STAR METRICS:

Science and Technology for America's Reinvestment:
Measuring the EffecTs of Research on Innovation, Competitiveness and Science

An Overview

Motivation

1. Good Government: Document Results
   - Who is supported by science funding?
   - What are the effects of science investments?

2. Good Management: Respond to Stakeholders
   - OMB/OSTP directives to Heads of Science Agencies
   - Congressional and public requests

3. Good Practices:
   - Analytical approach to complex problems
   - Utilization of cutting edge technologies

What is STAR METRICS?

STAR METRICS is a Federal and University partnership to document the outcomes of science investments to the public

•OSTP initiative partnering with NIH, NSF, DOE and EPA

•Phase I: uniform, auditable and standardized measures of the initial impact of ARRA and base budget science spending on job creation

•Phase II: collaborative development of measures of the impact of federal science investments on

- scientific knowledge (such as publications and citations)
- economic growth (through patents, firm start ups and other measures)
- workforce outcomes (through student mobility and employment)
- social outcomes (such as health and environment)

SCIENCE INNOVATION

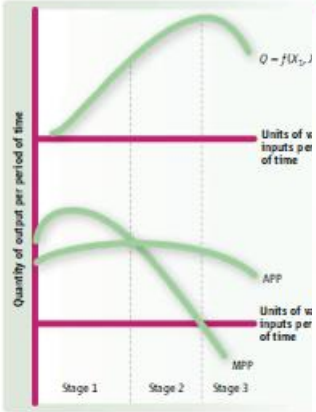
Assessing the Impact of Science Funding

Quantifying the outcomes of investment in science is not an easy task.

Julia Lane

Science supporters were rightly excited by the passage of the American Reinvestment and Recovery Act (ARRA, i.e., the stimulus package). Headlines in *Science* (*1*) and *Nature* (*2*) rejoiced at the new value placed on science as a basis for economic growth and associated job creation. Indeed, federal investment was at least partly based on a belief that the result would be more competitive firms and more, and better, jobs—and soon! (*3*). That belief was bolstered by advocacy groups: For example, a report by the Information Technology and Innovation Foundation (ITIF) estimated that an additional $20 billion investment in research in the stimulus package would create ~402,000 American jobs for 1 year.

Within 2 years, the public will want to be informed about the impact of the stimulus on the economic recovery. Were the estimates accurate? How can they be validated? And, in the longer term, what were the impacts of the reinvestment strategy on scientific knowledge, economic growth, and job creation? But we should also want to be informed about questions that go beyond the immediate accounting issues raised by ARRA. For example, what deeper understanding did we gain about the mechanisms whereby knowledge is created and how it contributes to both economic and social outcomes? Given the global nature of both economic and scientific activity, how did the science investments of other countries affect the United States? What new measures and indicators were developed to measure those contributions, and how can they be used to inform future investments and the response to future economic downturns? Answers to these questions will need to be based on theory and empirical evidence, as well as conveyed in a manner that is understandable to the general public.



A standard approach to linking inputs and outputs that used to study innovation. Output (Q) is shown as as a function of inputs (X_1, X_2, and X_3). The efficiency of different production is described by the relationship between average product (APP) and marginal physical product (MPP).

The public discussion about the "science stimulus," consistent with the apparent precision of the ITIF estimates, suggested that the outcomes of scientific investments were both certain and tied to economic growth. It is true that science policy in the United States and abroad is largely predicated on such beliefs. The United Kingdom's *Innovation*

Agenda identifies basic research as critical to productivity and employment growth (*5*), as does the Organization for Economic Cooperation and Development's (OECD's) innovation strategy (*6*). Saudi Arabia has invested $ billion to set up a new science and technology university (*7*), and the Japan Science and

researchers at the University of California at San Diego have been credited for the vibrant growth of San Diego, creating more than 40,000 jobs in life sciences and over 12,800 in electronics (*9*). The emergence of Google has been traced to National Science Foundation support of one of its founders, Sergey Brin, who was an NSF Graduate Research Fellow,

RESEARCH FUNDING

Measuring the Results of Science Investments

Science agencies and research institutions are building the infrastructure to evaluate results of federal funding of scientific research.

Julia Lane[1] and Stefano Bertuzzi[2]

Historically, federally funded basic and applied scientific research has promoted scientific knowledge, innovation, economic growth, and social well-being. However, there is increasing pressure to document the results of these research investments in a scientific manner (*1*, *2*) and to quantify how much of the work is linked to innovation (*3*).

Is it possible to create a system in which the effects of scientific research can be described? If so, what would be the inputs, outputs, and structure of the system? What scientific disciplines should inform the formulation of such a model? Creating a system in which the effects of scientific research can be described on an ongoing basis—without increasing the burden on research institutions and principal investigators—is difficult.

The current scientific data infrastructure is based on identifying, funding, and managing high-quality science, not on understanding its impact. The main sources of data on research and development in the United States—the Survey of Federal Funds for Research and Development (the federal funds survey) and the Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions—were designed to describe the types and levels of science investments, not their impact or effects (*4*). There are systems available to capture outcomes (for example, various health and economic information systems) but they do not link inputs with outputs and outcomes. Historically, there have been limited resources

The challenge is not limited to the United States; other countries have been developing systematic ways of describing the results of science investments. Since 1986, the Higher Education Funding Councils in England has assessed research with its Research Assessment Exercises (now a Research Assessment Framework) intended to assess the quality, impact, and vitality of funded research. Their lessons are salutary: Although the exercises did help to improve research quality, the process of producing the data was burdensome and complex (*7*). In 2009, the European Union EUFORDIA conference, which examined the impact of the Framework Programme (FP) 6, included, as a major recommendation, of building a database of project results for future FPs, noting that "getting robust data on the FPs in terms of participation and results is the foundation for any evaluation" (*8*). In 2011, the Japanese government is creating a program to advance the science of science and innovation.

A high-quality system should be based on describing the activities of scientists and clusters of scientists. Of course, the direct output of research is knowledge, which includes even research "failures," and is difficult to measure. Despite this, the system should include proximal measures of scientific output (such as publications, citations, and patents) and go well beyond simple publication

> *"The intent is to leverage revolutionary digital technology to capture the broad scientific, social, economic, and workforce impacts of science investments."*

Developing such a system and the associated data infrastructure will require financial and intellectual resources. Other efforts to put together a data infrastructure describing the outcomes of research and development (R&D) investments, both by the private and the public sectors, no longer function for a variety of reasons (*11*). The new focus on accountability, combined with new technology and the broad-based commitment of key stakeholders, may result in a better outcome.

Currently, key data elements are dispersed across federal agencies and research institutions or are in third-party databases. For example, information about what science is being funded is often neither in structured format nor systematically shared across agencies; administrative information about the students supported by federal funding is housed at research institutions, but not by the

Published by AAAS

OPINION

Let's make science metrics more scientific

Stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says Julia Lane — how else shall we know what to reward?

Measuring and assessing academic performance is now a fact of scientific life. important discovery from left-field. It is true that good metrics are difficult to develop, but

SUMMARY

- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

Identifier (DOI) protocol, which has become the international standard for identifying unique documents. The ORCID (Open Researcher and Contributor ID) project, for example, was launched in December 2009 by parties including Thompson Reuters and Nature Publishing Group. The engagement of international funding agencies would help to push this movement towards an international standard.

