low coverage data typically leads to low quality results.
- Studies performed using proprietary tools and/or proprietary data are hard to replicate.




Papers, maps, cyberinfrastructures, talks, press are linked from
*http://cns.slis.indiana.edu*

# Science of Science Management

## Knowledge Discovery and Management Topics

October 2-3, 2008

NIH Campus

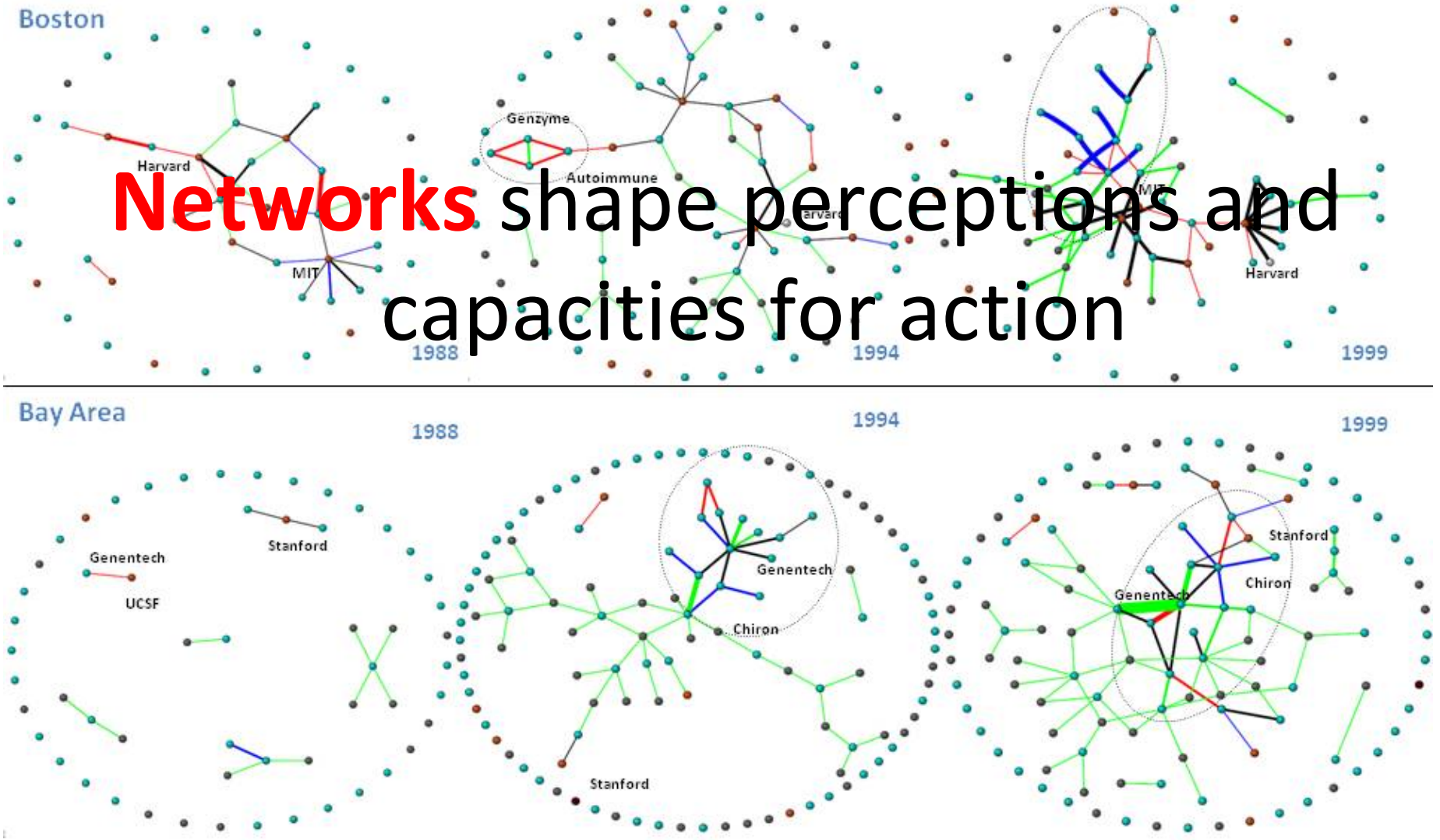
Bethesda, MD

# Networks & Institutions in the Utilization of Science

Jason Owen-Smith

University of Michigan

# Boston and Bay Area Networks, 1988, 1994, 1999




**Networks** shape perceptions and capacities for action

Owen-Smith et al 2002; Owen-Smith & Powell 2004, 2006

# Networks

**Definition:** Concrete patterns of interaction among participants in a recognizable field of endeavor

**Mechanisms of Action:** Status Signaling (Podolny 1993; Phillips & Zuckerman 2001), Resource/Information Transfer (Granovetter 1985; Burt 1992), Social Influence (Coleman et al 1966; Mizruchi 1992)

**Life Science Examples:** Flows of research materials among scientists, flows of students/post-docs across labs, Co-authorship, overlapping