

TODAY'S PUBLIC LECTURE

Measuring the results of research investments

Dr Julia I. Lane

Program Director, Science of Science and Innovation Policy
US National Science Foundation (NSF)

Thursday 22 March 2012 12.30 - 1.30pm with a light lunch provided

Weston Theatre, JG Crawford Bldng #132, Lennox Crossing, ANU

Presented by
the HC Coombs Policy Forum,
Crawford School of Economics and Government, ANU
in partnership with the Melbourne Institute of Applied Economic
and Social Research, University of Melbourne

Building an Evidence Basis for S&T Policy

Overview

- Background
- The STAR METRICS program
 - Workforce
 - Grants
 - Outputs/Outcomes

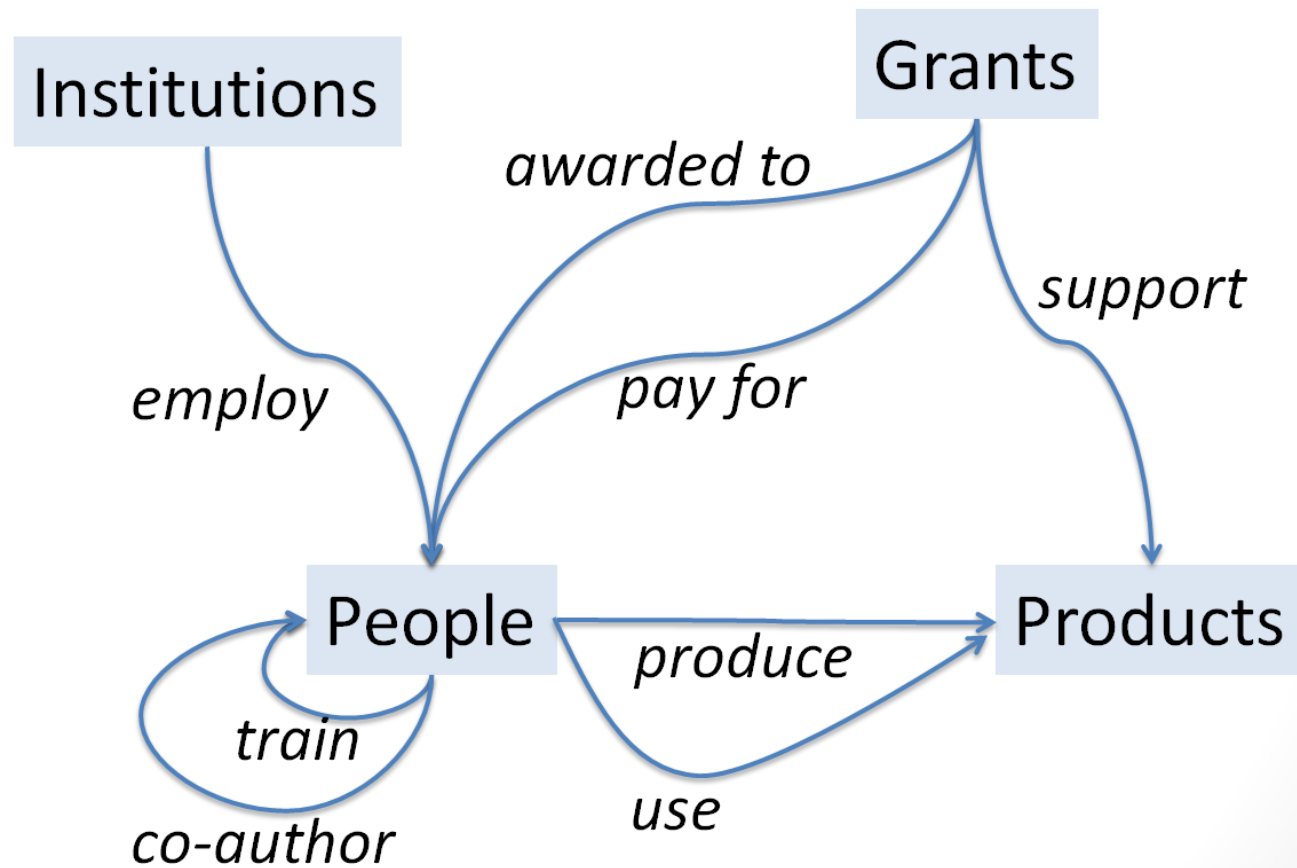
Brief Background: Evolution

- 2005: Jack Marburger points out there is no “science” of science policy
- 2006: Science of Science Policy established – NSTC interagency group plus NSF SciSIP program
- 2008: Roadmap finds “data inadequate for decision-making” across all science agencies
- 2009: ARRA reporting requirements demonstrated uneven capacity across agencies; STAR pilot
- 2010: STAR METRICS established
- 2012: 6 federal agencies, 85 research institutions (~40% of NSF/NIH extramural research portfolios)

Brief Background: Why this matters



Brief Background: Why this is difficult to do

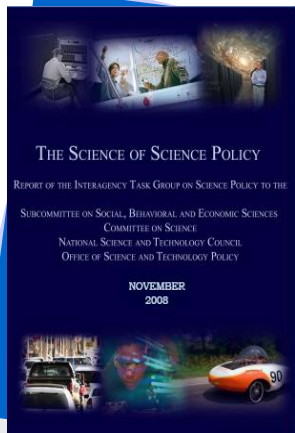


Brief Background: the data don't exist

The ITG undertook a literature review to determine the state of the science to date. A questionnaire was also circulated to Federal agencies to ascertain what methods are currently in use for programmatic investment decision making, as well as to ask what tools and resources are needed by Federal agencies that are currently unavailable. The ITG found that:

- There is a well developed body of social science knowledge that could be readily applied to the study of science and innovation.
- Although many Federal agencies have their own communities of practice, the collection and analysis of data about the science and scientific communities they support is heterogeneous and unsystematic.
- Agencies are using very different models, data and tools to understand their investments in science and technology.
- The data infrastructure is inadequate for decision-making.

THE SCIENCE OF SCIENCE POLICY: A FEDERAL RESEARCH ROADMAP

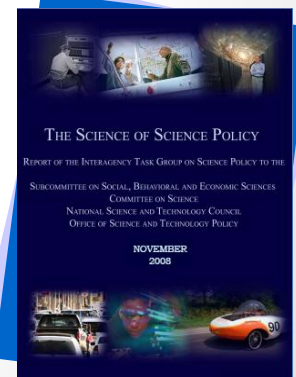


The data don't exist for good reason

- No systematic documentation of inputs (who is supported by science funding)
- No systematic links between inputs and outputs
- Heavy reliance on manual reporting
- No systematic ability to capture outputs and outcomes beyond the period of an award
- Balkanized agency systems => impossible to get overview of science investments



The data infrastructure is inadequate for decision-making



Brief Background: The Approach

Understanding

develop usable
knowledge and theories

Measurement

improve and expand
science metrics, datasets
and analytical models
and tools

Community of Practice

cultivate a community of
practice focusing on
SciSIP across the
academy, the public
sector and industry

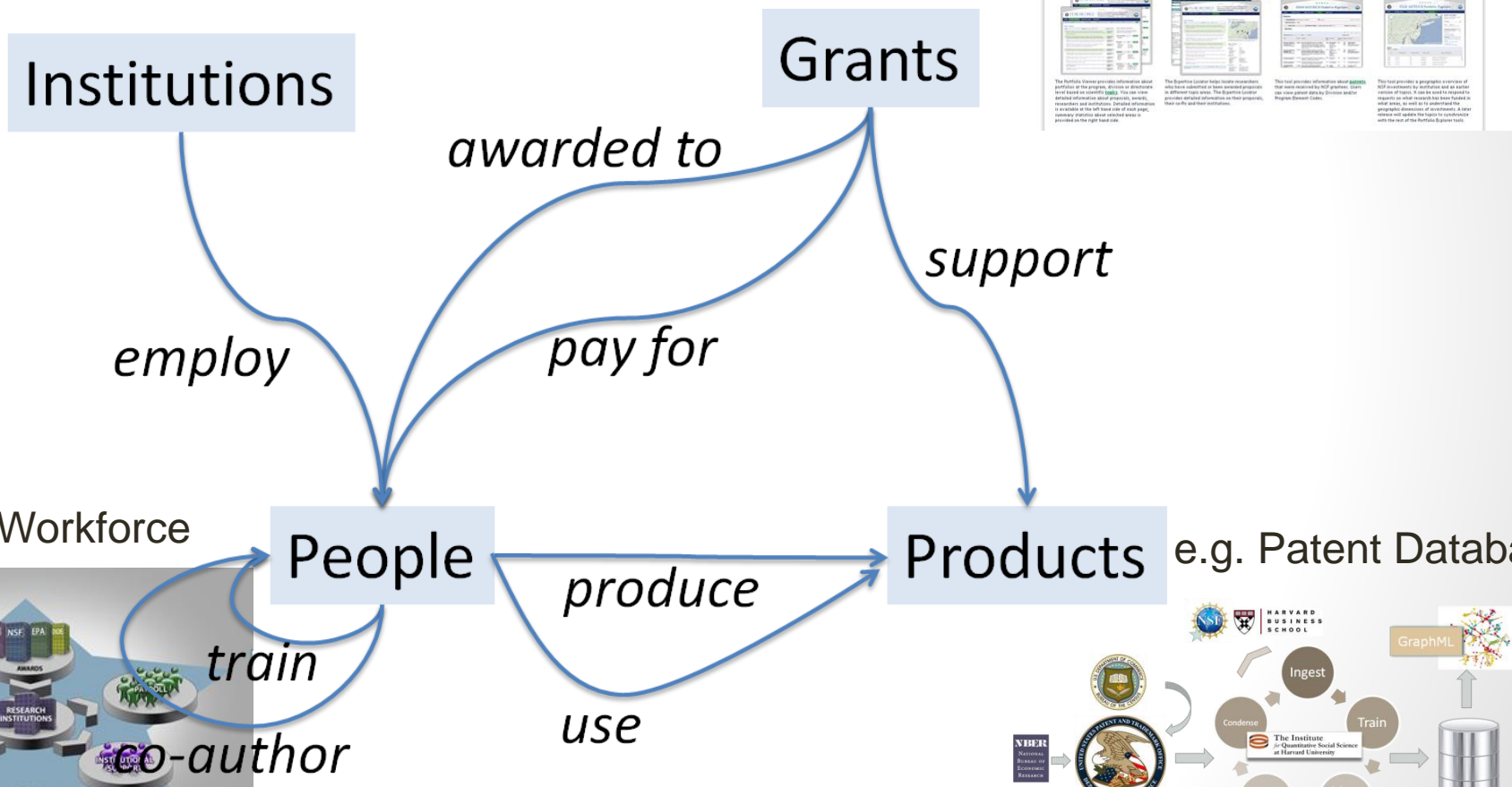
The Nascent System in the US

Goal: Provide a better empirical basis for science policy:

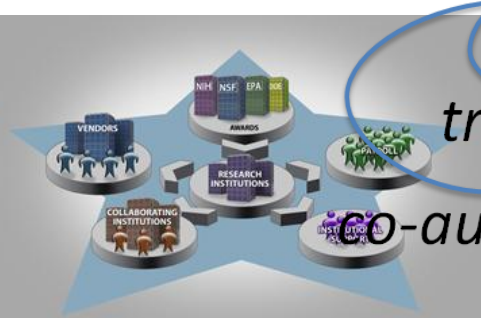
- providing an open and automated data infrastructure that can be used by federal agencies, research institutions, and researchers,
- documenting federal investments in science, and
- analyzing the resulting relationship between inputs, outputs, and outcomes.

Approach: automatically capture data about the conduct of science – inputs, outputs and the connections between the two