Similarly, if all funding agencies used a universal template for reporting scientific achievements, it could improve data quality and reduce the burden on investigators. In January 2010, the Research Business Models Subcommittee of the US National Science and Technology Council recommended the Research Performance Progress Report (RPPR) to standardize the reporting of research progress. Before this, each US science agency required different reports, which burdened principal investigators and rendered a national overview of science investments impossible. The RPPR guidance helps by clearly defining what agencies see as research achievements, asking researchers to list everything from publications produced to websites

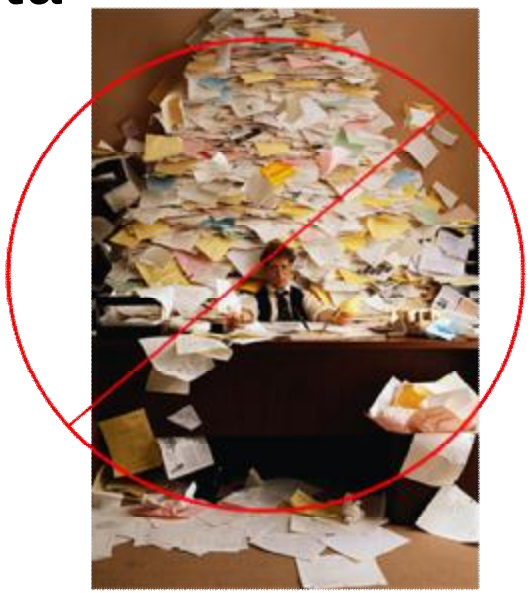
Basic Approach

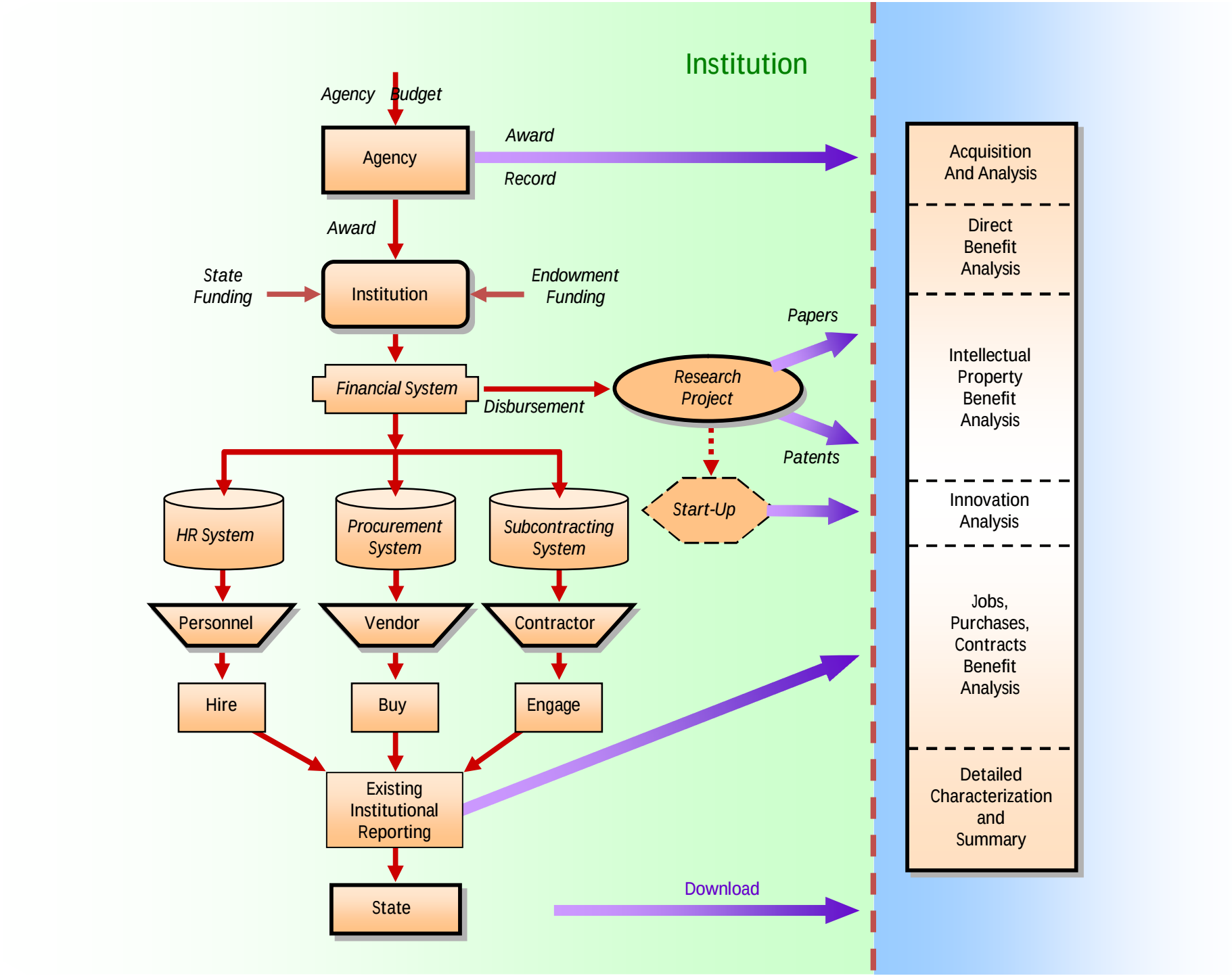
Creating the Frame

- Start with basic unit of analysis
  - Science is done by scientists. Need to identify universe of individuals funded by federal agencies (PI, co-PI, RAs, graduate students etc.)
- Capture Inputs using existing data

Measuring outcomes

- Scientific
- Social
- Economic
- Workforce




Institution
Agency Budget
Agency
Award
Record
Award
State Funding
Institution
Endowment Funding
Papers
Financial System
Disbursement
Research Project
Patents
Start-Up
HR System
Procurement System
Subcontracting System
Personnel
Vendor
Contractor
Hire
Buy
Engage
Existing Institutional Reporting
State
Download
Acquisition And Analysis
Direct Benefit Analysis
Intellectual Property Benefit Analysis
Innovation Analysis
Jobs, Purchases, Contracts Benefit Analysis
Detailed Characterization and Summary

Phase I: What is requested

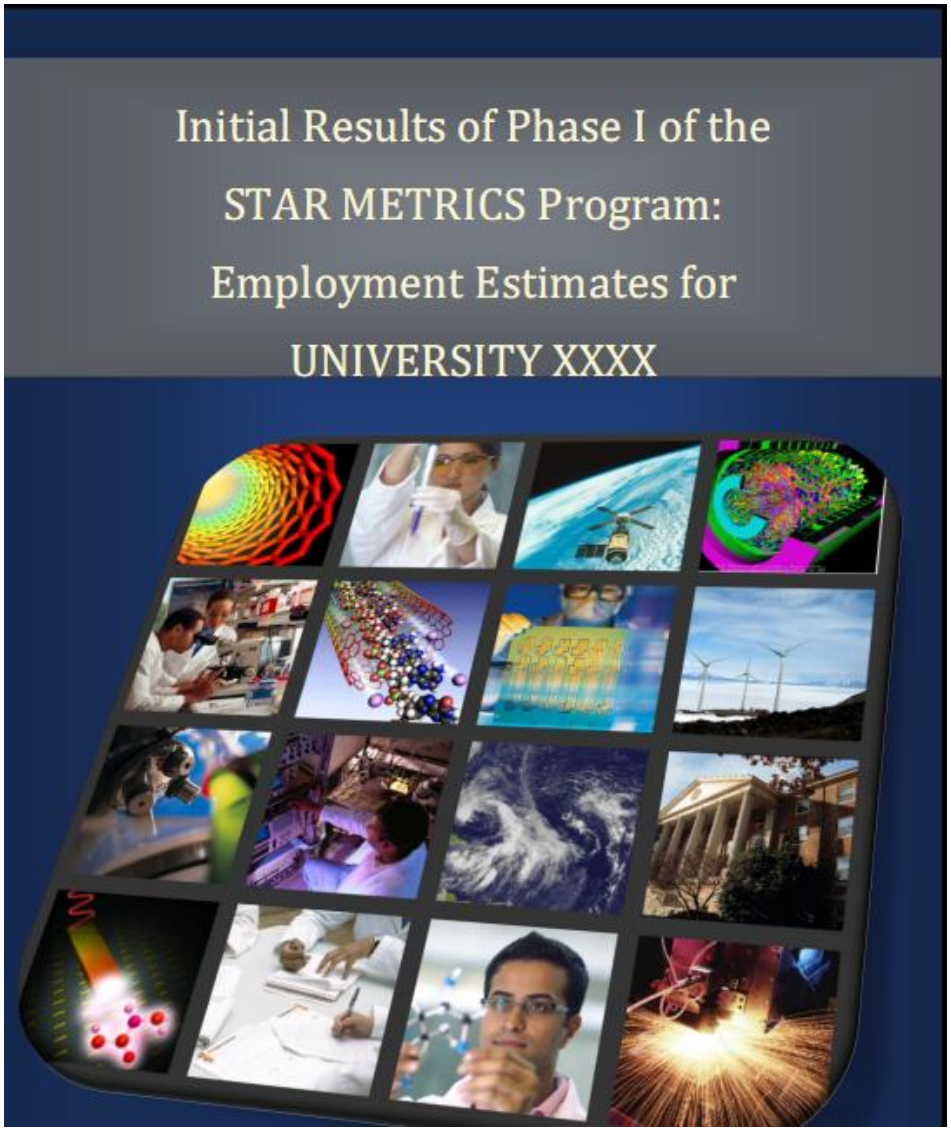
14 administrative data elements from awards, grants, HR or finance systems are provided to STAR Metrics on a quarterly basis…

