National Patterns of R&D Resources: Key Aspects of Data, Methods, and Reports

Mark Boroush
Senior Analyst, R&D Statistics Program

CNSTAT Workshop on Future Directions for the
NSF National Patterns of R&D Reports

September 6, 2012

National Science Foundation
National Center for Science and Engineering Statistics
www.nsf.gov/statistics/

National Patterns of R&D – the Big Picture

- U.S. total R&D expenditures and sources of funding (annual, domestic performance)
- Covers all major U.S. R&D performers and sources of funding
- Data timeline: 1953 to current (2009 – 2010/11 in late fall)
- Derives primarily from compiling and combining the published data from the NCSES annual surveys of R&D expenditures by U.S. R&D performers
- Generally comparable to the national R&D totals reported by the other major countries
- Published by NCSES as InfoBrief and Data Update (annual – for the most). Also included as part of NSF's biennial *Science and Engineering Indicators*

Main Outputs of National Patterns

■ CORE

• U.S. total R&D expenditures and sources of funding

• With disaggregations for:

- Performers: businesses, federal government (intramural, FFRDCs), universities & colleges, other nonprofit organizations

- Funders: businesses, federal government, nonfederal government, universities & colleges, other nonprofit organizations

- Character-of-work: basic research, applied research, development

• Some 168 variables tracked. 1953 to current reporting year. Current and inflation-adjusted dollars. Calculated shares and comparative growth rates.

■ SUPPLEMENTARY

• U.S. aggregate R&D/GDP ratios – most recent year, trend over time

• U.S. total R&D expenditures compared to that of other large countries

• U.S. total R&D expenditures disaggregated by state (performers, funding)

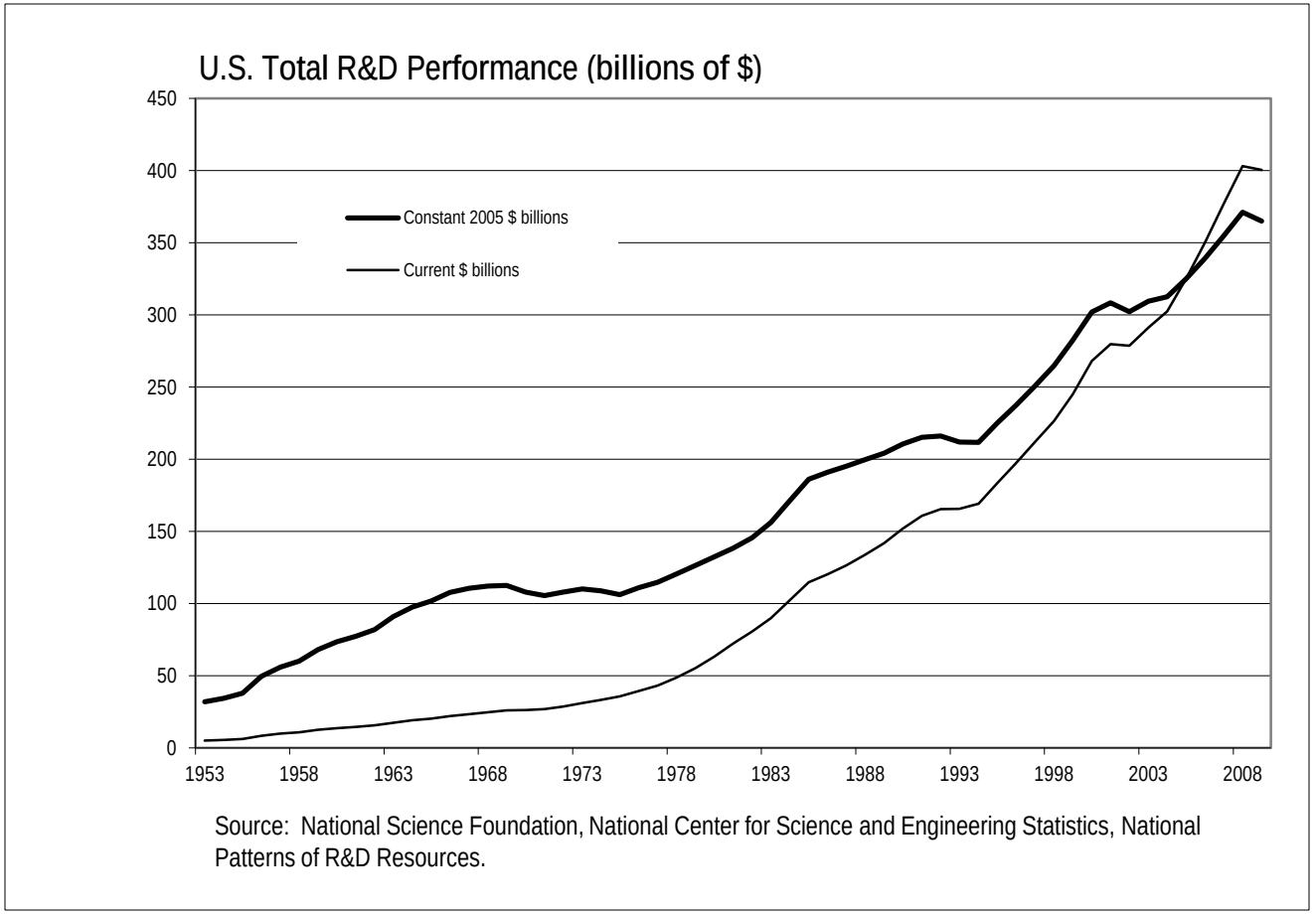
Nat Pat Tables and Charts – Illustration 1

U.S. R&D expenditures, by performing sector and source of funding: 2004–09

Sector	2004	2005	2006	2007	2008	2009
	Current $millions					
All performing sectors	302,503	324,993	350,162	376,960	403,040	400,458
Business	208,301	226,159	247,669	269,267	290,681	282,393
Federal government	37,685	39,568	41,611	43,906	44,674	46,151
Federal intramural	24,898	26,322	28,240	29,859	29,839	30,901
FFRDCs	12,788	13,246	13,371	14,047	14,835	15,250
Industry-administered	2,485	2,601	3,122	5,165	6,346	6,446
U&C-administered	7,659	7,817	7,306	5,567	4,766	4,968
Nonprofit-administered	2,644	2,828	2,943	3,316	3,724	3,835
Universities and colleges	43,122	45,190	46,955	49,010	51,650	54,382
Other nonprofit organizations	13,394	14,077	13,928	14,777	16,035	17,531
All funding sectors	302,503	324,993	350,162	376,960	403,040	400,458
Business	191,266	207,680	227,057	246,679	258,626	247,357
Federal government	91,656	96,276	100,768	105,822	117,611	124,432
Universities and colleges	7,936	8,578	9,285	9,959	10,707	11,436
Nonfederal government	2,883	2,922	3,021	3,265	3,518	3,675
Other nonprofit organizations	8,761	9,538	10,031	11,235	12,578	13,559
	Constant 2005 $millions					
All performing sectors	312,548	324,993	339,202	354,864	371,184	364,951
Business	215,218	226,159	239,917	253,484	267,706	257,355
Federal government	38,937	39,568	40,308	41,332	41,143	42,059
Federal intramural	25,724	26,322	27,356	28,109	27,480	28,161
FFRDCs	13,212	13,246	12,953	13,224	13,663	13,897
Industry-administered	2,568	2,601	3,024	4,862	5,844	5,875
U&C-administered	7,913	7,817	7,078	5,241	4,389	4,528
Nonprofit-administered	2,732	2,828	2,851	3,121	3,429	3,495
Universities and colleges	44,554	45,190	45,485	46,137	47,568	49,561
Other nonprofit organizations	13,839	14,077	13,492	13,911	14,767	15,977
All funding sectors	312,548	324,993	339,202	354,864	371,184	364,951
Business	197,617	207,680	219,950	232,220	238,184	225,425
Federal government	94,700	96,276	97,614	99,619	108,315	113,399
Universities and colleges	8,200	8,578	8,995	9,375	9,861	10,422
Nonfederal government	2,979	2,922	2,926	3,074	3,240	3,349
Other nonprofit organizations	9,052	9,538	9,717	10,576	11,584	12,356

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Patterns of R&D Resources.

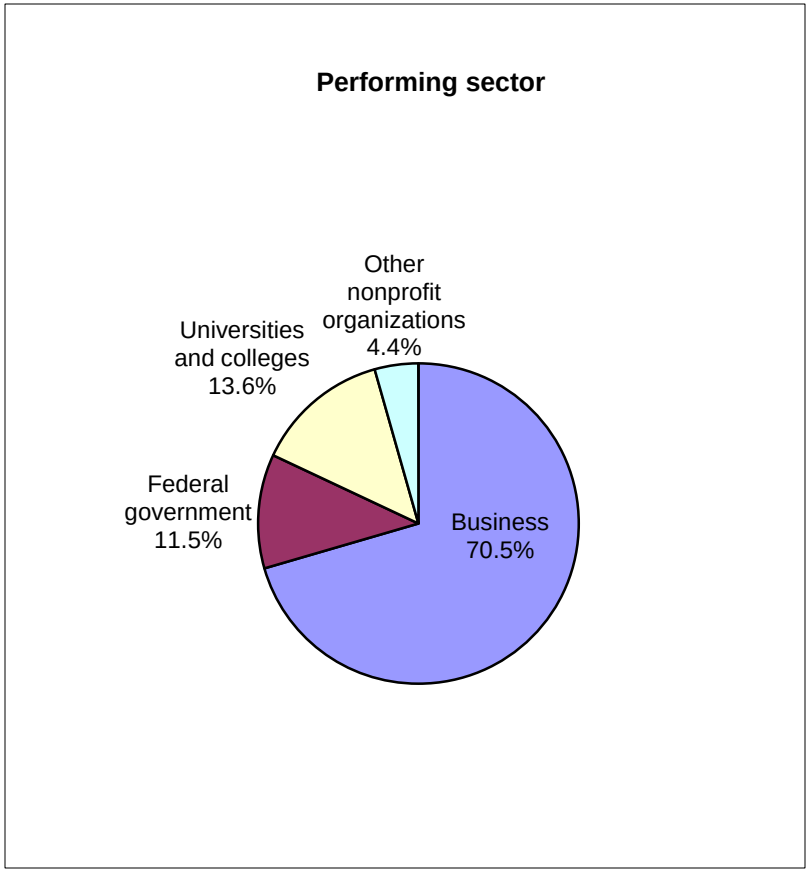
Nat Pat Tables and Charts – Illustration 2

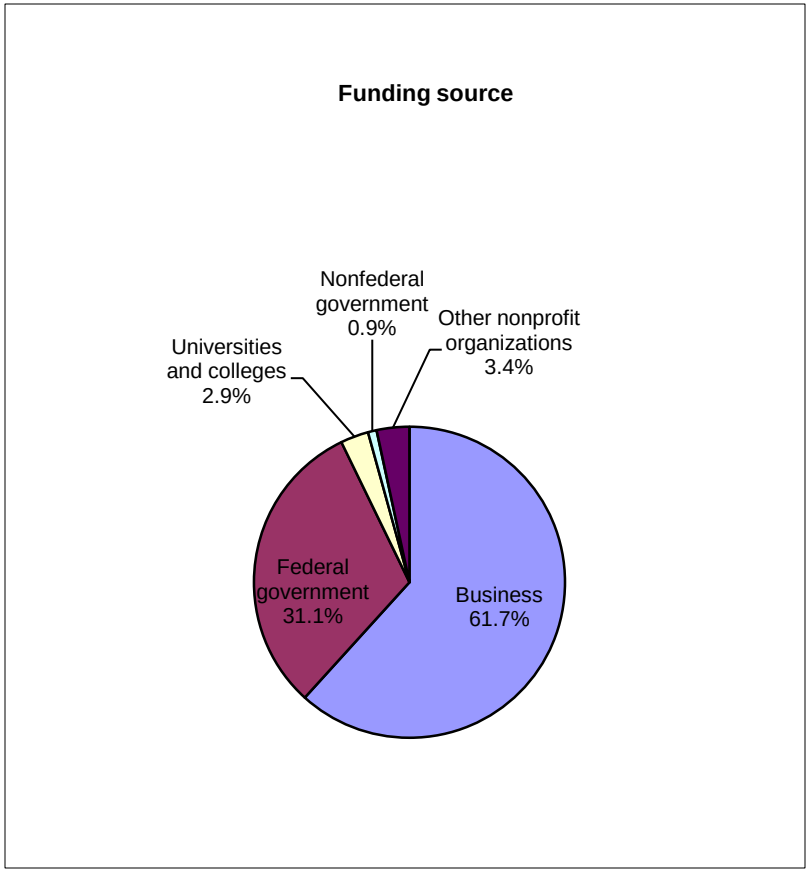


Source: National Science Foundation, National Center for Science and Engineering Statistics, National Patterns of R&D Resources.

Nat Pat Tables and Charts – Illustration 3

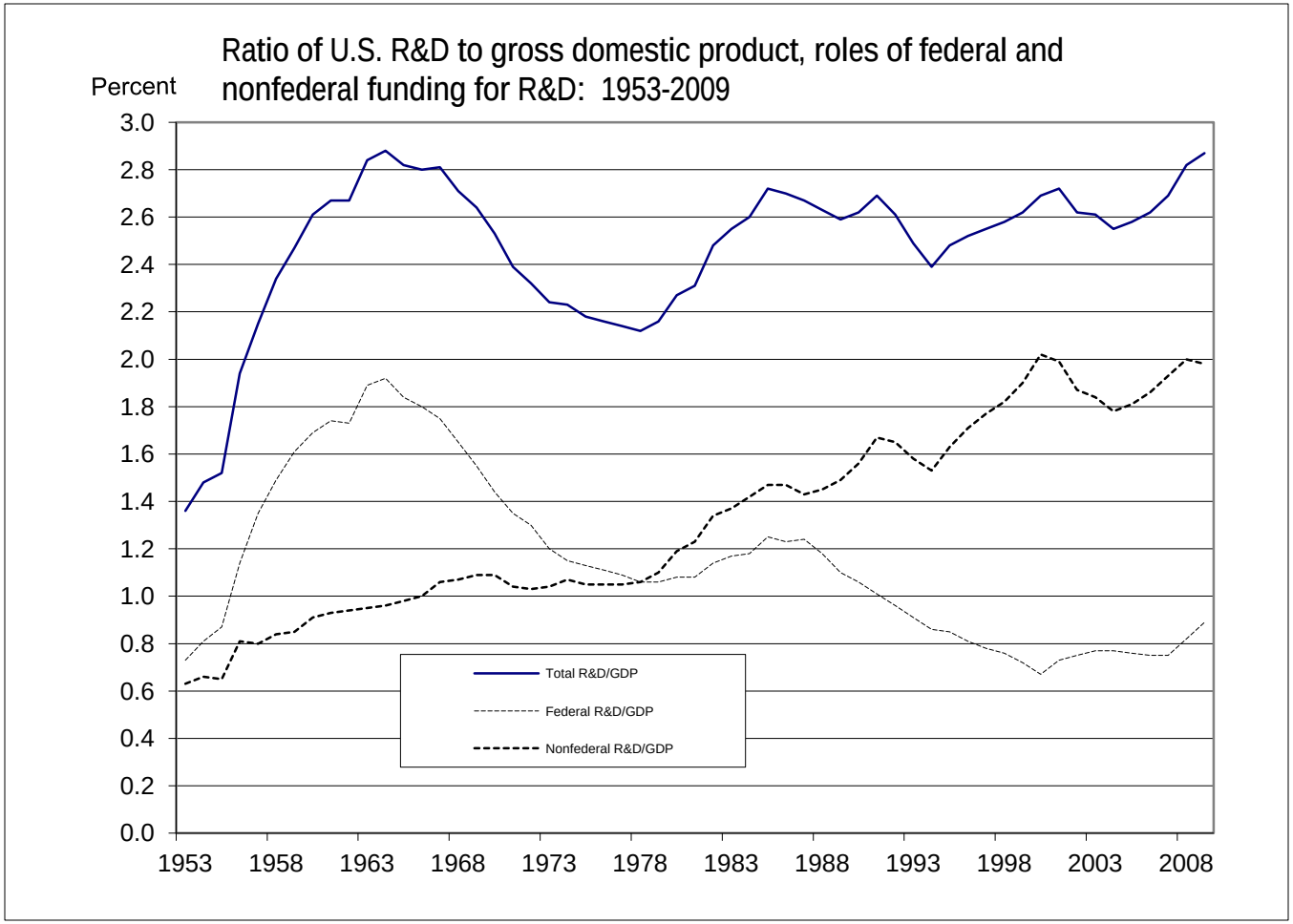
Shares of U.S. total R&D Expenditures, by performing sector and funding source: 2009





SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Patterns of R&D Resources.

Nat Pat Tables and Charts – Illustration 4



SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Patterns of R&D Resources.

Nat Pat Tables and Charts – Illustration 5

International comparisons of gross domestic expenditures on R&D and R&D share of gross domestic product, by selected country/economy: 2009 (or most recent year)

Region/country-economy	GERD (PPP $millions)	GERD/ GDP (%)
North America		
United States	401,576.5	2.88
Canada	24,551.3	1.92
Mexico (2007)	5,719.6	0.37
South America		
Brazil (2008)	21,649.4	1.08
Argentina (2007)	2,678.8	0.51
Chile (2004)	1,227.7	0.68
Europe		
Germany	82,730.7	2.78
France	47,953.5	2.21
United Kingdom	40,279.5	1.85
Italy	24,752.6	1.27
Spain	20,496.4	1.38
Sweden	12,494.9	3.62
Netherlands	12,273.8	1.82
Switzerland (2008)	10,512.7	3.00
Austria	8,931.3	2.75
Belgium	7,684.9	1.96
Finland	7,457.8	3.96
Denmark	6,283.8	3.02
Norway	4,734.1	1.76
Poland	4,874.9	0.68
Portugal	4,411.0	1.66
Czech Republic	4,094.8	1.53
Ireland	3,164.6	1.79
Ukraine	2,485.7	0.86
Hungary	2,333.8	1.15
Romania	1,471.5	0.47
Greece (2007)	1,867.9	0.59
Belarus	813.3	0.65
Slovenia	1,043.6	1.86
Croatia	743.1	0.84
Luxembourg	708.5	1.68
Slovak Republic	595.5	0.48

Country/economy	GERD (PPP $millions)	GERD/ GDP (%)
Middle East		
Israel	8,810.1	4.28
Turkey	8,681.2	0.85
Iran (2008)	6,465.2	0.79
Africa		
South Africa (2008)	4,689.3	0.93
Egypt	997.3	0.21
Morocco (2006)	765.1	0.64
Tunisia	1,048.5	1.21
Central Asia		
Russian Federation	33,368.1	1.24
South Asia		
India (2007)	24,439.4	0.76
Pakistan	2,055.2	0.46
East, Southeast Asia		
Japan	137,908.6	3.33
China	154,147.4	1.70
South Korea (2008)	43,906.4	3.36
Taiwan	21,571.8	2.93
Singapore	5,626.5	2.35
Malaysia (2006)	2,090.9	0.64
Thailand (2007)	1,120.8	0.21
Australia, Oceania		
Australia (2008)	18,755.0	2.21
New Zealand (2007)	1,422.5	1.17
Selected country groups		
EU-27	297,889.6	1.90
OECD (2008)	965,629.1	2.33
G-20 countries	1,181,263.7	2.01

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, National Patterns of R&D Resources.

Main Audience for National Patterns Reports and Data

- Federal policymakers, agencies, congressional staff

  -Including OMB, White House OSTP, Committees with jurisdiction for R&D funding and agency S&E programs, federal agency leadership and managers, CRS, GAO

- International

  - OECD, EU, UNESCO, country national statistics offices

- Media

  - National, international; mainstream news, public policy, business press

- Broader science and technology policy community

  - Including AAAS/Science, professional societies, policy NGOs, academia

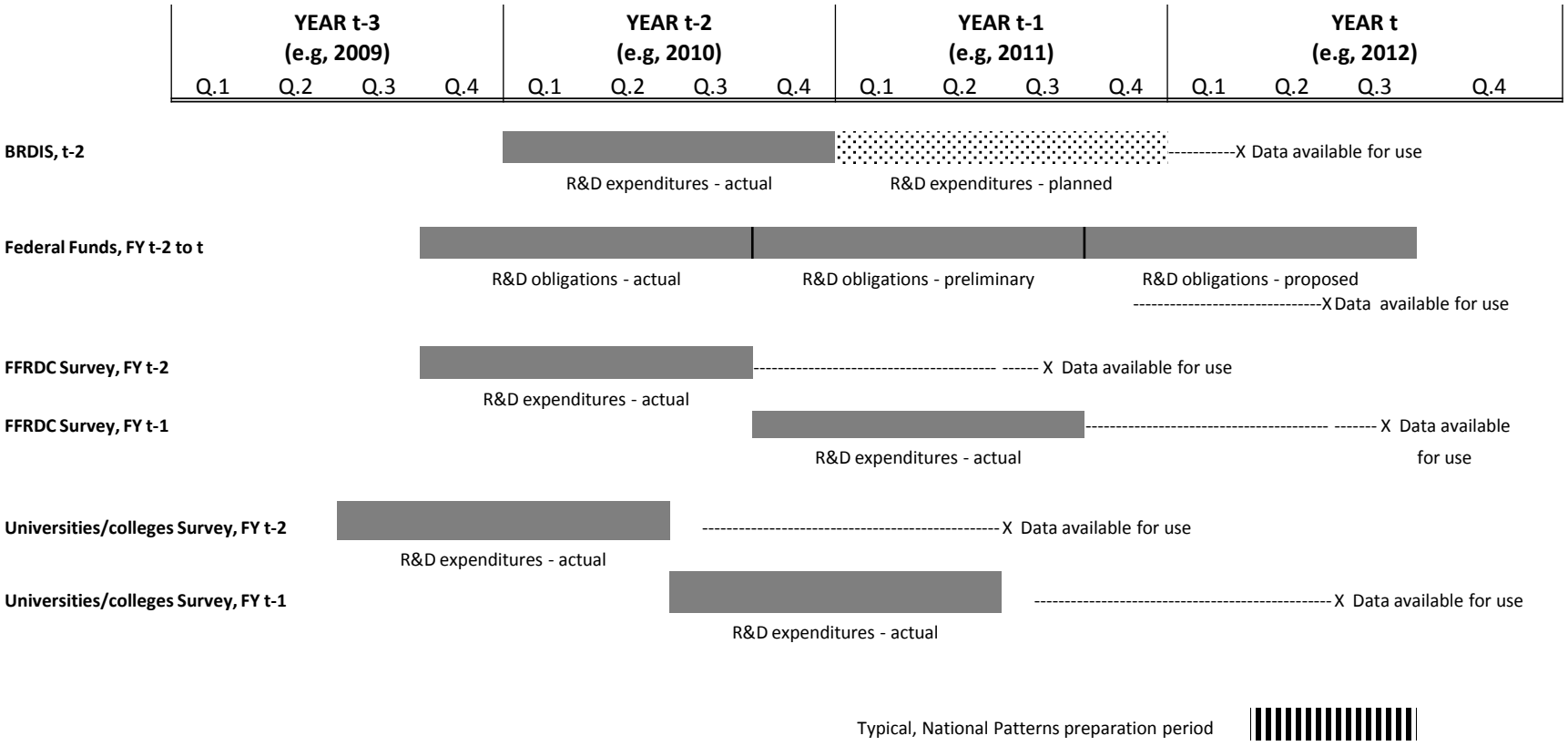
- Business managers and analysts

- Other

Main Data Sources for National Patterns

Performer/Funding Source	NCSES Survey
Business sector	-Business R&D and Innovation Survey (annual, 2008 and on) -Survey of Industrial R&D (annual, 2007 and earlier)
Federal government	- Survey of Federal Funds for R&D (annual, federal FY)
FFRDCs	- Survey of R&D Expenditures by FFRDCs (census, annual since federal FY 2001; in earlier years, part of other surveys)
Academia	-Survey of R&D Expenditures at Universities and Colleges (annual, academic FY, 2009 and earlier) - Higher Education R&D Survey (annual, beginning with academic FY 2010)
Other nonprofit organizations	-Survey of R&D Funding and Performance by Nonprofit Organizations (most recent is for 1996/97; previously, conducted periodically) -Survey of Federal Funds for R&D (data on ONP R&D funded by federal government; annual, federal FY)
State agencies	-Survey of State Agency R&D Expenditures (data to date for FYs 2006, 2007, 2009)

Survey Data Timing Key to Managing a National Patterns Report Cycle



Data Resources for Supplementary National Patterns Outputs

- R&D to GDP ratios
  - National Patterns data on R&D; U.S. economic data from Dept of Commerce, Bureau of Economic Analysis
- R&D allocation to the U.S. States
  - Relies on geographic breakdowns in the R&D expenditure surveys
- International Comparisons
  - U.S. National Patterns data adjusted to include R&D plant in federal intramural R&D performance and federally funded R&D plant in other nonprofit R&D performance
  - Other country data primarily from the OECD's *Main Science and Technology Indicators* series. Some from UNESCO's science and technology data series.