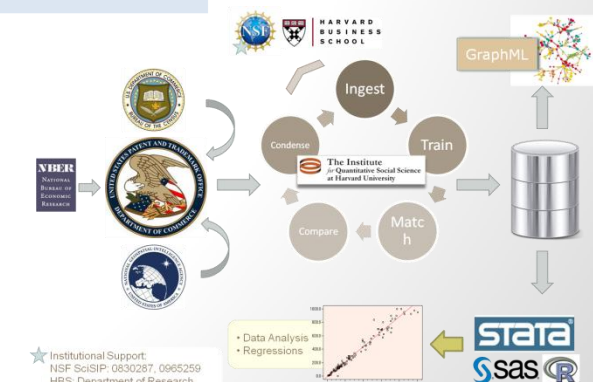
Portfolio Explorer



Scientific Workforce



e.g. Patent Database



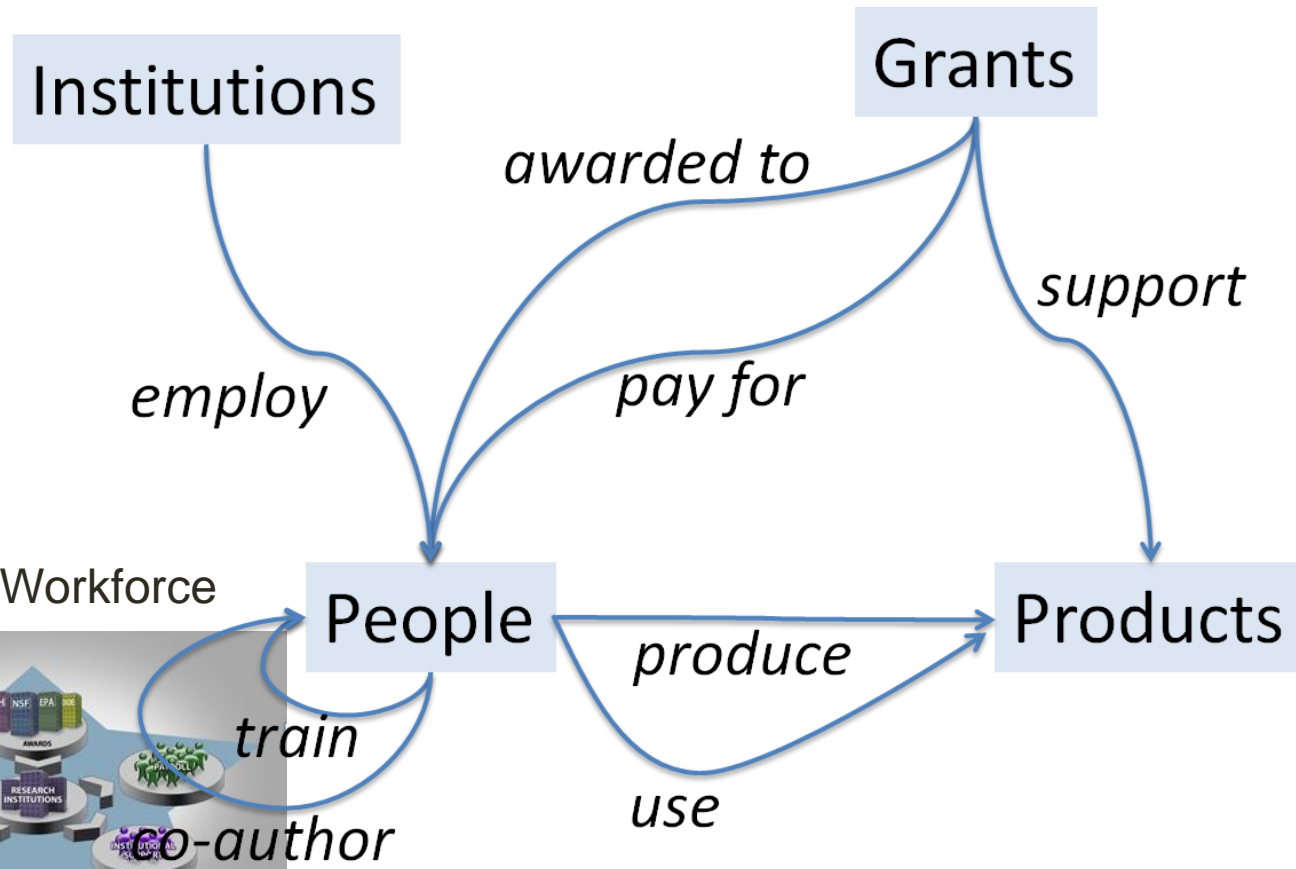
Source: Ian Foster University of Chicago



Basic Approach

- **Focus:** Build coherent data infrastructure (not shiny tools) leverage existing systems – and minimize manual input
- **Unit of Analysis:** Individual - Senior personnel – AND postdocs, graduate students and undergraduate students
- **Incentives:** Create value for all stakeholders to ensure capture of right measures and data quality

Approach: automatically capture data about the conduct of science – inputs, outputs and the connections between the two

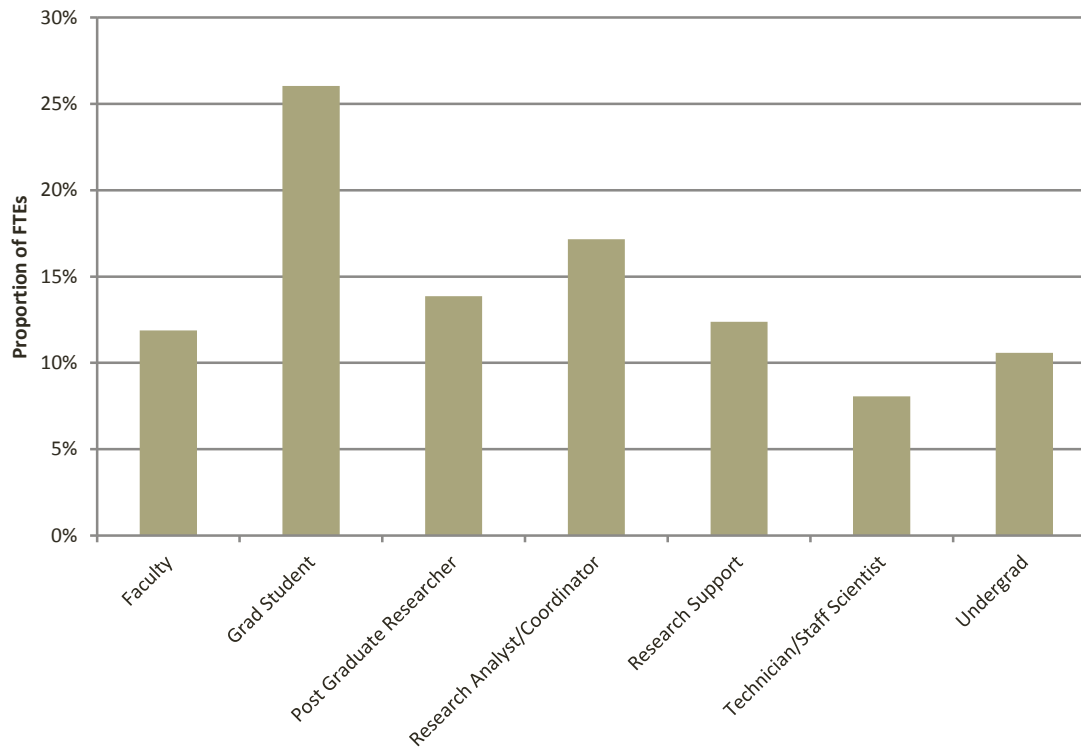


Source: Ian Foster University of Chicago



Information About Individuals Tells Us About Future Supply

Figure 2: The Distribution of Occupations



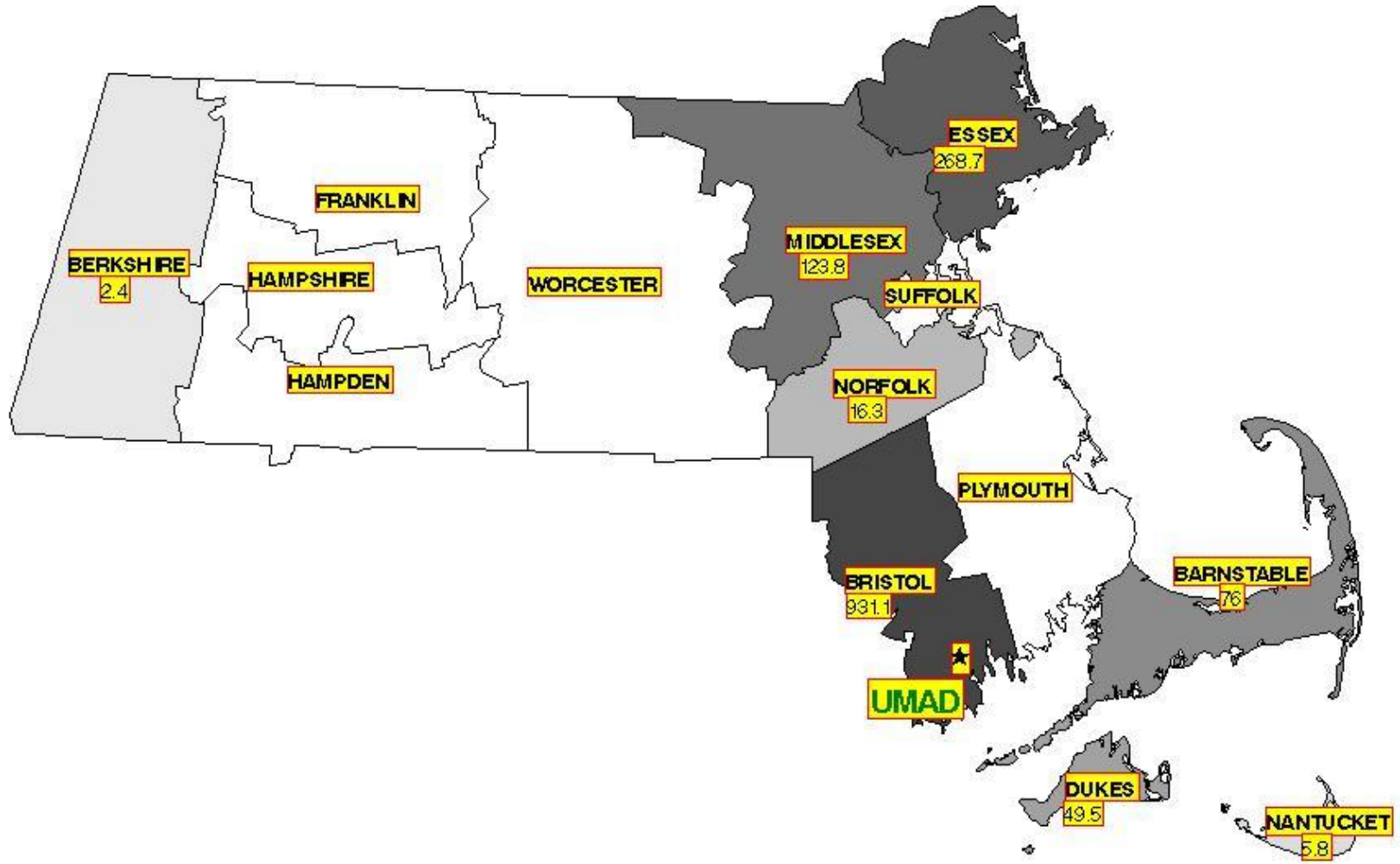
And In Which Areas

Table 3: Number of Individuals Working in Top 5 Research Areas

Research Area	Type	Post-Docs	Graduate Students	Undergraduate Students	All Occupations
Astronomical Sciences	Individuals	101	268	106	1203
	Award	52	97	55	216
Mathematical Sciences	Individuals	11	222	54	596
	Awards	11	99	32	244
Cybersecurity	Individuals	17	177	95	521
	Awards	9	70	38	150
Gene Expression	Individuals	61	107	95	498
	Awards	31	54	30	102
Magnetic Field	Individuals	24	140	67	348
	Awards	14	57	24	84

Local Economic Impact

for UNIVERSITY OF MASSACHUSETTS DARTMOUTH
Total Jobs (SIMULATED DATA)