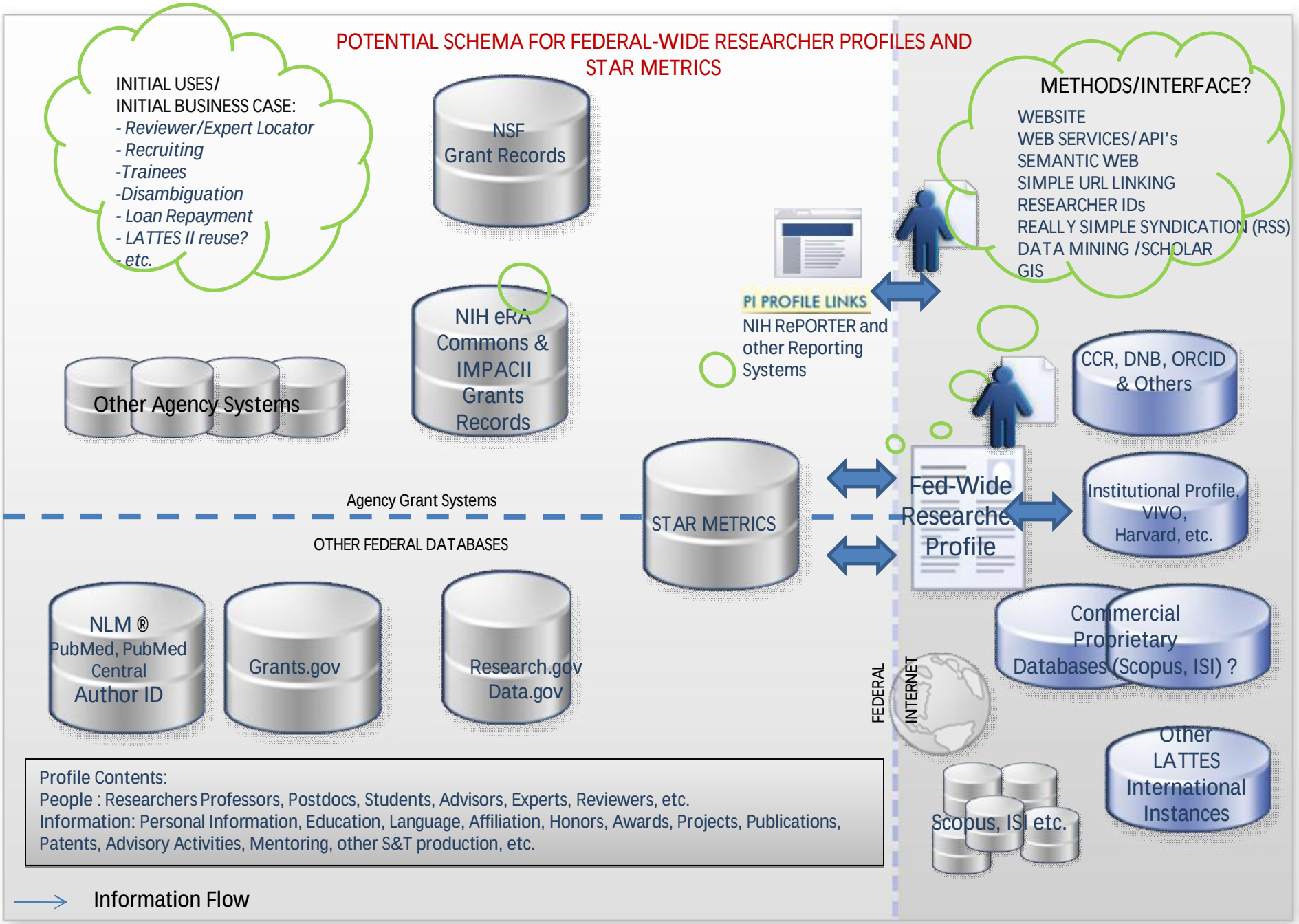
- Award data
- Payroll Staff Information
- Non-Payroll Charges
- Sub-awards
- Indirect Cost Rate Proposal