membership on study section panels

**Key Idea:** Much of what accounts for the success of a new discovery is the networks in place at the time of its diffusion

**Institutions** define rules, roles and incentives

# Institutions

**Definition:** Shared expectations that define goals, rewards, and standards of appropriate behavior in a recognizable field of endeavor.

**Mechanisms of Action:** assumptions (Zucker 1977), norms (Meyer & Rowan 1977), regulations (DiMaggio & Powell 1983)

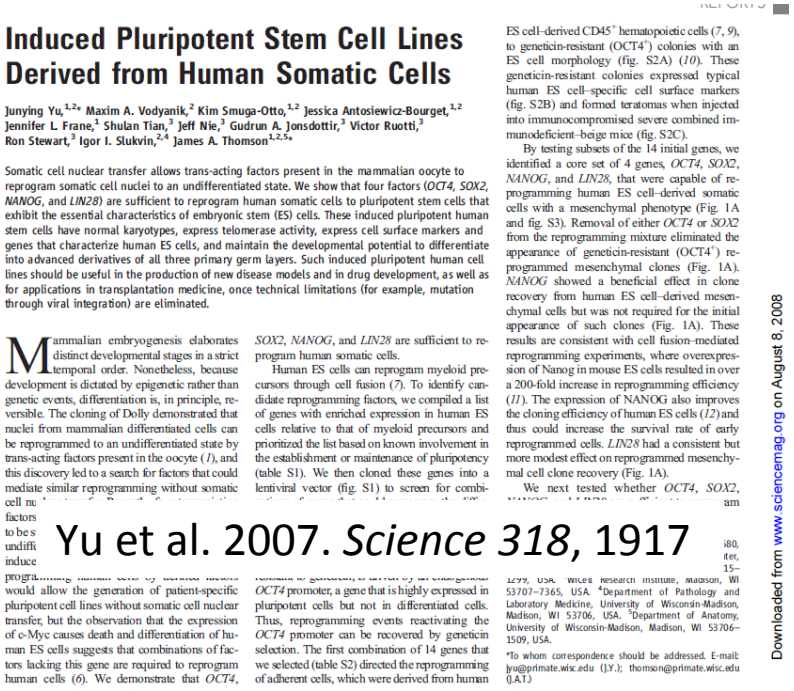
**Life Science Examples:** Peer Review, Authorship standards v. Inventorship rules, record keeping practices (lab notebooks), COI/COC policies

**Key Idea:** Institutions are the formal and informal rules of the game. What makes different stakeholders distinct is their orientation toward disparate institutions.

**Institutions** help determine why people want information on new discoveries

**Networks** help determine how they access and understand that information

Understanding knowledge utilization and diffusion requires that we understand the joint effects of networks and institutions.



## Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells

Junying Yu,[1,2]* Maxim A. Vodyanik,[2] Kim Smuga-Otto,[1,2] Jessica Antosiewicz-Bourget,[1,2] Jennifer L. Frane,[1] Shulan Tian,[3] Jeff Nie,[3] Gudrun A. Jonsdottir,[3] Victor Ruotti,[3] Ron Stewart,[3] Igor I. Slukvin,[2,4] James A. Thomson[1,2,5]*

Somatic cell nuclear transfer allows trans-acting factors present in the mammalian oocyte to reprogram somatic cell nuclei to an undifferentiated state. We show that four factors (*OCT4*, *SOX2*, *NANOG*, and *LIN28*) are sufficient to reprogram human somatic cells to pluripotent stem cells that exhibit the essential characteristics of embryonic stem (ES) cells. These induced pluripotent human stem cells have normal karyotypes, express telomerase activity, express cell surface markers and genes that characterize human ES cells, and maintain the developmental potential to differentiate into advanced derivatives of all three primary germ layers. Such induced pluripotent human cell lines should be useful in the production of new disease models and in drug development, as well as for applications in transplantation medicine, once technical limitations (for example, mutation through viral integration) are eliminated.

Mammalian embryogenesis elaborates distinct developmental stages in a strict temporal order. Nonetheless, because development is dictated by epigenetic rather than genetic events, differentiation is, in principle, reversible. The cloning of Dolly demonstrated that nuclei from mammalian differentiated cells can be reprogrammed to an undifferentiated state by trans-acting factors present in the oocyte (*1*), and this discovery led to a search for factors that could mediate similar reprogramming without somatic cell nu… factors to be s… undiffe… induce… progra… would allow the generation of patient-specific pluripotent cell lines without somatic cell nuclear transfer, but the observation that the expression of c-Myc causes death and differentiation of human ES cells suggests that combinations of factors lacking this gene are required to reprogram human cells (*6*). We demonstrate that *OCT4*, *SOX2*, *NANOG*, and *LIN28* are sufficient to reprogram human somatic cells.

Human ES cells can reprogram myeloid precursors through cell fusion (*7*). To identify candidate reprogramming factors, we compiled a list of genes with enriched expression in human ES cells relative to that of myeloid precursors and prioritized the list based on known involvement in the establishment or maintenance of pluripotency (table S1). We then cloned these genes into a lentiviral vector (fig. S1) to screen for combi… *OCT4* promoter, a gene that is highly expressed in pluripotent cells but not in differentiated cells. Thus, reprogramming events reactivating the *OCT4* promoter can be recovered by geneticin selection. The first combination of 14 genes that we selected (table S2) directed the reprogramming of adherent cells, which were derived from human ES cell–derived CD45$^+$ hematopoietic cells (*7*, *9*), to geneticin-resistant (OCT4$^+$) colonies with an ES cell morphology (fig. S2A) (*10*). These geneticin-resistant colonies expressed typical human ES cell–specific cell surface markers (fig. S2B) and formed teratomas when injected into immunocompromised severe combined immunodeficient–beige mice (fig. S2C).

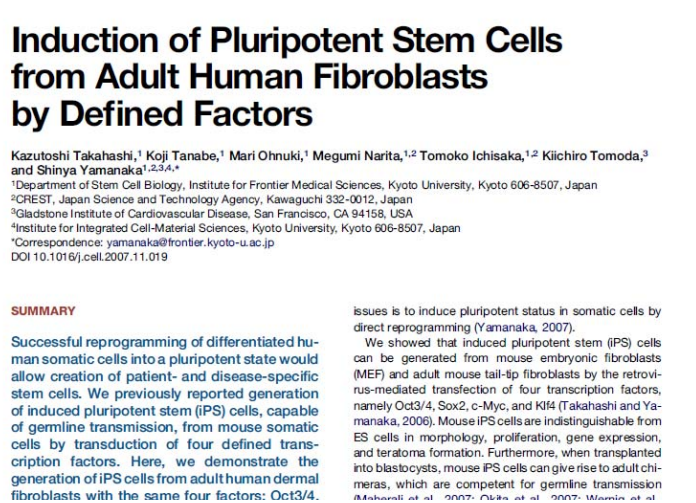
By testing subsets of the 14 initial genes, we identified a core set of 4 genes, *OCT4*, *SOX2*, *NANOG*, and *LIN28*, that were capable of reprogramming human ES cell–derived somatic cells with a mesenchymal phenotype (Fig. 1A and fig. S3). Removal of either *OCT4* or *SOX2* from the reprogramming mixture eliminated the appearance of geneticin-resistant (OCT4$^+$) reprogrammed mesenchymal clones (Fig. 1A). *NANOG* showed a beneficial effect in clone recovery from human ES cell–derived mesenchymal cells but was not required for the initial appearance of such clones (Fig. 1A). These results are consistent with cell fusion–mediated reprogramming experiments, where overexpression of Nanog in mouse ES cells resulted in over a 200-fold increase in reprogramming efficiency (*11*). The expression of NANOG also improves the cloning efficiency of human ES cells (*12*) and thus could increase the survival rate of early reprogrammed cells. *LIN28* had a consistent but more modest effect on reprogrammed mesenchymal cell clone recovery (Fig. 1A).