Issues for National Patterns – as perceived by NCSES

- Content and format of the National Patterns reports
- Data on R&D performance by Other Nonprofit Organizations
- Improving the timeliness of National Patterns data reporting
- Data additions to improve international comparability
- Availability of data on state agency intramural R&D expenditures
- Expanded treatment of R&D performance by S&E fields and socioeconomic objectives

Concluding Comments

- National Patterns reporting has a decades long history. The process has continued to incorporate changes arising from improvements in its methodology and from advances in the underlying NCSES R&D expenditure surveys.
- We look forward to your observations and recommendations on how we can further improve the National Patterns data and publicly accessible information products

Supplemental Slides – as needed for further discussion

Issues for National Patterns – as perceived by NCSES

- Content and format of the National Patterns reports
  - What improvements should we consider in the published National Patterns documents we now regularly provide?
- R&D performance by Other Nonprofits
  - Clearly, more recent data on Other Non-Profit R&D are needed
  - Short of mounting a comprehensive new survey, is there a path to improved estimates we should consider?
- Improving the timeliness of National Patterns data reporting
  - Year t-2 as the concluding datapoint in the National Patterns report released in Q.3 or 4 of year t is noticed by our external data users
  - Gap could perhaps be reduced with greater reliance on estimated/projected R&D expenditure values – particularly for the business sector. How should we view the tradeoff between final survey data vs. more recent figures with estimated values?

Issues for National Patterns – as perceived by NCSES (continued)

- Data additions to improve international comparability
  - Estimates for “departmental research” in academic R&D. (Which are included by many OECD members, but not the U.S.)
  - Capital expenditures for R&D (Which have not been a part of the National Patterns data historically, but are now available in the BRDIS and HERD surveys)

Note: The new HERD survey (starting with data year 2010) explicitly includes non-Science & Engineering R&D in the academic R&D performance total. BRDIS (starting with data year 2008) includes social science research in the business R&D performance total. The National Patterns time series will be revised accordingly (including revisions back to 2003 for the academic data).

Issues for National Patterns – as perceived by NCSES (continued)

•Availability of data on state agency intramural R&D expenditures

- New NCSES survey. Currently have data for 2006, 2007, and 2009.
- Is a small R&D component – some $300 million for U.S. as a whole in 2007. Should this new data be a priority for inclusion in National Patterns?

•Expanded treatment in National Patterns of R&D performance by S&E fields and socioeconomic objectives

- Long recommended by the OECD's Frascati Manual
- Only partially available in the current NCSES R&D expenditure surveys.
- Should this be a data expansion priority for National Patterns?

Integrating the Survey R&D Performance Data by Sector

Performer	Essential mechanics – current process (National Patterns report published in late year t, with t-2 the concluding year in time series)
Businesses	Data from BRDIS, CY t-2: domestic R&D expenditures, total and splits by character of work; -data used without adjustment
Federal intramural	Data from Federal Funds Survey, federal FY t-2 and t-1: intramural R&D obligations, total and splits by character of work; - data in federal FY adjusted to CY equivalent
FFRDCs	Data from FFRDC Survey, t-2 R&D expenditures, total and splits by character of work; -Data in federal FY adjusted to CY equivalent (which requires estimate for FY t-1 Q.1)

Integrating the Survey R&D Performance Data by Sector (continued)

Performer	Essential mechanics – current process (National Patterns report published in late year t, with t-2 the concluding year in time series)
Academia	Data from U&C/HERD Survey, t-2 R&D expenditures, total and splits by character of work; - Data adjusted to remove pass-through double counting - Data in academic FY adjusted to CY equivalent (which requires estimate for FY t-1 Q1 and Q2)
Other nonprofit organizations	Most recent nonprofit survey was in 1998 for 1996/97; thus, much of the ONP R&D time series is currently estimated -ONP R&D performance for that funded by ONP and businesses estimated based on elasticity relationships from the 1996/97 survey data - ONP R&D performance for that funded by the federal government based on current data from the Federal Funds survey

National Patterns – Further Details

Estimating R&D Performance by Nonprofit Organizations

Year	Total ONP R&D (billions of $)	a) ONP R&D w/federal funding	b) ONP R&D w/business funding	c) ONP R&D w/ONP funding
1995	$5.83	$2.85	$0.67	$2.31
2000	9.73	4.51	1.02	4.20
2005	14.08	6.55	1.11	6.43
2007	14.78	5.98	1.26	7.54
2009	17.53	7.13	1.26	9.14
	Sum of a), b), c) to right	From Federal Funds Survey	Estimated *	Estimated *

*Based on elasticity ratios from 1996/97 Survey of R&D Funding and Performance by Non-profit Organizations. Business funded ONP R&D related to business funding of business R&D. ONP funded ONP R&D related to ONP funding for academic R&D.

National Patterns – Further Details

R&D Performance Detail by Field-of-Science

Performing Sector	Field-of-Science Data Status
Business R&D	R&D expenditure detail in the SIRD (2007 and earlier) and BRDIS (2008 and on) is primarily by industry sector (NAICS) and company size. BRDIS has provided detail on R&D each year so far for selected applications and technology focus areas (such as defense, energy, environment, agriculture, software, medical clinical trials, biotech, nanotech) -- but not a comprehensive FOS breakdown.
Academic R&D	Survey of University/College R&D has provided detail on expenditures for R&D and R&D equipment by FOS. This detail continues in the new HERD Survey (2010 and on).
Federal Intramural R&D	Survey of Federal Funds for R&D provides FOS detail for research (but not development) by agency. This FOS detail not further disaggregated by performer, except for universities.
FFRDC R&D	Breakdowns by FOS not provided by the FFRDC R&D survey.
Other Nonprofit R&D	Breakdowns by FOS not provided in the R&D estimates.

NOTE: NCSES publishes a data series on federal R&D funding by OMB budget function categories. But this covers only federal R&D support and is based on annual federal budget authority figures. NCSES also provides somewhat similar statistics, detailed along Eurostat socioeconomic objectives categories, to the OECD for its GBAORD indicator series.

National Patterns of R&D Resources: Purposes and Uses

Kei Koizumi

Assistant Director for Federal R&D

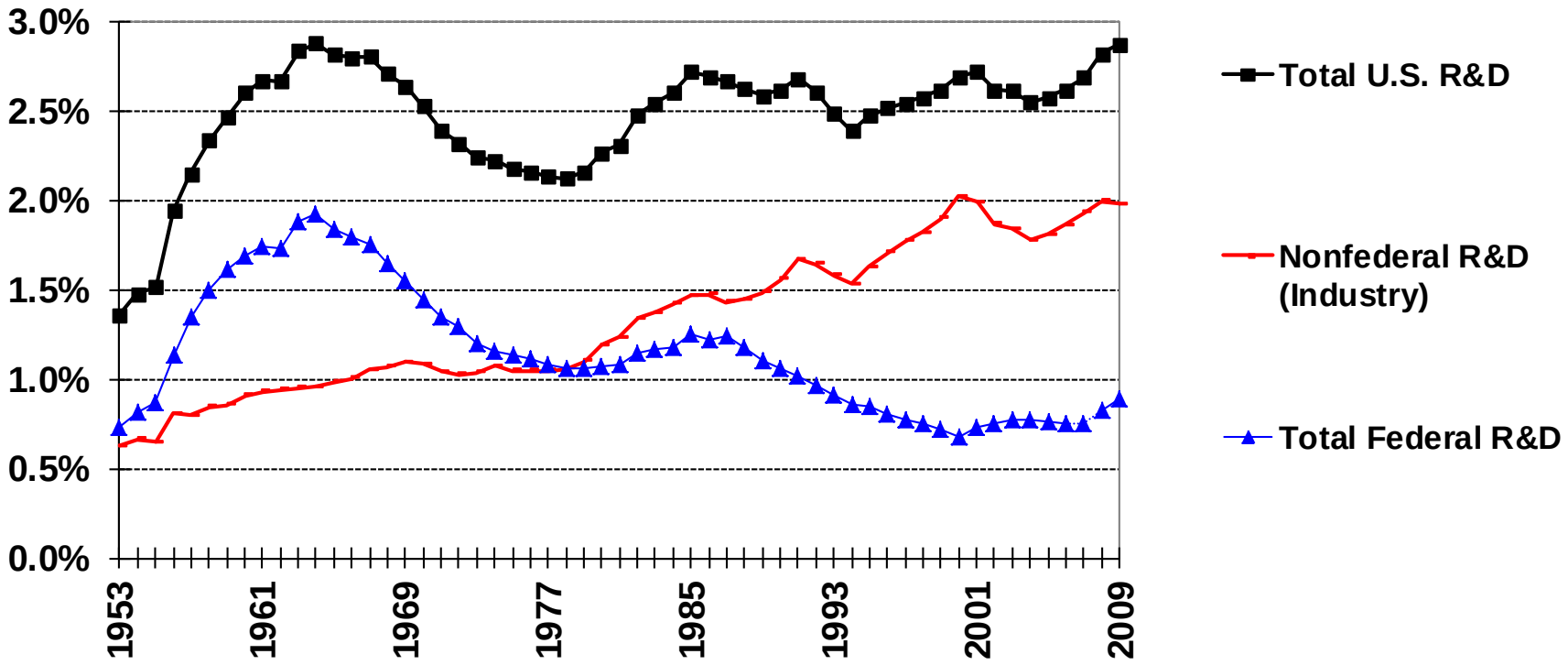
White House Office of Science & Technology Policy

“So I'm here today to set this goal: We will devote more than 3 percent of our GDP to research and development. We will not just meet, but we will exceed the level achieved at the height of the space race, through policies that invest in basic and applied research, create new incentives for private innovation, promote breakthroughs in energy and medicine, and improve education in math and science.”

- President Barack Obama
April 27, 2009

U.S. R&D as Percent of Gross Domestic Product

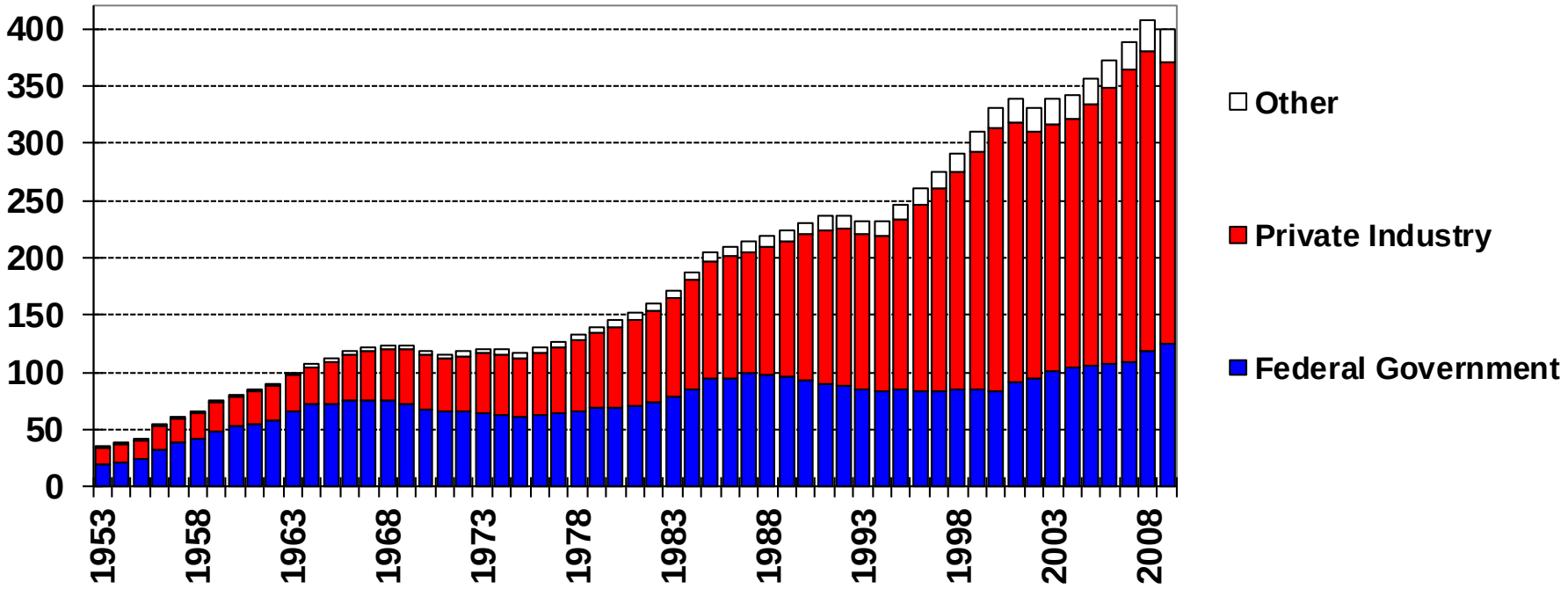
Total, Industrial, and Federal R&D - 1953-2009



NSF, Division of Science Resources Statistics.
OSTP FEBRUARY '12

U.S. R&D Funding by Source, 1953-2009

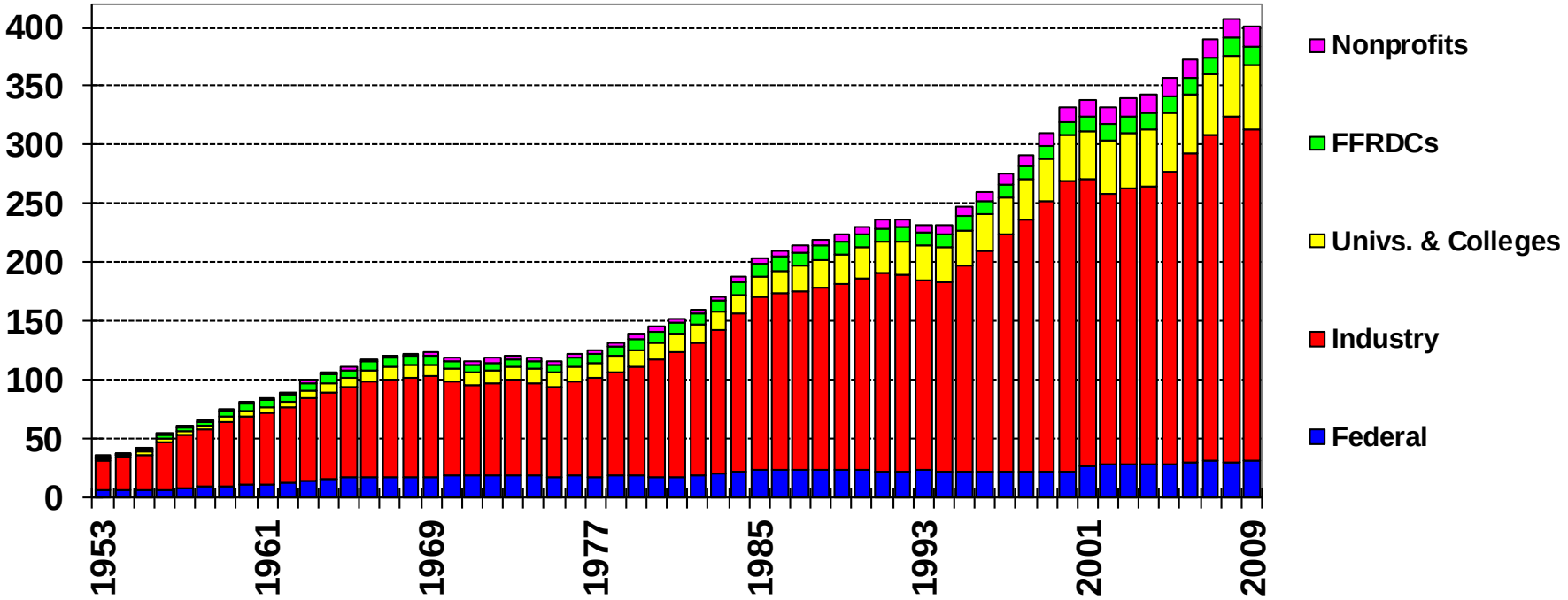
expenditures in billions of constant 2009 dollars



Source: National Science Foundation.
FEB. '12 OSTP

U.S. R&D Funding by Performer, 1953-2009

expenditures in billions of constant 2009 dollars



Source: National Science Foundation.
FEB. '12 OSTP

A STRATEGY FOR AMERICAN INNOVATION

Securing Our Economic Growth and Prosperity

National Economic Council, Council of Economic Advisers,
and Office of Science and Technology Policy

FEBRUARY 2011

THANK YOU

Kei Koizumi
Kei_Koizumi@ostp.eop.gov
www.ostp.gov

An academic user's perspective on *National Patterns*

David C. Mowery

Haas School of Business, U.C. Berkeley & NBER

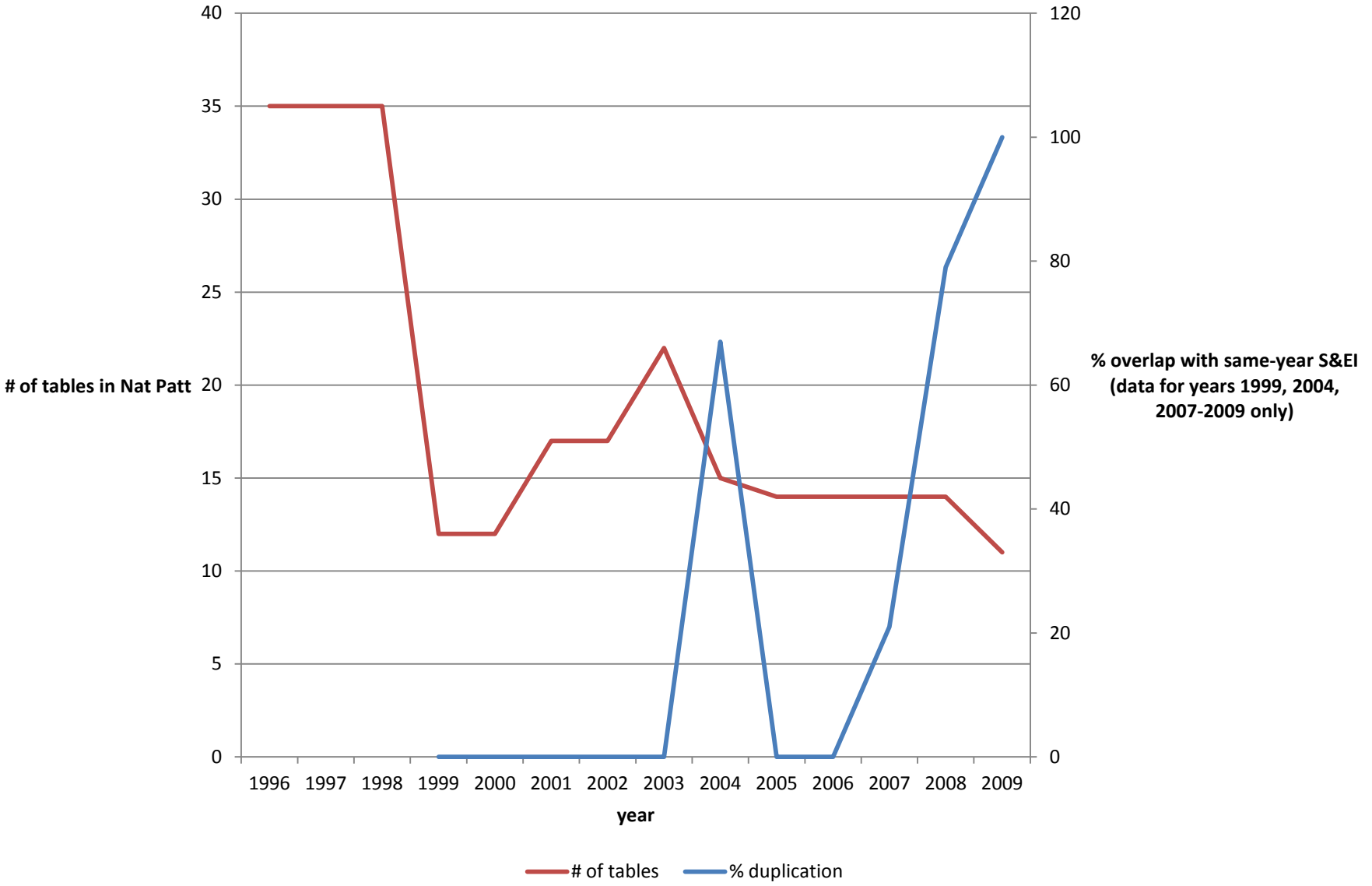
My perspective

- Economist & historian interested in S&T policy issues.
- Some recent uses of data from *NatPatt*:
  - Long-term trends in industry support of US university research (share in 1957 > 2005)
  - Change in the shares of nonfederally funded R&D performed by different size classes of firms, 1980 – 2008 (large firms' share declines from 60% to 30%).
  - Growth in self-finance by universities of academic R&D, 1952 - 2008 (from 12% to 22%).
- What makes *NatPatt* data especially useful?
  - Long timeseries, easy to access and work with.
  - Focus on intersectoral funding, performance relationships arguably enables aggregation to offset some noise in data.
- *NatPatt* data also enable one to probe beneath R&D/GDP ratios to highlight structural features that are at least as important for long-term innovative performance as the ratio.
- *NatPatt* is less useful for tracking emerging areas of R&D, innovation (a frequent topic of NRC panels):
  - Photonics
  - Nanotechnology

Recent releases of *NatPatt* are light on data, and duplicate *S&EI*

- 2012 release of *NatPatt* ("Data Update," June 30, 2012) contains 11 tables, all of which appear in *S&EI* published in January 2012.
- Since late 1990s, *NatPatt* has contained less data (fewer tables) while duplication with *S&EI* has grown.
  - Compare 1998 *NatPatt* report (released in 1999) with 2012 "Update."
    - 1998 : 31 tables, including extensive data on employment of S&Es in R&D.
    - 2003 (released in 2005): 22 tables.
    - 2012: 11 tables, all of which duplicate S&EI, and all of which focus exclusively on funding.
- "R&D resources" include people as well as $$.
  - HR data (e.g., *SED*) may provide clues to emerging areas of academic R&D.