…will yield these Quarterly pre-calculated reports…

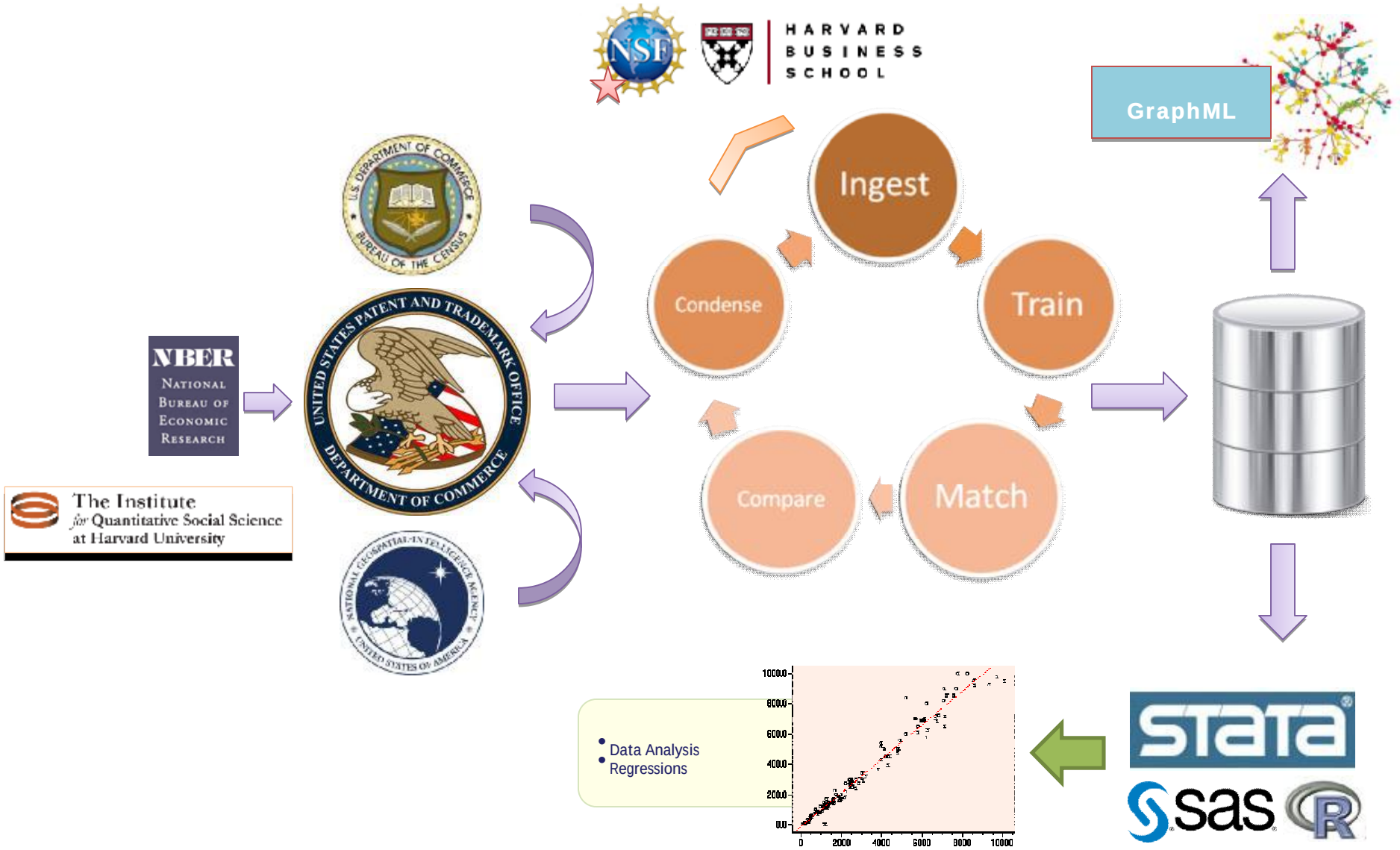
- Ø Stimulus FTE Jobs (ARRA) – with and without Overhead Job calculations
- Ø FTE Jobs and Positions – All awards (with and without Overhead)
- Ø FTE Sub-awards – All awards (with and without Overhead)
- Ø Vendor FTE's (Jobs) – All awards
- Ø Overhead Jobs (calculated from Indirect Costs)

What is Returned ?



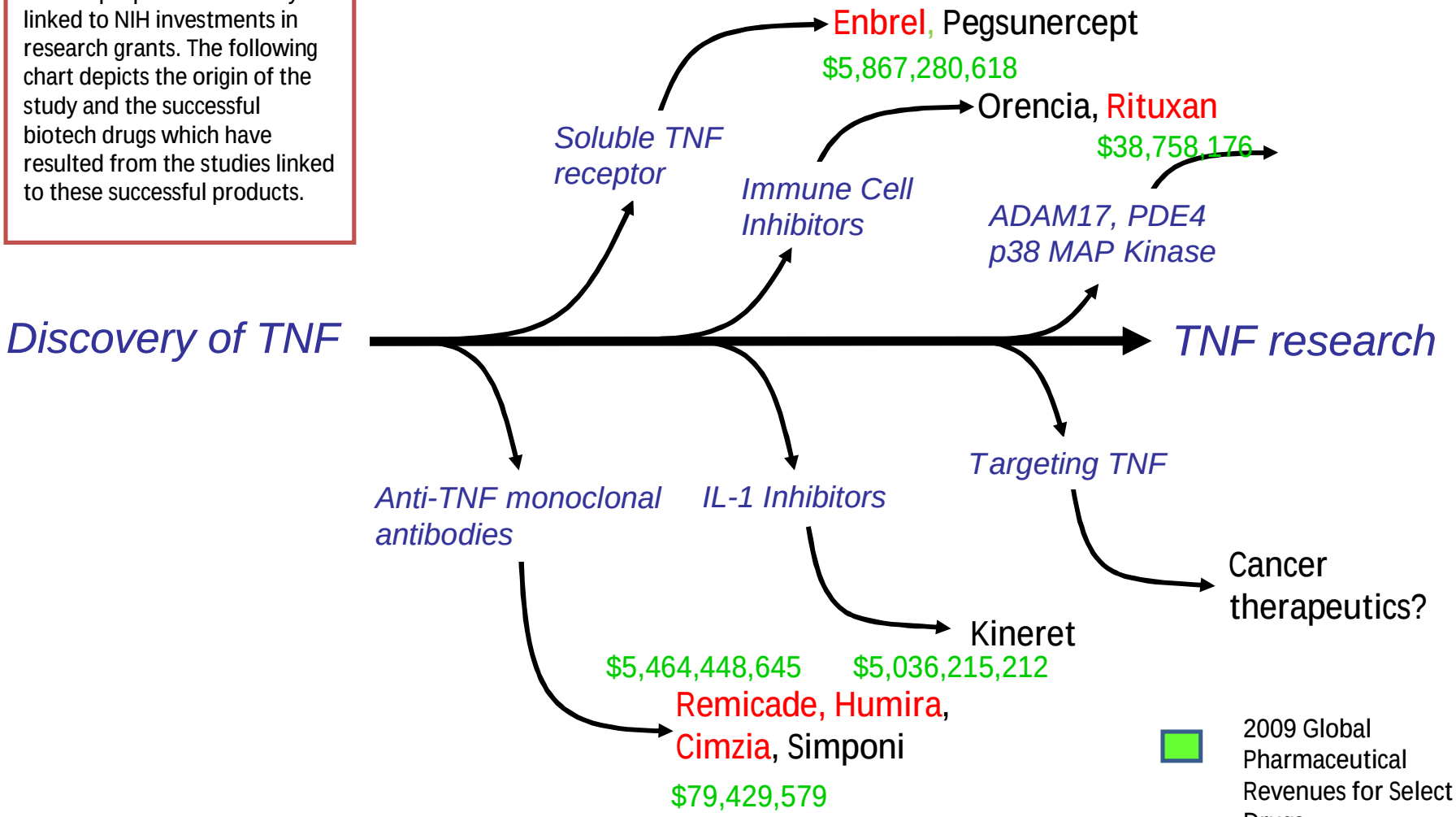

POTENTIAL SCHEMA FOR FEDERAL-WIDE RESEARCHER PROFILES AND STAR METRICS
INITIAL USES/
INITIAL BUSINESS CASE:
- Reviewer/Expert Locator
- Recruiting
-Trainees
-Disambiguation
- Loan Repayment
- LATTES II reuse?
- etc.
NSF Grant Records
METHODS/INTERFACE?
WEBSITE
WEB SERVICES/API's
SEMANTIC WEB
SIMPLE URL LINKING
RESEARCHER IDs
REALLY SIMPLE SYNDICATION (RSS)
DATA MINING /SCHOLAR
GIS
PI PROFILE LINKS
NIH RePORTER and other Reporting Systems
NIH eRA Commons & IMPACII Grants Records
Other Agency Systems
CCR, DNB, ORCID & Others
Agency Grant Systems
STAR METRICS
Fed-Wide Researcher Profile
Institutional Profile, VIVO, Harvard, etc.
OTHER FEDERAL DATABASES
NLM ® PubMed, PubMed Central Author ID
Grants.gov
Research.gov Data.gov
Commercial Proprietary Databases (Scopus, ISI) ?
FEDERAL INTERNET
Other LATTES International Instances
Scopus, ISI etc.
Profile Contents:
People : Researchers Professors, Postdocs, Students, Advisors, Experts, Reviewers, etc.
Information: Personal Information, Education, Language, Affiliation, Honors, Awards, Projects, Publications, Patents, Advisory Activities, Mentoring, other S&T production, etc.
Information Flow