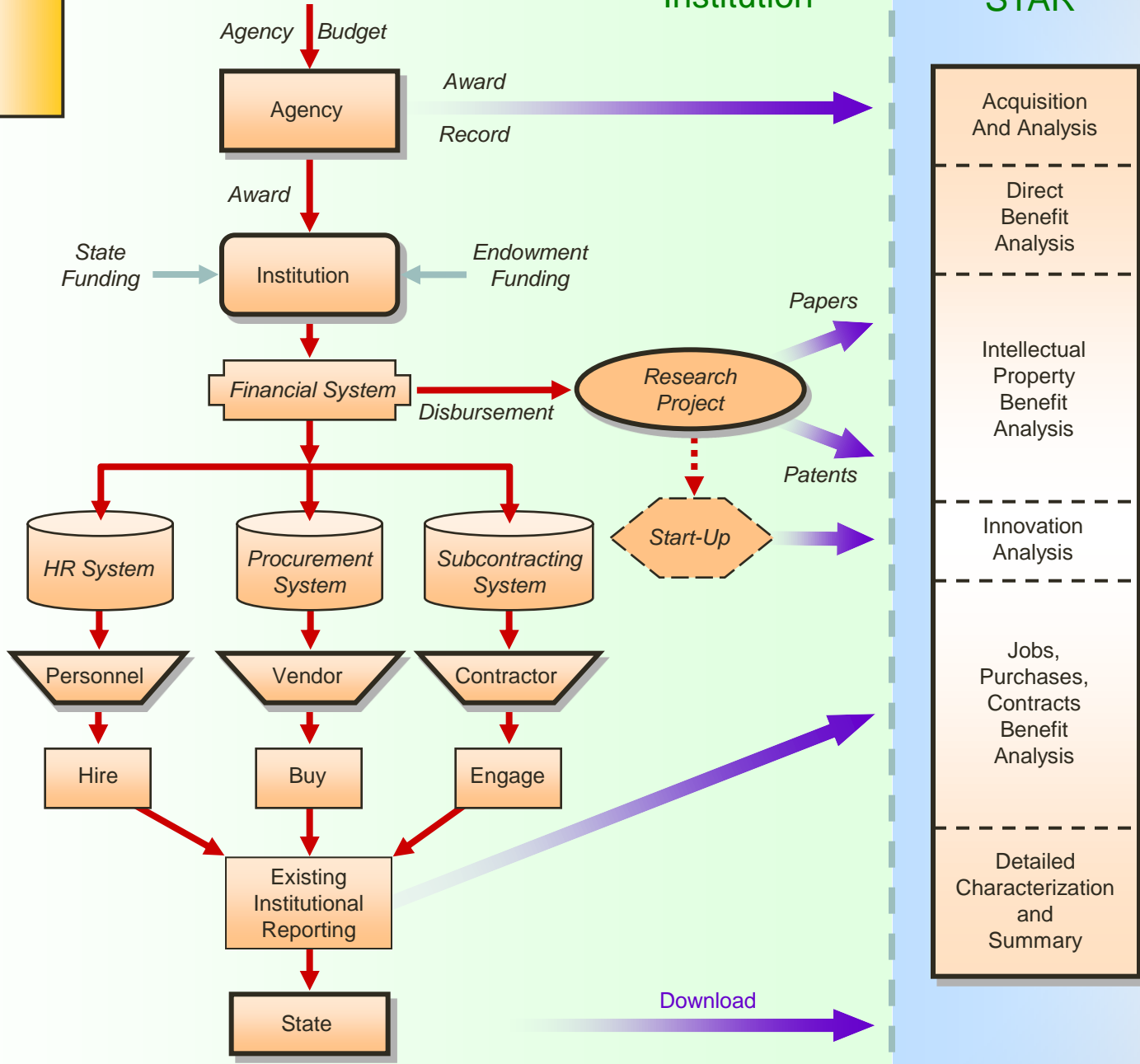


Source: STAR Metrics - Jobs

STAR Pilot Project

Institution

STAR



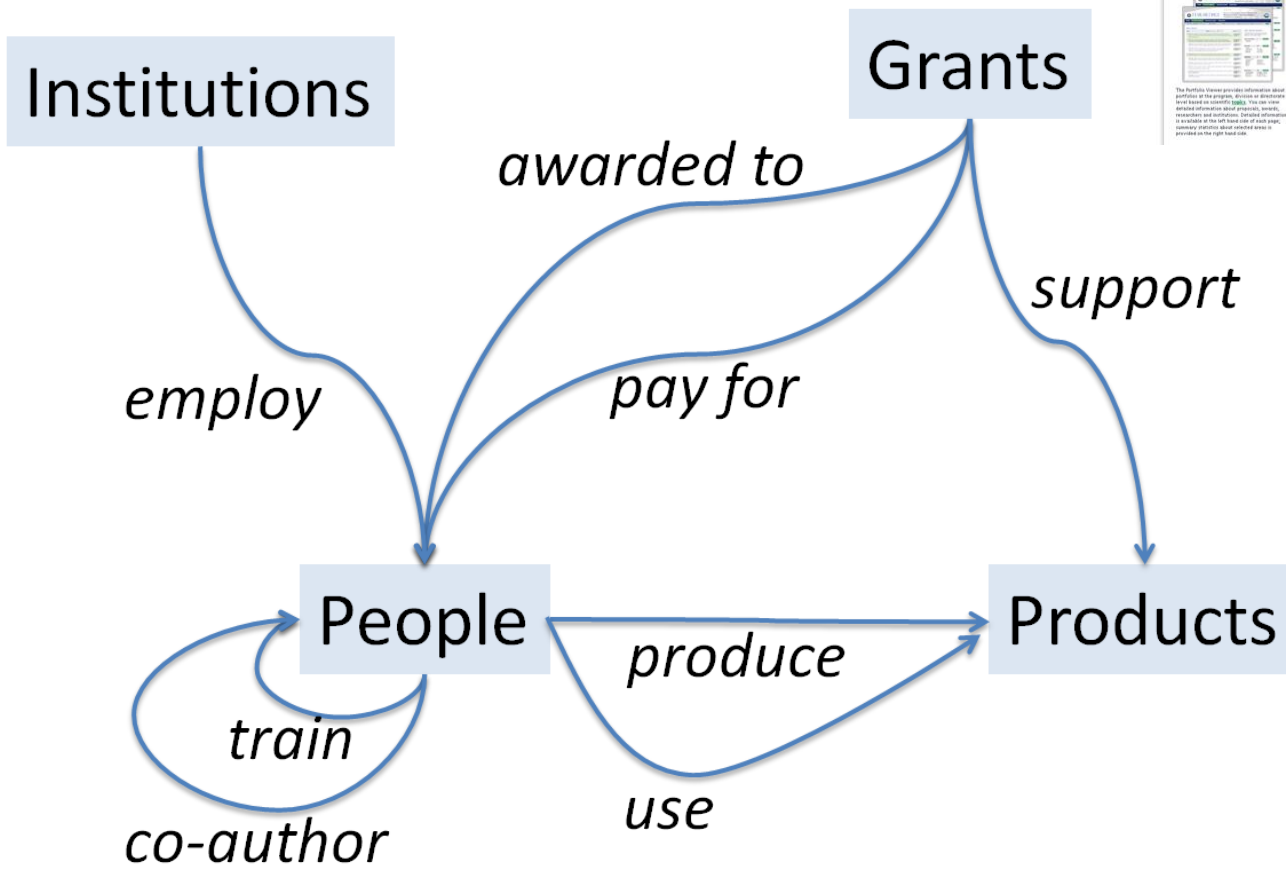
Star Metrics Level I – 14 Requested Data Elements

*Each file has a period start and end date that reflects the quarter or time period in which the transaction occurs, NOT the start and end date of the award being reported.

Description	Element ID	Item
Information on Scientists and Awards	1	Unique Award #
	2	Recipient Account #
	3	Overhead charged
	4	De-identified Employee ID #
	5	Occupational Classification
	6	Proportion of earnings allocated to award
	7	FTE status
Subcontracts and subawards	1	Unique Award #
	2	Recipient Account #
	8	Sub Award Recipient Duns #
	9	Sub Award Payment Amount
Payments to Vendors	1	Unique Award #
	2	Recipient Account #
	10	Vendor Duns #
	11	Vendor Payment Amount
Each file has a period start and end date that reflects the quarter or time period in which the transaction occurs	12	Period Start Date
	13	Period End Date
Information on Overhead	14	Proportion of overhead associated with salaries (from Indirect cost rate proposal)

Basic Approach

- **Focus:** Build coherent data infrastructure (not shiny tools) leverage existing systems – and minimize manual input
 - Used existing financial systems of research institutions
- **Unit of Analysis:** Individual - Senior personnel – AND postdocs, graduate students and undergraduate students
 - Used HR systems of research institutions
- **Incentives:** Create value for all stakeholders to ensure capture of right measures and data quality
 - Created reports of value to research institutions

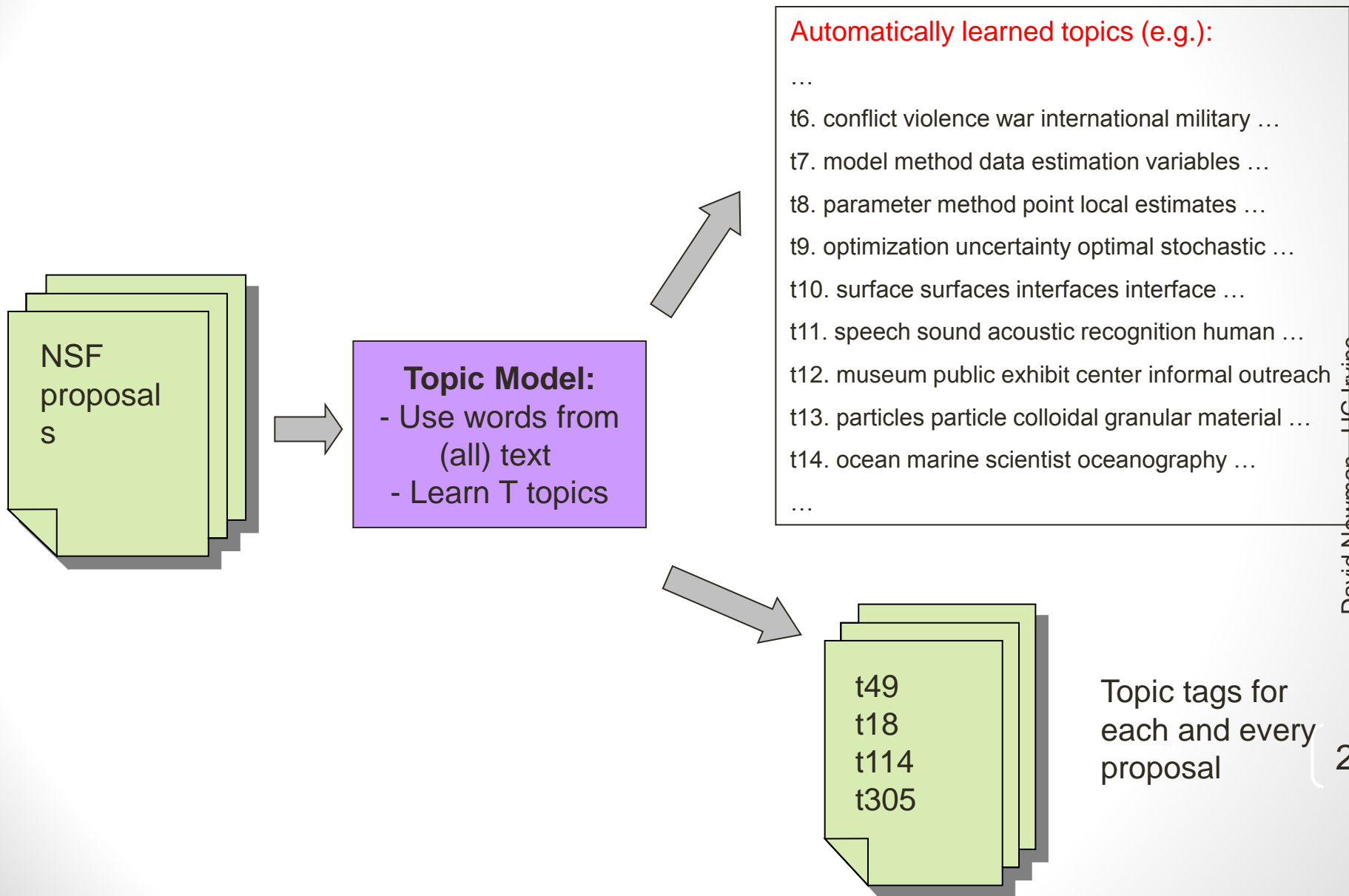


Source: Ian Foster University of Chicago

Information about grants tells us about portfolio using different lenses

- Divisions
- Programs
- Researchers
- Institutions
- Time
- Geography

Automated capture of scientific topics





STAR METRICS Portfolio Explorer

Limited Distribution
ALPHA-version

[Home](#)[Portfolio](#)[Expertise](#)[Patents](#)[Maps](#)[About](#)[Feedback](#)

Welcome

This site provides four tools that provide different views of scientific portfolios. The tools are provided by the [STAR METRICS](#) program; an interagency collaboration to provide a stronger empirical basis for science policy decisions.

Portfolio Viewer



The Portfolio Viewer provides information about portfolios at the program, division or directorate level based on scientific [topics](#). You can view detailed information about proposals, awards, researchers and institutions. Detailed information is available at the left hand side of each page; summary statistics about selected areas is provided on the right hand side.

Expertise Locator



The Expertise Locator helps locate researchers who have submitted or been awarded proposals in different topic areas. The Expertise Locator provides detailed information on their proposals, their co-Pis and their institutions.

Patent Viewer



This tool provides information about [patents](#) that were received by NSF grantees. Users can view patent data by Division and/or Program Element Codes.

Map Viewer



This tool provides a geographic overview of NSF investments by institution and an earlier version of topics. It can be used to respond to requests on what research has been funded in what areas, as well as to understand the geographic dimensions of investments. A later release will update the topics to synchronize with the rest of the Portfolio Explorer tools.

Send Us Feedback

We would love to hear from you! Please email us at PEfeedback@nsf.gov with any questions or feedback. If you have a problem to report, please include the url of the page you were on together with a description of what happened.

DIVISIONAL VIEW

<input type="checkbox"/>	Division of Civil, Mechanical & Manufacturing Innovation (CMMI)	3010	\$887.62M	▶
<input type="checkbox"/>	Division of Electrical, Communications & Cyber Systems (ECCS)	1547	\$458.56M	▶
<input type="checkbox"/>	Division of Engineering Education & Centers (EEC)	583	\$287.70M	▶
<input type="checkbox"/>	Office of Emerging Frontiers in Research & Innovation (EFRI)	98	\$149.05M	▶
<input type="checkbox"/>	Division of Industrial Innovation & Partnerships (IIP)	2960	\$704.69M	▶
<input type="checkbox"/>	Office of the Assistant Director (ENG)	32	\$11.06M	▶

Directorate for Geosciences

<input type="checkbox"/>	Division of Atmospheric and Geospace Sciences (AGS)	1729	\$1188.76M	▶
<input type="checkbox"/>	Division of Earth Sciences (EAR)	3130	\$873.27M	▶
<input type="checkbox"/>	Division of Ocean Sciences (OCE)	2579	\$1566.00M	▶
<input type="checkbox"/>	Office of the Assistant Director (GEO)	254	\$104.04M	▶

Directorate for Mathematical & Physical Sciences

<input type="checkbox"/>	Division of Astronomical Sciences (AST)	1161	\$1021.33M	▶
<input type="checkbox"/>	Division of Chemistry (CHE)	2529	\$1091.00M	▶
<input type="checkbox"/>	Division of Materials Research (DMR)	2389	\$1136.60M	▶
<input checked="" type="checkbox"/>	Division of Mathematical Sciences (DMS)	4955	\$1067.59M	▼

<p>Most Common Topics (by # of Awards)</p> <ul style="list-style-type: none"> (277) t455:Differential Equations (274) t451:Statistical Estimation (252) t817:Number Theory (234) t642:Conference (219) t149:Mathematics (204) t681:Algebraic Geometry (198) t99:Numerical Methods (195) t674:Algebra (187) t181:Mathematics/Geometry (184) t399:Topology 	<p>Most Common PGE Codes (by # of Awards)</p> <ul style="list-style-type: none"> (724) p1264:ALGEBRA,NUMBER THEORY,AND COM (656) p1281:ANALYSIS PROGRAM (627) p1266:APPLIED MATHEMATICS (504) p1271:COMPUTATIONAL MATHEMATICS (471) p1269:STATISTICS (380) p7335:WORKFORCE IN THE MATHEMAT SCI (370) p1267:TOPOLOGY (309) p1265:GEOMETRIC ANALYSIS (217) p7334:MATHEMATICAL BIOLOGY (186) p1263:PROBABILITY
---	--