We next tested whether *OCT4*, *SOX2*, …am

…i80, …ter, 15– 1299, USA. [2]WiCell Research Institute, Madison, WI 53707–7365, USA. [4]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA. [5]Department of Anatomy, University of Wisconsin-Madison, Madison, WI 53706–1509, USA.
*To whom correspondence should be addressed. E-mail: jyu@primate.wisc.edu (J.Y.); thomson@primate.wisc.edu (J.A.T.)

Downloaded from www.sciencemag.org on August 8, 2008

Yu et al. 2007. *Science 318*, 1917



## Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors

Kazutoshi Takahashi,[1] Koji Tanabe,[1] Mari Ohnuki,[1] Megumi Narita,[1,2] Tomoko Ichisaka,[1,2] Kiichiro Tomoda,[3] and Shinya Yamanaka[1,2,3,4,*]
[1]Department of Stem Cell Biology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan
[2]CREST, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan
[3]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
[4]Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 606-8507, Japan
*Correspondence: yamanaka@frontier.kyoto-u.ac.jp
DOI 10.1016/j.cell.2007.11.019

SUMMARY

Successful reprogramming of differentiated human somatic cells into a pluripotent state would allow creation of patient- and disease-specific stem cells. We previously reported generation of induced pluripotent stem (iPS) cells, capable of germline transmission, from mouse somatic cells by transduction of four defined transcription factors. Here, we demonstrate the generation of iPS cells from adult human dermal fibroblasts with the same four factors: Oct3/4, …tent cell-specific genes, and telomerase activity. Furthermore, these cells could differentiate into cell types of the three germ layers in vitro and in teratomas. These findings demonstrate that iPS cells can be generated from adult human fibroblasts.

issues is to induce pluripotent status in somatic cells by direct reprogramming (Yamanaka, 2007).

We showed that induced pluripotent stem (iPS) cells can be generated from mouse embryonic fibroblasts (MEF) and adult mouse tail-tip fibroblasts by the retrovirus-mediated transfection of four transcription factors, namely Oct3/4, Sox2, c-Myc, and Klf4 (Takahashi and Yamanaka, 2006). Mouse iPS cells are indistinguishable from ES cells in morphology, proliferation, gene expression, and teratoma formation. Furthermore, when transplanted into blastocysts, mouse iPS cells can give rise to adult chimeras, which are competent for germline transmission (Maherali et al., 2007; Okita et al., 2007; Wernig et al., …

…from adult human somatic cells by optimizing retroviral transduction in human fibroblasts and subsequent culture conditions. These efforts have enabled us to generate iPS cells from adult human dermal fibroblasts and other human somatic cells, which are comparable to human ES cells in their differentiation potential in vitro and in teratomas.