Capturing Outputs and Outcomes: Harvard/NBER Disambiguated Patent Database

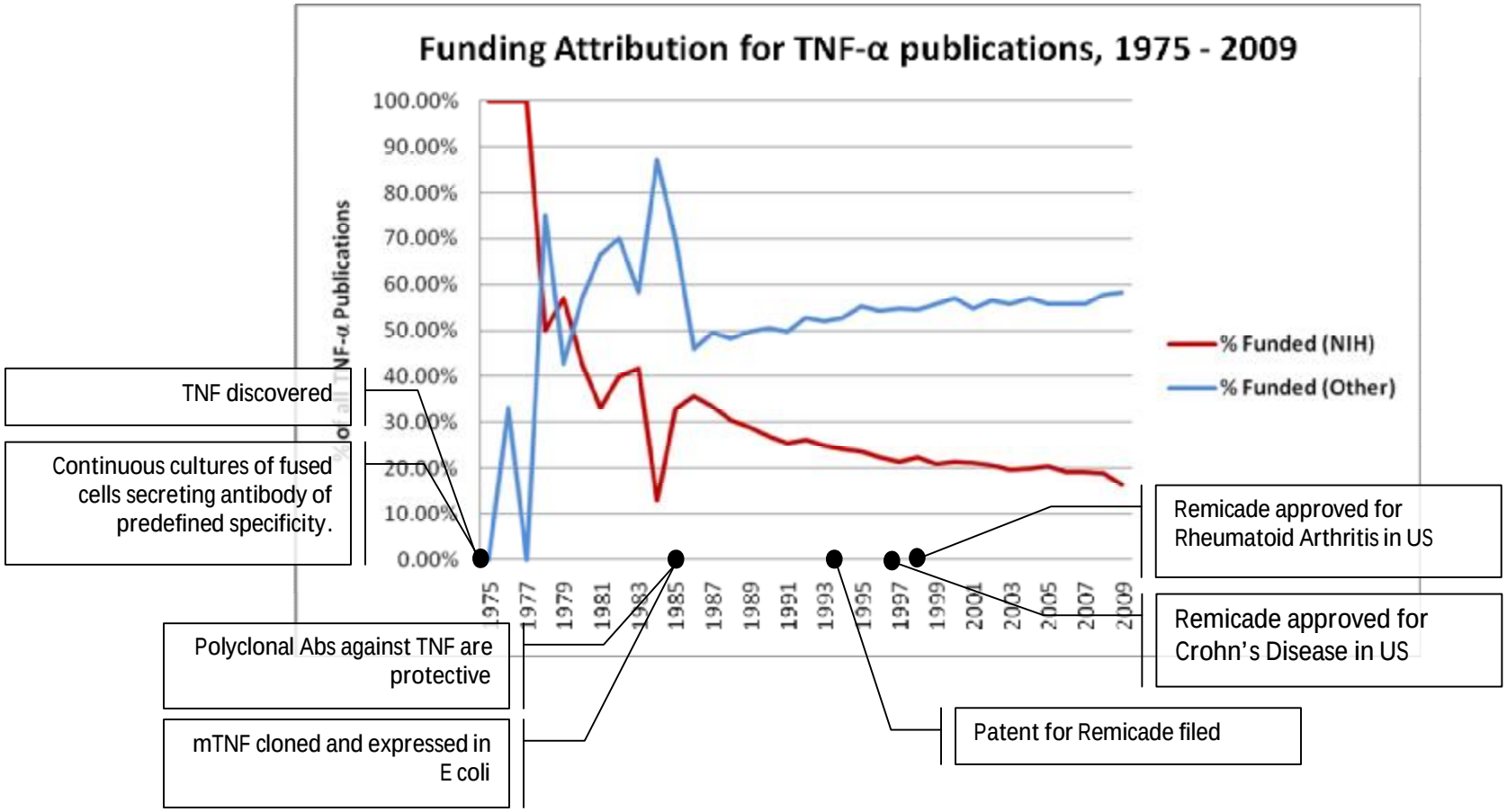


Evolution of Disease Modifying Anti-Rheumatic Drugs (DMARD) TNF- α drugs

The discovery of TNF and its related properties is directly linked to NIH investments in research grants. The following chart depicts the origin of the study and the successful biotech drugs which have resulted from the studies linked to these successful products.



Funding trends in TNF-α publications



Events leading to the approval of Enbrel

1975 Discovery of TNFα

1975 Hybridoma technology for mAbs

1985 Passive immunization with polys against TNFα are protective against endotoxin

1984 Chimeric mAbs

1993 Characterization of a human-mouse TNFα chimeric mAb

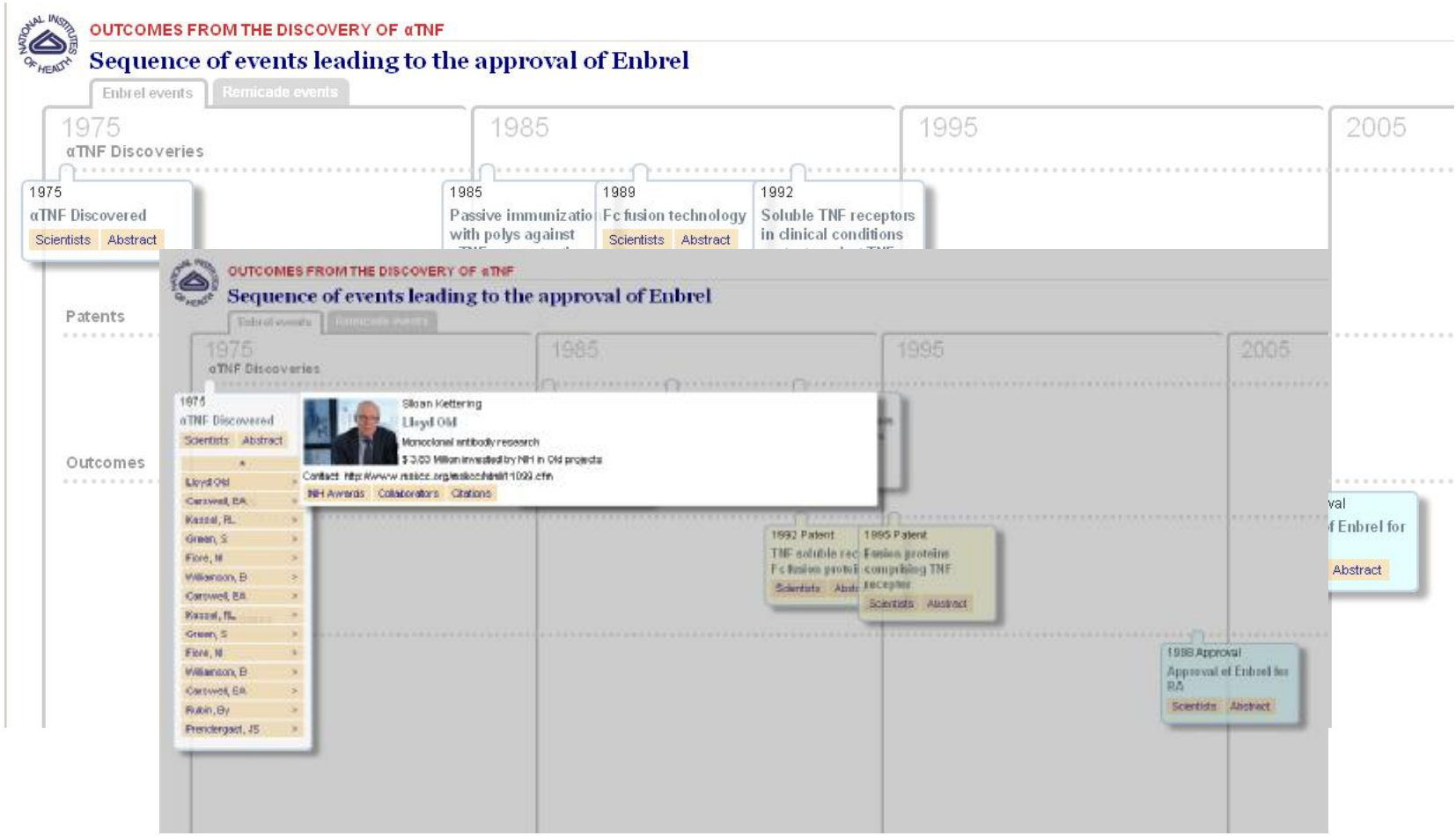
1997 Methods for treating Crohn's Disease using chimeric mAbs (Patent)