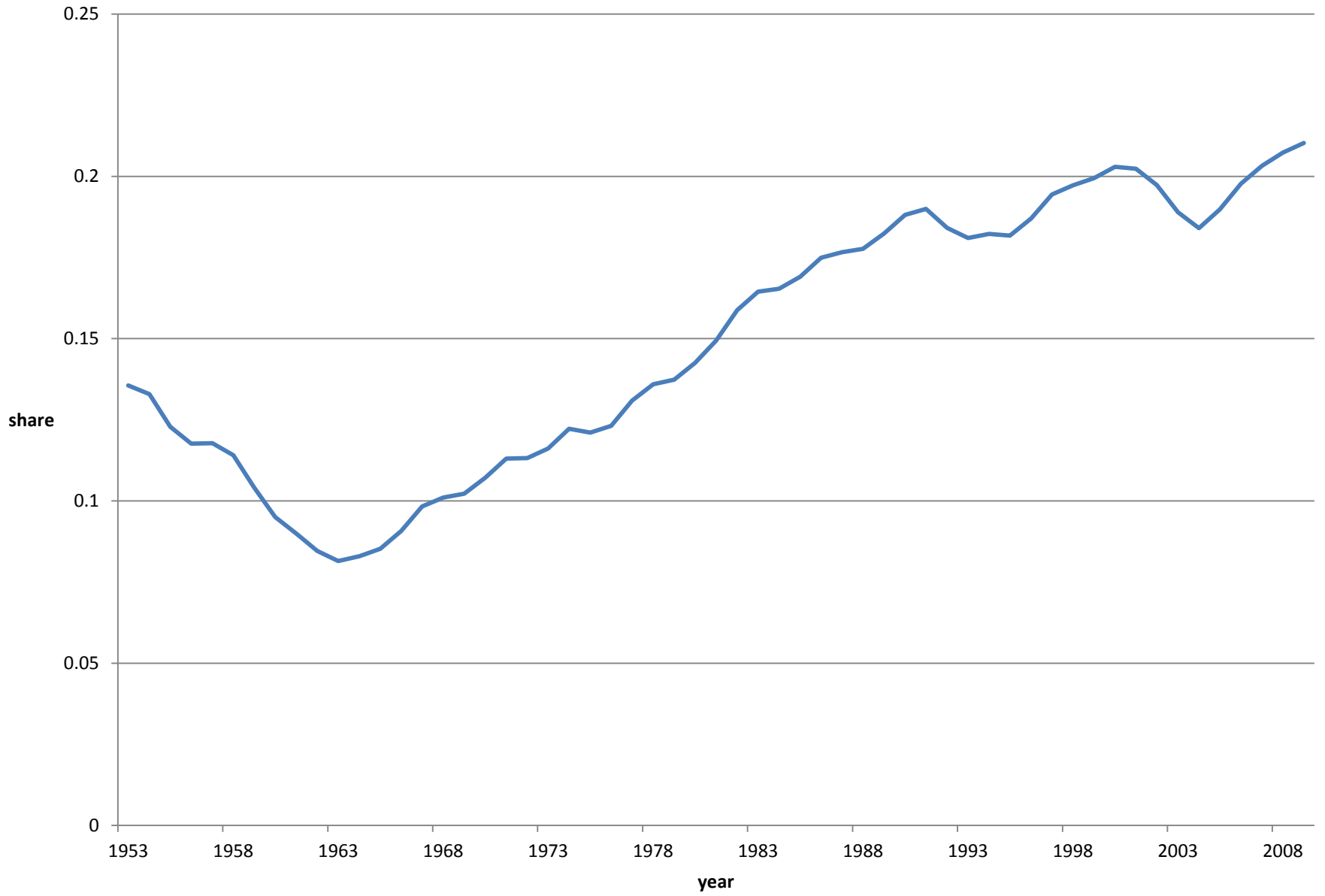
of tables and % duplication with same-year *S&EI*, 1996 - 2009 editions of *National Patterns*



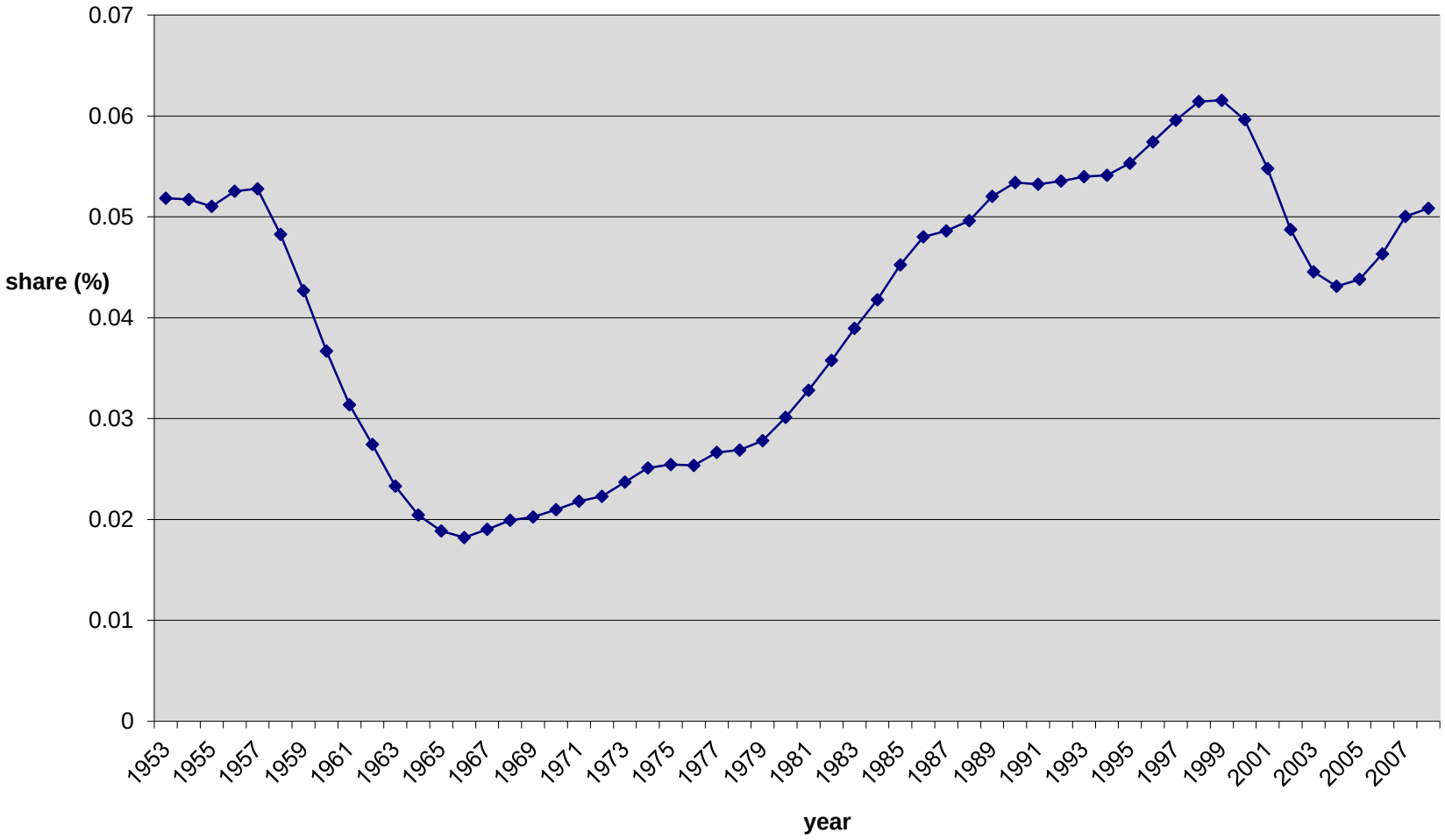
How might content and release of *NatPatt* be changed?

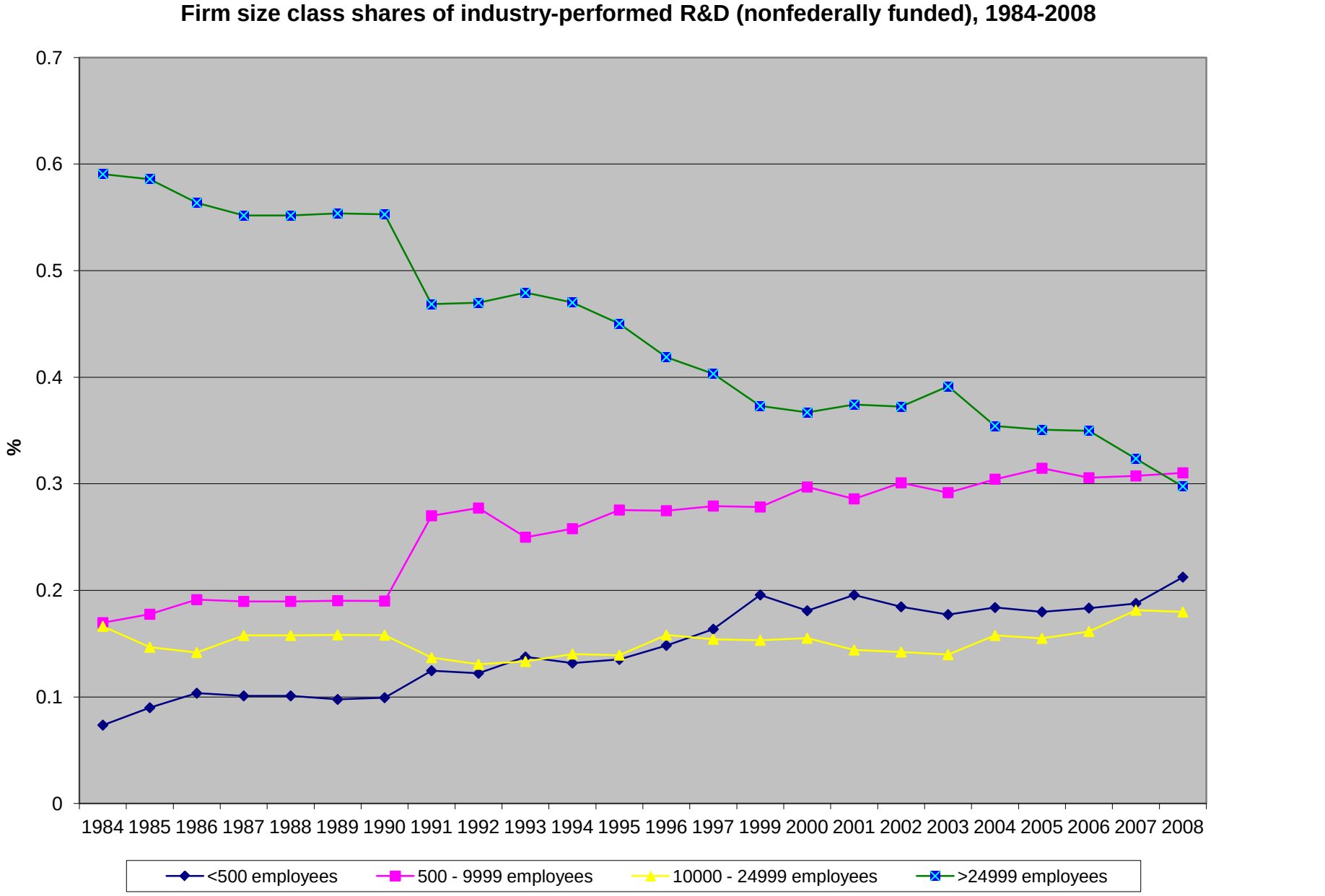
- Content (more data, please!!):
  - HR data are an invaluable complement to funding.
  - Industry breakdown by funding sources, including a disaggregated breakdown of sources of R&D $$ in nonmfg industry.
- Release "strategy": Focus on complementing *S&EI*
  - Rather than duplicating *S&EI* in years of its publication, release a richer (more data tables) biennial *NatPatt* in years without *S&EI*?
  - Consider working with other entities (e.g., other funding agencies, contractors) to assemble "satellite assessments" of emerging fields on a selective basis for release in non-*S&EI* years?

Appendix slides


U&C-funded share of U&C-performed R&D, 1953 - 2009 (exc. FFRDCs)
0.25
0.2
0.15
0.1
0.05
0
share
1953
1958
1963
1968
1973
1978
1983
1988
1993
1998
2003
2008
year

Industry-funded share of total academic R&D, 1953 - 2008 (exc. FFRDCs)




Firm size class shares of industry-performed R&D (nonfederally funded), 1984-2008
%
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008
<500 employees
500 - 9999 employees
10000 - 24999 employees
>24999 employees

National Patterns Dataset as Baseline Data for Battelle/R&D Magazine R&D Funding Forecast

Martin Grueber
Research Leader
Battelle

Workshop on Future Directions for the NSF National Patterns of Research and Development
September 6-7, 2012

Battelle
The Business of Innovation

R&D

2012 Global R&D Funding Forecast

Background

- 2012 Forecast was R&D Magazine's 54th and Battelle's 44th
  - Forecast for upcoming year, published in December
  - 2012 Forecast PDF available at: http://www.battelle.org/docs/default-document-library/2012_global_forecast.pdf
- Collaborative effort between Battelle and R&D Magazine—joint efforts began in 1996
- Synthesis, Analysis, Estimation, and Forecast Process
  - Secondary Data: NSF, OSTP, AAAS, Federal Agencies, OECD, EU, IMF, Trade/Technical Associations, Third Party Providers, Media Feeds/Alerts, etc.
  - Primary Data: R&D Magazine Reader Surveys & Battelle Internal Survey and Global Researcher Survey
  - Company Data: Annual Reports and SEC Filings

Use of NatPat Baseline Data for 2012 Forecast

2008 Prelim Data Most Recent Available in Fall 2011

TABLE 1. U.S. research and development expenditures, by performing sector and source of funds: 1953–2008

Year	Performer: Funding source:	All performers All sources	Federal Federal	Industry Total	Industry Federal[a]	Industry Industry[d]	Industry FFRDCs[a] Total	U&C[b] Total	U&C[b] Federal	U&C[b] Other government[e]	U&C[b] Industry	U&C[b] U&C	U&C[b] Other nonprofit	U&C FFRDCs Total	Other nonprofit organizations Total	Other nonprofit organizations Federal[c]	Other nonprofit organizations Industry	Other nonprofit organizations Other nonprofit	Nonprofit FFRDCs[c] Total
		Current $millions																	
2000		267,298	17,917	199,961	17,117	182,844	2,001	30,705	17,727	2,247	2,174	6,232	2,326	5,742	9,506	4,447	1,118	3,941	1,465
2001		277,366	20,426	202,017	16,899	185,118	2,020	33,743	19,784	2,397	2,190	6,827	2,546	6,225	10,743	5,289	1,132	4,322	2,192
2002		276,022	21,499	193,868	16,401	177,467	2,263	37,215	22,395	2,557	2,160	7,344	2,758	7,102	11,756	5,731	1,084	4,941	2,319
2003		288,324	22,752	200,724	17,798	182,926	2,458	40,484	25,129	2,742	2,129	7,650	2,833	7,301	12,111	5,686	1,118	5,307	2,494
2004		299,201	22,844	208,301	20,266	188,035	2,485	43,128	27,173	2,883	2,190	7,937	2,946	7,659	12,140	5,695	1,151	5,294	2,644
2005		322,104	24,470	226,159	21,909	204,250	2,601	45,197	28,260	2,922	2,323	8,579	3,113	7,817	13,032	5,932	1,253	5,846	2,828
2006		347,048	25,556	247,669	24,304	223,365	3,122	46,983	28,815	3,021	2,515	9,307	3,325	7,306	13,469	5,992	1,374	6,103	2,943
2007		372,535	25,858	269,267	26,585	242,682	5,165	49,021	29,328	3,249	2,748	9,993	3,703	5,567	14,341	5,954	1,497	6,890	3,316
2008 preliminary		397,629	27,000	289,105	25,795	263,310	6,337	51,163	30,177	3,453	2,908	10,600	4,024	4,717	15,606	5,982	1,629	7,995	3,701

- Various 2011 ***InfoBriefs*** and ***2009 Academic R&D Expenditures*** data provided additional more recent info
- However, 2009 was a “perfect storm” data year for examining R&D expenditures
  - Recessionary impacts, ARRA, corporate R&D restatements and Q4 investment spike, etc.
  - Difficult to launch estimation process starting with this unique year
- NCSES staff provided insights and “sanity checks” as appropriate

Assumptions for 2012 U.S. R&D Forecast

U.S. R&D Situation

- Guarded optimism and continued stability
- Continued tightening of federal R&D budgets
- Increased expectations for R&D ROI
- Revised federal expenditure reporting
- Continued effect of ARRA

Underlying assumption that R&D funding has a certain level of inertia

2012 U.S R&D Forecast

U.S. R&D Situation

The Source-Performer Matrix

Estimated Distribution of U.S. R&D Funds in 2012
Millions of Current U.S. Dollars (Percent Change from 2011)

Source	Performer					
	Federal Gov't.	FFRDC	Industry	Academia	Non-Profit	Total
Federal Government	$29,152 -2.51%	$14,666 -3.69%	$37,577 -2.42%	$37,440 0.93%	$6,817 -2.29%	$125,652 -1.61%
Industry		$202 2.20%	$273,487 3.37%	$3,868 26.49%	$2,129 8.89%	$279,685 3.75%
Academia				$12,318 2.85%		$12,318 2.85%
Other Government				$3,817 2.72%		$3,817 2.72%
Non-Profit				$3,491 2.70%	$11,055 2.70%	$14,546 2.70%
Total	$29,152 -2.51%	$14,868 -2.36%	$311,063 2.63%	$60,934 2.85%	$20,001 1.55%	$436,018 2.07%

Source: Battelle, *R&D Magazine*

* When adjusted for a projected 2-3% inflation for 2012, little or no "real" growth

2012 U.S R&D Forecast

U.S. R&D Situation

The Source-Performer Matrix

Estimated Distribution of U.S. R&D Funds in 2012
Millions of Current U.S. Dollars (Percent Change from 2011)

Source	Performer					
	Federal Gov't.	FFRDC	Industry	Academia	Non-Profit	Total
Federal Government	$29,152 -2.51%	$14,666 -3.69%	$37,577 -2.42%	$37,440 0.93%	$6,817 -2.29%	$125,652 -1.61%
Industry		$202 2.20%	$273,487 3.37%	$3,868 26.49%	$2,129 8.89%	$279,685 3.75%
Academia				$12,318 2.85%		$12,318 2.85%
Other Government				$3,817 2.72%		$3,817 2.72%
Non-Profit				$3,491 2.70%	$11,055 2.70%	$14,546 2.70%
Total	$29,152 -2.51%	$14,868 -2.36%	$311,063 2.63%	$60,934 2.85%	$20,001 1.55%	$436,018 2.07%

Source: Battelle, *R&D Magazine*

2012 U.S R&D Forecast

U.S. R&D Situation

The Source-Performer Matrix

Estimated Distribution of U.S. R&D Funds in 2012
Millions of Current U.S. Dollars (Percent Change from 2011)

Source	Performer					
	Federal Gov't.	FFRDC	Industry	Academia	Non-Profit	Total
Federal Government	$29,152 -2.51%	$14,666 -3.69%	$37,577 -2.42%	$37,440 0.93%	$6,817 -2.29%	$125,652 -1.61%
Industry		$202 2.20%	$273,487 3.37%	$3,868 26.49%	$2,129 8.89%	$279,685 3.75%
Academia				$12,318 2.85%		$12,318 2.85%
Other Government				$3,817 2.72%		$3,817 2.72%
Non-Profit				$3,491 2.70%	$11,055 2.70%	$14,546 2.70%
Total	$29,152 -2.51%	$14,868 -2.36%	$311,063 2.63%	$60,934 2.85%	$20,001 1.55%	$436,018 2.07%

Source: Battelle, *R&D Magazine*

2012 U.S R&D Forecast

U.S. R&D Situation

The Source-Performer Matrix

Estimated Distribution of U.S. R&D Funds in 2012
Millions of Current U.S. Dollars (Percent Change from 2011)

Source	Performer					
	Federal Gov't.	FFRDC	Industry	Academia	Non-Profit	Total
Federal Government	$29,152 -2.51%	$14,666 -3.69%	$37,577 -2.42%	$37,440 0.93%	$6,817 -2.29%	$125,652 -1.61%
Industry		$202 2.20%	$273,487 3.37%	$3,868 26.49%	$2,129 8.89%	$279,685 3.75%
Academia				$12,318 2.85%		$12,318 2.85%
Other Government				$3,817 2.72%		$3,817 2.72%
Non-Profit				$3,491 2.70%	$11,055 2.70%	$14,546 2.70%
Total	$29,152 -2.51%	$14,868 -2.36%	$311,063 2.63%	$60,934 2.85%	$20,001 1.55%	$436,018 2.07%

Source: Battelle, *R&D Magazine*

NatPat Baseline Data

2009 Prelim Data Most Recent Available in Fall 2012

TABLE 2. U.S. research and development expenditures, by performing sector and source of funds: 1953–2009

	Performer:	All performers	Federal	Industry			Industry FFRDCs[a]	U&C[b]						U&C FFRDCs	Other nonprofit				Nonprofit FFRDCs[c]
Year	Funding source:	All sources	Federal	Total	Federal[a]	Industry[d]	Total	Total	Federal	Other government[e]	Industry	U&C	Other nonprofit	Total	Total	Federal[c]	Industry	Other nonprofit	Total
		Current $millions																	
2000		267,983	18,374	199,961	17,117	182,844	2,001	30,705	17,727	2,247	2,174	6,232	2,326	5,742	9,734	4,510	1,020	4,204	1,465
2001		279,755	22,374	202,017	16,899	185,118	2,020	33,743	19,784	2,397	2,190	6,827	2,546	6,225	11,182	5,488	1,029	4,666	2,192
2002		278,744	23,798	193,868	16,401	177,467	2,263	37,215	22,395	2,557	2,160	7,344	2,758	7,102	12,179	5,778	998	5,404	2,319
2003		291,239	24,982	200,724	17,798	182,926	2,458	40,484	25,129	2,742	2,129	7,650	2,833	7,301	12,796	5,945	1,020	5,831	2,494
2004		302,503	24,898	208,301	20,266	188,035	2,485	43,122	27,168	2,883	2,190	7,936	2,945	7,659	13,394	6,537	1,041	5,816	2,644
2005		324,993	26,322	226,159	21,909	204,250	2,601	45,190	28,254	2,922	2,323	8,578	3,113	7,817	14,077	6,545	1,107	6,425	2,828
2006		350,162	28,240	247,669	24,304	223,365	3,122	46,955	28,810	3,021	2,509	9,285	3,329	7,306	13,928	6,044	1,182	6,702	2,943
2007		376,960	29,859	269,267	26,585	242,682	5,165	49,010	29,351	3,265	2,741	9,959	3,694	5,567	14,777	5,980	1,257	7,541	3,316
2008		403,040	29,839	290,681	36,360	254,321	6,346	51,650	30,341	3,518	3,004	10,707	4,080	4,766	16,035	6,236	1,301	8,498	3,724
2009		400,458	30,901	282,393	39,573	242,820	6,446	54,382	31,575	3,675	3,279	11,436	4,418	4,968	17,531	7,133	1,258	9,141	3,835

- Full 2009 NatPat data will reset estimates, and hence, impact 2013 forecast to some degree
- Historical corrections make for better current data and hopefully better forecasts—important for both understanding R&D structure and policy

Limitations, Challenges, & Issues

➢ Timeliness of Data

- ➢ 2013 U.S. Forecast will again require “estimating” three years of data to generate a fourth year “forecast” for 2013.
- ➢ Though R&D expenditures have a fairly stable long term growth trajectory, dynamic nature and policy context suggest higher priority and budget for NatPat (and all NCSES R&D survey efforts)

➢ Source-Performer Component Coverage

- ➢ Increasing number of potential players in both funding and performance. How important are missing or underestimated cells in understanding R&D policy and performance?
- ➢ How might more detailed extensions beyond S&E fields in Higher Ed survey need to be accounted for among other sources and performers?
- ➢ Should NatPat explicitly account for and monitor trends in foreign sources and performers of R&D in the U.S.?

Innovation and Competitiveness—New Initiatives for National Patterns

Charles F. Larson
President Emeritus
Industrial Research Institute

Workshop on Future Directions
for National Patterns of R&D
September 6-7, 2012
Washington, DC

Introduction

- U.S. has led global competitiveness many years
- NSB warned that our edge was slipping and U.S. leadership would be challenged
- Recent National Academies' report urged renewing investments in our "pillars of innovation" to ensure economic growth and national security
- National Patterns can help by becoming broader, deeper and more timely

A Look Back at U.S. Competitiveness

- Threat of Japan led to emphasis on management of technology in the 1980s and 1990s
- Universities implemented programs on MOT and Entrepreneurship
- Data are needed on success factors in management practice for R&D and innovation
- Government implemented studies to stimulate innovation
- U.S. rose to top of most competitive list in 1995
- Criteria for competitiveness should be analyzed

R&D and Innovation

- R&D is not innovation and innovation is not R&D; both are risky and costly
- Apple, Google, 3M, GE and Microsoft lead the global innovation list, but no correlation with R&D investment due to strategic alignment and corporate culture
- Roche, Pfizer, Novartis, Microsoft and Merck lead the global R&D list; only half of top 10 are U.S. firms
- More data are needed on factors making R&D more effective in stimulating innovation

Global 10 Most Innovative Companies
(Data for 2010 from Booze & Company)

Apple	$1.8 B on R&D	R&D/Sales = 2.7%
Google	$3.8 B on R&D	R&D/Sales = 12.8%
3M	$1.4 B on R&D	R&D/Sales = 5.4%
General Electric	$3.9 B on R&D	R&D/Sales = 2.6%
Microsoft	$8.7 B on R&D	R&D/Sales = 14.0%
IBM	$6.0 B on R&D	R&D/Sales = 6.0%
Samsung	$7.9 B on R&D	R&D/Sales = 5.9%
Procter & Gamble	$2.0 B on R&D	R&D/Sales = 2.5%
Toyota	$8.5 B on R&D	R&D/Sales = 3.9%
Facebook	N/A	N/A

Conclusion: It is not how much is spent on R&D, but how it is spent!

Top 10 Global Business R&D Investors
(Data for 2010 from Booze & Company)

1.	Roche	$ 9.7 billion
2.	Pfizer	9.4 billion
3.	Novartis	9.1 billion
4.	Microsoft	8.7 billion
5.	Merck	8.6 billion
6.	Toyota	8.5 billion
7.	Samsung	7.9 billion
8.	Nokia	7.8 billion
9.	General Motors	7.0 billion
10.	Johnson & Johnson	6.8 billion

Entrepreneurship and Impact of Government Regulations

- The Economist recently suggested Europe's growth crisis due to few ambitious entrepreneurs
- Government can create the right climate for innovation in business
- Risk-taking and freedom to fail are critical, along with seed capital and labor law
- Economic freedom shifted to Asia, with U.S. 9th
- New data are needed on impact of risk, seed capital and labor regulations on investment in innovation

New Initiatives in R&D and Innovation

- Other nations striving to improve economic growth through R&D and innovation
- U.S. has the advantage of strong R&D investment, including basic research and IP protection
- Higher R&D investment suggested, along with higher return on the investment
- New data needed on measuring return on R&D investment and innovation
- New data needed on success factors in a form more conducive for use by business

Summary of Data Needs

- National Patterns should be made broader, deeper and more timely
- Data are needed on success factors in management practice for R&D and innovation
- Criteria enabling competitiveness should be analyzed and publicized for U.S. benefit
- Data are needed on factors in making R&D more effective in stimulating innovation

Summary of Data Needs

- Data are needed on impact of risk, seed capital and labor regulations on investment in innovation
- Data are needed on measuring return on investment in R&D and innovation
- Data are needed on factors of R&D and innovation success in a form that can be used more easily by business

Conclusions

- U.S. still leads in R&D and innovation, but this lead is slipping
- Government policies have a key role in promoting innovation and competitiveness
- Creativity and execution are keys for success
- New data from National Patterns, combined with data from BRDIS, can be of great value to policy makers and business

ADVANCES IN INTERNATIONAL COMPARABILITY OF NATIONAL PATTERNS DATA AND REPORTS

Fernando Galindo-Rueda
Head of Science and Technology Indicators Unit
Economic Analysis and Statistics Division
OECD Directorate for Science, Technology and Industry

The author acknowledges contributions from colleagues and national experts.
Any errors are the author's sole responsibility. Please do not quote or cite without the author's consent.