<input type="checkbox"/>	Division of Physics (PHY)	1835	\$1396.57M	▶
<input type="checkbox"/>	Office of the Assistant Director (MPS)	7	\$11.90M	▶

Directorate for Social, Behavioral & Economic Sciences

<input type="checkbox"/>	Division of Social and Economic Sciences (SES)	2695	\$486.71M	▶
<input type="checkbox"/>	Division of Behavioral and Cognitive Sciences (BCS)	2950	\$473.32M	▶

Topic Data

Clear Selection

Show 50 entries

Keyword Filter:

Export as CSV

Select	Topic	Awarded	Awarded Amt.
<input checked="" type="checkbox"/>	t455: Differential Equations - equation solution nonlinear existence PDE estimate initial_data soliton asymptotic global space wave hyperbolic singular limit wave_equation linear boundary_condition decay conservation_law ...	277	\$56.18M
<input checked="" type="checkbox"/>	t451: Statistical Estimation - estimator function regression estimation nonparametric asymptotic multivariate parameter procedure covariate inference estimate conditional functional time_series statistic bootstrap distribution empirical parametric ...	274	\$43.94M

PGE Code (Total:5)	Division	Awarded	Awarded Amt.
p1269:STATISTICS	DMS	268	\$42.99M
p1271:COMPUTATIONAL MATHEMATICS	DMS	2	\$0.15M
p1260:INFRASTRUCTURE PROGRAM	DMS	2	\$0.10M
p7334:MATHEMATICAL BIOLOGY	DMS	1	\$0.55M
p1263:PROBABILITY	DMS	1	\$0.15M

Co-occurring Topics: t451:Statistical Estimation, t613:Bayesian Statistics, t4:Stochastic Processes, t146:Census, t820:Probability, t70:Statistics

<input checked="" type="checkbox"/>	t817: Number Theory - conjecture prime arithmetic representation theorem number_theory function integer subgroup series curve Galois L_function elliptic_curve modular_form p_adic abelian mod formula weight ...	252	\$47.95M
<input type="checkbox"/>	t642: Conference - speaker talk participant Chair organizer topic GRC graduate_student invited_speaker poster held poster session reader conference_attendee junior invited ...	234	\$6.20M

Port

The b
select
expan
analy

Inst

Res

Awa

Tota

D

t1

Drill down into specific awards

Which researchers – and in which areas?

Star Metrics | readidata.nitrd.gov/star/explore.php

▼ Portfolio Viewer

Awards | **Researchers** | Institutions

Show 50 entries | Keyword Filter: | Export as CSV

Name	Institution	Department	Count*	IDs**	Details
Ellen J Maycock	American Mathematical Society	Meetings & Professional Services	7	0649635, 0705394, 0751449, 0933983, 1007980, 1045050, 1112903	
Matthew G Boylan	University of South Carolina at Columbia	Mathematics	6	0703547, 0901732, 1101301, 0801150, 0901068, 1001553	

Matthew G Boylan:
boylan@math.sc.edu
 8039172551

Mathematics
 University of South Carolina at Columbia

Awarded

Grant: **1101301**
 Status: Awarded \$11,223 on 2011/08/01
 Title: Palmetto Number Theory Series/SouthEast Regional Meeting On Numbers
 NSF Division: Division of Mathematical Sciences (DMS)
 Program Element: Algebra, Number Theory, and Combinatorics (1264)
 Topics: t817: Number Theory, t642: Conference, t149: Mathematics

Grant: **1001553**
 Status: Awarded \$13,423 on 2010/08/01

Total: 7
 Awarded: 7 (\$0.28M)
 Date First: 6/1/2006
 Date Last: 7/1/2011

Patents: 0

Co-PIs

- Michael A Filaseta
- Kevin L James
- Hui Xue
- James L Brown

Researchers by State

1 141

State	Count
California	141
Texas	93
New York	77
North Carolina	72
Illinois	63
Pennsylvania	61
Massachusetts	55
New Jersey	48
Michigan	47

AGENCY VIEW

What is being done in a particular topic across the entire agency?



STAR METRICS Portfolio Explorer

ALPHA-version distribution



- Home
- Portfolio
- Expertise**
- Patents
- Maps
- About
- Feedback

▼ Topics
Filter

Expertise Locator

This tool allows you to locate researchers based on their expertise. Using the "ANY" option in the Inclusion dropdown yields researchers with awards in ANY of the Topics you select (check) in the list below. Using the "ALL" inclusion limits your selection to those researchers with awards in ALL of the topics you select.

Topic options: From to [Update Topic List](#)

Topic Data [Clear Selection](#)

Show entries Keyword Filter: [Export as CSV](#)

Inclusion*:

Select	Topic	Awarded	Awarded Amt.	Declined	Requested Amt.
<input checked="" type="checkbox"/>	t445: Tsunamis - wave tsunami breaking wave_energy height interaction amplitude wave_propagation bottom surface_wave speed ocean propagation propagate depth standing tank propagating direction characteristic ...	60	\$20.06M	89	\$65.20M

Showing 1 to 1 of 1 entries (filtered from 941 total entries)

[First](#) [Previous](#) [1](#) [Next](#) [Last](#)

Topic Relevance: each proposal is assigned a total of 4 Topics. The Topic set is either based on a) the top/most relevant topic for each proposal, or b) based on all four topics assigned to proposals.

Please note that the co-occurring topics noted within each Topic's details view represent those Topics that co-occur most frequently together on proposals. While a good indicator of expertise, please note that using the ALL inclusion method yields researchers with Topics assigned to any of their proposals (not just those with selected Topics assigned to individual proposals). **Internal NSF users:** the ALL inclusion method only regards Awarded proposals (Declines and Others are excluded).

Researcher Summary

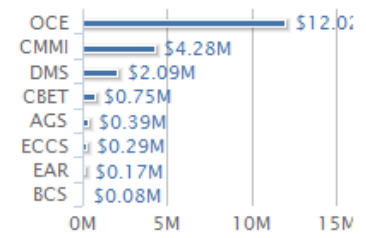
The below reflects a summary of the Topics you select/ed on the left. Click the triangle controls for expanded summaries or click the 'Explore' button to analyze your selection deeper.

Researchers (136) [Explore](#)

Awarded (60)

Total Funding: \$20.06M

Funding by Division (top 8)





STAR METRICS Portfolio Explorer

ALPHA-version



Home

Portfolio

Expertise

Patents

Maps

About

Feedback

Topics
Filter

1 of 941
Timing: 2007 - t445: Tsunamis

Change Selection

Expertise Locator

Researchers

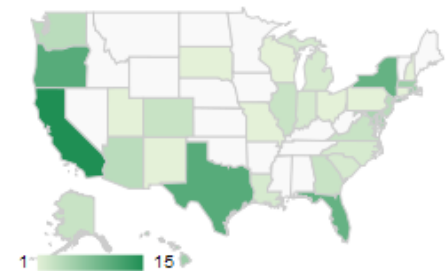
Show 50 entries

Keyword Filter:

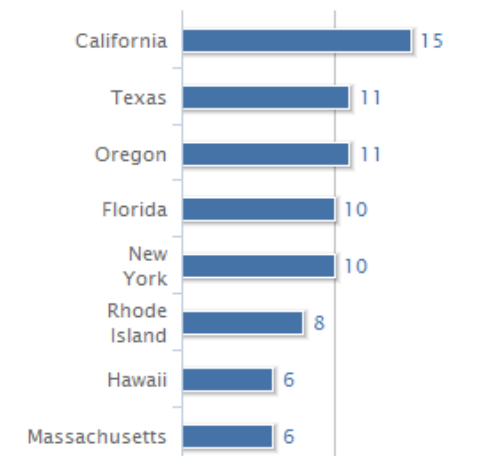
Export as CSV

PI ID	Name	Institution	Department	Count*	IDs**	Details
000032701	Philip L Liu	Cornell University	Civil & Environmental Engineering	10	0710751 , 0751079 , 0756327 , 0925711 , 0960512 , 1041541 , 0723578 , 0828552 , 0800372 , 0967003	▶
269824301	Robert Weiss	Virginia Polytechnic Institute and State University	Geosciences	8	1136534 , 1137611 , 0934681 , 1134926 , 1135027 , 0956094 , 1056467 , 1124295	▶
000013965	Harry H Yeh	Oregon State University	Civil & Cnstr. Engrg.	7	0742806 , 1135768 , 0723578 , 0828552 , 0935933 , 1041531 , 1129767	▶
000163080	Solomon C Yim	Oregon State University	Civil and Construction Engineering	6	0723277 , 0800822 , 1005740 , 1037861 , 0830365 , 0936642	▶
269744680	Hermann M Fritz	Georgia Tech Research Corporation	GT-Savannah / CEE	5	1034886 , 0936603 , 1000694 , 1105577 , 1135768	▶
000052129	James H Duncan	University of Maryland College Park	Department of Mechanical Engineering	4	0962107 , 0751853 , 0928318 , 0728770	▶
000160142	Costas E Synolakis	University of Southern California	Dept. of Civil Engineering	4	1000694 , 1105577 , 1135768 , 1034886	▶
000193450	Stephan T Grilli	University of Rhode Island	Department of Ocean Engineering	4	0927014 , 0940398 , 0928293 , 0830365	▶
000204919	Daniel T Cox	Oregon State University	Civil Engineering	4	0723277 , 0800822 , 1005627 , 1134971	▶
000225937	Tetsu Hara	University of Rhode Island	Graduate School of Oceanography	4	0824906 , 0927014 , 0940398 , 0820872	▶

Researchers by State



1 15

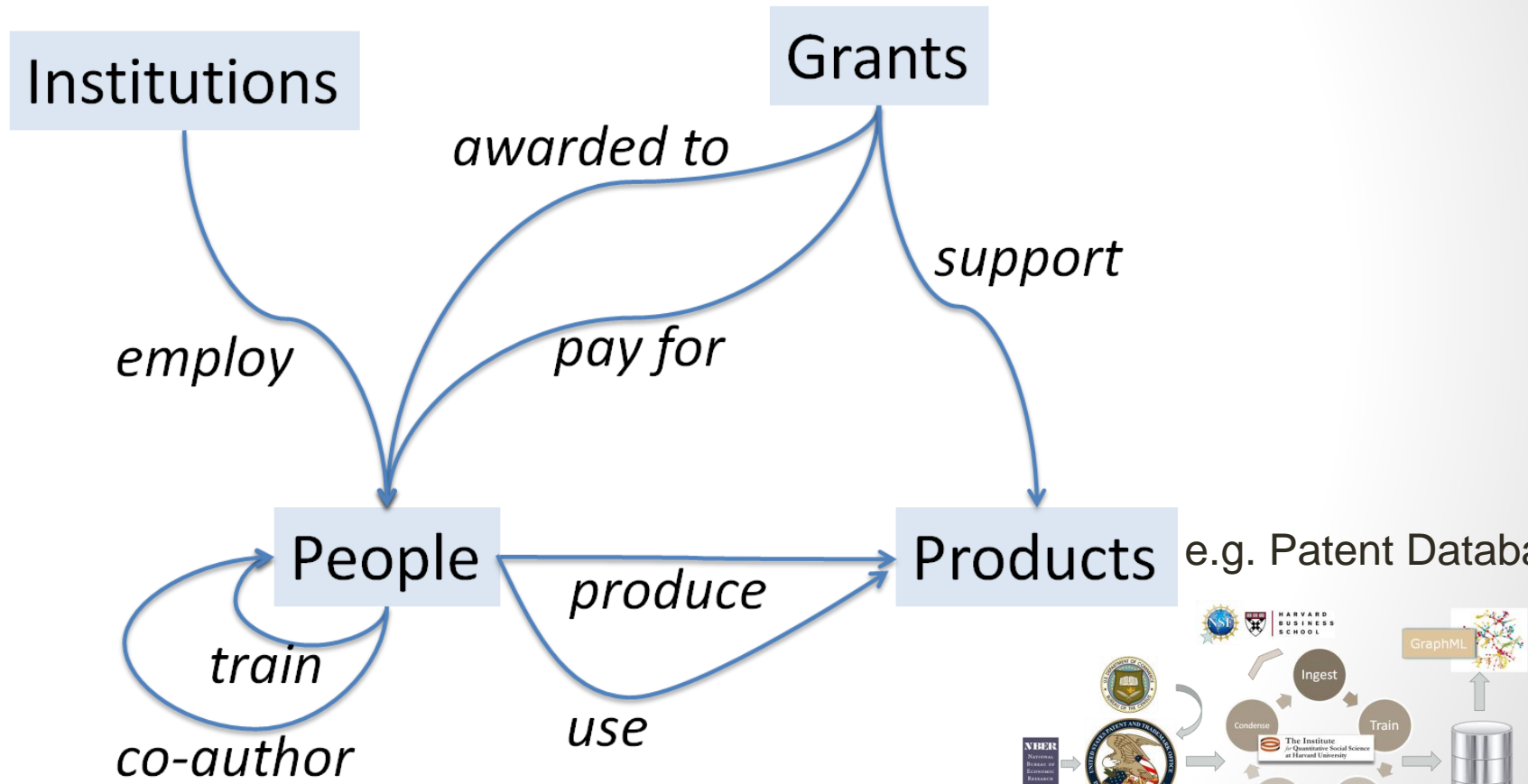


Basic Approach

- **Focus:** Build coherent data infrastructure (not shiny tools) leverage existing systems – and minimize manual input
 - Worked within agency to pull from internal production systems
- **Unit of Analysis:** Individual - Senior personnel –
 - No individual data on postdocs, graduate students and undergraduate students
- **Incentives:** Create value for all stakeholders to ensure capture of right measures and data quality
 - Created tools of value to program managers, division directors