Takahashi et al. 2007. *Cell 131*, 861

**An Example: Induced Pluripotent Stem Cells**

| Stakeholder | Question/Motivation | Access/Network |
|---|---|---|
| Academic Scientist | Should I shift bodies and bench space? | Colleagues, co-authors, etc. |
| Physician | What do I tell my patients? | Medline, news media |
| Patient | When will there be a cure? | News media, rumor, physicians |
| Legislator | How should I vote on stem cell funding? | Briefings, news media, staff |
| Biotech CEO | What does this mean for my IP position? | Scientific advisory board, legal counsel |

We know too little about how different stakeholders access and understand new discoveries

Involving diverse stakeholders directly in scientific networks may speed utilization but could shift the dynamics of discovery

Keeping stakeholders separate from discovery may protect science but widen the gap between bench and bedside

# Science of Science Management

## Knowledge Discovery and Management Topics

October 2-3, 2008

NIH Campus

Bethesda, MD

# Knowledge Discovery & Management (Public Health)

Nathaniel Osgood
University of Saskatchewan
Department of Computer Science
& School of Public Health

# Ideas for Moving Forward…

## Gaps

- Fragmented, diffuse public health information
- Engineered systems but not incentives
- Scattered understanding of intervention prospects
- Limited understanding of ecosystem dynamics
- Insufficient attention to strategic flexibility

## Possible Responsive Research Avenues

- Integrated public health observatories
- Information technology fostering incentives
- Knowledge/prediction markets
- Dynamic ecosystem models
- Real options

# Ideas for Moving Forward…

## Gaps

- Fragmented, diffuse public health information
- Engineered systems but not incentives
- Scattered understanding of intervention prospects
- Limited understanding of ecosystem dynamics
- Insufficient attention to strategic flexibility

## Possible Responsive Research Avenues

- ✸ **Integrated public health observatories**
- Information technology fostering incentives
- Knowledge/prediction markets
- Dynamic ecosystem models
- Real options