OECD
BETTER POLICIES FOR BETTER LIVES

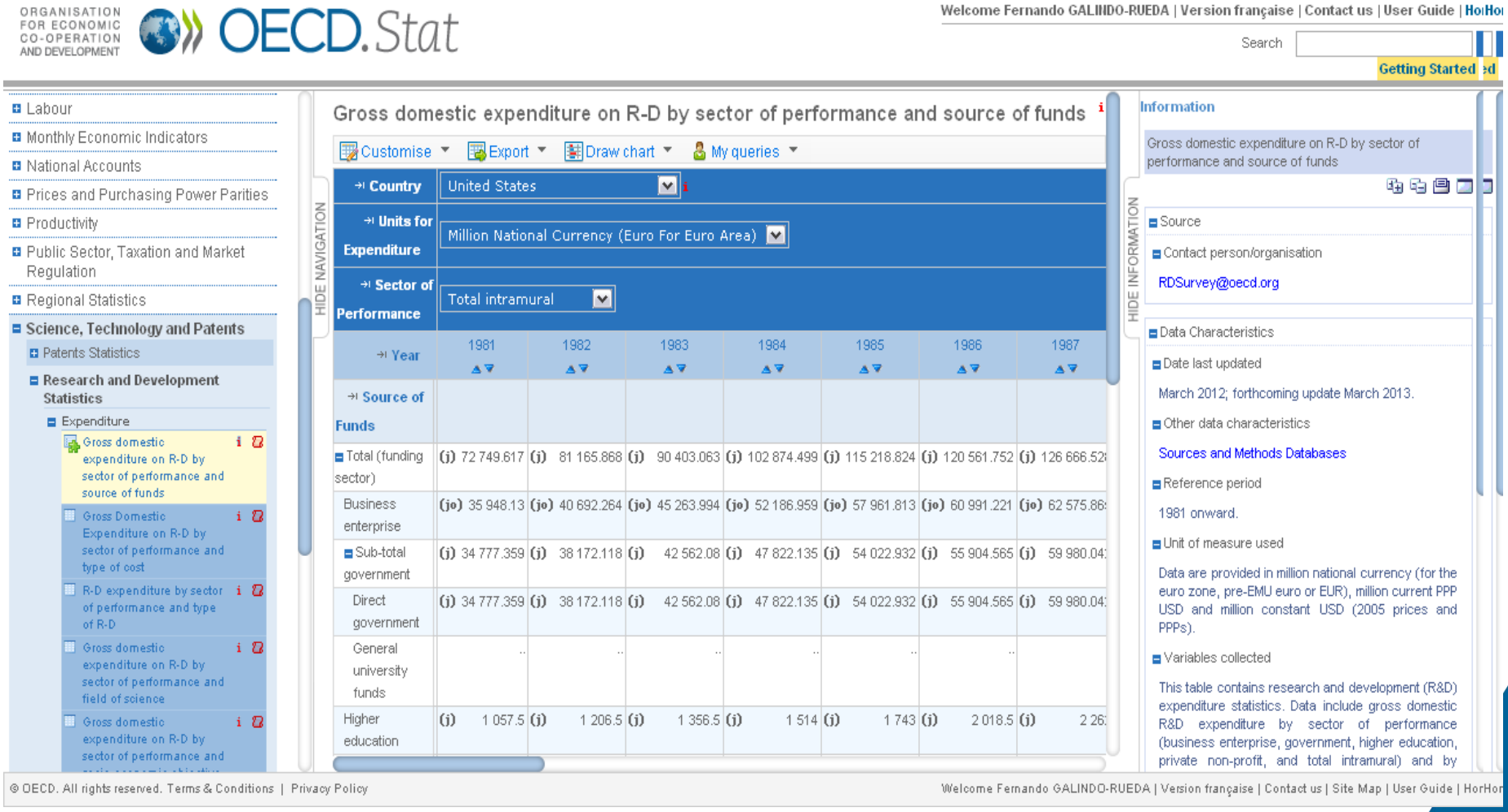
Presentation outline

1. Introduction to OECD and R&D statistics
2. National R&D Patterns – examples from counterpart publications worldwide
3. The international comparability of US National Patterns – main differences
4. Concluding remarks

1. Introduction to OECD and R&D statistics

- Working Party of National Experts on S&T Indicators (NESTI).
  - 34 members, several observers (RUS, IND, BRA, ZAF,...)
  - NCSES/NSF represent the United States
- R&D, innovation and S&T measurement standards.
  - Frascati Manual www.oecd.org/sti/frascatimanual - 6th ed- review
- International R&D data collection and publication
  - Joint OECD/Eurostat questionnaire. (UIS run separate collection)
    - Two collections: Feb/March and July/August. NCSES replies for USA.
  - R&D Statistics database (updated March each year, electronically only). www.oecd.org/sti/rds Data since 1981. Breaks.
  - R&D statistics metadata. http://webnet.oecd.org/rd_gbaord_metadata/default.aspx
  - Derived industry R&D database (ANBERD). www.oecd.org/sti/anberd
  - Moving toward open standards for data/md collection. http://sdmx.org/

Extracting R&D statistics from the OECD website



OECD: Publication of R&D-based indicators

- Main S&T Indicators. www.oecd.org/sti/msti
  - Based on R&D expenditures, R&D personnel, R&D budgets, patents, trade in "high tech" sectors, tech balance of payments
  - June MSTI(year t/1) and following January MSTI(year t/2); paper and online versions
  - Input to national publications (international comparisons section, including NatPat)
  - Key economic indicators (GDP, GDP deflator), exchange rates and PPPs
- STI Scoreboard www.oecd.org/sti/scoreboard (SEI-"like"), others...

2. International experience in R&D statistics production and publication

- Usual contents of headline R&D publications:
  - Mapping of performing (institutional) sectors and funding sectors
  - Time series, current and constant prices, totals and as % of GDP
  - Character of R&D (Basic/applied/ experimental development) breakdowns
  - Type of cost breakdown (employment costs/other current/capital)
  - Regional breakdowns
  - International comparisons, drawing on OECD data
- Common challenges and features:
  - A statistical product for policy usage, feature of national statistical framework
  - Based on combination of performer and funder-based data
  - Few countries can produce breakdowns for total GERD by S&E fields and SEO
- Some differences
  - Different producers across and within countries. Units within national statistical agencies – stat units within ministries or science & innovation agencies.
  - Different balance of survey and admin data sources

Example: Canada

- Gross Domestic Expenditures on Research and Development in Canada (GERD), and the Provinces Published by Statistics Canada on Jan 2012, national estimates 2001 to 2011, provincial estimates 2005 to 2009
- The annual survey of Research and Development in Canadian Industry is the source of the **business enterprise** sector's R&D expenditure data, combined with info from R&D tax incentives
- **HERD** estimation model. Uses info from Canadian Association of University Business Officers (CAUBO) Financial Information on Universities and Colleges (FIUC) survey. Link to review recommendations.
- **Federal government intramural** R&D expenditures are estimated from the Annual Federal Science Expenditure and Personnel survey
- **Provincial governments**' intramural R&D expenditures derived from annual provincial surveys of scientific activities (StatsCan support recently discontinued)
- Annual survey of the Research and Development Activities of **Provincial Research Organizations** is the source of expenditure data displayed in the column for provincial research organizations.
- The annual survey of Research and Development in **Private Non-Profit Organizations** provides national R&D expenditure data for this sector.

Text table 1 Gross domestic expenditure on research and development (GERD) matrix - Canada

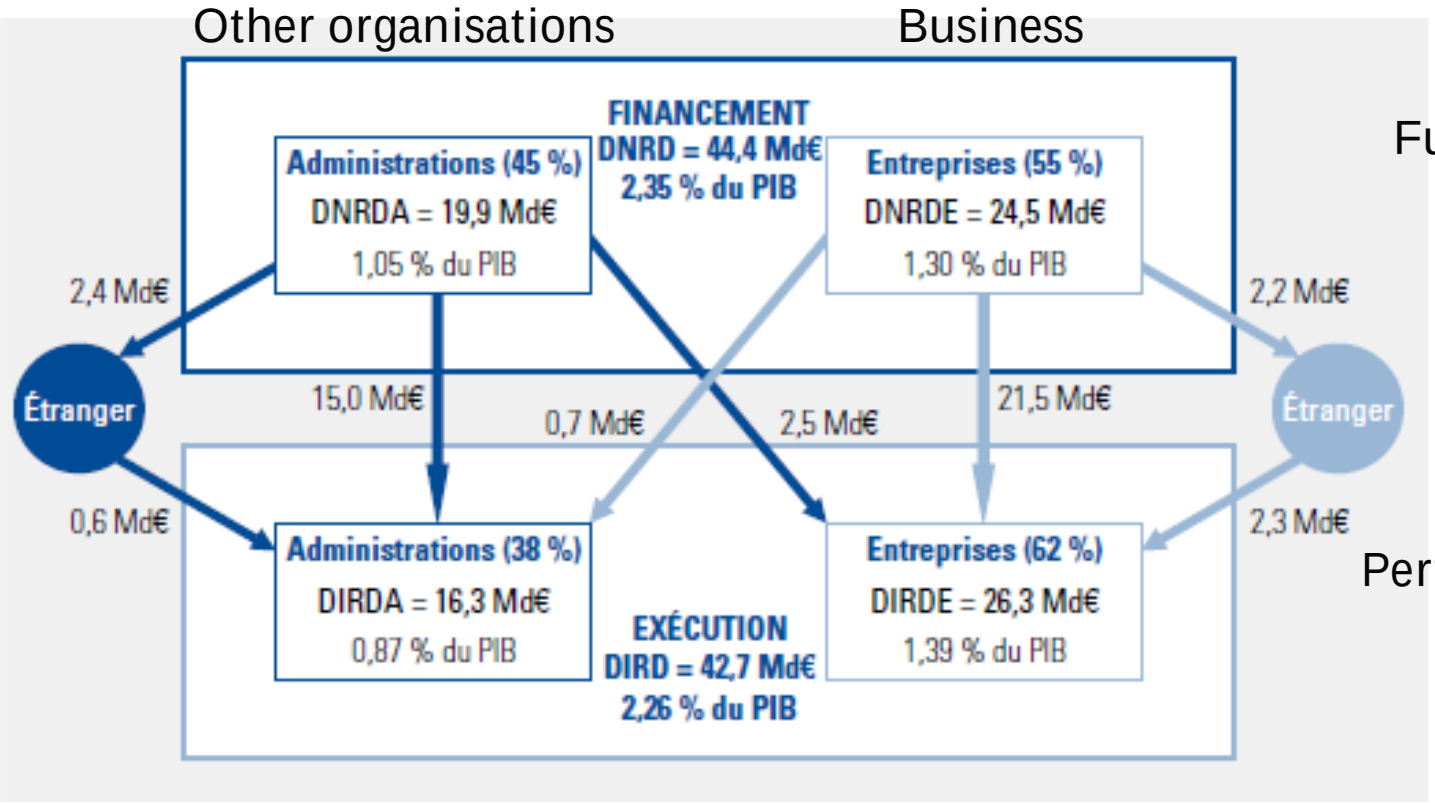
Funding sector	Performing sector Total intramural (domestic) research and development performed by: Federal government	Provincial governments	Provincial research organizations	Business enterprise	Higher education	Private non-profit organizations[1]	Total
	millions of dollars						
Total	Federal government intramural (domestic) R&D expenditures provided by this performing sector and identifying the funding sector	Provincial governments intramural (domestic) R&D expenditures provided by this performing sector and identifying the funding sector	Provincial research organizations intramural (domestic) R&D expenditures provided by this performing sector and identifying the funding sector	Business enterprise intramural (domestic) R&D expenditures provided by this performing sector and identifying the funding sector	Higher education intramural (domestic) R&D expenditures provided by this performing sector and identifying the funding sector	Private non-profit organizations intramural (domestic) R&D expenditures provided by this performing sector and identifying the funding sector	**GERD is total intramural (domestic) R&D expenditures provided by the performing sector**
Federal government							**Federal government**
Provincial governments							**Provincial governments**
Provincial research organizations							**Provincial research organizations**
Business enterprise							**Business enterprise**
Higher education							**Higher education**
Private non-profit organizations							**Private non-profit organizations**
Foreign[2]							**Foreign**

1.These data are not distributed provincially.

2. Foreign enterprises within same group or other foreign governments or international organizations.

Data on sources of funds from performers given priority

Example: France – mapping funder/performer flows



- Produced by Education and Research ministry. 6-page information note with commentary
- Spreadsheets ; Methodology document

Source: Note d'Information – Oct 2011. Expenditures on R&D for France, 2009 (and preliminary estimates for 2010) « Dépenses de recherche et développement en France en 2008 - Premières estimations pour 2009 », *Note d'Information Enseignement supérieur & Recherche 11.01, MESR-SIES, février 2011*. Site REPÈRES : http://cisad.adc.education.fr/reperes www.enseignementsup-recherche.gouv.fr Md=Billions

Example: United Kingdom

- UK Gross Domestic Expenditure on R&D, 2010.
  - Annul pub: Spring year t, data for t-2
  - *National Statistic* "tag", specific rules. Produced by stat office (ONS) with own sources and inputs from various parties
  - PDF + spreadsheet tables.
  - Quality report.
- Contents
  - Performer/funder matrix. As per FM, with separate HE funding body.
  - Sep. performer & funder series. Civil and defence breakdown.
  - No now-/forecasting.
  - Estimation for non profit sector from other sources. HE estimation by HE funding council stats. BE: standard survey. GOV: Returns to survey of Govt departments.
  - Publication of revisions in main document

Example: EU-Eurostat

- European Commission's Directorate General for Statistics
- Covers EU member and associate states
- EU Regulation, impact on timetable and data collected.
- Quality profile and quality report
- Reference metadata http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/en/rd_esms.htm

3. Main differences with respect to OECD data published for other countries

- Capital expenditures for R&D
- Sources of funds reporting
  - Most countries (BE, GOV, HE, PrivNP / Abroad)
  - Countries differ on how they reconcile performer and funder-based data.
- Social sciences/humanities R&D
  - No uniform practice across OECD countries. SSH typically covered in HE, not so in BE.
- Inclusion R&D performed by state (regional) and local government institutions (?)
- No data for R&D personnel / only R&D S&E personnel
  - Problematic measurement of FTE units.
- Timeliness
  - Quality / timeliness trade off addressed differently

Specificities of US NatPat data – A: Capital expenditures

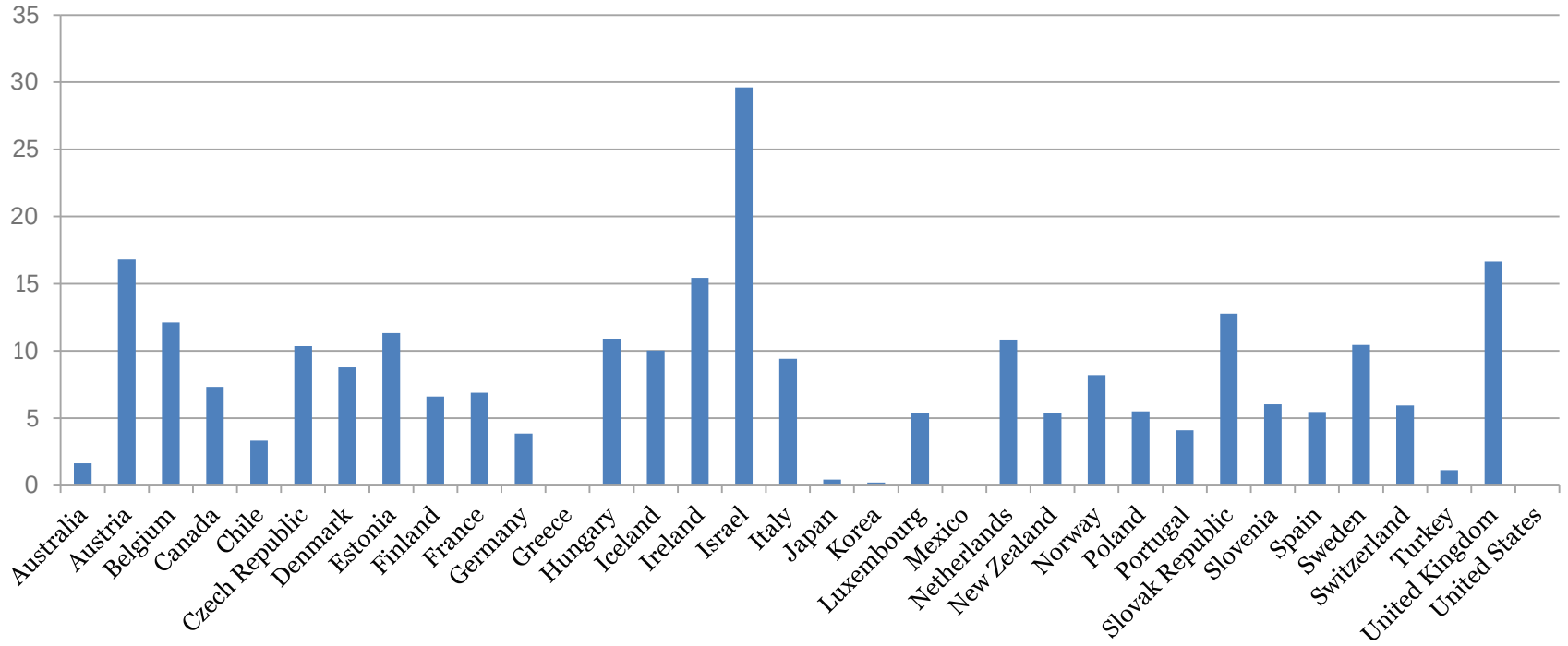
- Treatment of capital expenditures
  - OECD FM= Current + Capital
  - NCSES=Current + Capital depreciation (historical)
  - BEA = R&D capital services:

    User cost of capital used for R&D

    = net return + *depreciation* - revaluation
- *New sources (BRDIS) include capital expenditures. (Sources of funds =?)*

B. Sources of funds

Percentage of Gross domestic expenditures on R&D funded by the rest of the world, 2009

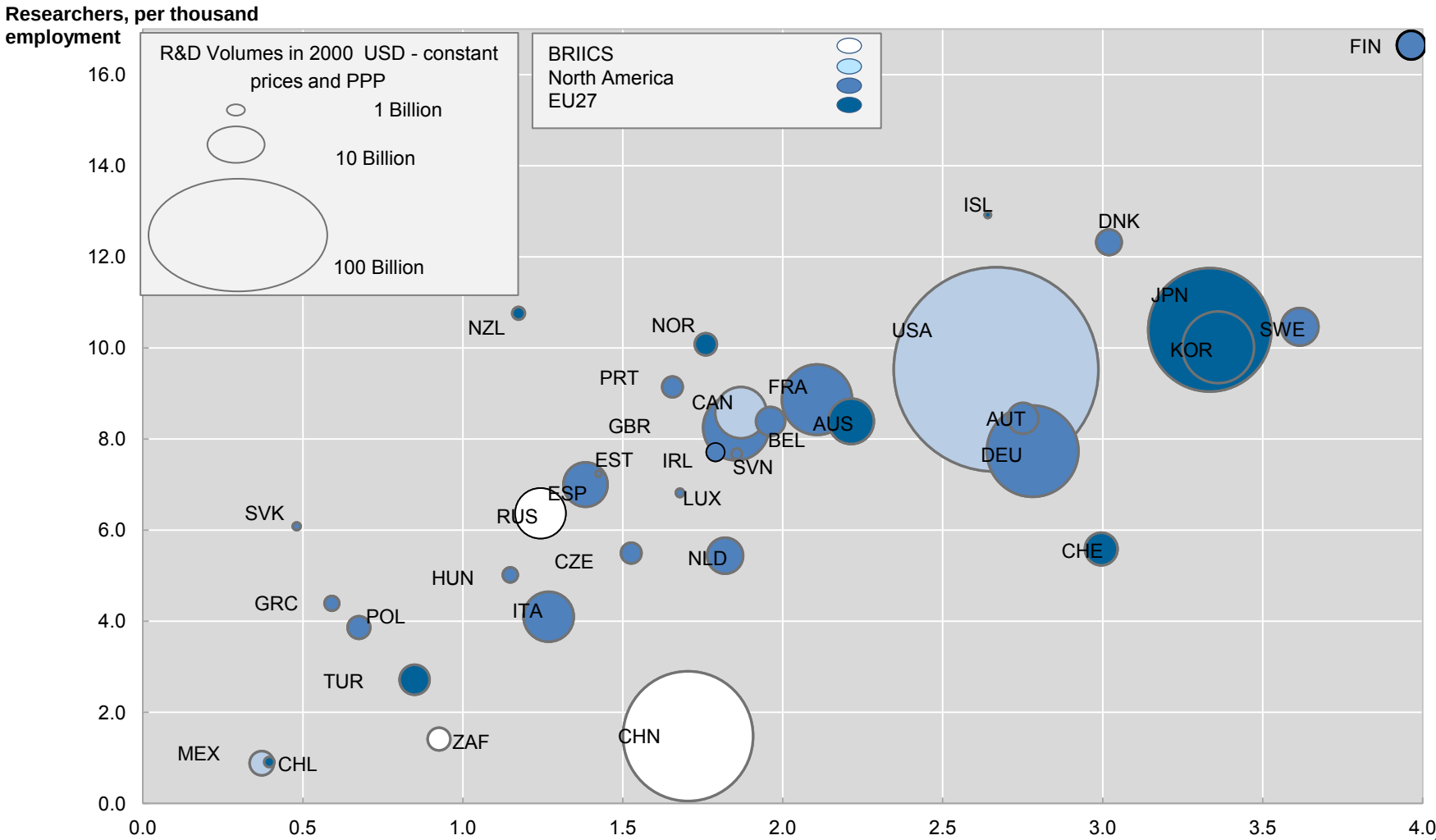


Source: OECD, MSTI 2012/1.
Note: 2008 data for Australia, Chile, Iceland, Israel and Switzerland
Missing data for Greece, Mexico, and United States

C. Other differences/ Beyond financial R&D sources

- R&D personnel and researchers
  - Sectoral researcher data demanded by users as a normalising factor (e.g. scientific productivity measures).
  - Inclusion in re-designed surveys – quality, time consistency issues
- Most countries do not provide funder/performer matrix by type (character) of R&D (basic/applied/development) but provide info by type of cost

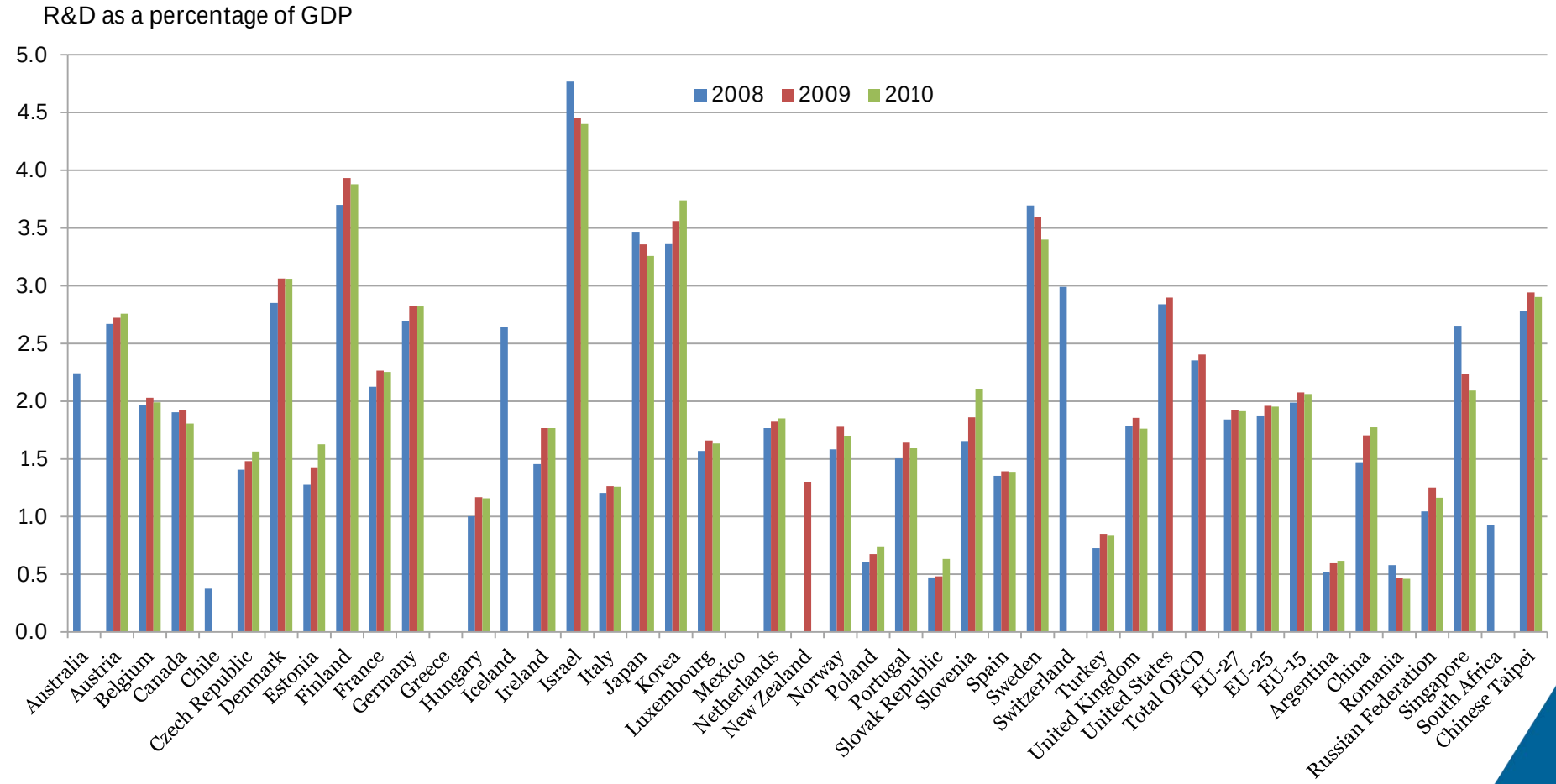
Combining researcher and expenditure data



For Australia (2008), Canada (2008), Chile (2008), France (2008), Greece (2007), Iceland (2008), Korea (2008), Mexico (2007), New Zealand (2007), South Africa (2008), Switzerland (2008) and the United States (2007). Source: OECD STI Scoreboard 2011

D. Timeliness in an international context

- International comparisons only as timely as the least timely data source, infeasible to estimate OECD total values without results for United States.