1997 Methods for treating RA using chimeric mAbs (Patent)

1998 Approval of Remicade for Crohn's Disease

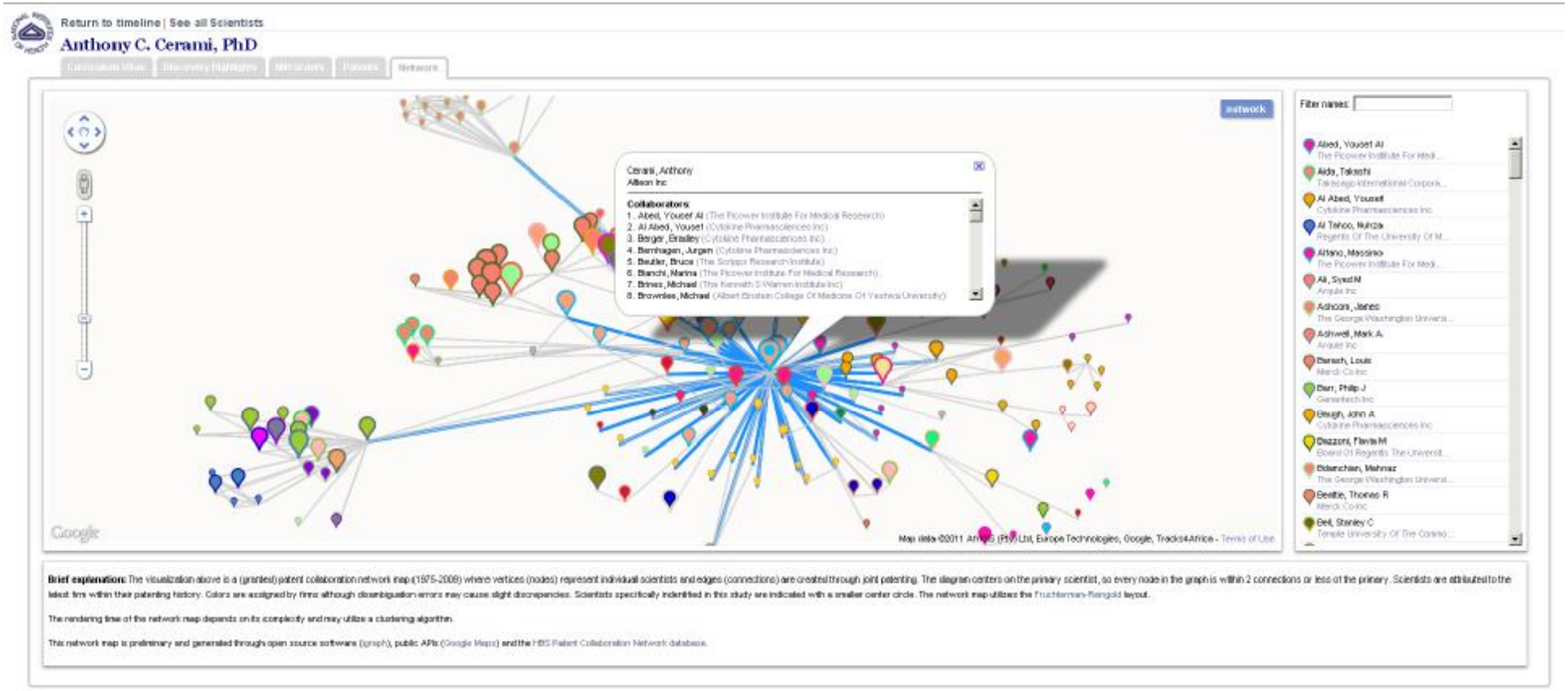
NIH Funded Studies data overlay –
pubs
grants
citations
Notable PI's