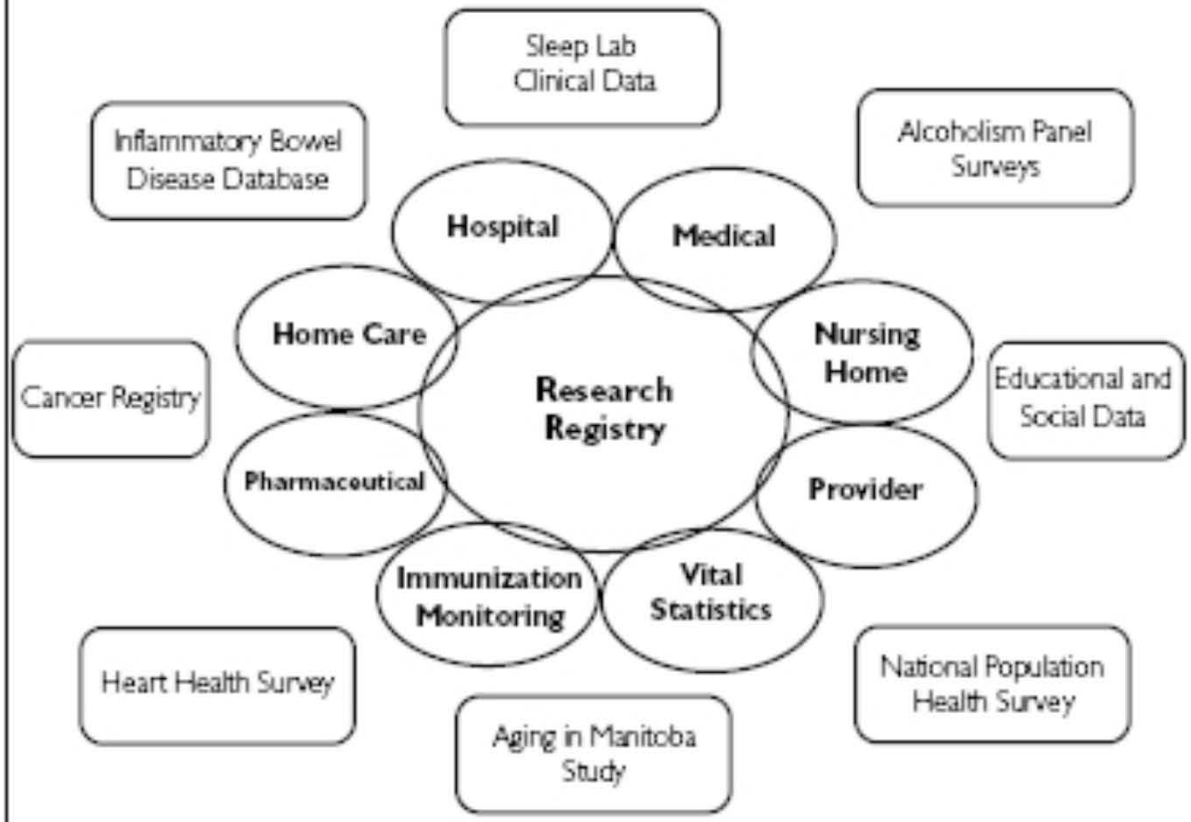
# Public Health Observatories

- Cross-linking & annotating multi-level data
  - Multiple cross-sectional & longitudinal survey instruments (health status, risk factors)
  - Administrative data (e.g. diagnostic codes, pharmaceutical & healthcare utilization, vital statistics, cost & resource use, education, justice, housing, …)
- Critical components
  - Confidentiality
  - Federated data access at different levels of resolution & authority
  - Rich & consistent metadata
  - Case ascertainment algorithms
  - Data cleaning
- Strongly empowers
  - Systems modeling efforts
  - Derivation of surrogate measures for conditions
  - Context-rich cross-sectional & longitudinal analysis
- Priorities
  - Incorporation of
    - Contextual information
      - (De-identified) & cross-linked social/ family network data
      - Determinants of health (e.g. Socio-economic status)
      - Intervention history
    - Transparent systems models
    - Data cleaning & analysis algorithms
    - EMA data
    - Community-contributed metadata
  - Rewarding contributions
  - Operational support
  - Voluntary privacy waivers?

# Example Data Repository

FIGURE 1.

The Manitoba Health Research Data Repository*




*A similar version of this figure appeared in Roos et al. 2004.

Lisa Lix, 2008

# Ideas for Moving Forward…

## Gaps

- Fragmented, diffuse public health information
- Engineered systems but not incentives
- Scattered understanding of intervention prospects
- Limited understanding of ecosystem dynamics
- Insufficient attention to strategic flexibility