Source: OECD, MSTI (2012/1) June 2012

4. Concluding remarks – international comparability

- Under current form: Clarify to users difference between Table 11 GERD data and US R&D expenditure data reported elsewhere
  - R&D plant adjustment solely for Federal R&D. Worth having different series?
- Address main differences:
  - Some OECD recommendations would also be relevant to domestic users of *NatPat*, especially role of "abroad" sector.
  - Key elements to improve comparability are present since HERD and BERD surveys re-engineering.
- Would a slightly different *NatPat* reduce NCSES burden of reporting to OECD?
  - Historical series one of its strongest assets – time series v cross-sectional comparability?
- *NatPat* are widely-used resource worldwide - a global statistical public good
  - key component of OECD R&D stats - % of OECD totals
  - findings from surveys re-engineering contributing to review of OECD guidelines
  - valuable resource in its own right for analysts worldwide, including OECD staff

Advances in International Comparability of National Patterns R&D Data and Reports

John Jankowski
Program Director
Research & Development Statistics

CNSTAT Workshop on Future Directions for the
NSF National Patterns R&D Reports
September 6, 2012

National Science Foundation
National Center for Science and Engineering Statistics
www.nsf.gov/statistics/

US-Collected R&D Totals Differed from Frascati Guidance

- Include R&D for the social sciences and humanities
- Include capital R&D and exclude depreciation costs
- Provide detailed sources of funding
  - Own source
  - Businesses within the same group
  - Other businesses
  - Government sector
  - Private nonprofit sector
  - Higher education sector
  - Abroad
- Personnel employed in R&D
  - Headcounts and Full-time equivalents
  - Researchers , technicians, support staff

Redesign of Underlying R&D Surveys

- 2008 Business R&D and Innovation Survey (BRDIS)
  - Social sciences R&D
  - Capital R&D
  - More detailed R&D funding data
  - Expanded R&D personnel data
- 2010 Higher Education R&D Survey (HERD)
  - Humanities and other non-S&E R&D
  - Cost components of R&D
  - More detailed R&D funding data
  - Expanded R&D personnel data

Inclusion of R&D Performance in the Social Sciences and Humanities (SSH)

Economic sector / Survey	S&E R&D ($millions)	Non-S&E R&D ($millions)	Non-S&E as percent of total
Higher Education / HERD 2010*	58,338	2,897	4.7%
Business / BRDIS 2009**	281,872	522	0.2%
Two sector total (mixed years)	340,209	3,419	1.0%

*For HERD, S&E includes social sciences R&D; non-S&E includes R&D in education, business, law, social work, humanities, etc.

** For BRDIS, non-S&E includes only social sciences, but not humanities or market research

Measuring Capital R&D in the Business Sector

2009 BRDIS	Current R&D costs * ($millions)	Capital R&D costs ($millions)
Total R&D performance, of which	282,393	
Depreciation costs	12,137	
Capital R&D, total funding for…		28,335
Structures		2,385
Equipment		11,896
Software and other		5,613

*Includes R&D performance funded from own sources and from external sources

Measuring Capital R&D in the Higher Education Sector

Higher Education R&D	2010 Current R&D costs ($millions)	2009* Capital R&D costs ($millions)
Total R&D*, of which	61,235	
Indirect costs (including unknown depreciation costs)	15,097	
Capitalized equipment and software	2,209	
Capital R&D, total funding for…		5,212
Repairs and renovation of facilities**		1,508
New construction of facilities**		3,703

*From 2010 HERD survey

** One half of amounts reported on the 2008-2009 Facilities survey

Frascati vs US Surveys: Sources of Business R&D Funding

- Own sources of funding
- Frascati-recommended extramural sources of funding
  - Businesses within the same group
  - Other businesses
  - Government sector
  - Private nonprofit sector
  - Higher education sector
  - Abroad
- Sources of industry R&D funding collected by NSF/Census
  - Total R&D performance
  - Amount funded by the Federal Government
  - Derived nonfederal sources combined (own company plus other companies plus other sources treated as equivalent to company funds)

Frascati Definition of R&D – Performance Based

- *R&D (Research and Experimental Development)* is creative work undertaken on a systematic basis aimed at discovering new knowledge… and the use of this stock of knowledge to devise new applications (including new or significantly improved goods or services)

- Intramural expenditures are all expenditures for R&D performed within a statistical unit or sector of the economy during a specific period, whatever the source of funds

FAS-2 Definition of R&D – Expense Based

- Statement of Financial Accounting Standards No. 2 – Accounting for Research and Development Costs – 1974
  - Included: “The costs of services performed by others in connection with the research and development activities of an enterprise, including research and development conducted by others in behalf of the enterprise” (that is, purchased contract R&D services are included in expenses)
  - Excluded: “Accounting for the costs of research and development activities conducted for others under a contractual arrangement is a part of accounting for contracts in general and is beyond the scope of this Statement. Indirect costs that are specifically reimbursable under the terms of a contract are also excluded from this Statement.”

Relationship between Frascati and FAS2

- Total R&D performance (Frascati):

  R&D performed and paid for by the company

  + R&D performed by the company that is paid for by others

- FAS2 R&D Expense: R&D that the company "Pays For"

  R&D performed and paid for by the company

  + R&D Contract Services (Outsourced R&D)

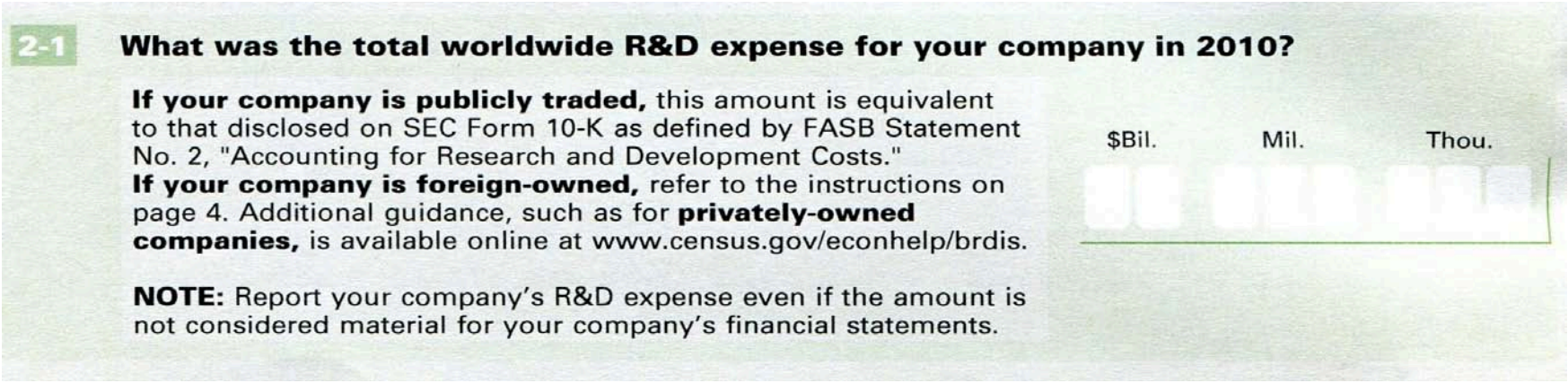
Types of Business R&D Relationships

- Company to Company: Contractual
- Company to Company: Collaborative Partnership
- Intra-Company transfers
- Company and Governments
- Company and Other (e.g., academia, nonprofits)

- Domestic – Domestic
- Domestic – Foreign
- Foreign – Foreign (especially for multi-national companies)

How BRDIS Measures R&D Performed by the Company That is Paid for Out of Own Funds

Data collection begins with accounting “expense” concept



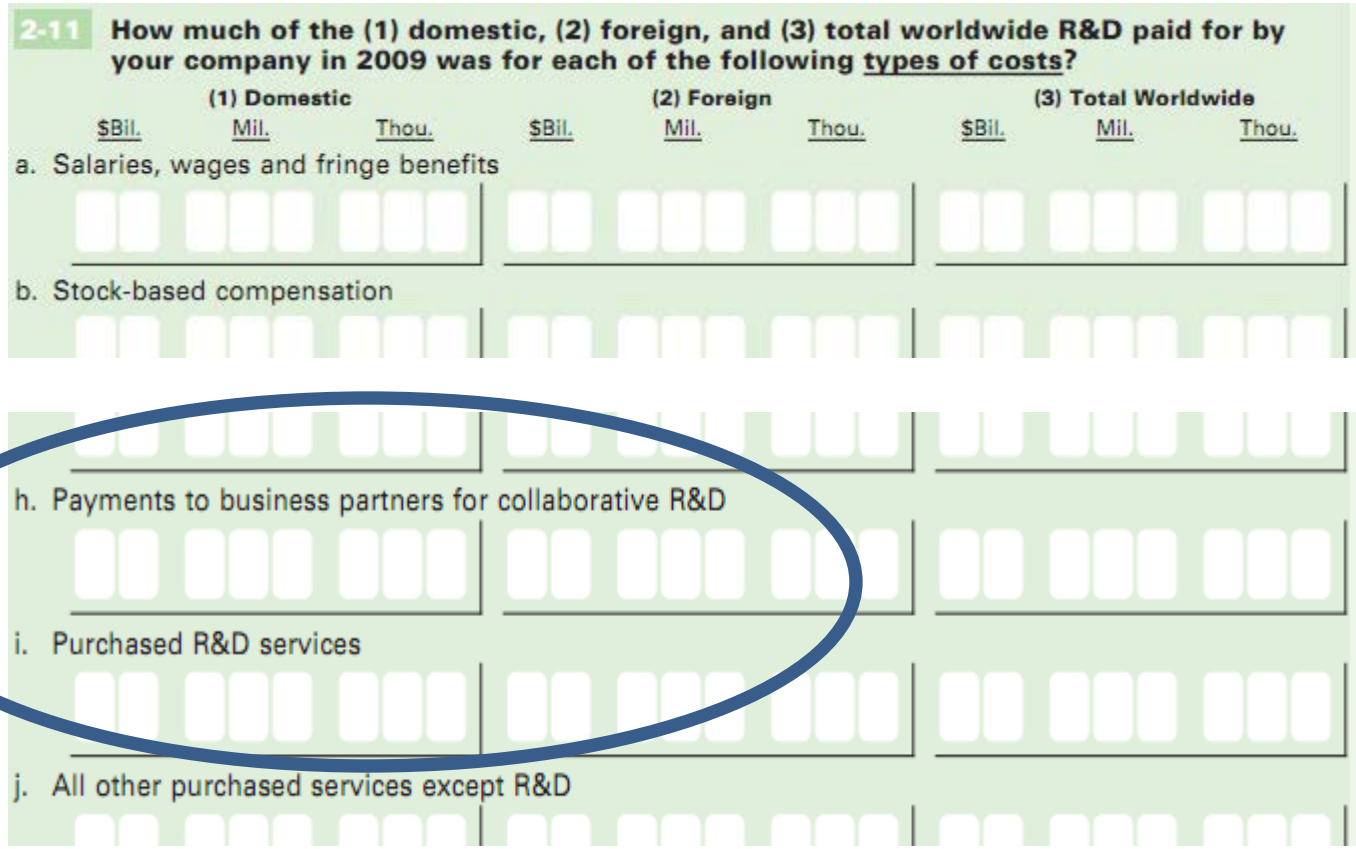
2-1 What was the total worldwide R&D expense for your company in 2010?

If your company is publicly traded, this amount is equivalent to that disclosed on SEC Form 10-K as defined by FASB Statement No. 2, "Accounting for Research and Development Costs."
If your company is foreign-owned, refer to the instructions on page 4. Additional guidance, such as for **privately-owned companies,** is available online at www.census.gov/econhelp/brdis.

NOTE: Report your company's R&D expense even if the amount is not considered material for your company's financial statements.

$Bil. Mil. Thou.

How BRDIS Measures R&D Relationships – Outsourced R&D (1)



2-11 How much of the (1) domestic, (2) foreign, and (3) total worldwide R&D paid for by your company in 2009 was for each of the following types of costs?

	(1) Domestic			(2) Foreign			(3) Total Worldwide		
	$Bil.	Mil.	Thou.	$Bil.	Mil.	Thou.	$Bil.	Mil.	Thou.
a. Salaries, wages and fringe benefits									
b. Stock-based compensation									
h. Payments to business partners for collaborative R&D									
i. Purchased R&D services									
j. All other purchased services except R&D									

How BRDIS Measures R&D Relationships – Outsourced R&D (2)

2-23 **How much of the amount reported in Question 2-22, was performed by the following types of organizations?**

	$Bil.	Mil.	Thou.
a. Companies located inside the United States			
b. Companies located outside the United States			
c. U.S. federal government agencies or laboratories			
d. U.S. state and local government agencies or laboratories			
e. Foreign government agencies or laboratories			
f. Universities, colleges, and academic researchers (including principal investigators) located inside the United States			
g. Universities, colleges, and academic researchers (including principal investigators) located outside the United States			
h. All other organizations			
i. **Total R&D performed by others** (equals Question 2-22)			

How BRDIS Measures R&D Performed by the Company That is Paid for by Others

3-1 **What were your company's total worldwide costs (both direct and indirect) in 2009 for the following that were funded, paid for, or reimbursed by others not owned by your company?**

Exclude:
- Costs that were paid for by your company such as those reported in Question 2-4
- Payments in excess of the actual cost of the work performed (such as profit or fees)

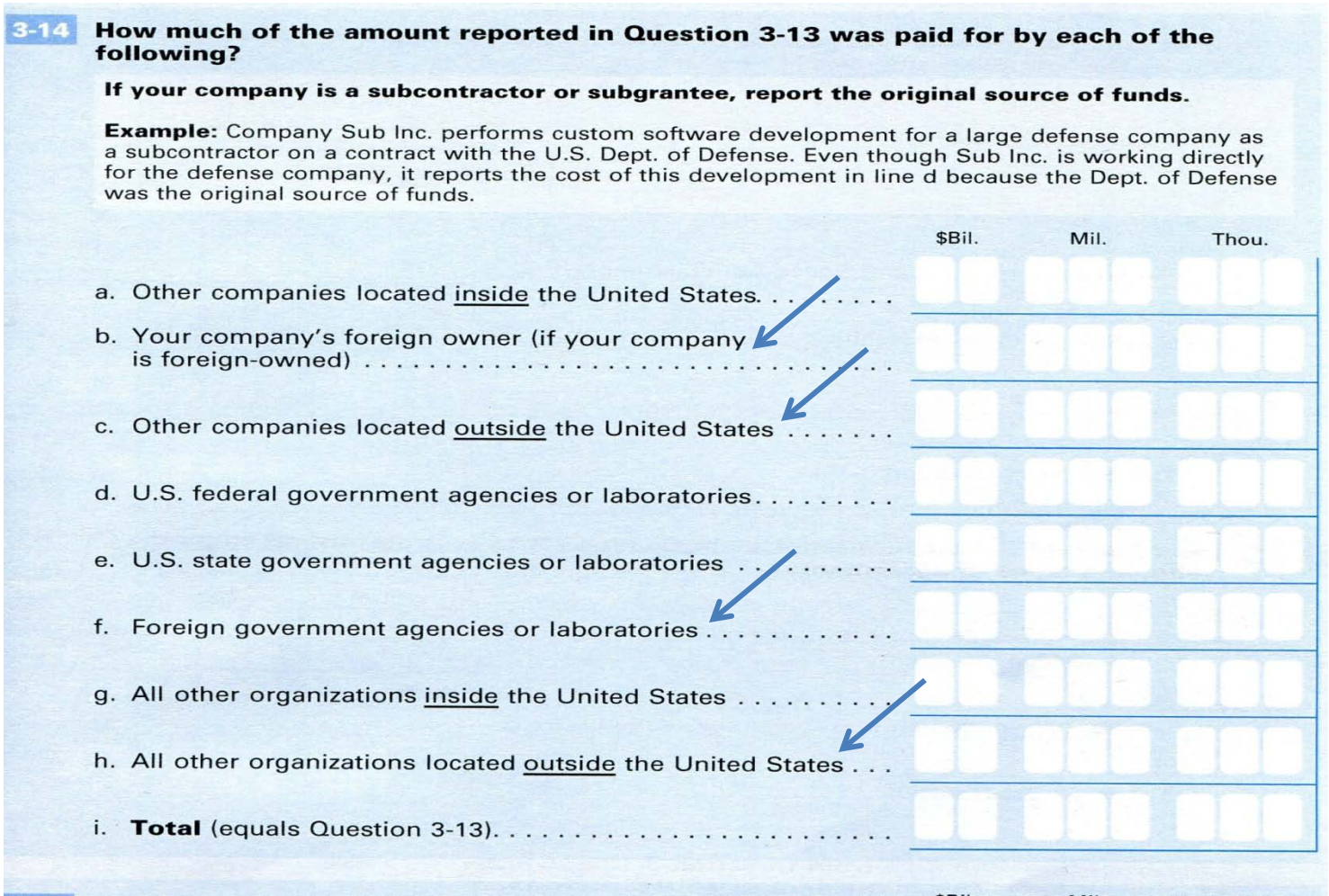
	$Bil.	Mil.	Thou.
a. Collaborative research and development that was reimbursed by business partners, such as through cost-sharing agreements			
b. R&D paid for by government or private foundation grants			
c. Defense R&D services provided to the government and/or government contractors			
d. Medical nonclinical R&D services provided to others not owned by your company			
e. Phase I-III medical clinical trial services provided to others not owned by your company (include pass-through costs)			
f. Non-defense custom software development and/or computer systems design for others not owned by your company			

Exclude:
- Software development that does not depend on a scientific or technological advance such as adding functionality to existing application programs, debugging systems, and adapting existing software

g. Developing, producing, and testing prototypes of customer's products prior to their introduction to the market (excluding defense-related prototyping reported in line c)			
h. All other R&D services provided to others not owned by your company			
i. **Total**			

Collecting R&D Performance Funded by Sources from Abroad (1)

Expanded question detail first included on the 2010 BRDIS



3-14 **How much of the amount reported in Question 3-13 was paid for by each of the following?**

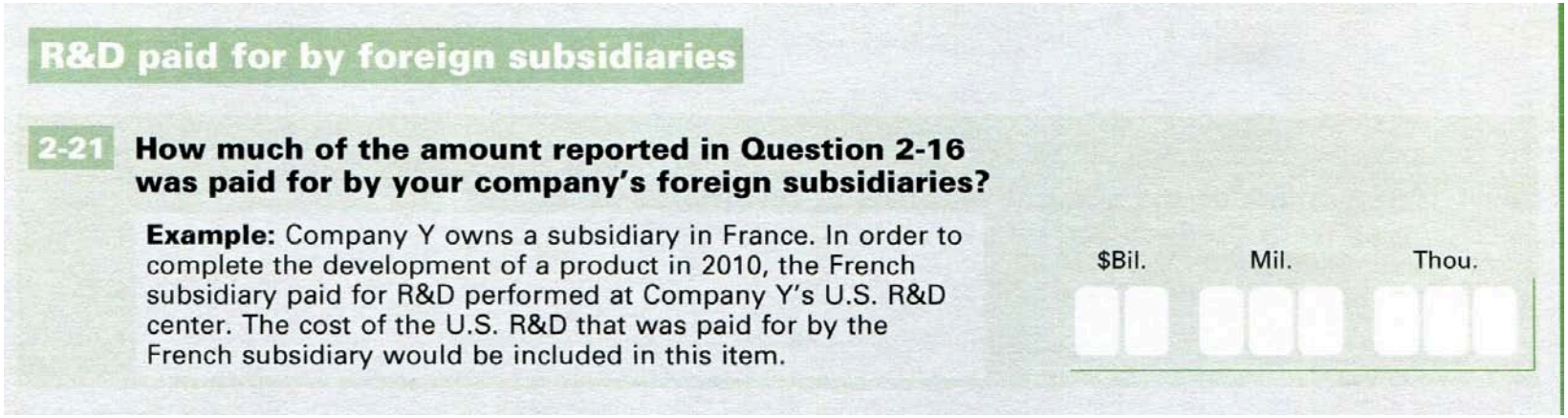
If your company is a subcontractor or subgrantee, report the original source of funds.

Example: Company Sub Inc. performs custom software development for a large defense company as a subcontractor on a contract with the U.S. Dept. of Defense. Even though Sub Inc. is working directly for the defense company, it reports the cost of this development in line d because the Dept. of Defense was the original source of funds.

	$Bil.	Mil.	Thou.
a. Other companies located inside the United States			
b. Your company's foreign owner (if your company is foreign-owned)			
c. Other companies located outside the United States			
d. U.S. federal government agencies or laboratories			
e. U.S. state government agencies or laboratories			
f. Foreign government agencies or laboratories			
g. All other organizations inside the United States			
h. All other organizations located outside the United States			
i. **Total** (equals Question 3-13)			

Collecting R&D Performance Funded by Sources from Abroad (2)

Question first included on the 2009 BRDIS



R&D paid for by foreign subsidiaries

2-21 How much of the amount reported in Question 2-16 was paid for by your company's foreign subsidiaries?

Example: Company Y owns a subsidiary in France. In order to complete the development of a product in 2010, the French subsidiary paid for R&D performed at Company Y's U.S. R&D center. The cost of the U.S. R&D that was paid for by the French subsidiary would be included in this item.

$Bil. Mil. Thou.

Collecting R&D Performance Funded by Sources from Abroad (3)

Question first included on the 2010 HERD survey*

Question 2. **How much of the total R&D expenditures reported in Question 1 came from foreign sources?**

- Include foreign governments, businesses, universities, nonprofit organizations, and any other entity sending funds to the U.S. from a location outside the U.S. and its territories.
- Projects sponsored by a U.S. location of a foreign company are **not** considered foreign.
- Include international governmental organizations located in the U.S., such as the United Nations, the World Bank, and the International Monetary Fund.

	R&D expenditures (Dollars in thousands)
Total R&D expenditures from foreign sources	$_____

*HERD question does not collect information on the originating foreign sector.

Measuring R&D Performance Funded by Sources "From Abroad"

Economic sector / Survey	Current R&D costs ($millions)	Funding from abroad ($millions)	Foreign sources as percent of total (%)
Higher Education / HERD 2010	61,235	653*	1.1%
Business / BRDIS 2009	281,872	11,590	4.1%
Foreign subsidiaries		3,816	
Foreign parent		Not separately collected in 2009	
[Unaffiliated] foreign-located companies		7,647	
All other foreign sources		127	

*For HERD, includes all foreign sources. Do not have information from which "domestic" sector should be assigned the foreign source.

**For BRDIS, beginning in 2010, foreign parents and unaffiliated foreign-located companies collected separately.

US-Reported R&D Personnel Data Differs from Frascati Guidance

Frascati Recommendation

- Headcounts and Full-time equivalents
- Separate totals for researchers (scientists and engineers), technicians, support staff
- Combined three “occupations” equate to “R&D personnel”

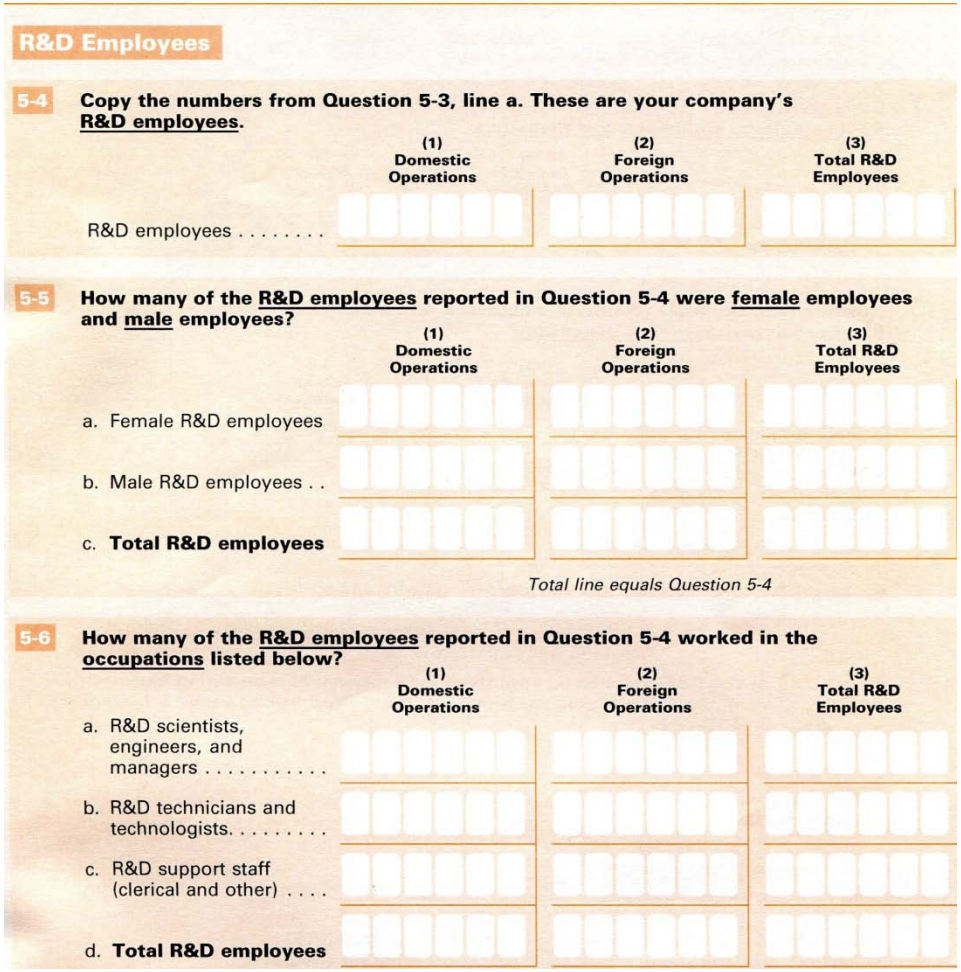
Past U.S. Situation

- No R&D personnel data collected for higher education sector on R&D surveys
- Full-time equivalent domestic researchers collected for the business sector on R&D surveys
- BRDIS and HERD redesigns attempt to capture Frascati-recommended details

Survey results, to date, have not been very satisfying

BRDIS R&D Personnel

2008 BRDIS began collecting (worldwide) R&D personnel headcounts
Continued to collect FTE domestic researchers



R&D Employees

5-4 **Copy the numbers from Question 5-3, line a. These are your company's R&D employees.**

	(1) Domestic Operations	(2) Foreign Operations	(3) Total R&D Employees
R&D employees			

5-5 **How many of the R&D employees reported in Question 5-4 were female employees and male employees?**

	(1) Domestic Operations	(2) Foreign Operations	(3) Total R&D Employees
a. Female R&D employees			
b. Male R&D employees . .			
c. **Total R&D employees**			

Total line equals Question 5-4

5-6 **How many of the R&D employees reported in Question 5-4 worked in the occupations listed below?**

	(1) Domestic Operations	(2) Foreign Operations	(3) Total R&D Employees
a. R&D scientists, engineers, and managers			
b. R&D technicians and technologists.			
c. R&D support staff (clerical and other)			
d. **Total R&D employees**			

BRDIS R&D Personnel

2011 BRDIS began collecting R&D personnel FTEs

Domestic full-time equivalents (FTEs)

5-7 **Of the domestic R&D employees reported in Question 5-4, column 1, what was the number of full-time equivalents (FTEs) for R&D activity for full-time R&D employees, other full-time employees not working solely on R&D, and part-time employees?**

Number

a. FTEs for full-time R&D employees

Count the number of full-time employees who work only on R&D. . . . .