The Network of Scientists that did it



Collaboration & Innovation Network Maps

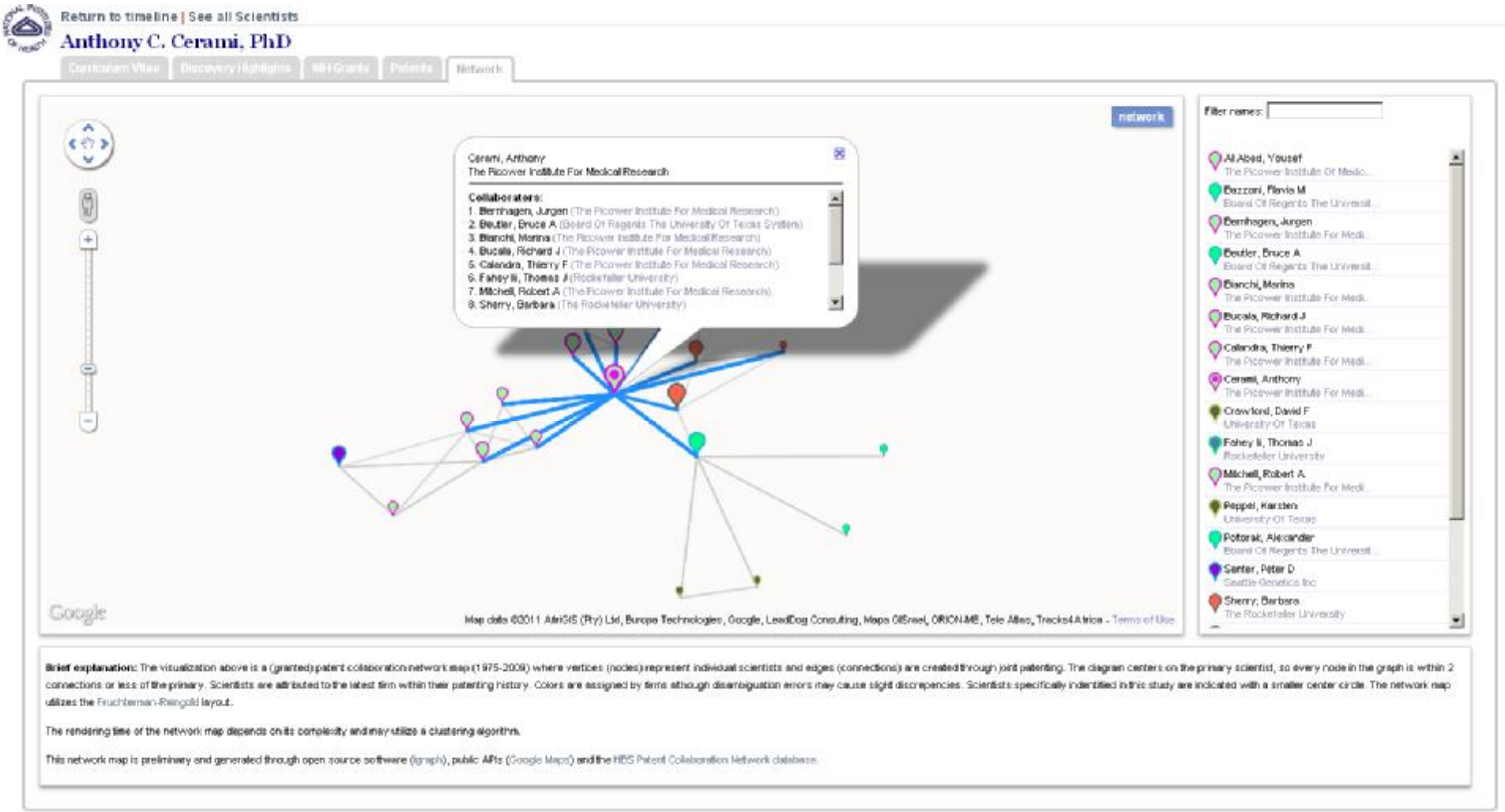
The ripple effect of NIH funding



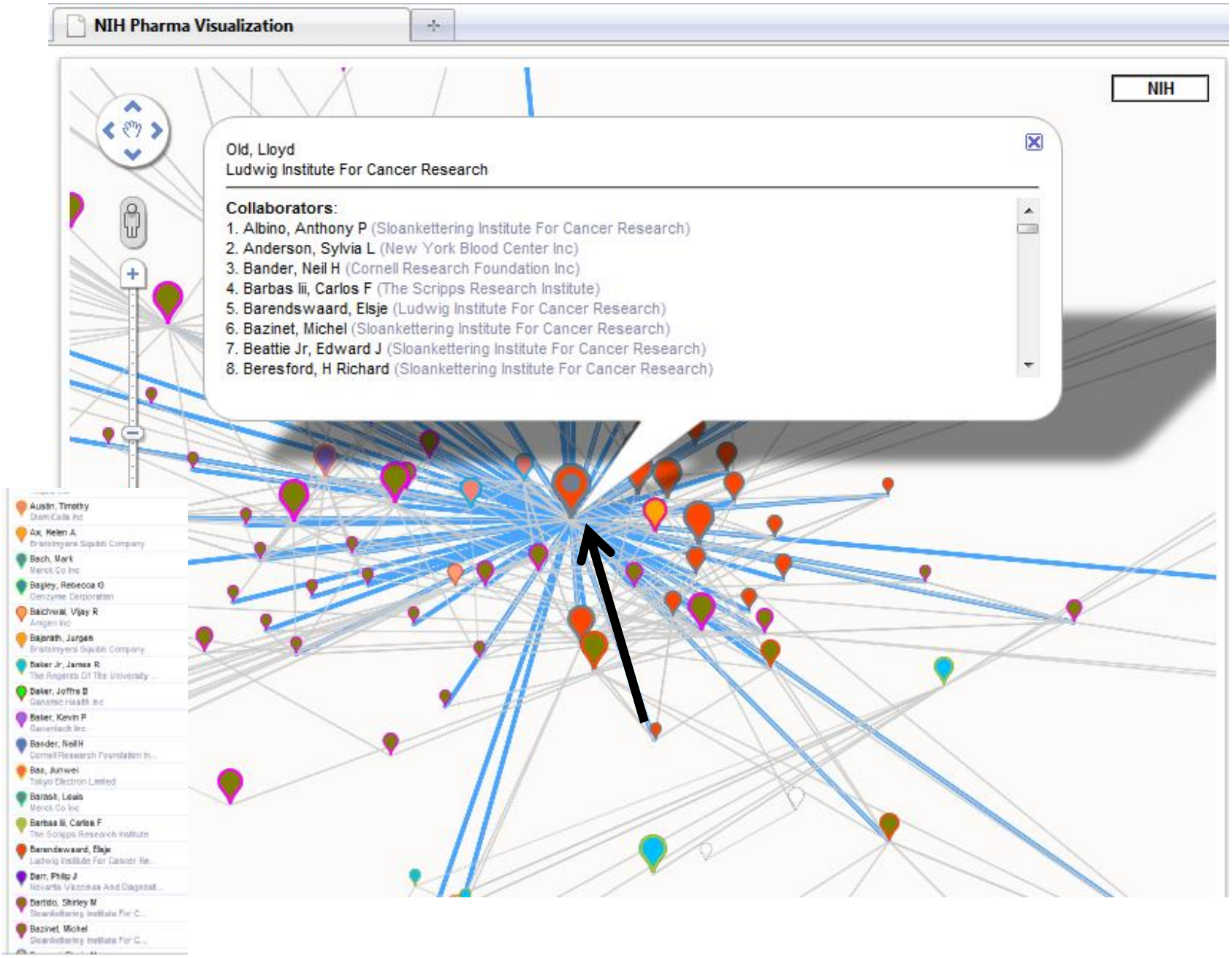
Collaboration & Innovation Network Maps

The ripple effect of NIH funding

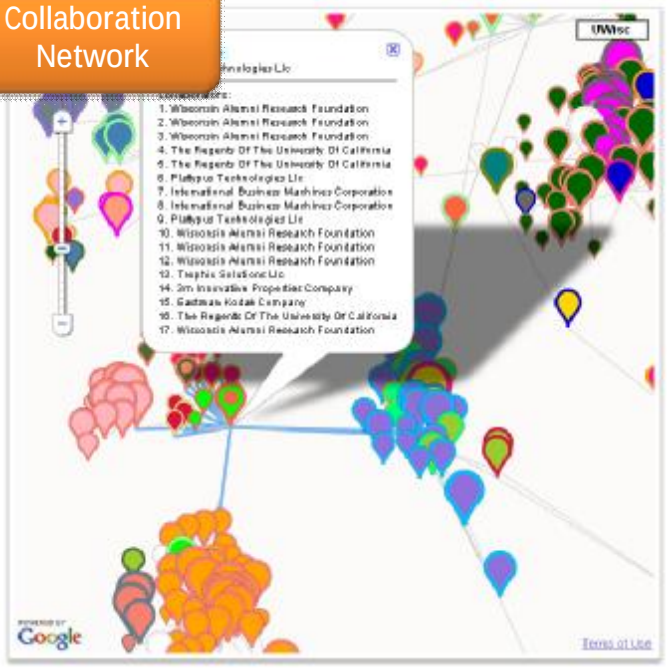
The network of *Enbrel/Remicade-relevant patents* authored by Dr. Anthony Cerami and his immediate collaborators:



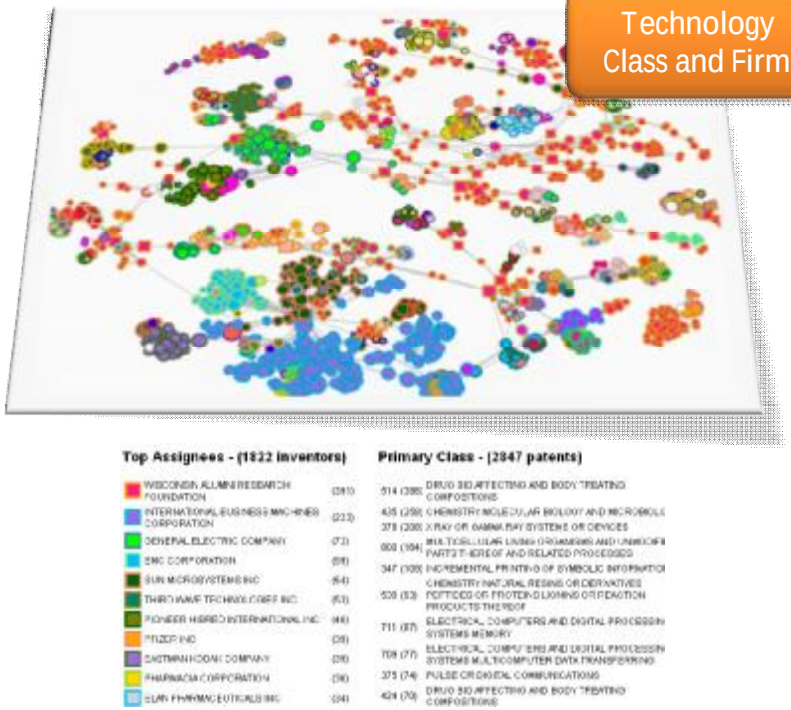
Network Visualization



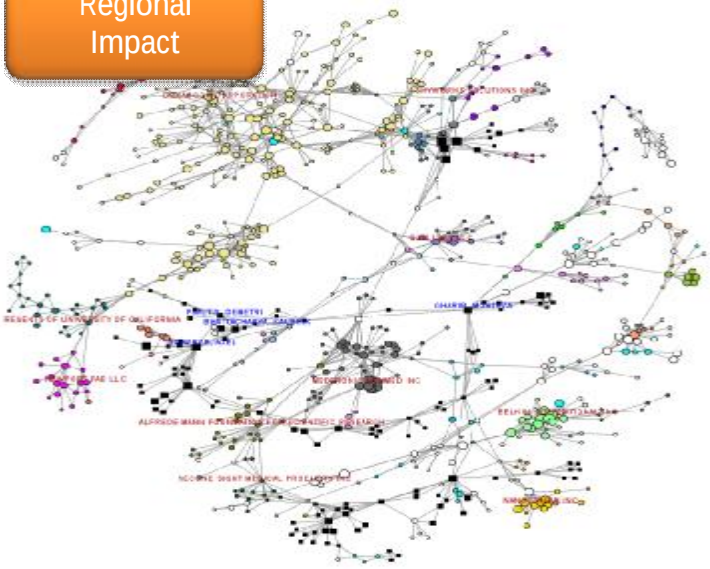
Collaboration Network



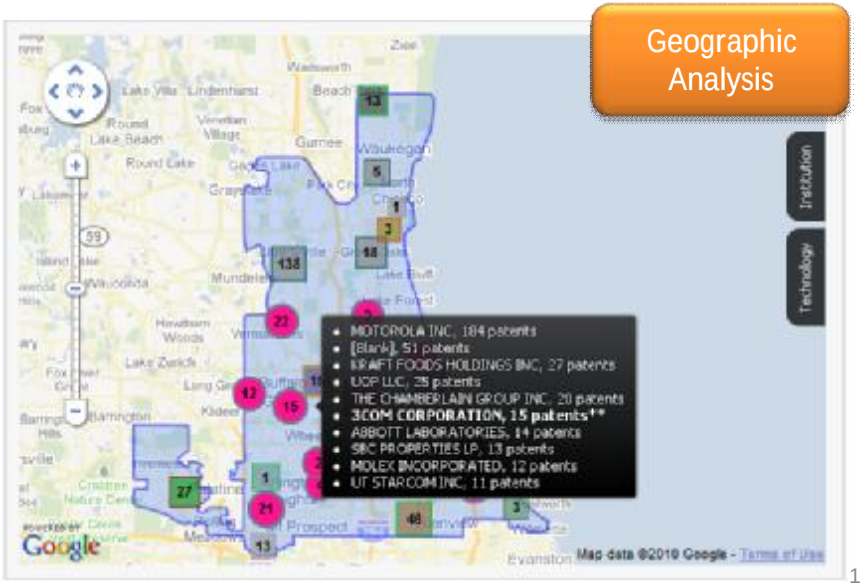
Technology Class and Firm



Regional Impact



Geographic Analysis

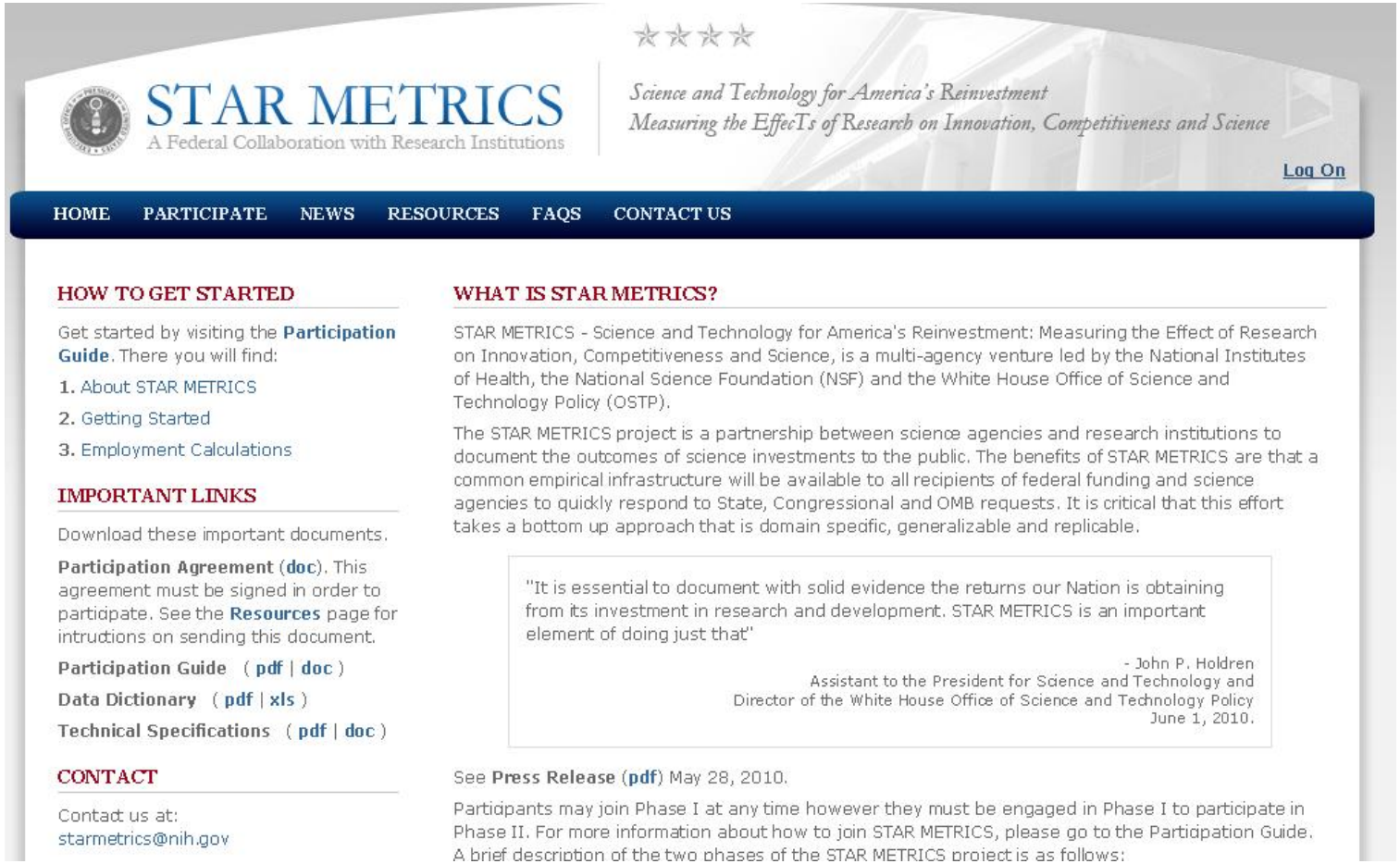


Next Steps:

- Continue Phase I
- Technical meetings to develop the key features of the Phase II platform
- Meeting with FDP Steering Committee
- Presentation at FDP Meeting
- Implement a pilot project with a set of institutions
- Regional Workshops

More Information:

https://www.starmetrics.nih.gov/



For More Information – and to provide input

- Contact
- Stefano Bertuzzi (stefano.bertuzzi@nih.gov)
- Kei Koizumi (Kei_Koizumi@ostp.eop.gov)
- Julia Lane (jlane@nsf.gov)
- Mya Strauss (strauss.mya@epa.epamail.gov)
- Bill Valdez (bill.valdez@science.doe.gov)
- https://www.starmetrics.nih.gov