PATENTS

Approach: automatically capture data about the conduct of science – inputs, outputs and the connections between the two



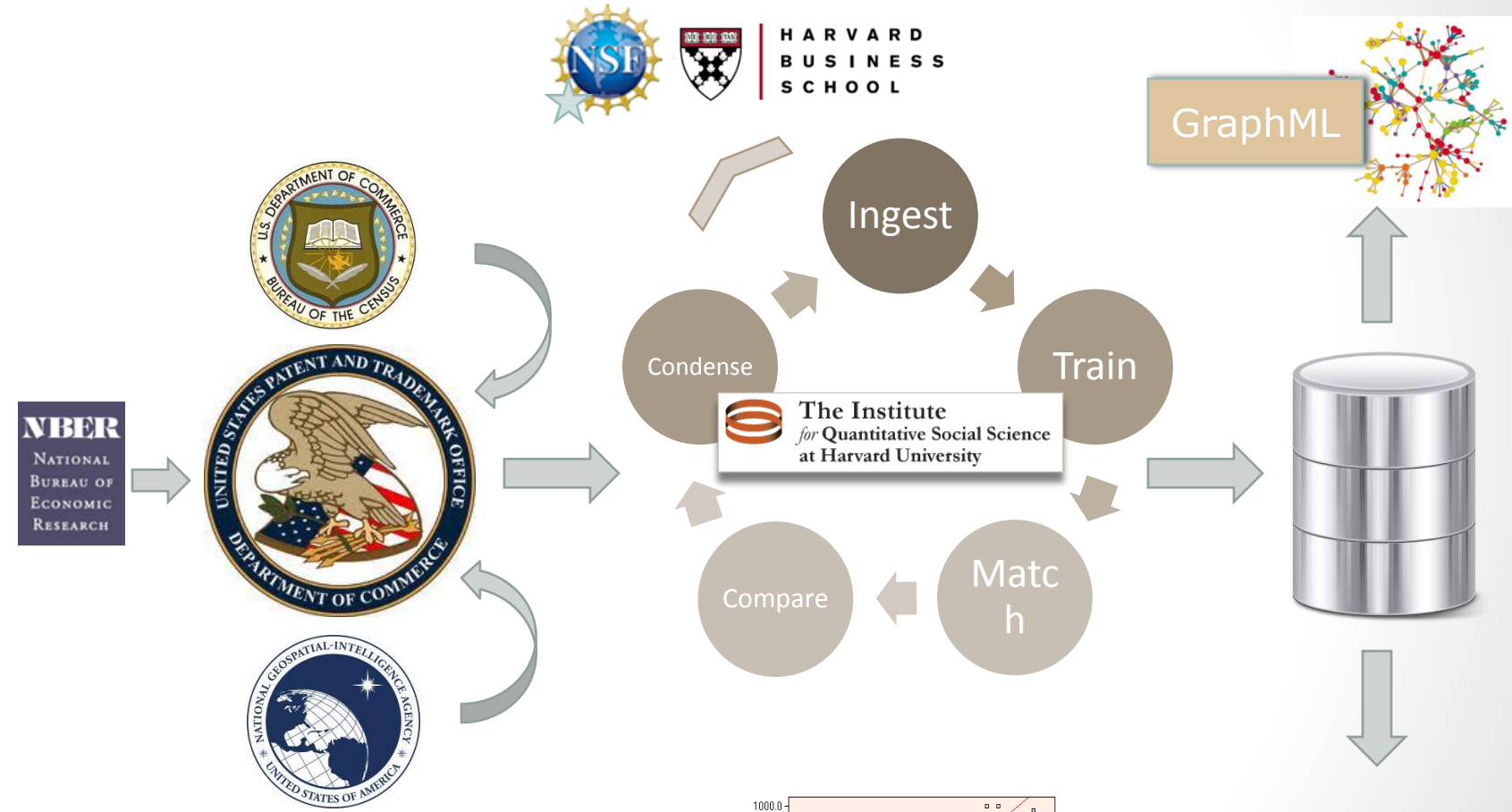
Source: Ian Foster University of Chicago

NSF
HARVARD BUSINESS SCHOOL
The Institute for Quantitative Social Science at Harvard University
Ingest
Condense
Compare
Match
Train
GraphML
STATA
sas

★ Institutional Support:
NSF SciSIP: 0830287, 0665259
HBS: Department of Research

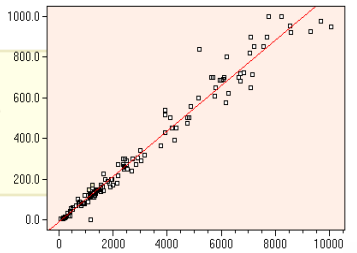
• Data Analysis
• Regressions

Automated capture of economic outcomes

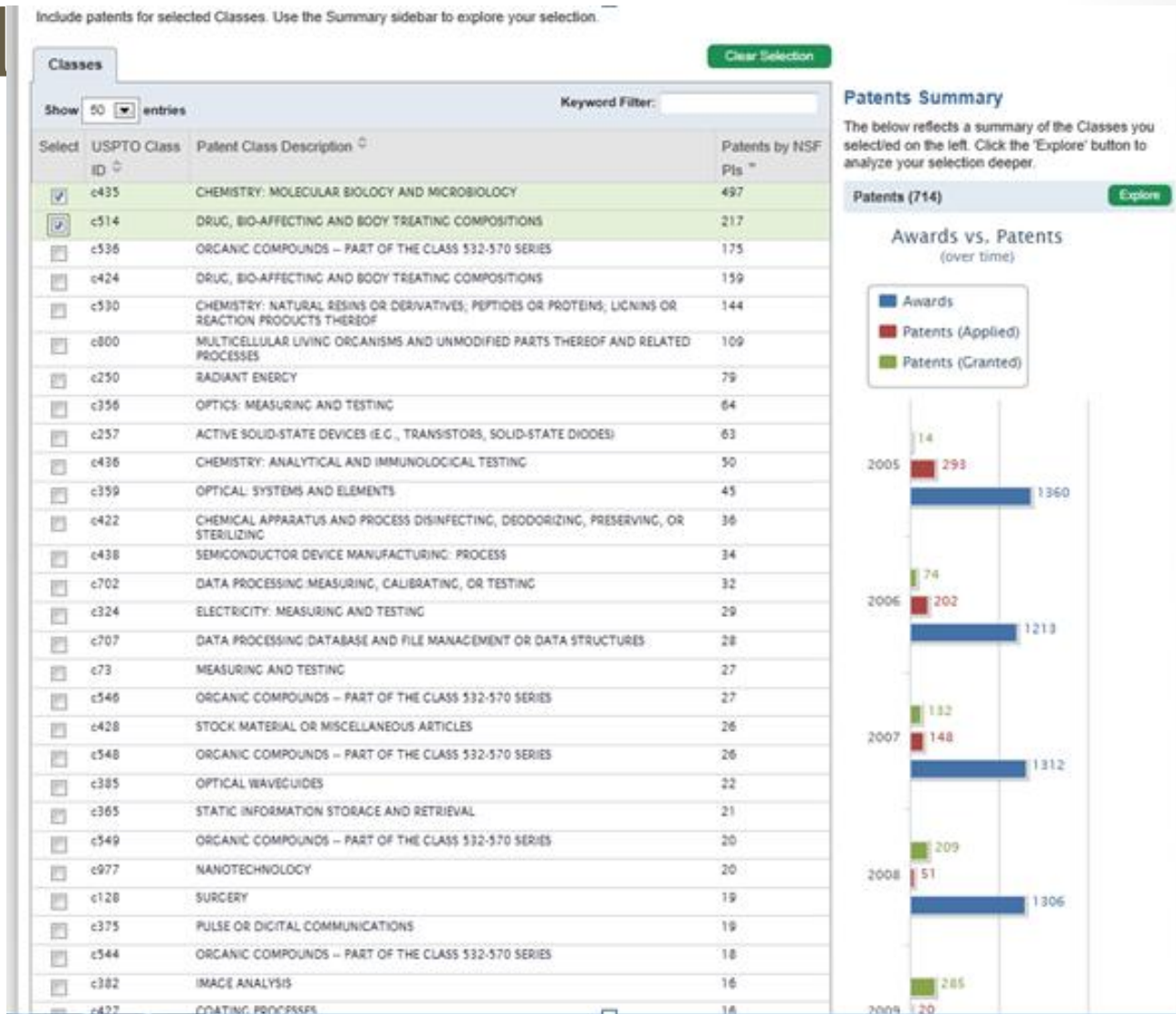


★ Institutional Support:
 NSF SciSIP: 0830287, 0965259
 HBS: Department of Research

- Data Analysis
- Regressions



What patents have been granted to NSF funded researchers – in which technol



Drilling down into details

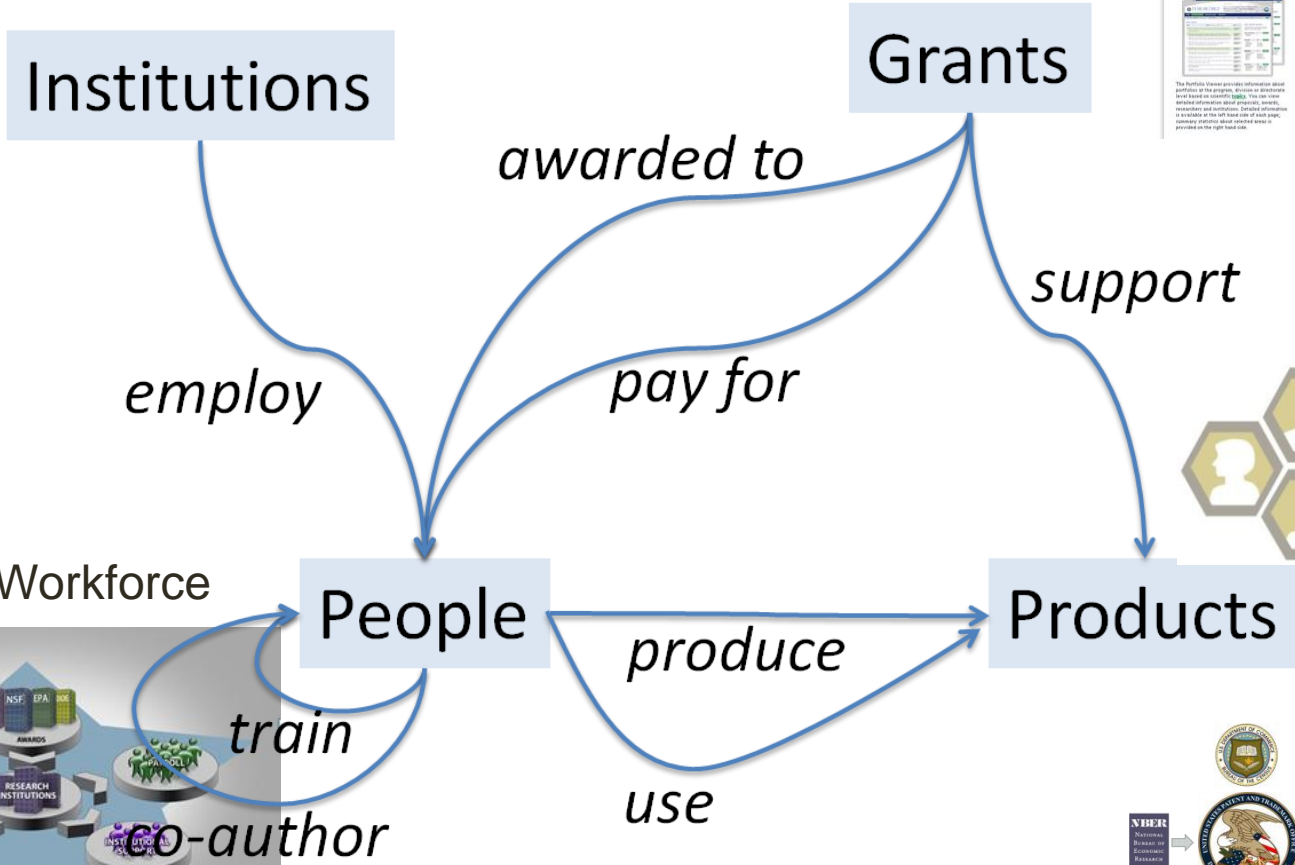
Home	Portfolio	Expertise	Patents	Maps	About	Feedback	
▶ Select	Divisions: (5 of 5) MCB, DBI, IOS, DEB, EF PGE Codes: All		Timing: Last 5 years		Change Selection		
▶ Classes Filter	2 of 243		e435: CHEMISTRY: MOLECULAR BIOLOGY AND MICROBIOLOGY e514: DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS		Change Selection		
▼ Patents Viewer							
Patents							
Show 50 entries		Keyword Filter: <input type="text"/>			Export as CSV		
Patent (uspto)	Title	Patent Classes	Appl. Year	Grant Year	Inventors	Assignee	Details
7311039	β-sheet mimetics and use thereof as inhibitors of biologically active peptides or proteins	e548: ORGANIC COMPOUNDS – PART OF THE CLASS 532-570 SERIES e514: DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS	2006	2009	MICHAEL KAHN (000128969) CYPRIAN O OCEBU MASAKATSU ECUCHI HWA OK KIM PATRICK D BOATMAN JR	MYRIAD GENETICS INC	▼
There are disclosed β-sheet mimetics and methods relating to the same for imparting or stabilizing the β-sheet structure of a peptide, protein or molecule. In one aspect, β-sheet mimetics are disclosed having utility as protease inhibitors in general and, more specifically, as serine protease inhibitors such as thrombin, elastase and Factor X inhibitors. In one embodiment, the β-sheet mimetic is a thrombin inhibitor.							
7670818	β-mannanase from coffee berry borer, hypothermus hampei, and uses thereof	e435: CHEMISTRY: MOLECULAR BIOLOGY AND MICROBIOLOGY	2009	2010	JOCELYN ROSE (259677040) RICARDO ACUNA ZORNOSA	FEDERACION NACIONAL DE CAFETEROS DE COLOMBIA	▶
7737132	β-cyclodextrin derivatives as antibacterial agents	e514: DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS e536: ORGANIC COMPOUNDS – PART OF THE CLASS 532-570 SERIES	2006	2010	NOUREDDINE FAHMI FRANK WERNER SCHMIDTMANN SIDNEY HECHT (000045475)	PINNACLE PHARMACEUTICALS	▶
7831457	β-Cyclodextrin derivatives	e514: DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS e435: CHEMISTRY: MOLECULAR BIOLOGY AND MICROBIOLOGY	2005	2010	SIDNEY HECHT (000045475) VLADIMIR KARCINOV NOUREDDINE FAHMI	INNOVATIVE BIOLOGICS INC	▶
7851457	β-Cyclodextrin derivatives	e514: DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS e435: CHEMISTRY: MOLECULAR BIOLOGY AND MICROBIOLOGY	2005	2010	SIDNEY HECHT (000045475) VLADIMIR KARCINOV NOUREDDINE FAHMI	INNOVATIVE BIOLOGICS INC	▶
7666840	α-Conotoxin peptides	e514: DRUG, BIO-AFFECTING AND BODY TREATING COMPOSITIONS e536: CHEMISTRY: NATURAL RESINS OR DERIVATIVES, PEPTIDES OR	2007	2010	MAREN VATKINS BALDOMERO M OLIVERA (000119304)	COCNETIX INC	▶