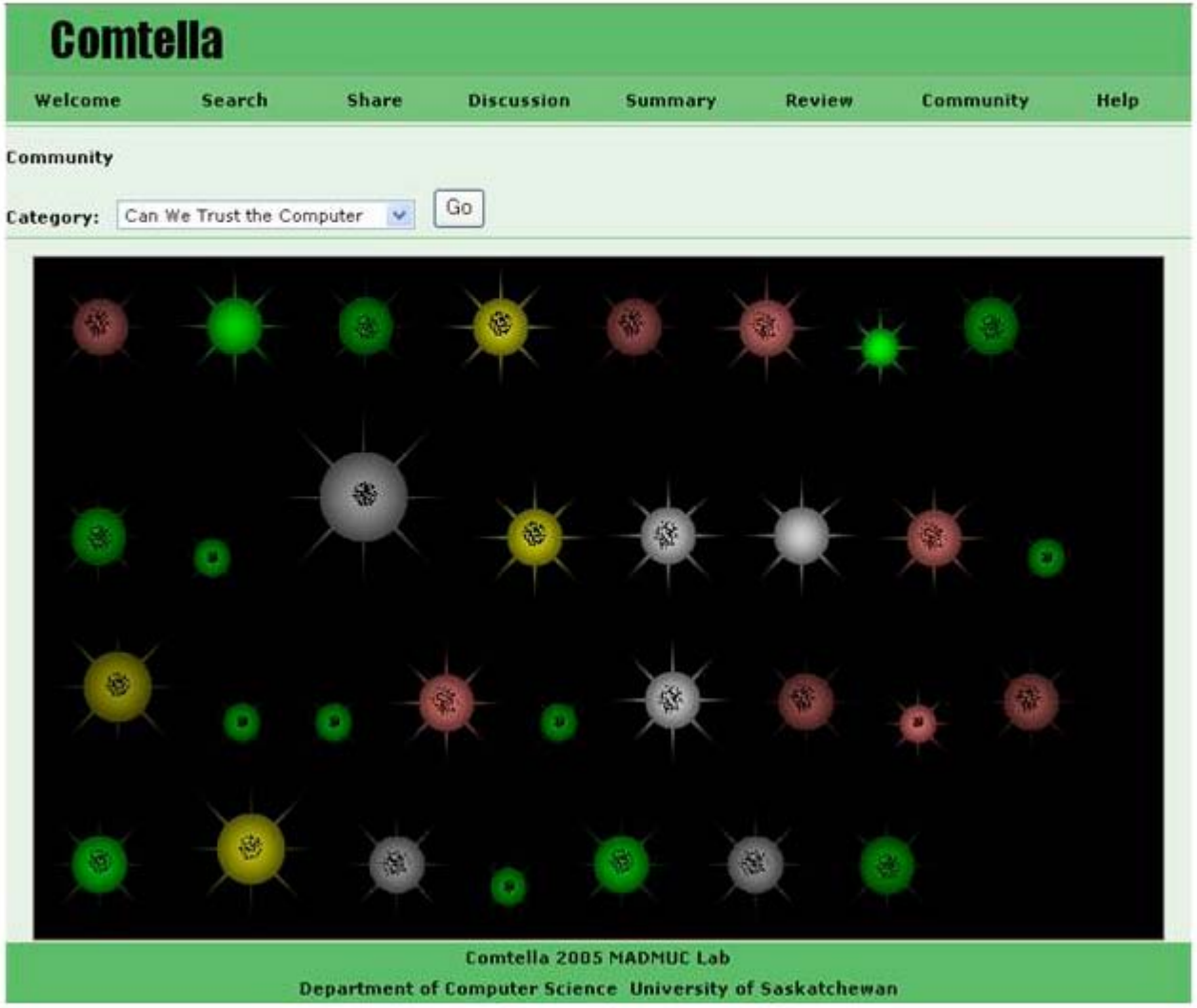
## Possible Responsive Research Avenues

- Integrated public health observatories
- ✸ **Information technology fostering incentives**
- Knowledge/prediction markets
- Dynamic ecosystem models
- Real options

# Learning from Management of Learning Communities

- Issue: IT infrastructure for sharing not matched by incentive structures
  - Enhancing research incentives may improve contributions
    - Reviews, sharing negative results, result sharing & reproduction…
- Example: Comptella (J. Vassisleva)
- High sensitivity to even simplest incentive schemes
  - Non-monetary incentives highly effective in promoting "culture change" towards sharing
  - Incentives for reviews helps highlight quality contributions
  - Gaming of incentives common
- Overcoming local optima: Injection of stochastics can help reduce lock-in to local minima

# Example Learning Community




J. Vassisleva

# Example: Bioinformatics Community

- Enormously rich annotated cross-linked databases
  - Federal support (e.g. PDB, SwissProt, EBI, GENBANK, …)
  - Cross-linking key to use & realized value
- Cultural norms & incentives value shared contributions
  - Reputation accrues through sharing
  - Sharing encouraged/required through
    - Funding guidelines
    - Publication policy
    - Society guidelines & consensus statements
  - Shared data fairly standard
  - Shared code encouraged
- Programmatic data access & services (via web services)

# Ideas for Moving Forward…

## Gaps

- Fragmented, diffuse public health information
- Engineered systems but not incentives
- Scattered understanding of intervention prospects
- Limited understanding of ecosystem dynamics
- Insufficient attention to strategic flexibility

## Possible Responsive Research Avenues

- Integrated public health observatories
- Information technology fostering incentives
- ✸ **Knowledge/prediction markets**
- Dynamic ecosystem models
- Real options

# Knowledge Markets

- Prediction markets for lines of research, with added sharing incentives
- Exploit market information
  - Aggregation
  - Dissemination
- Components
  - Market incentives can help overcome vested interests in statements of likely outcome
  - Market signals reflect confidence of beliefs
  - Reward exchange of information for trades
- Challenges: Market design, market gaming & insider trading, adequate market size

# Science of Science Management

## Knowledge Discovery and Management Topics

October 2-3, 2008

NIH Campus

Bethesda, MD