Example:
50 full-time R&D employees worked only on R&D = 50 FTEs

b. FTEs for other full-time employees not working solely on R&D

Use the proportion of the time they work on R&D to calculate the number of FTEs. . . . . .

Example:
60 full-time employees averaged one-fourth of their time on R&D = 15 FTEs

c. FTEs for part-time employees working on R&D

Use the portion of a full-time week (such as 40 hours) that they work on R&D to calculate the FTEs. . . . . .

Example:
20 part-time employees averaged 20 hours a week on R&D activities = 10 FTEs

d. **Total FTEs** . . . . . .

HERD R&D Personnel

2010 HERD began collecting researcher and R&D personnel headcounts*

Question 16. How many principal investigators and other personnel (headcount) were paid from the R&D salaries, wages, and fringe benefits you reported in Question 13, row a?

- A **principal investigator (PI)** is designated by your institution to direct the R&D project or program and be responsible for the scientific and technical direction of the project. Co-investigators (co-PIs) may be designated for this role and should also be included in column 1.
- Count each person only once.
- If a person serves as a PI or co-PI on one project and other personnel on another project, count that person as a PI.
- Include all personnel and students paid from R&D accounts regardless of how much they received.

	(1) Principal investigators	(2) All other personnel	(3) Total[1]
Number of people (headcount)	_____	_____	TOTAL

*HERD redesign investigation indicated collecting FTE data would be extremely problematic. Collected PI data looks rather reasonable; “all other R&D personnel” data look suspect.

Thoughts on Non-Profit R&D Estimation

Workshop on National Patterns

CNSTAT Staff

September 6-7, 2012

Current Approach to Estimation of Non-Profit R&D Funds

- NCSES has not collected direct information on several major components of non-profit funding of R&D since 1997. However, a considerable amount of information is available on business and federal government funding to non-profits from BRDIS and from the Federal Funds survey, respectively. But NCSES has had to estimate industry funding of R&D to non-profits and non-profit funding of R&D to other non-profits using the following models.

Current Approach to Estimation of Non-Profit R&D Funds

- Denote by X_T industry funding of R&D by industry, denote by Y_T industry funding of R&D by non-profits, denote by Z_T non-profit funding of R&D by academia, and denote by W_T non-profit funding of R&D by (likely other) non-profits.

Current Approach to Estimation of Non-Profit R&D Funds

- Then NCSES takes the following ratio:

$$\frac{\frac{Y_{1997} - Y_{1996}}{Y_{1996}}}{\frac{X_{1997} - X_{1996}}{X_{1996}}}$$

the percentage change in industry to non-profit funding between1996 and 1997 divided by the same for industry to industry funding during the identical period.

Current Approach to Estimation of Non-Profit R&D Funds

- Given that the denominator is known for the time periods *T* and *T*+1, multiplying that value by the above ratio gives an estimate of the percentage change in industry to non-profit R&D funding for the current time period.

Primary Assumption

- That is we are hoping that:

$$\frac{\frac{Y_{T+1}-Y_T}{Y_T}}{\frac{X_{T+1}-X_T}{X_T}} = \frac{\frac{Y_{1997}-Y_{1996}}{Y_{1996}}}{\frac{X_{1997}-X_{1996}}{X_{1996}}}$$

- It is a simple matter to take the estimate for the previous year to get the current estimate.

Primary Assumption

- Similarly, to estimate non-profit R&D funding to non-profits, we take the analogous ratio, this time of non-profit R&D funding of non-profits over academic R&D funding to non-profits in 1996-1997.

$$\frac{\frac{W_{1997} - W_{1996}}{W_{1996}}}{\frac{Z_{1997} - Z_{1996}}{Z_{1996}}}$$

Primary Assumption

- Again, to estimate the current non-profit to non-profit R&D funding, we would multiply this ratio by the current percentage difference for academic funding to non-profits collected now on HERD, and then it is a simple matter to take the previous estimate to produce the current estimate of non-profit to non-profit R&D funding.

What Else Could Be Done?

- While there are alternatives to this approach, there are no obviously better alternatives. For example, one could argue (1) that the above model should be fit using homogeneous clusters of types of non-profits, and (2) that the ratio being estimated should be smoothed over time.
- But it is unclear how to form such clusters and whether the variables to do so are available and their stability over time. Second, due to the irregular history of collecting data on non-profits, one can't form a time series since data have only been collected in 1973, 1996, and 1997.

The current method is unlikely to provide quality estimates

- Further, we repeat a point made in the 2006 NCSES methodology report --- that the 1996 -1997 data itself are incomplete due to nonresponse, accounting for between 80 and 90% of this sector's R&D total. So these percentage differences from 1996-1997 are possibly not well-estimated in the first place.

- But much more important that that is the reliance on what must be viewed as an heroic assumption that the ratio of percentage yearly change in one sector of R&D funding divided by that for another sector would remain stable for even a few years, let alone 17 years. That is the key concern.

SRI International

Obtaining Quality Data on NPO R&D Activities

Opportunities for Leveraging Administrative Records and Other Sources

Jeffrey Alexander, Ph.D.
Center for Science, Technology & Economic Development
6 September 2012

Discussion Topics to Cover

- Nature of non-profit R&D activities
- Changes in the NPO sector since 1996
- Evaluating potential non-survey data collections on NPO R&D (update to 2006 SRI report to NCSES)
- Key concerns and future opportunities

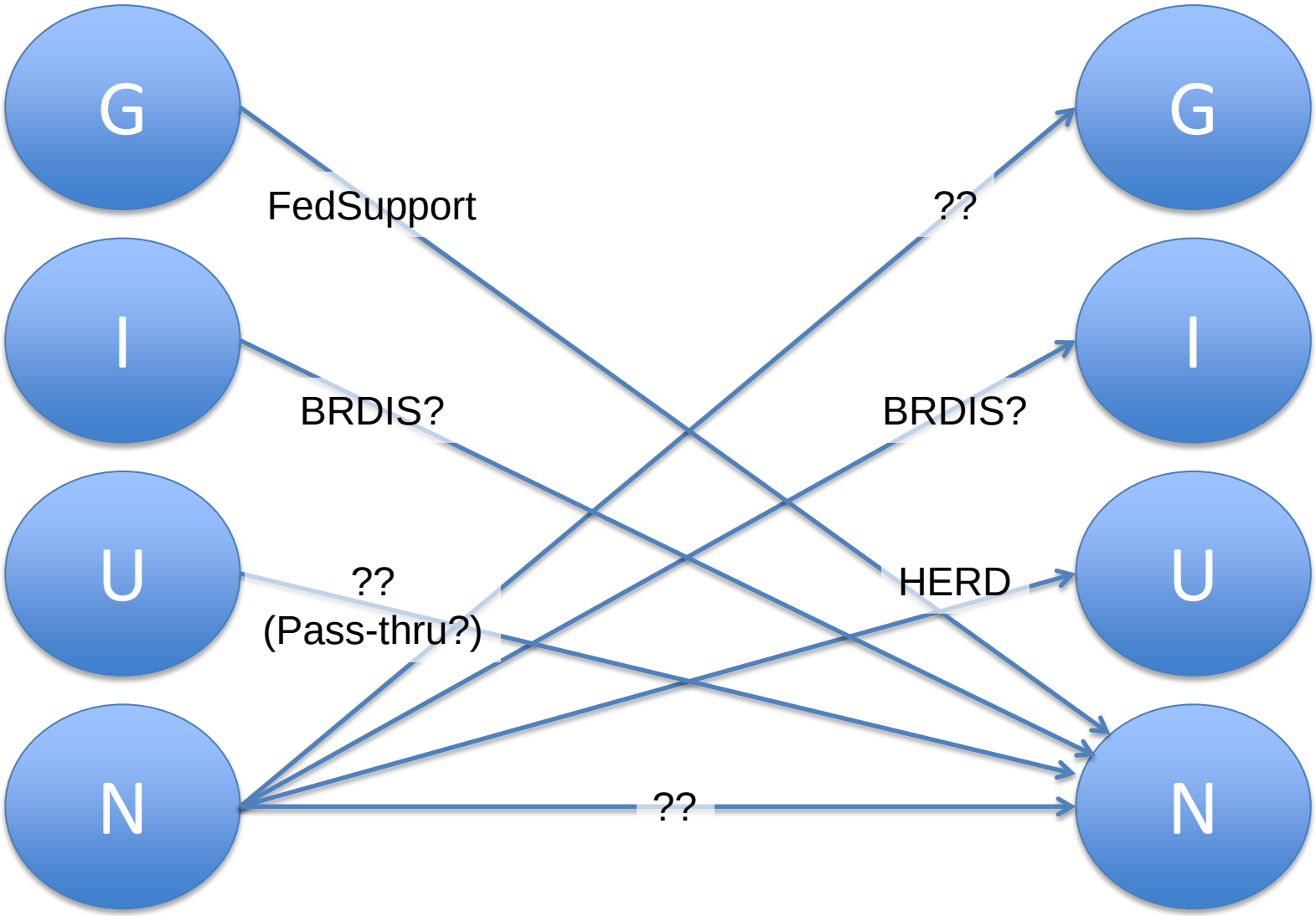
Key Findings

- The non-profit sector has changed both quantitatively and qualitatively since 1996, and those trends probably affect non-profit R&D activities.

- The scale of non-profit R&D activity is relatively small compared to other sectors, but is likely to have disproportionate impact in particular fields (e.g. biomedical research).

- Non-survey sources for data on non-profits exist, but are severely limited in their ability to provide estimates of total non-profit R&D activity.

Basic Problem Statement



A Starting Point…

"There's no point in being precise if you don't even know what you're talking about."

John von Neumann
Pioneer in computer science research

SRI International

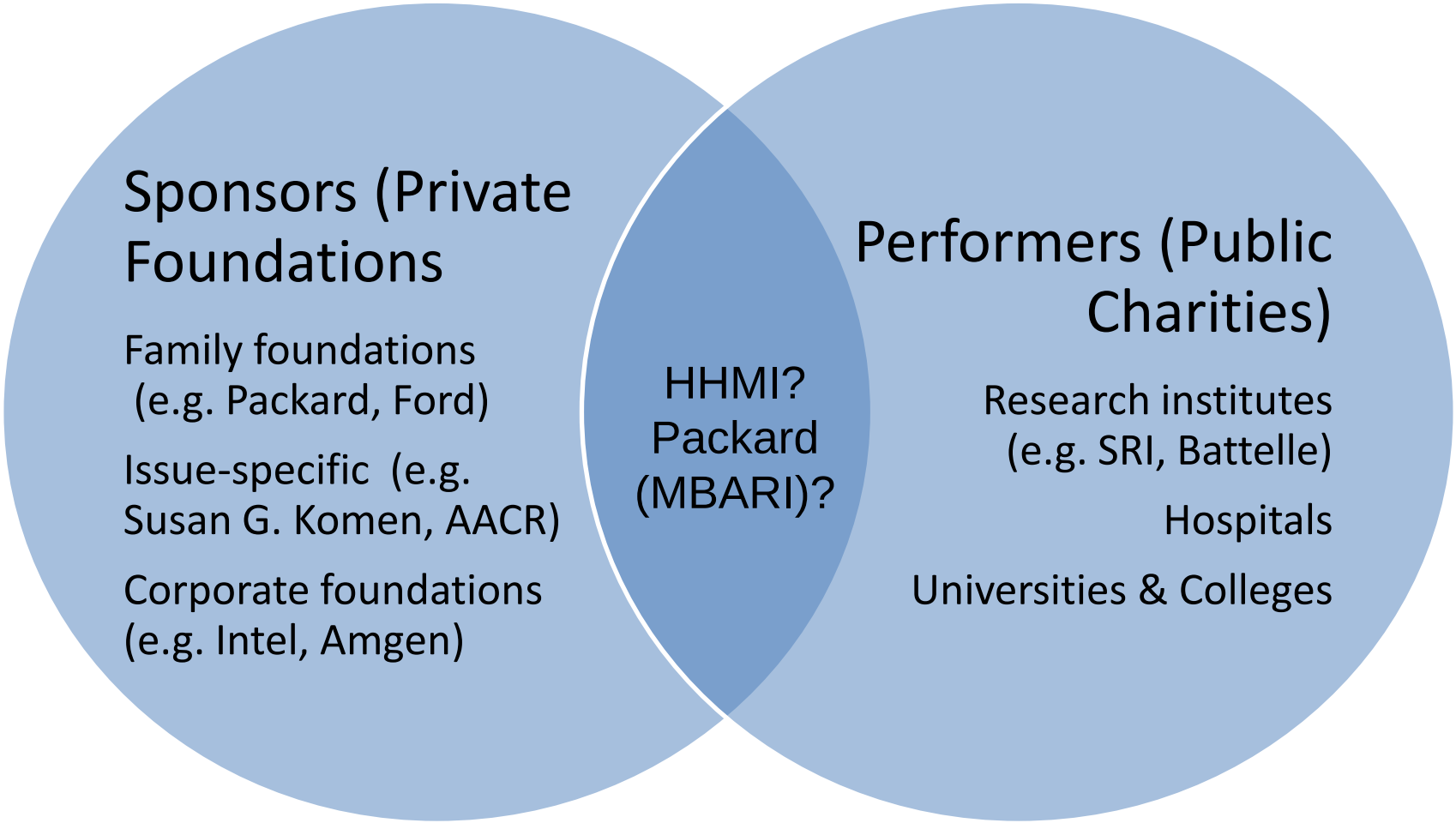
Nature of NPO R&D Activities

Nature of NPO R&D Activities

Classifications of the NPO sector

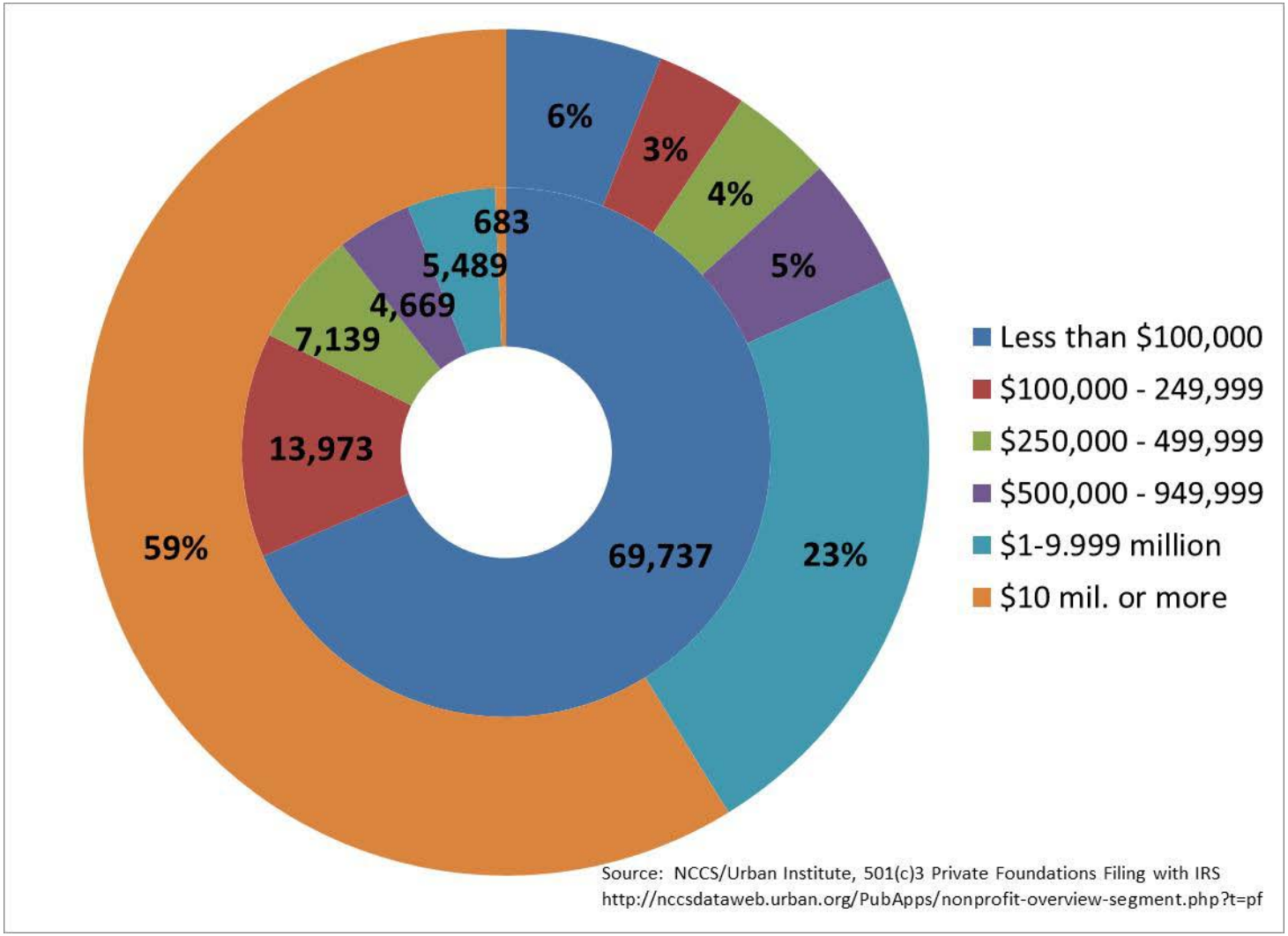
- IRS code—classification by purpose
  - "The exempt purposes set forth in section 501(c)(3) are charitable, religious, educational, scientific, literary, testing for public safety, fostering national or international amateur sports competition, and preventing cruelty to children or animals." *(IRS exemption requirements for 501(c)3 organizations)*
  - Includes charitable organizations (primarily recipients of charitable contributions) and private foundations (grant-makers)
  - Other non-profits (political advocacy groups, unions & trade associations, etc.) covered under different sections of the IRS code

Nature of NPO R&D Activities



Nature of NPO R&D Activities

Total grants paid (outer circle) by private foundations by organization annual expense (inner circle)



Nature of NPO R&D Activities
Classifications of the NPO sector

- NTEE taxonomy—classification by organizational mission
  - 26 major groups
    - Denotes primary interest, activities and focus areas
    - E.g. E=Health Care, G=Diseases, Disorders & Medical Disciplines, H=Medical Research, U=Science & Technology
  - Each group has 2-digit subcategories
    - 01 thru 19: 7 subcategories of "management & planning" purposes
      - E.g. 01=Alliances & Advocacy, 05=Research Institutes & Policy Analysis, 12=Fundraising & Fund Distribution
    - 20 thru 99: varying areas of specialization
      - E.g. U20=General Science, U34=Mathematics, U99=S&T not elsewhere classified
  - Code assigned by IRS examiner for tax purposes based on application for exemption
    - NPOs also self-identify for other purposes (e.g. entries in Guidestar database)

Nature of NPO R&D Activities

Sample NTEE Codes for Specific R&D Non-Profits

- U20: Organizations that focus broadly on scientific research and inquiry or which engage in interdisciplinary science activities (96 orgs found)
  - Institute for Advanced Study ($656,458,901 in assets)
  - Research Triangle Institute ($441,855,940)
  - SRI International ($322,849,000)
  - Institute for Defense Analyses ($275,646,442)
- H30: Organizations that conduct research which can be used to improve the prevention, diagnosis and treatment of cancer (245 orgs found)
  - Fred Hutchinson Cancer Research Center ($555,798,313 in assets)
  - American Cancer Society Inc, East Syracuse ($168,181,446)
  - American Cancer Society Inc, Austin ($159,211,398)
- V05: Organizations whose primary purpose is to conduct research and/or public policy analysis within the Social Science major group area. (37 found)
  - American Enterprise Institute for Public Policy Research ($138,205,460 in assets)
  - The Milken Institute ($48,210,278)
  - Pew Research Center ($45,788,087)

Nature of NPO R&D Activities
Oceanography Research

Leading Sponsors of Ocean/Marine Research in $thousands

Organization	Estimated Funding
NSF	$ 373,533
DOD/Navy	$ 243,934
NOAA	$ 132,100
NASA	$ 36,557
Packard/MBARI	$ 35,300
EPA	$ 20,732
Moore	$ 18,000
DOD/OSD	$ 15,459
Other federal	$ 11,924

- David & Lucile Packard Foundation funds numerous projects to increase the impact of science on policy for oceans and the marine environment, but grants to MBARI are its only direct research funds in this field
- Gordon & Betty Moore Foundation provides research funds in marine science to a number of universities and non-profit performers, including MBARI

Nature of NPO R&D Activities

Observations on NPO classification

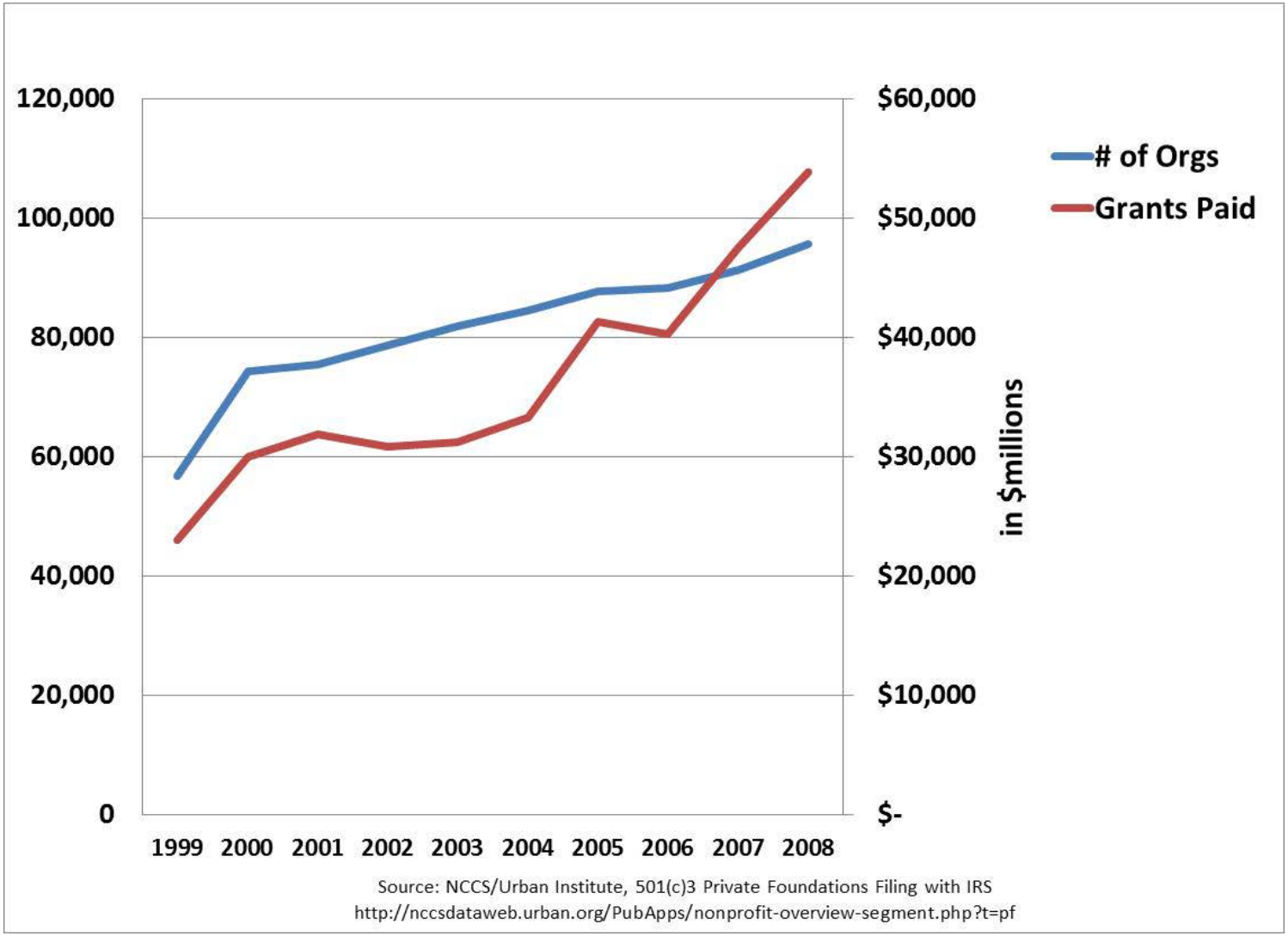
- NPOs tend to be either sponsors of R&D or performers of R&D, but some play both roles.
- A relatively small number of NPOs sponsors are most likely responsible for the bulk of NPO funding for R&D, which may facilitate data collection
- NPO funders of R&D tend to focus on specific applications/topics, which may aid in collection & analysis
  - However, the classification of NPO activities by topic can be difficult, and coding is unreliable (more on this later)