Basic Approach

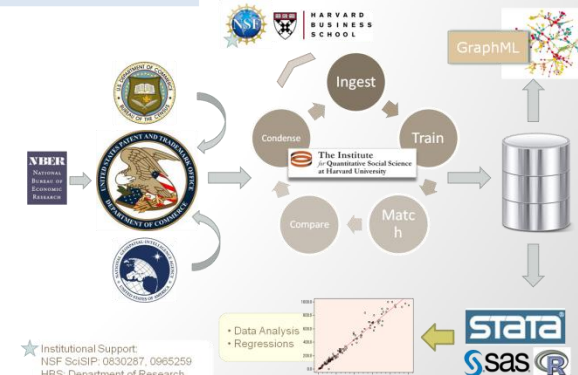
- **Focus:** Build coherent data infrastructure (not shiny tools) leverage existing systems – and minimize manual input
 - Worked with research community to build outputs/outcomes of interest
- **Unit of Analysis:** Individual - Senior personnel
 - Automated link to USPTO data
- **Incentives:** Create value for all stakeholders to ensure capture of right measures and data quality
 - Developing incentives for researchers – burden reduction? Name in lights? ORCID?

Approach: automatically capture data about the conduct of science – inputs, outputs and the connections between the two

Portfolio Explorer



Scientific Workforce



Source: Ian Foster University of Chicago



Example of the capacity in Australia

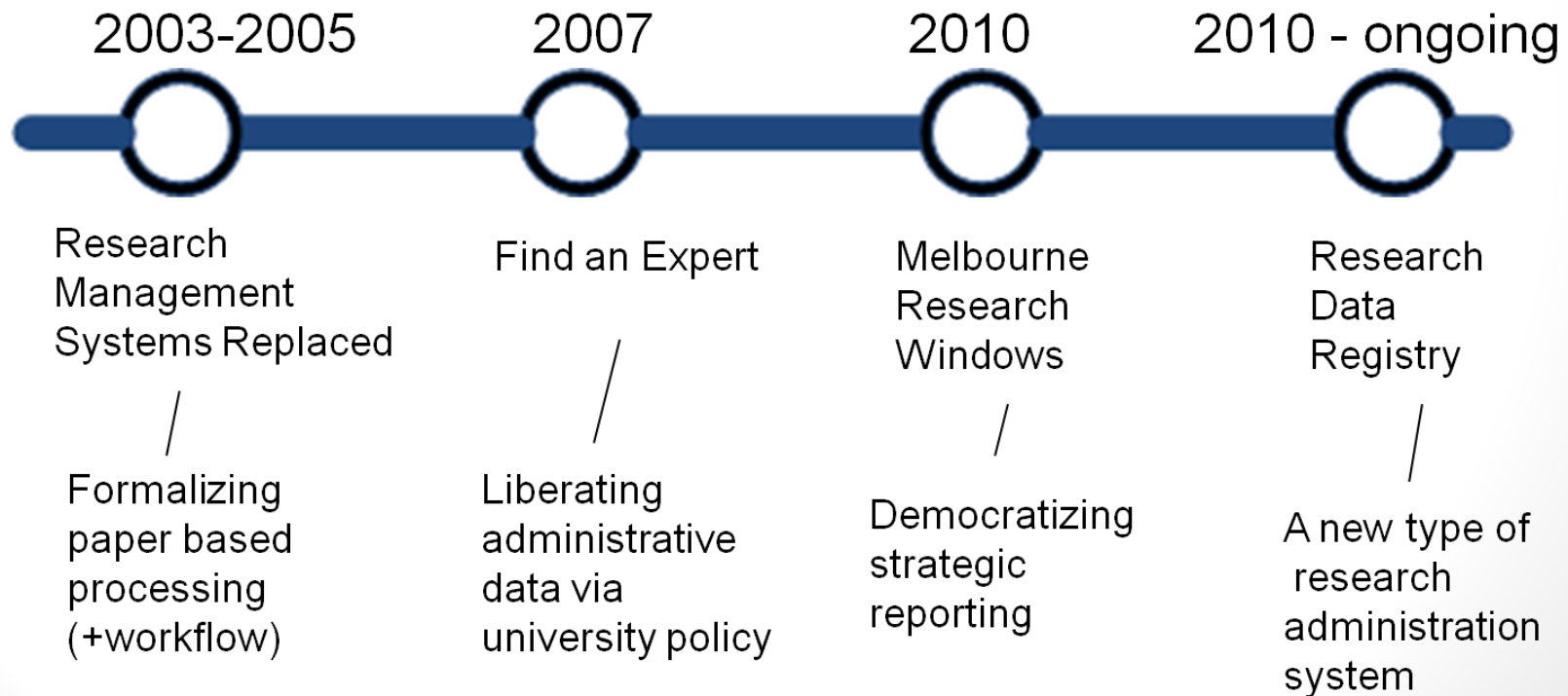


**How many ways would you like your
research metadata served?**

Simon Porter – Strategic Research Information
Analyst, the University of Melbourne

Existing systems

- A journey of meta data reuse at Melbourne....



Existing Data

Find an Expert

PROF LIZ SONENBERG

Contact Details

Organization: Melbourne Research Information Systems
 Position: PRO VICE-CHANCELLOR (RESEARCH COLLABORATION), CHAIR OF INFORMATION SYSTEMS
 Email: liz.sonenberg@unimelb.edu.au
 Homepage: <http://www.ris.unimelb.edu.au/staff/liz>
 Work: 9055 8819
 Fax: 9347 9326
 Room: 5.06
 Level: 05
 Building: Alan Gilbert Building
 Campus: Parkville

Biography

I am a professor in the Department of Information Systems and since August 2005 part-time role of Pro Vice-Chancellor (Research Collaboration) in Melbourne Res

The integrating theme of my research is the conceptualisation and constructor distributed, and intelligent information systems. Much of the work focuses on interacting autonomous software entities. Using the agent metaphor can allow s is useful for modelling complex tasks and environments, and in building software events. An important aspect of the research is the requirement of the human-m development of computational mechanisms to support decision-making in comple * Multi-agent systems - especially collaboration and teamwork
 * Automated negotiation and decision support

Qualifications, Honours, Fellowships and Other Awards

Qualifications

Title	Institution	Date Awarded	Abbreviation
Bachelor of Science	Monash University	21-Apr-1972	
Graduate Diploma in Computing Studies	University of Melbourne	31-Dec-1982	
Doctor of Philosophy	Monash University	31-Mar-1976	

Memberships

Membership Type	Membership Body	Description	Start Date	End Date
Member	ACS	Member	01-Jan-2006	

Government Research Classifications

Research Fields, Courses and Discipline Classifications

- Information Storage, Retrieval and Management (INFORMATION SYSTEMS) (280103)
- Computer-Human Interaction (INFORMATION SYSTEMS) (280104)
- Interfaces & Presentation (excl. Computer-Human Interaction) (INFORMATION SYSTEMS) (280105)
- Interorganizational Information Systems (INFORMATION SYSTEMS) (280106)
- Decision Support and Group Support Systems (INFORMATION SYSTEMS) (280109)
- Conceptual Modelling (INFORMATION SYSTEMS) (280111)
- Information Systems Development Methodologies (INFORMATION SYSTEMS) (280112)
- Information Systems not elsewhere classified (INFORMATION SYSTEMS) (280199)
- Text Processing (ARTIFICIAL INTELLIGENCE AND SIGNAL AND IMAGE PROCESSING) (280205)
- Intelligent Robotics (ARTIFICIAL INTELLIGENCE AND SIGNAL AND IMAGE PROCESSING) (280209)
- Simulation and Modelling (ARTIFICIAL INTELLIGENCE AND SIGNAL AND IMAGE PROCESSING) (280210)

Research Expertise and International Linkages

Research Expertise

Research Interest	Key Words	Country of Expertise
Multi-agent systems-collaboration and teamwork, Automated negotiation and decision support, Context-aware computing.	Computational modelling of Human problem solving	Australia

International Linkages

Country	Establishment	Collaboration
Netherlands	Utrecht University	Exchange, Research
United Arab Emirates	The British University in Dubai	Research

Grants and Contracts

Research Grants, Contracts and Consultancies awarded to the University of Melbourne as the administering institution (since 2003) as recorded in Themis Agreements.

Grants

Title	Role	Funding Source	Scheme	Award Date
Agent-based frameworks for coordinated activities in E-Business: supply chain management & workflow applications	Chief Investigator	AUST RESEARCH COUNCIL	Linkage Projects	01/01/2002
Automated interaction technologies for mobile hand-held devices	Chief Investigator	AUST RESEARCH COUNCIL	Linkage International Awards	01/01/2004
Interest-based negotiation: theory and practice	Chief Investigator	AUST RESEARCH COUNCIL	Discovery Projects	01/01/2005
Dynamic personalisation for assisted navigation of information rich, physical environments	Chief Investigator	AUST RESEARCH COUNCIL	Discovery Projects	01/01/2007
Cross-Community Information Systems: Understanding Technology-	Chief	AUST RESEARCH	Discovery Projects	01/01/2008

Publications

Publications produced at the University of Melbourne and reported in the Annual Publications Collection and Research Report since 2001. The Themis Publications module, released in November 2006, allows additional publications from previous institutions and publications from past years to be entered.

Publications in 2011

- Journal Articles
- An empirical study of interest-based negotiation([show details](#))

Publications in 2010

- Conference Publications/Papers
- A Composite Task Meta-model as a Reference Model([show details](#))
 - Smarttrumps: A Negotiated Relative Advantage([show details](#))

Publications in 2009

- Book Chapters
- Coordination in Adaptive Organisations: Extending Shared Plans with Knowledge Cultivation([show det](#))
 - Design and analysis of organization adaptation in agent systems([show details](#))
 - Exploiting Hierarchical Goals in Bilateral Automated Negotiation: Empirical Study([show details](#))

Journal Articles

- A Formal Analysis of Interest Based Negotiation([show details](#))

Example of the capacity in Japan

Researchmap Project, building an “ecosystem” of the research information

Noriko H. Arai
National Institute of Informatics,
Japan

Researchmap is a Science2.0 service for the researchers



The screenshot shows the Researchmap profile page. On the left is a sidebar with categories like 'Communities', 'Publications', 'Patents', etc. The main area has a list of categories such as 'Career', 'Prizes', 'Papers', 'Bibliography', etc. A dialog box titled 'Loading data from other databases' is open, listing sources like 'CiNii Article', 'CiNii Book', 'PubMed', 'LitSec', 'arXiv', and 'Amazon'. Below the dialog, a 'Web of Science, Patent' section is visible, showing a list of papers with details like titles and authors.

+ Web of Science, Patent

Author identification



Researchmap is a Sciendo2.0 service for the researchers



Research & Academic Activities

(Writing research papers, Experiments, Lectures, ...)

- University Researchers Database
- University Assessment Database
- R&D management system
- Research group homepage
- Science Portal
- J-rein ...



Research Products

- Electrical journals
- Database
- Experimental data
- Reports ...
- Institutional repository
- Patent

Primary Database Provider



API

Integrated Database Provider

▶ researchmap

API

Researchmap is a Science2.0 service for the researchers

▶ researchmap

Building the capacity in the US



Allows a user to "see" his/her profile as a member of the public, etc.



Thurston Howell, III

Executive Director, The Minnow Group
 Suite 229
 1234 S Main St
 Grovers Corners, NH

[edit public profile](#)

Views ▾

[Help / How do I...?](#)

Add Data Source

Create CV

"Action area"

Linked accounts/Data sources

Name	Type	Last Update
OpenID	Login	Today
Portfolio Explorer	Citations	02/11/2012
Scopus	Citations	01/05/2012
NSF	Grants	Today
Cornell VIVO	Citations	Today

[Manage linked accounts/Data sources](#)

Data sources, inputs to the system

Education

Washington University; St. Louis, MO
 Doctor of Philosophy in Immunology
 Dissertation: T cell Receptor function

Employment

University of Delaware; Newark, DE
 Bachelor of Science
 Major: Biology
 Minor: Music

Awards

[Manage content](#)

Bibliography

Source	Items	Last Update
All Items	272	Today
Scopus	112	01/05/2012
Manual Entry	14	08/29/2011
PubMed	146	Today

[Manage Bibliography](#)

SciENCV Collaborators

Name	Institution
John Doe	Polytechnic U.
Norville Rogers	Cal Berkeley
Daphne Blake	UCSD

[Manage Collaborators](#)

Smaller data objects are found here

CV Views

Name	Privacy	Last
CV1	Private	Today
CV2	Private	02/14/2012
Biosketch	Public	02/14/2012
Messaround	Private	07/22/2011

[Manage CVs](#)

Account History

Action	Date
Logged in via Google	Today
Made CV1 Private	Today
Edited CV2	02/14/2012
Created Biosketch	02/14/2012
Logged in via OpenID	01/05/2012

CV views and Account activity/history

The pitfalls

- Need to paint full picture of scientific outcomes
 - => engagement of scientific community critical
 - => Open and transparent process
- Data misuse
 - => careful presentation of results
- Data quality
 - => full collaboration
 - => extensive use of pilots
- Confidentiality
 - => researcher, institution and agency controls

Ultimate Goals

- Fully fledged academic field
- Fully fledged analytical tool set across government agencies
 - Science policy in same analytical tier as tax policy
- Common empirical infrastructure available to all universities and science agencies to quickly respond to State, Congressional and OMB requests
- Common scientific infrastructure for researchers to develop and study science policy

Thank you

- Comments and questions?

Background information

Dave Newman, UCI – and from
Melbourne!

What is the topic model?

- The topic model is an algorithm that automatically learns topics (themes) from a collection of documents
 - It works by observing words that tend to co-appear in documents, for example *gene* and *protein*, or *climate* and *warming*
 - The topic model assumes each document exhibits multiple topics
 - The topic model learns topics directly from the text

Latent Dirichlet allocation (LDA)

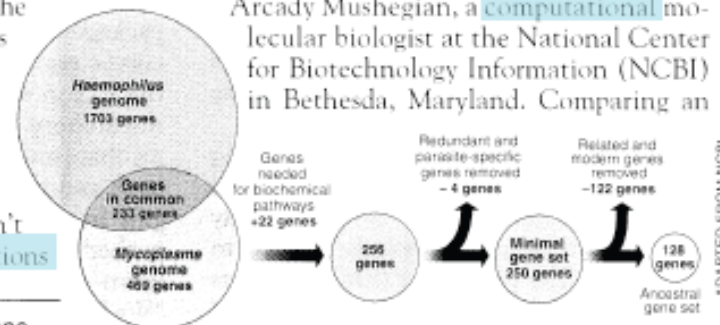
Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

“are not all that far apart,” especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. “It may be a way of organizing any newly sequenced genome,” explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

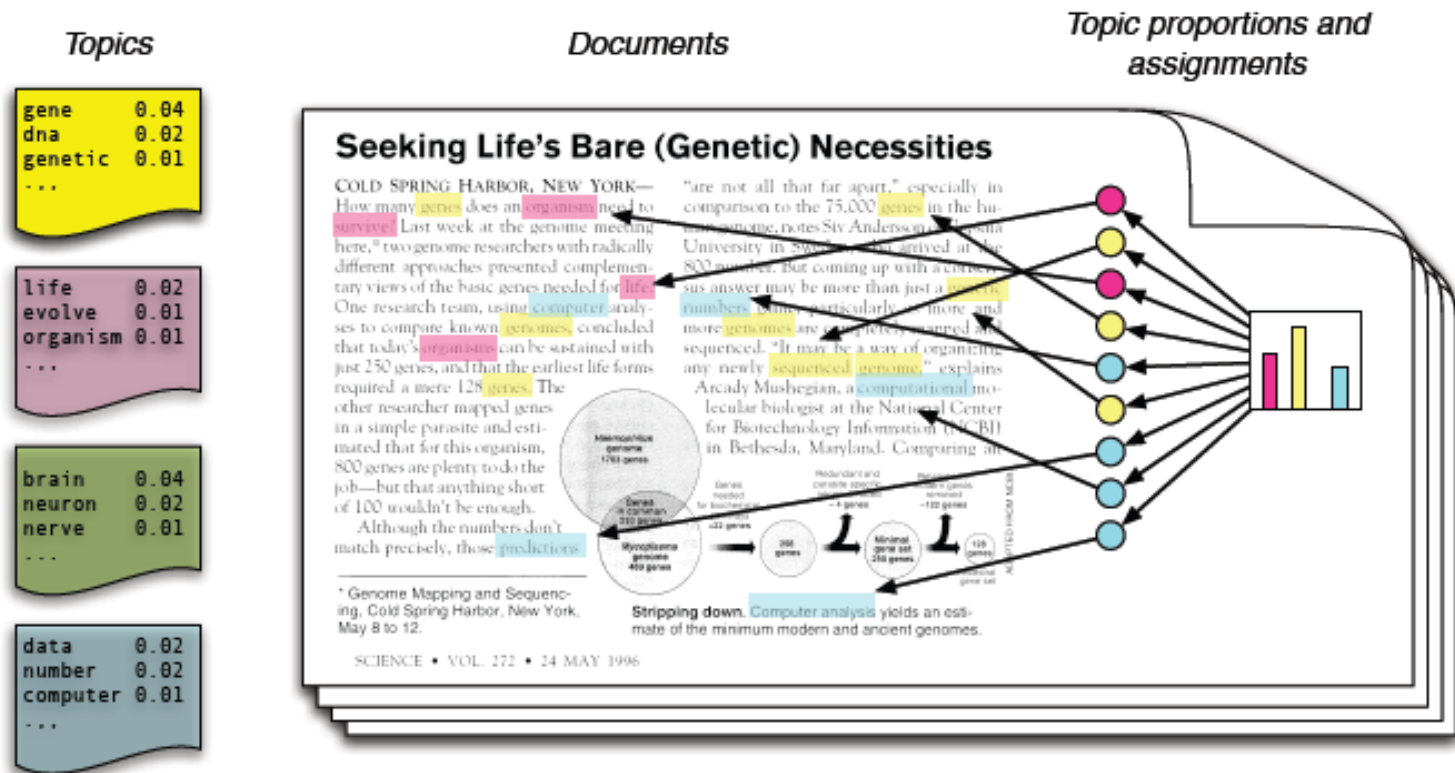


Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Simple intuition: Documents exhibit multiple topics.

Generative model for LDA



- Each **topic** is a distribution over words
- Each **document** is a mixture of corpus-wide topics
- Each **word** is drawn from one of those topics

Why is it useful?

- Topics can help organize, search and understand a document collection. One can ask:
 - What topics are in this document?
 - What documents are related to this topic?

Use topics to characterize things

- One proposal or award
- A set of proposals (in a program)
- An NSF division
- An investigator
- A reviewer
- An institution or center
- ... etc
- *And track changes over time*

Topics vs. Keywords

- Topics are complementary to keywords, so use both!
- But only topics give you dimensionality reduction (represent proposals by top-4 topics)
- Topics are like *keyword expansion*

Topic Modeling Process

Documents
(proposals,
awards,
etc.)

**Topic Modeling
Algorithm (LDA)**

Automatically Learned Topics

...

- t6. conflict violence war international military ...
- t7. model method data estimation variables ...
- t8. parameter method point local estimates ...
- t9. optimization uncertainty optimal stochastic ...
- t10. surface surfaces interfaces interface ...
- t11. speech sound acoustic recognition human ...
- t12. museum public exhibit center informal outreach
- t13. particles particle colloidal granular material ...
- t14. ocean marine scientist oceanography ...
- t15. atmospheric chemistry ozone air organic ...

...

top-4:

t49
t18
t114
t305

Topic tags for
every word in
every document

What is a topic?

- The topic model learns a set of topics
- Each topic is displayed by showing its top-20 words, for example:
 - dark_matter cosmological cosmology universe dark_energy lensing survey CMB redshift cosmic mass galaxy scale galaxies gravitational measurement power_spectrum parameter observation structure ...
 - This is a topic about *Dark Matter, Dark Energy and Cosmology*
 - Note: The ellipsis indicates that the list of words continues

Example of topics in an award

0910908: Intelligent Tracking Systems that Reason about Group Behavior

Top-4 topics in this award:

- (topic 400) Computer Vision
- (topic 512) Small Mammals
- (topic 319) Tracking Systems
- (topics 94) Complex Systems

The ability to reason about the complexity of living organisms in diverse environments is one of the hallmarks of intelligence. In this project the PI and her interdisciplinary team of investigators will design computer vision algorithms for intelligent tracking of large groups of living individuals in three-dimensional space. She will develop specific systems for tracking groups of microorganisms, bats, birds, and humans. And she will formulate machine learning methods for analyzing group behavior, specifically the conditions for formation and dispersal of groups, and the interactions of individuals within a group. An important innovative aspect of this research is the systematic and comprehensive approach to reasoning about the motion of large groups of living organisms observed in video data, independently of whether they happen to be humans, animals, or cells ...

Topic Labels (example)

Finance	[t300] financial bank credit tax loan insurance payment debt money finance market income
Risk Management	[t301] risk risk_assessment mitigation hazard risk_management potential risk_analysis co
Salmon and Trout	[t302] salmon pacific_northwest Pacific trout Brook River Boise rainbow MMI salmonid S
Plasma Physics	[t303] plasma discharge plasmas gas diagnostic pulsed electron species electrode tempe
PDEs	[t304] equation mathematical numerical solution mathematical_model differential_equa
Sequence Alignment	[t305] sequence alignment matching string match length matche position aligned matche
Halogens	[t306] fluorinated halogen chlorine fluorine fluoride halide iodine chemistry CF3 bromine
Great Lakes	[t307] Lake great_lake freshwater aquatic Erie water basin superior shore ecosystem dep
Surveys	[t308] survey respondent sample questionnaire survey_data responses population attitu
Analog Circuits	[t309] signal analog circuit digital ADC amplifier receiver output bandwidth frequency co
Fear Conditioning	[t31] conditioning fear conditioned tone contextual behavioral acquisition associative sti
News Media	[t310] media public radio newspaper coverage article audience magazine television pres

Topic Labels

- Topic labels are provided purely for convenience
- Our experience: Topic labels caused some confusion
- Pros
 - Simple and easier to display than *t387* or full 20-term topic
 - e.g. one can refer to the “Halogens” topic
 - Have expert(s) label topic
- Cons
 - Lose the richness and detail captured in the topic
 - Experts don’t agree
- Changing the topic label doesn’t change the topic assignment

Research and Non-Research Topics (sample)

1	[t222] lab laboratories equipment curriculum undergraduate exercises exercise laboratory_course
0	[t223] primate monkey chimpanzee species female lemur animal baboon ape sample wild fruit food
0	[t224] code simulation disk hydrodynamic calculation star accretion radiation gas angular_momentum
1	[t225] countries USA Europe European world global Germany Sweden Netherland Japan Canada France
0	[t226] sequencing genome sequence read genomic Illumina assembly sequenced coverage EST
0	[t227] oxidation radical reaction oxygen oxidant H2O2 oxidized product species oxidative
1	[t228] women gender family marriage sexual sex couple female men_women woman gendered male
0	[t229] environmental pollution pollutant contaminant EPA contamination health environment toxic
1	[t230] percent period hour twenty average twelve half thirty fifteen hundred fifty fourth forty
1	[t231] prairie grassland vole CEDAR grasshopper TPS Plain great_plain Wichita Hay MOM Flint Konza
1	[t232] Canada Canadian Grand Cook Ontario Toronto british_columbia North_American Vancouver
1	[t233] null *** null *** null *** null *** null *** null
0	[t234] fault diagnosis diagnostic failure faulty detection fault_tolerant fault_tolerance
0	[t235] wood asphalt pavement binder moisture construction mixture bamboo TDR aggregate
0	[t236] online blog Facebook Internet social social_network social_networking web user social_media
0	[t237] reconstruction scan resolution scanning tomography reconstructed scanner image reconstruct
0	[t238] diatom phytoplankton bloom culture nutrient marine dinoflagellate species cell cyanobacteria

Research and Non-Research Topics (sample)

- Different classes of non-research topics
 - NSF Admin Categories (STEM, REU, IGERT, etc.)
 - Workshop/conference
 - NSF broader impacts categories (women, underrepresented groups, etc)
 - Words used to describe research
 - Other “junk” topics (topics of inst. names, places, etc.)

Topics that mention “memory”

[t36] thread parallel transaction execution parallelism synchronization **memory** concurrent c

[t46] **memory** memories retrieval recall encoding recognition information episodic remembe

[t161] recovery smart recover recovered healing SMP self_healing **shape_memory** recovering
cycle ...

[t180] cache processor **memory** architecture hardware thread workload multicore compiler
execution ...

[t183] cluster computing node HPC processor high_performance server storage computation
comp

[t275] operating_system kernel hardware virtualization virtual_machine Linux software virtu

[t362] circuit logic gate output architecture bit function logic_gate signal digital CMOS QCA .

[t403] transformation transform transformed SMA transforming **shape_memory** NiTi phase_
SMAs [t541] cognitive psychology cognition mental cognitive_science psychological mind int
abilities

[t719] parallel processor GPU code computing computation MPI **memory** GPUs architecture
high_performance .