SRI International

Changes to the NPO Sector Since 1996

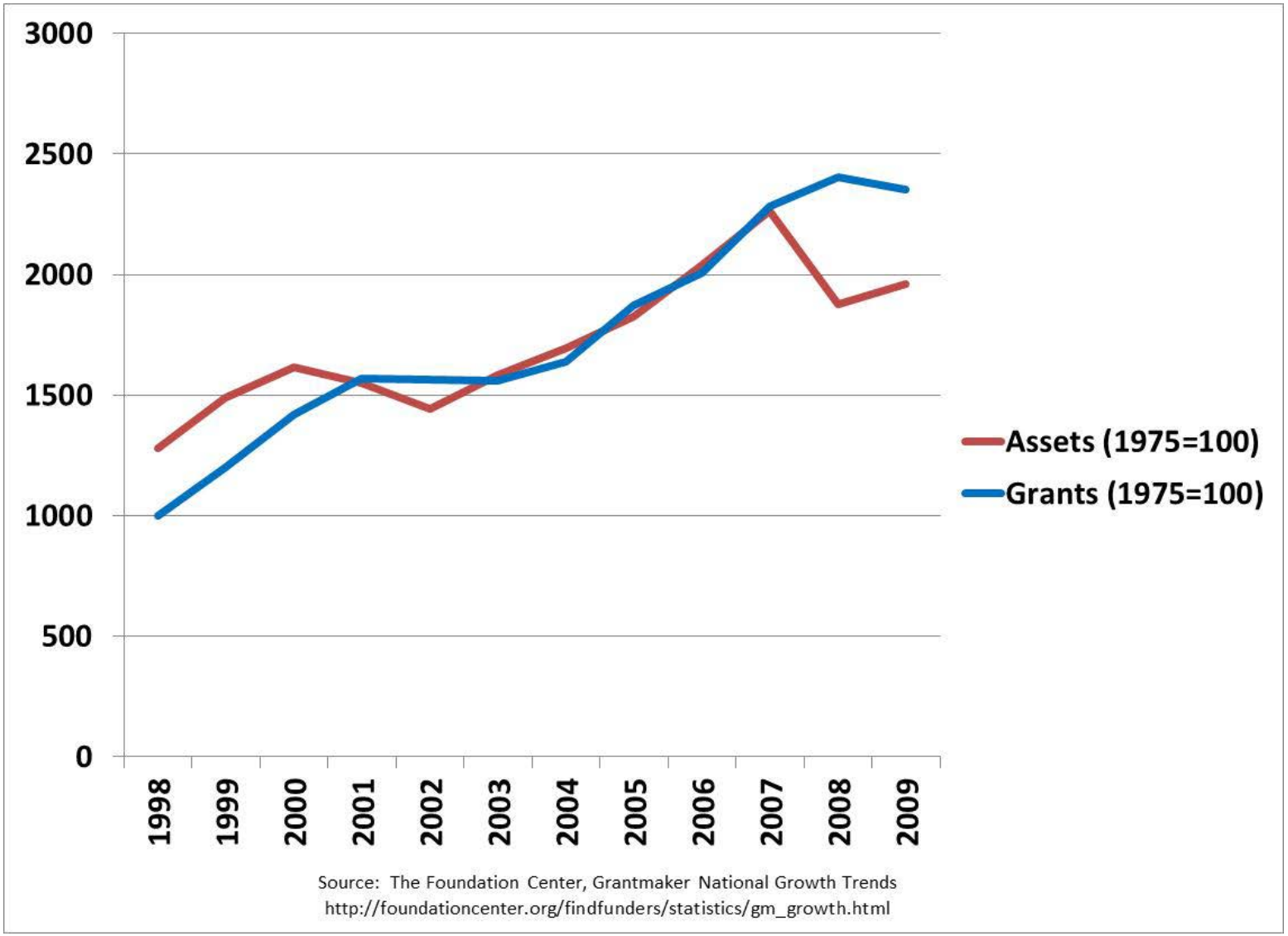
Changes in the NPO Sector Since 1996

Growth in NPOs and grants paid ($current)



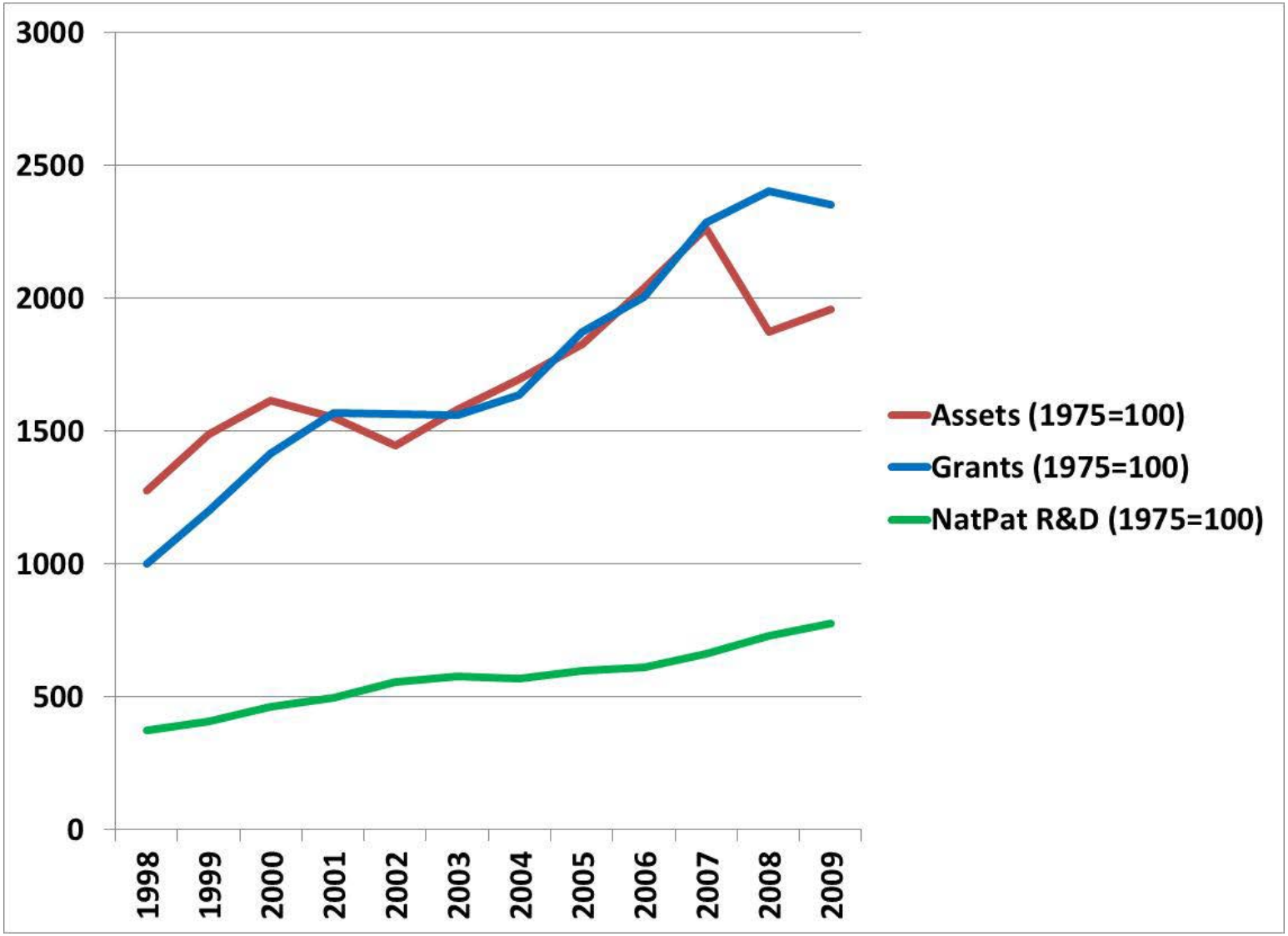
Changes in the NPO Sector Since 1996

Assets and grants indexed to 1975 constant dollars



Changes in the NPO Sector Since 1996

Experimental comparison to NatPat estimates



Changes in the NPO Sector since 1996

Top 10 Grant-Making NPOs in "S&T," 1998 vs. 2010

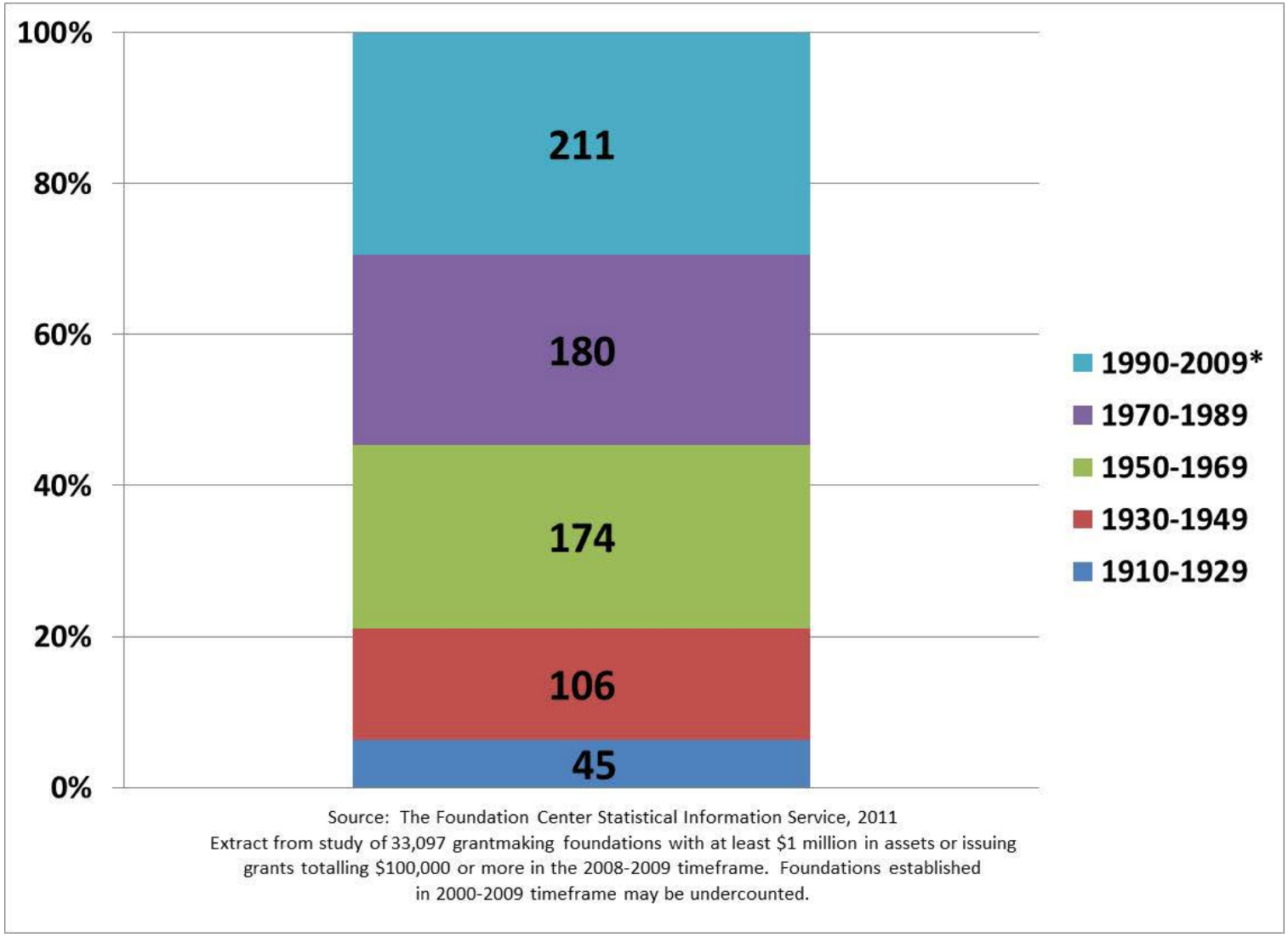
Org	Amount
The David and Lucile Packard Foundation	81,651,751
Lilly Endowment Inc.	23,748,930
The Robert A. Welch Foundation	21,729,021
W. M. Keck Foundation	15,279,446
F. W. Olin Foundation, Inc.	12,137,315
The Whitaker Foundation	11,521,864
W. K. Kellogg Foundation	11,121,620
Donald W. Reynolds Foundation	10,873,590
John D. and Catherine T. MacArthur Foundation	10,340,700
The Kresge Foundation	9,771,666

Org	Amount
Eli & Edythe Broad Foundation	115,025,000
Silicon Valley Community Foundation	51,980,260
Hall Family Foundation	50,000,000
Gordon and Betty Moore Foundation	45,616,849
The David and Lucile Packard Foundation	37,329,280
Alfred P. Sloan Foundation	27,874,053
Amgen Foundation, Inc.	20,926,484
The Robert A. Welch Foundation	14,375,000
The Tabasgo Foundation	10,993,200
The Marcus Foundation, Inc.	9,565,946

Source: The Foundation Center Statistical Information Service, http://foundationcenter.org/findfunders/statistics/gs_subject.html

Changes in the NPO Sector since 1996

Grantmaking foundations with <$100M in assets by founding year



Source: The Foundation Center Statistical Information Service, 2011
Extract from study of 33,097 grantmaking foundations with at least $1 million in assets or issuing grants totalling $100,000 or more in the 2008-2009 timeframe. Foundations established in 2000-2009 timeframe may be undercounted.

Changes in the NPO Sector since 1996

The rise of "venture philanthropy"

- Beginning in the late 1990s, technology entrepreneurs entered the world of philanthropy with ideas that differed from "established" foundations
  - Greater emphasis on outcomes and performance evaluation of grantees
  - Closer relationship between foundations and grantees
  - Targeted "social investment" strategies (such as the Gates Foundation's strategy towards reducing childhood diseases in developing nations)

- Venture philanthropy has changed traditional philanthropy
  - Multiple initiatives to improve outcome evaluation
    - Urban Institute "Outcome Indicators Project"
  - Efforts to measure grantee satisfaction with funding process
  - More focus on "innovation" in programs and service delivery
    - E.g. Janelia Farm project at HHMI

Changes in the NPO Sector Since 1996

Changes in the nature of NPOs

- NPOs are facing pressures similar to those felt throughout the economy
  - After-effects of the Great Recession on asset values
  - Rising health care and pension obligations
  - Difficulties in recruiting talent
  - Increased calls for tax authorities to scrutinize NPO activities
- NPOs may undergo structural change
  - Growing use of “program-related investments” in place of grants
  - Emergence of dedicated venture philanthropy “funds”
  - Legislation to create “for-benefit corporations”

Changes to the NPO Sector Since 1996

Observations on trends affecting R&D activities

- NPO funding of R&D is more concentrated and changing in focus
- Shift to outcomes-oriented investment strategies may complicate efforts to isolate the "R&D" component of grant-making programs
- NPO R&D performers are likely to diversify into new areas of activity, possibly leading to complications
  - E.g. spin-off of Mitretek Systems (now Noblis) from MITRE

SRI International

Evaluating Non-Survey Data Sources on Non-Profits

Evaluating Non-Survey Data Sources on Non-Profits

Potential sources of non-survey (administrative) data

- Federal data repositories
  - USASpending.gov (formerly FAADS)
  - Federal Audit Clearinghouse
  - Internal Revenue Service Master Business File
- Third-party data providers
  - Organizational profile repositories (Guidestar, NCCS)
  - Sector-wide surveys (Foundation Center, Council on Foundations, Center for Civil Society Studies)
  - Grant solicitation repositories (GrantStation, Web of Science)
- NPO self-reporting
  - Annual reports and financial statements
  - IRS Form 990

Evaluating Non-Survey Data Sources on Non-Profits

- Evaluation criteria for statistical data (based on *Statistics Canada Quality Guidelines*, October 2009)
  - Coverage: does the information provide a comprehensive view?*
  - Accessibility: how easily can the information be obtained?
  - Accuracy: does the information correctly describe the phenomena it was designed to measure?
  - Timeliness: what is the delay between the reference point to which the information pertains, and the date on which the information becomes available?
  - Interpretability: is there supplementary information and metadata to understand the information and use it appropriately?
  - Coherence: can the information be brought together with other statistical data to analyze broader concepts and trends?
  - Relevance: does the information shed light on the issues of most importance to users?

Evaluating Non-Survey Data Sources on Non-Profits

Evaluation of federal data repositories

	USASpending.gov	FAC	IRS
Coverage	NPOs receiving federal grants	NPOs receiving federal grants	Qualifying exempt organizations
Accessibility	Easy, bulk download	Easy, download by org	Easy, possible bulk download
Accuracy	Very good and improving	Inconsistent	Authoritative
Timeliness	Timely & improving	Inconsistent	Annual & reliable
Interpretability	Low (lacking in helpful detail)	Varies widely with organization	Moderate depending on org and data record
Coherence	Improving (EIN matching)	Low (PDF format)	Improving (due to e-filing)
Relevance	High (focus on R&D grants to NPOs)	Moderate (requires significant screening)	Moderate (accurate but lacks detail)

Evaluating Non-Survey Data Sources on Non-Profits

Evaluation of third-party data repositories

	Guidestar, NCCS (org profiles)	Foundation Center, CoF, etc. (surveys)	GrantStation, WoS (solicitations)
Coverage	High (over 1.5M profiles)	High but varies	Unknown
Accessibility	Easy	Easy with proprietary access	Varies
Accuracy	Very good (focus on data quality)	Good	Unknown
Timeliness	Timely	Acceptable (12-18 month lag	Very timely
Interpretability	High (detailed metadata)	Moderate	Low (depends on each entry)
Coherence	High (strong standards)	Low (varies by survey)	Low (poor standardization)
Relevance	Moderate (lacks specific detail)	Low (no focus on R&D activities)	Moderate (depends on data source)

Evaluating Non-Survey Data Sources on Non-Profits
NPO self-reporting

- Some NPOs publish annual reports, but not all

- IRS Form 990 can provide some detail on grants
  - Primarily found in "Program Service Accomplishments" statements (Part III of Form 990, also in supplemental statements in Form 990-PF)
  - Level of detail is generally shallow
    - Organizations offer varying degrees of information in supplemental schedules and statements
  - Schedules have field for "Nonprofit Program Classification" code from NCCS
    - Ex: G02.04 – Cancer: Programs that support the prevention, treatment and cure of neoplasms, abnormal formations of tissue, such as a tumor or growth, which serve no useful function, but grow at the expense of the healthy organism. See also: Cancer Research (H02.04)
    - NPC has never been required by IRS and is rarely used

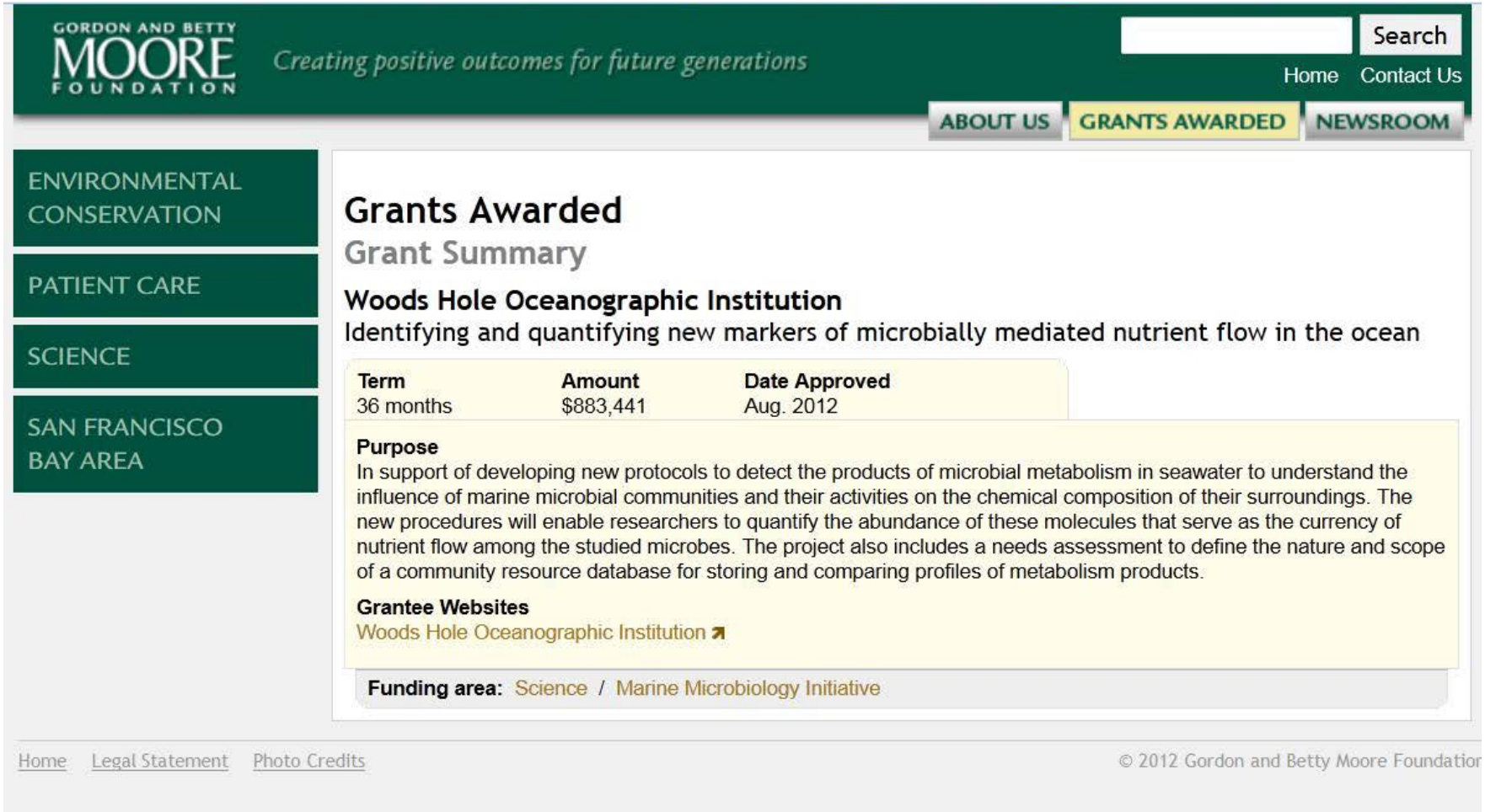
Evaluating Non-Survey Data Sources on Non-Profits

Example of self-report: RTI Schedule O PSA

HEALTH RESEARCH is our largest single field of study, encompassing research that ranges from studies of the human genome and the development of new drug compounds to national surveys of health behaviors and the implementation of global health programs Our activities in DRUG DISCOVERY AND DEVELOPMENT include working with pharmaceutical companies and government agencies to bring new medicines to market, and to ensure the safety and efficacy of those in the marketplace As part of a comprehensive EDUCATION AND TRAINING RESEARCH program, our experts conduct rigorous studies used to assess and improve the quality of early development, education, and employment programs in the United States and many other countries We also provide training solutions for forensic scientists, the military, and law enforcement RTI is a leader in SURVEY RESEARCH and SURVEY STATISTICS With a staff of highly skilled researchers using advanced data collection systems and techniques, we offer broad-based survey services for public and private sector studies Our SURVEY RESEARCH SERVICES include study design, instrument development and evaluation, pretests and pilot studies, mail surveys, telephone surveys, face-to-face field surveys, Web surveys, records abstraction, collection of biological specimens, mixed-mode surveys, subject tracing, focus groups, and health registries Our STATISTICS RESEARCH experts conduct complex statistical analyses to support wide ranging research programs in both laboratory and social sciences as well as ensure the quality, validity, and reliability of our research products and results by applying quality control and assurance procedures from sampling design through data collection and analysis Our multidisciplinary ECONOMIC & SOCIAL RESEARCH team includes experts in economics, demography, health, education, urban planning, and public financing We provide independent, objective information and analyses that help to inform and improve public policy discourse and advance economic and social development decision making in technology economics and policy, crime and justice, and food and agricultural policy We provide ADVANCED TECHNOLOGY RESEARCH & DEVELOPMENT services through the identification, development, application, and transfer of leading-edge technologies in areas including Materials Science, Electronics, Microfabrication, Information Technology, Innovation Advising, and Technology Economics & Policy Our ENERGY RESEARCHERS develop technologies that seek to generate cleaner and more efficient sources of energy Through these efforts, we support national and worldwide goals of energy security and creation of a reliable, sustainable, and economically viable energy supplies We are a leading provider of ENVIRONMENTAL RESEARCH SERVICES for the U S Environmental Protection Agency and other government agencies We conduct large-scale environmental management projects for both government and industry, helping them make critical policy and regulatory decisions Our LABORATORY & CHEMISTRY SERVICES support RTI research with state-of-the-art instrumentation, technologies, and processes Our lab specifications and procedures emphasize quality control and excellence at every level of activity to meet the expectations and needs of our clients