[t723] hardware FPGA architecture processor reconfigurable embedded_system platform FP

[t750] execution trace code bug software debugging error check call runtime overhead **mem**

[t834] stack stacking stacked tea ribbon structure TAC pattern thermoacoustic mechanism p

Top topics in a proposal

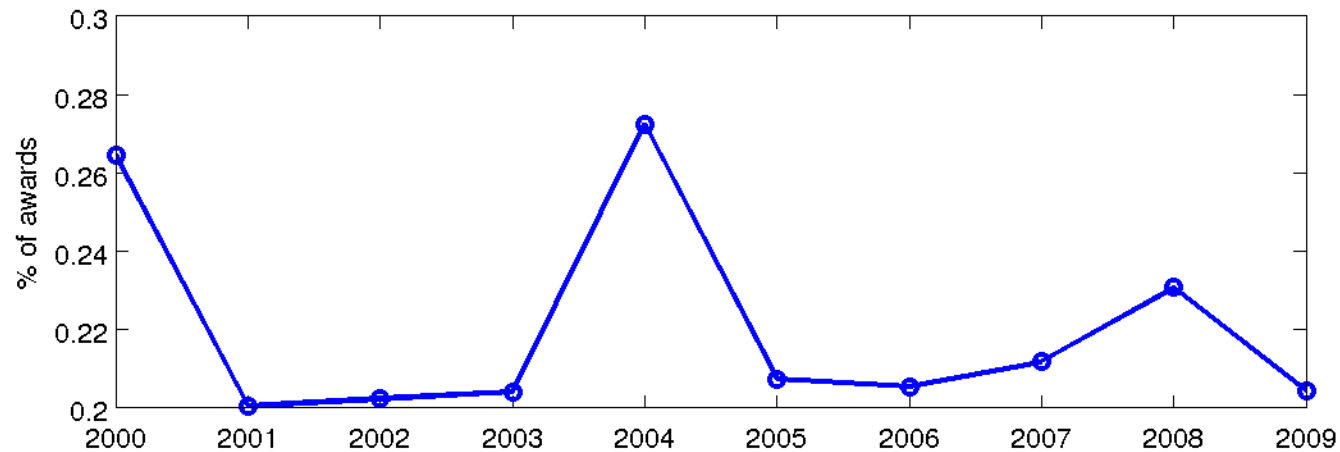
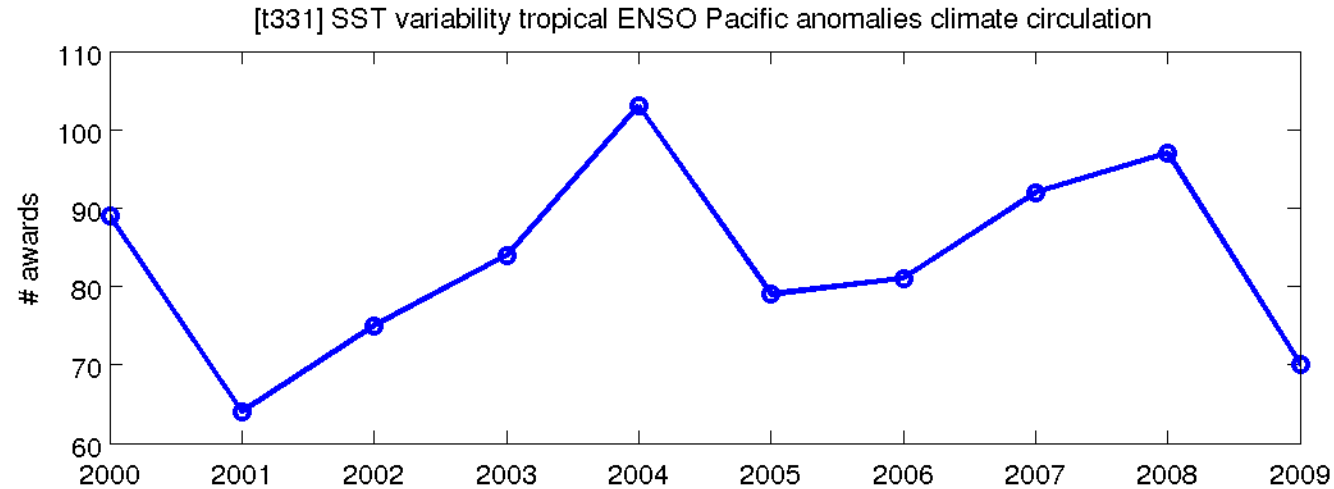
1103562

- (38%) [t851] permafrost arctic tundra warming alpine vegetation boreal soil alaska peat peatland depth meadow ecosystem ground active_layer northern thaw snow temperature ...
- (10%) [t930] fire vegetation disturbance wildfire grazing grassland burned burn grass burning fuel regime ecosystem shrub savanna woody landscape grasses rainfall plot ...
- (6%) [t783] native invasion plant species invasive plot invasive_species exotic seed non_native spread native_species plant_species establishment restoration seedling invader control invaded invasive_plant ...
- (6%) [t421] fellowship position postdoctoral_fellowship phd foreign imperative anticipated informed country check correspondence earned postdoctoral host_institution receipt registration fellow institution_including doctoral_degree citizenship ...
- (4%) [t384] modeling parameter modeled model_parameter model_development validation modeling_approach predict simulate empirical validated capture output simulation mathematical_model prediction incorporating modeling_effort realistic existing_model ...

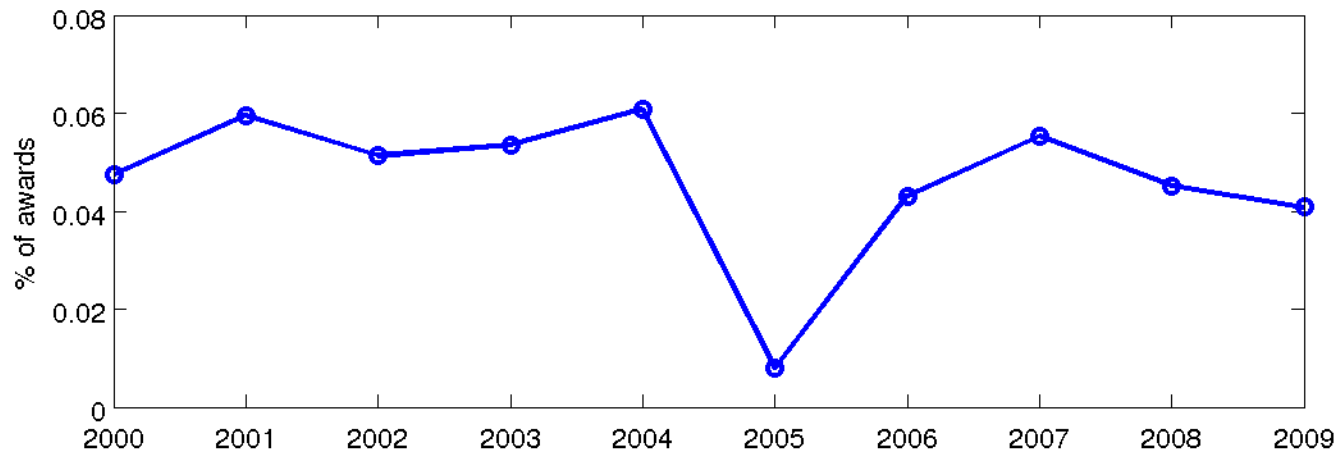
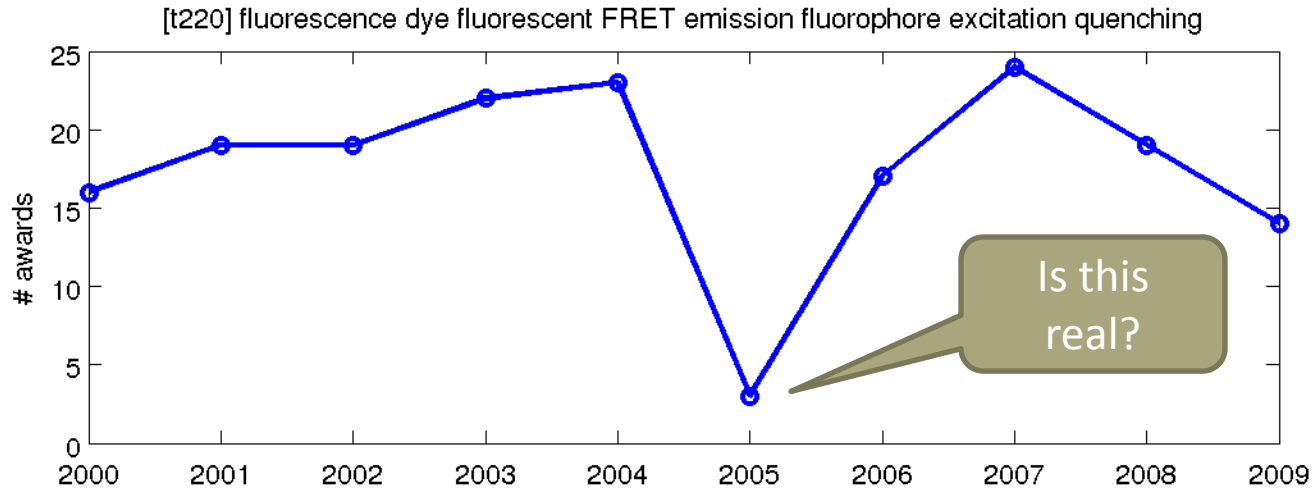
Analytical tools for PMs (example)

	Awarded	Declined
Overrepresented topics	<ul style="list-style-type: none">•••	<ul style="list-style-type: none">•••
Underrepresented topics	<ul style="list-style-type: none">•••	<ul style="list-style-type: none">•••

Topic trends 2000-2009



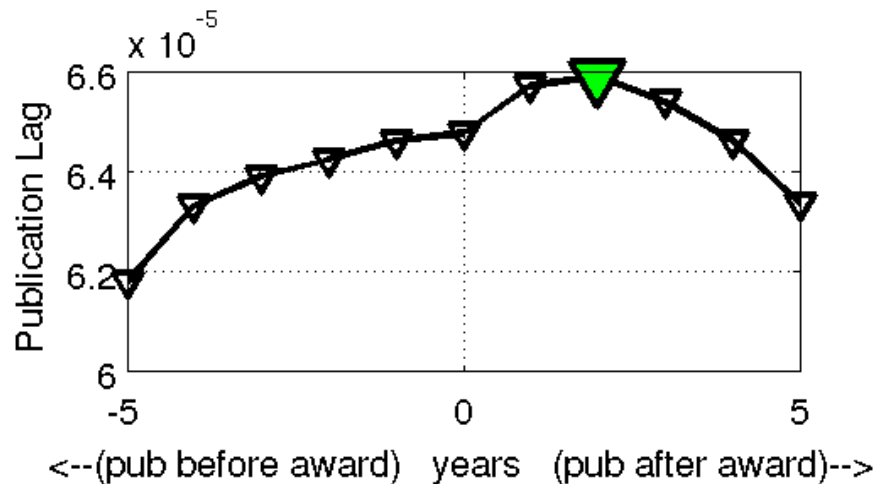
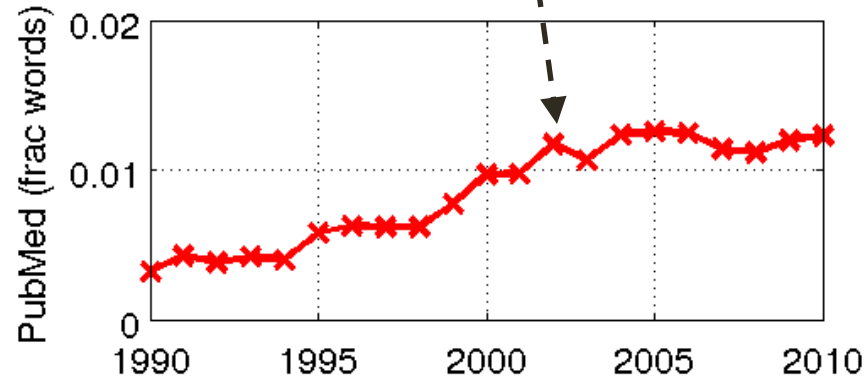
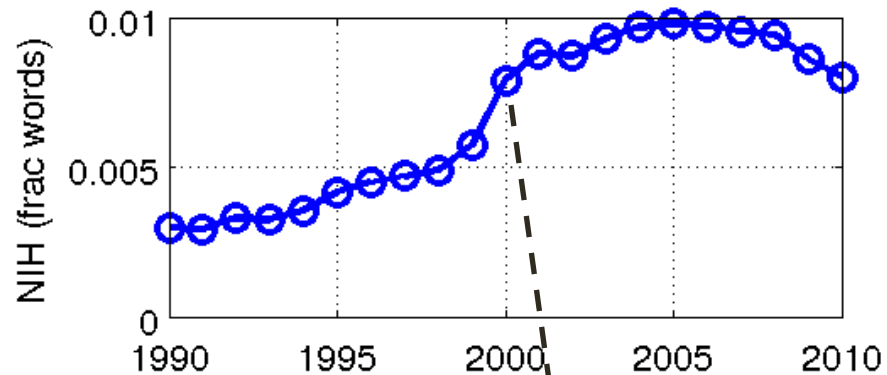
Topic trends 2000-2009



Research investments vs. outcomes

NIH investments in this angiogenesis topic correlated with publications 2 years later

Topic: angiogenesis tumor endothelial_cell vascular VEGF angiogenic vessel endothelial ...





If you are in government, academia, business or civil society and are interested in the links between government policy and research,

then please register with ANIPP

(Australian National Institute for Public Policy)

<http://publicpolicy.anu.edu.au>

to receive updates on thematic policy studies, events, research, activities and initiatives.

The 'HC Coombs Policy Forum' and the 'Australian National Institute of Public Policy' receive Australian Government funding under the *Enhancing Public Policy Initiative*.