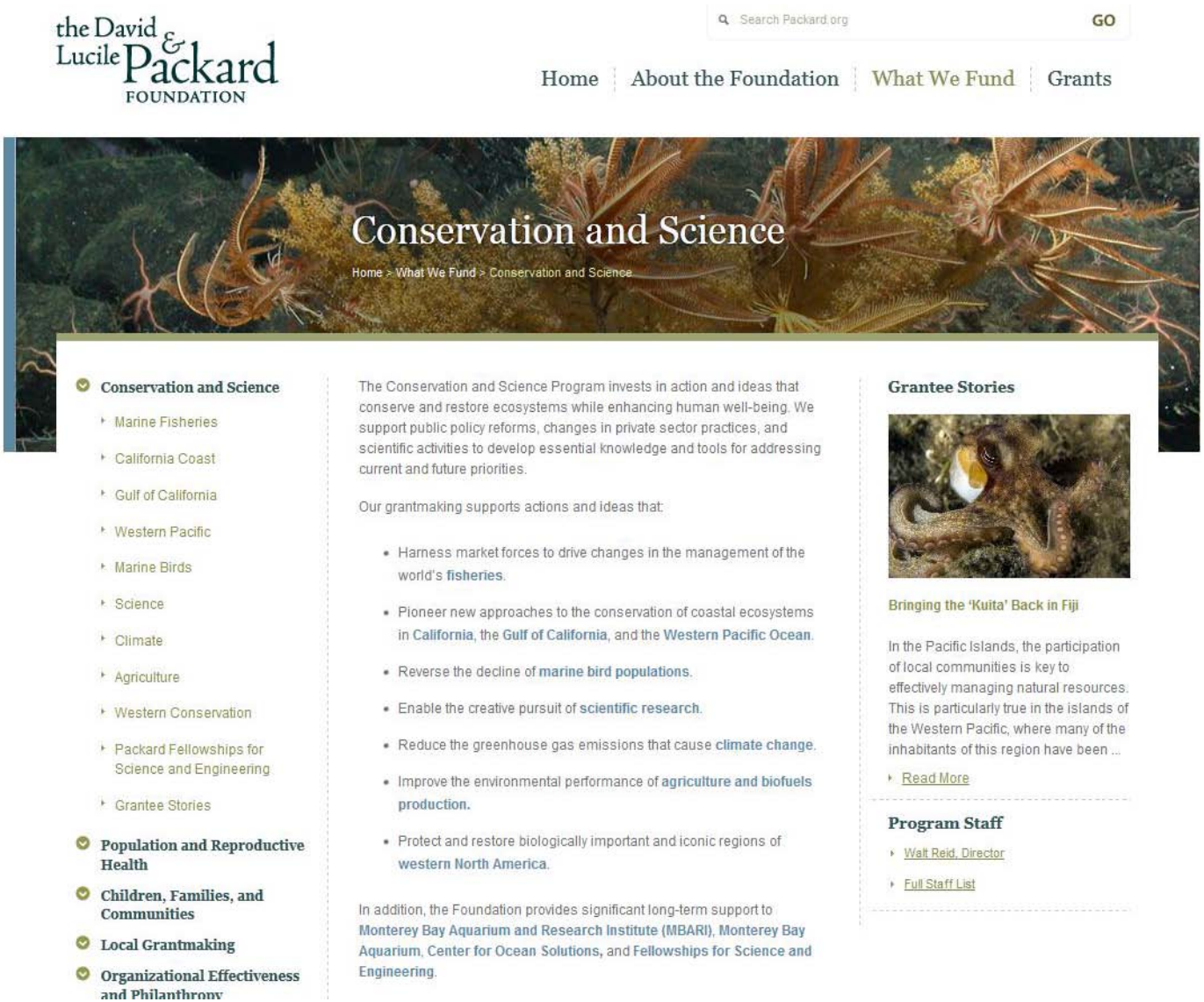
Evaluating Non-Survey Data Sources on Non-Profits

Example of self-report: Moore Foundation website



Evaluating Non-Survey Data Sources on Non-Profits

Example of self-report: Packard Foundation program overview



SRI International

Key Concerns & Future Opportunities

Key Concerns

- Strong likelihood that NatPat estimates are inaccurate
  - Not clear if it is overestimating or underestimating NPO R&D funding
  - Most recent NatPat estimates NPOs funded ~$12.6B in R&D in 2008, while NCCS reports private foundation grantmaking of $53.8B in 2008
    - Is it reasonable to assume that over 20% of grantmaking went to R&D activities?

- Administrative data collections are **not** sufficient to collect *consistent & comprehensive* information on NPO R&D activities
  - Research activities are often embedded in non-research programs (e.g., how to estimate the role of research in vaccine delivery in Africa?)
  - Classification systems do not treat R&D consistently, and are not easily aligned with NCSES classification systems (such as FOSE)
  - NPOs have little incentive to provide more detailed reporting of R&D activities
  - Data are not delivered in machine-readable format
    - Most IRS Form 990s are published online as image PDFs

Future Opportunities

- More administrative data on NPO activities are becoming available in electronic format
  - Move to e-filing of IRS Form 990
  - Possible move to XML standards for reporting financial data
  - Individual NPOs are putting more information on the Web
- Increased standardization of reporting processes and content
  - NCCS-led efforts to improve quality and timeliness of IRS submissions
  - Sector is generally supportive of efforts in transparency (e.g. Publish What You Fund)
- Possible options for using text analytics to assist in classification of R&D activities
  - Assumes some level of detail on project-level records

Key Findings

- The non-profit sector has changed both quantitatively and qualitatively since 1996, and those trends probably affect non-profit R&D activities.

- The scale of non-profit R&D activity is relatively small compared to other sectors, but is likely to have disproportionate impact in particular fields (e.g. biomedical research).

- Non-survey sources for data on non-profits exist, but are severely limited in their ability to provide estimates of total non-profit R&D activity.

Thank You

jeffrey.alexander@sri.com

Menlo Park Headquarters

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025-3493
650.859.2000

Washington, D.C.

SRI International
1100 Wilson Blvd., Suite 2800
Arlington, VA 22209-3915
703.524.2053

Additional U.S. and international locations

www.sri.com

Capturing Change:

Improving Measures of Science, Technology, and Innovation

Kaye Husbands Fealing

National Research Council

National Patterns Workshop

September 6, 2012

Battelle (2011) – *2012 Global R&D Funding Forecast*

		R&D Performance				
	Share of All R&D	Industry	Federal Government	Academia	Non-Profit	Total
Basic Research	18%	20%	7%	60%	13%	100%
Applied Research	22%	72%	8%	13%	7%	100%
Development	60%	91%	6%	1.5%	1.5%	100%
All R&D	100%	72%	8%	16%	4%	-

Source: NSF, 2008 National Patterns

The Source-Performer Matrix

Estimated Distribution of U.S. R&D Funds in 2012
Millions of Current U.S. Dollars (Percent Change from 2011)

Source	Performer					
	Federal Gov't.	FFRDC	Industry	Academia	Non-Profit	Total
Federal Government	$29,152 -2.51%	$14,666 -3.69%	$37,577 -2.42%	$37,440 0.93%	$6,817 -2.29%	$125,652 -1.61%
Industry		$202 2.20%	$273,487 3.37%	$3,868 26.49%	$2,129 8.89%	$279,685 3.75%
Academia				$12,318 2.85%		$12,318 2.85%
Other Government				$3,817 2.72%		$3,817 2.72%
Non-Profit				$3,491 2.70%	$11,055 2.70%	$14,546 2.70%
Total	$29,152 -2.51%	$14,868 -2.36%	$311,063 2.63%	$60,934 2.85%	$20,001 1.55%	$436,018 2.07%

Source: Battelle, *R&D Magazine*

Reliability

Due to our reliance on the NSF's National Patterns data as the foundation of our estimates, **significant adjustments to these data affect the levels and directions of our forecast**. For example, after the release of our 2011 forecast, the NSF issued an *InfoBrief* entitled *Department of the Air Force Revises R&D Data for FY 2000–07*. This report detailed upward adjustments to the Air Force's R&D obligations, ranging from $3 billion in 2000 to nearly $14 billion in 2007. In the context of our forecast, this historical change will also reset the federal R&D funding baseline by nearly $14 billion starting in 2007.

Battelle (2011) – *2012 Global R&D Funding Forecast*

Timeliness

"This detailed forecast of U.S. R&D investment is built upon data derived from the National Science Foundation's (NSF's) ***National Patterns of R&D Resources***, a longitudinal database of U.S. R&D funding and performance. **The most recent complete release of this database includes *estimated* data through 2008**. Additional NSF data from more recent survey releases, including initial data from the Business R&D and Innovation Survey (BRDIS) and recent *InfoBriefs,* are incorporated to develop estimates through 2011. The 2012 R&D spending forecast involves information from various sections of this report, including information about federal R&D budgets, corporate R&D expenditures and plans, and the general condition of the U.S. and global economies."

Battelle (2011) – *2012 Global R&D Funding Forecast*

Subnational

"To date, the National Patterns data series has **not tracked specific state and local funding to industrial R&D performers**. As a result, this funding—which could come from economic development incentives, innovation grants like the State of Ohio's Third Frontier program, etc.—is not represented."

Battelle (2011) – *2012 Global R&D Funding Forecast*

Outline

I. The panel's charge
II. Changing environment and related challenges
III. Dual conceptual frameworks
IV. Lessons learned
V. Next steps

Definitions

- **Statistical data** refers to data from a survey or administrative source used to produce statistics.
- **Statistical indicator** is a statistic, or combinations of statistics, that provide information on some aspect of the state of a system, or of its change over time (trend).
  - For example Gross Domestic Product provides information on the level of value added in the economy, and its change over time is an indicator of the economic state of the nation.

Definitions (continued)

- **Innovation -** an *innovation* is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations.
  - A new or improved **product** is implemented when it is introduced on the market.
  - New **processes**, **marketing methods** or **organizational methods** are implemented when they are brought into actual use in the firm's operations.

I. The panel's charge

POLICY RELEVANT, INTERNATIONALLY COMPARABLE STI INDICATORS

Panel Members

- Andrew Wyckoff (co-chair; OECD)
- Robert Litan (co-chair; Bloomberg Government, previously Kauffman Foundation)
- Carter Bloch (U. Aarhus)
- Nicholas Chrisman (U. Laval)
- Carl Dahlman (Georgetown U.)
- Geoff Davis (Google)
- Katharine Frase (IBM)
- Barbara Fraumeni (U. Southern Maine)
- Richard Freeman (Harvard)
- Fred Gault (UNU MERIT)
- David Goldston (Natural Resource Defense Council)
- Michael Mandel (Wharton School)
- John Rolph (U. Southern California)
- Leland Wilkinson (SYSTAT)

Statement of Task

- **Measures**
  - Revised, refocused and newly developed indicators
  - Research, development, engineering, and innovation
  - Human resources
  - Financial
- **Perspective**
  - National, international, subnational, and sectoral
  - Retrospective analysis and prospective trends
  - Policy relevant
- **Deliverables**
  - Evidence, conclusions, recommendations, priorities, and implementation strategies

Problem Set

1. What are the user demands for STI indicators?
   - Policy makers
   - Academic researchers
   - Academic administrators
   - Business managers

Problem Set (cont.)

2. What is the best framework for setting priorities for developing indicators representing the system of STI activities?
   - NCSES's mandate as a statistical agency (external)
   - Data curation⇔ analytical statistics (internal)
3. **What are the Key National STI Indicators that NCSES should produce in the next 5-10 years?**
   - **Policy relevance**
   - **International comparability**

II. Changing environment and related challenges

EXTERNAL AND INTERNAL FORCES

Changing Environment

- Rapid technological changes that reshape the U.S. and global economies
- Globalization of technology and other economic activities
- Evolving methods of gathering data used to produce measures of capacities and trends in the global STI system
- Augmented NCSES mission

National Center for Science and Engineering Statistics (NCSES)

- Federal clearinghouse for the collection, interpretation, analysis, and dissemination of objective data on science, engineering, technology, and research and development.
- U.S. and international statistics relevant and useful to practitioners, researchers, policymakers, and the public, particularly:
  - research and development trends;
  - the science and engineering workforce;
  - U.S. competitiveness in science, engineering, technology, and research and development; and
  - the condition and progress of United States STEM education.
- Issue regular and special statistical reports on topics related to the national and international science and engineering enterprise.

Additional Challenges

- Shift in the nature of innovation to a more open and collaborative model
- Timeliness of the data/statistics – timeliness is another dimension of quality
- Local geography and confidentiality issues

III. Dual conceptual frameworks

POLICY MATTERS & ACCOUNTING

Frameworks

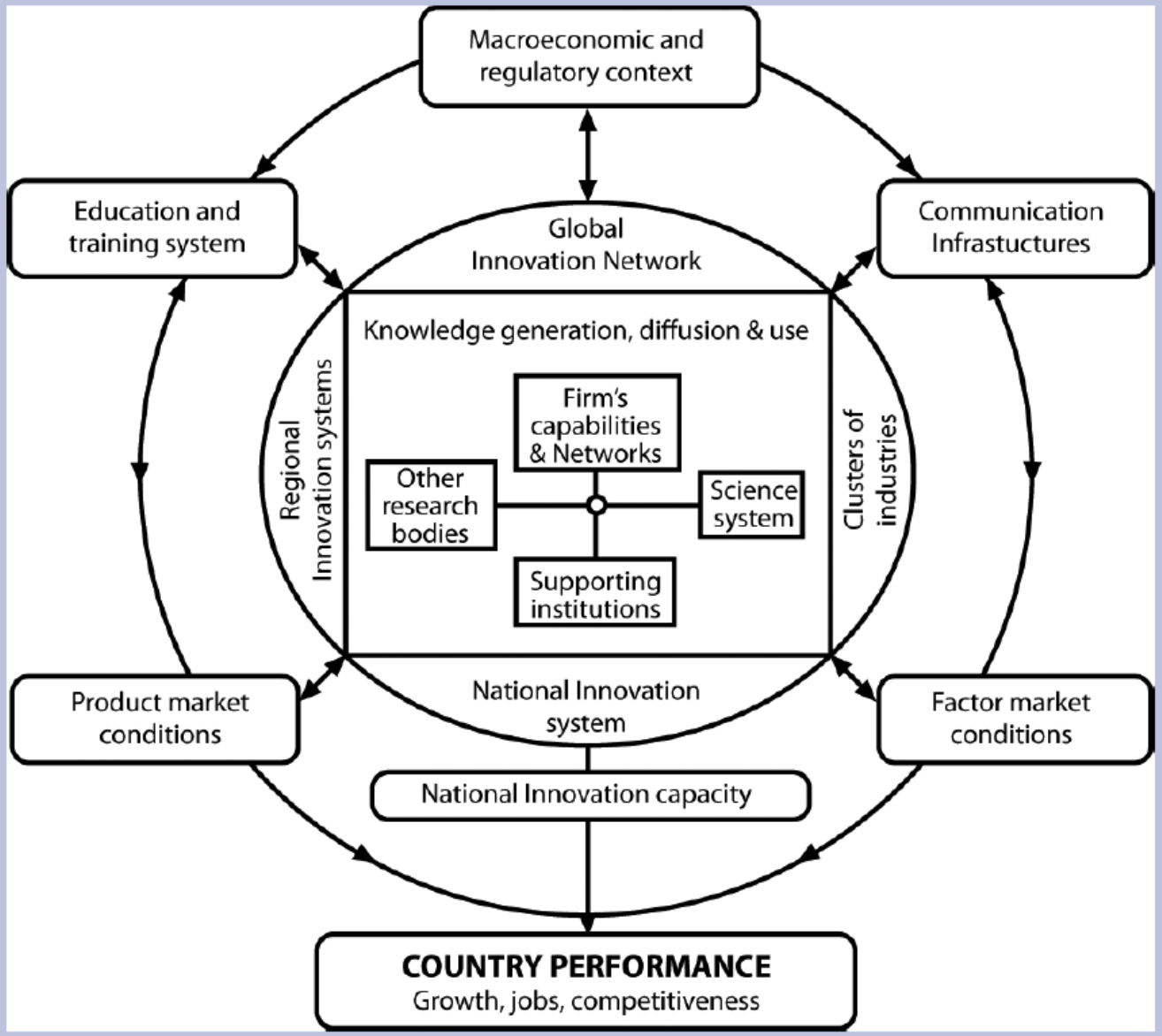
- Policy metrics
  - User-driven queries
- Growth/production function models
  - Inputs, outputs, outcomes, impacts
- National income accounting
  - Expenditures
- Evaluation
  - Project level activities

Policy-driven STI Indicators

Policy Issue	Indicator
What is the contribution of science, technology, and innovation (STI) activity to productivity, employment and growth?	Jobs associated with firms that are highly innovative; Sales from innovative products; Contribution of intangibles to GDP growth; Multifactor productivity decomposition.
What is the relative importance of technological innovation versus non-technological innovation for economic growth?	Contribution of intangibles to growth; Occupation composition of fast growing, innovative firms.
In which industries is there rapid growth and what are the characteristics of the firms?	Business demographics: births and deaths, age, lifetimes, growth rates measured by employment or revenue.
Does the location of R&D affect the creation of jobs in the US?	.....

- Knowledge generation, stocks & flows
- Innovation
- Human Capital
- Uses of tech. & practices
- Role of select actors
- Linkages btwn. actors
- Globalization
- Finance

The Many Interdependencies of a National Innovation System



Adapted from: OECD. *Managing National Innovation Systems*, 1999.

Source: NSB, *A Companion to SEI, 2012.*

IV. Lessons learned

SOME KEY FINDINGS

Interim Report Recommendations

- REC. 1: NCSES should explore methods of using existing longitudinal data on **labor force mobility** related to STI activities, including data that are collected by NCSES surveys *(and possibly the ACS).*
- Rec. 2: NCSES should develop new, internationally comparable indicators on **innovation**, based on data from *(a restructured)* BRDIS.

Interim Report Rec. (cont.)

- Rec. 3: NCSES should develop indicators of **firm dynamism**, by matching BRDIS data to data from ongoing surveys at the U.S. Census Bureau and the Bureau of Labor Statistics.
  - Enables the measures of activities by high-growth firms, and on births and deaths of businesses linked to innovation outputs.
  - Should be established by geographic and industry sectors and by business size and business age ("gazelles").

Interim Report Rec. (cont.)

- Rec. 4: NCSES should more fully use data from BRDIS to provide indicators on **payments and receipts for R&D services** between the United States and other countries *(continuing to collaborate with BEA in this area)*.
- Rec. 5: NCSES should host working groups to further develop indicators on **subnational STI activities**.
  - Focus on data reliability *(and confidentiality)* at subnational scales.

Interim Report Rec. (cont.)

- Rec. 6: NCSES should fund **exploratory activities** on frontier data extraction and development methods. Including:
  - research funding or prize competitions;
  - pilot programs or experiments; and
  - workshops of experts on multimodal data development *(including survey, administrative records and unstructured data)*.

Lessons Learned

- Here are some indicators that users suggested for further developed to **improve** policy relevance and international comparability:
  - Commercialization of research results (innovation)
    - Technology transfer
    - Interactions between actors
    - Turnover of innovative products (in relation to TFP)
  - Entrepreneurship, including institutional tools (e.g., R&D tax incentives)
  - Technology balance of payments (particular focus on MNE R&D)
  - Demographic breakdowns (including gender)
  - Career of doctorate holders (longitudinal)
  - Public attitude towards science

Lessons Learned

- Here are some **new** areas that users suggested for development of indicators:
  - Knowledge diffusion
  - Intangibles—including capitalization of R&D; design; software
  - Subnational R&D beyond states—product/enterprise breakdowns
  - User generated innovation
  - Non-R&D innovation (including organizational, managerial, etc.)
  - Public sector innovation
  - STEMM training
  - Economic and social impacts of STI
  - Institutions, laws and standards governing STI

Lessons Learned

- Need to develop the methodological underpinnings for the use of **"Big Data"** for indicators development.
  - Requires research on applicable statistical methodology

V. Next Steps

DELIVERABLES

Next Steps

- Final report preparation and review
  - There are several areas where NCSES is to be congratulated on data extraction, analysis and dissemination efforts.
  - There is no magic bullet that we discovered during our investigation to suggest that NCSES missed any major area of measurement that would be easily remedied at minimal cost.
- Final report release date expected beginning of 2013

Thank you!

kfealing@nas.edu

sti-indicators.ning.com

Using STAR METRICS TO IMPROVE NATIONAL PATTERNS

Workshop on National Patterns

CNSTAT Staff

September 6-7, 2012

Using STAR METRICS Jointly with Census Information

For a non-representative fraction of the census and survey respondents that support *National Patterns* reports, STAR METRICS values exist that may or may not be subject to less measurement error than the current responses to these censuses and surveys. ASSUMING THAT THEY ARE OF HIGHER QUALITY --- and NCSES has Chris Pece looking into that -- it would be useful to examine whether and how to utilize this information to improve the responses used in *National Patterns*. While such sources of information are very **unlikely** to exist anytime soon for BRDIS, this possibility is worth exploring for both federal fund grants for R&D, for most Academic R&D grants, and for Non-profit R&D.

STAR METRICS AND CENSUSES IN NATIONAL PATTERNS

Reasons to think census responses are preferable --- just the fact that any administrative records system can contain errors.

Reasons to think STAR METRICS might be preferable --- going to the actual documentation might avoid some response errors.

People here can probably do a better job of suggesting why either source might have advantages or disadvantages.

For current purposes, we assume that the administrative sources are of superior quality in comparison to the information received from respondents, though that may certainly turn out not to be true.

STAR METRICS AND CENSUSES IN NATIONAL PATTERNS

Editing. One obvious way this information could be used is through the editing of responses. Assume that a census response and a STAR METRICS response differ by more then *p*%, or that one was zero and the other non-zero. It would seem beneficial in such cases to contact the respondent to ask for a clarification.

A problem with this is that it could be labor intensive.

More importantly, as discussed soon, the type of differences that arise would need to be better understood to develop an effective editing routine – otherwise you are unnecessarily bothering your respondents and wasting staff time.

STAR METRICS AND CENSUSES IN NATIONAL PATTERNS

Imputation. In situations for which a clarification cannot be obtained, what could be done?

A difficulty with any proposed imputation model for census responses using STAR METRICS responses is that it explicitly or implicitly assumes a model for the correct responses --- often that the errors that appear in the censuses that don't appear in STAR METRICS are reasonably well-behaved (e.g., normal-shaped errors on a log scale).

STAR METRICS AND CENSUSES IN NATIONAL PATTERNS

But the errors might be quite different – for example they might include the error of changing awards to zero with some probability. Diagnosing such errors, and providing imputations for them when necessary, would require a very different model than that posited above. Therefore, as a first step towards developing such editing or imputation procedures, NCSES is undertaking an exploratory data analysis of the differences between STAR METRICS and census values.

(Precisely this type of research problem is now being carried out at the Census Bureau for a wide range of survey and census responses because of the increased availability of administrative records in a variety of contexts.)

EDA is Necessary First Step

And once this research has been completed, it may still be very challenging to design an effective editing or imputation routine.

Blue-Sky Notions

Further Use. Situations where you are missing either the census response or the STAR METRICS response.

Two possibilities exist. First, for values for which there have been no census responses, specifically for estimating the components of Non-Profit R&D discussed earlier, STAR METRICS responses --- assuming that non-profit R&D providers could be added to STAR METRICS --- could be used as surrogates for a survey that has not been collected in many years.

Absent a new census, one would be concerned that such data were not validated, but given the lack of a recent survey, one might still decide to go ahead. Coverage is obviously also a worry today.

Where data had been collected recently for census/STAR METRICS pairs, such as for HERD, using data for other institutions one might be able to develop a model-based imputation for missing census responses.

Blue-Sky Notions

For the reverse situation, where one had a census response but no STAR METRICS response. Assuming:

1) STAR METRICS data were shown to be more reliable AND
2) One discovered that STAR METRICS values had a stable relationship to the census values

then one might use such a model to adjust the census response even for STAR METRICS values that were not available, using the full set of STAR METRICS and census values.

In effect, one is imputing the STAR METRICS value using the census value and census and STAR METRICS paired responses for others.

Problems:

1) There would always be concerns about the quality of the model and validation might be hard.
2) In addition, since STAR METRICS is a voluntary program, one should also be concerned that the respondents not volunteering to join would be different than the cases for which the model was developed.

Any other ideas?

STAR METRICS and VIVO

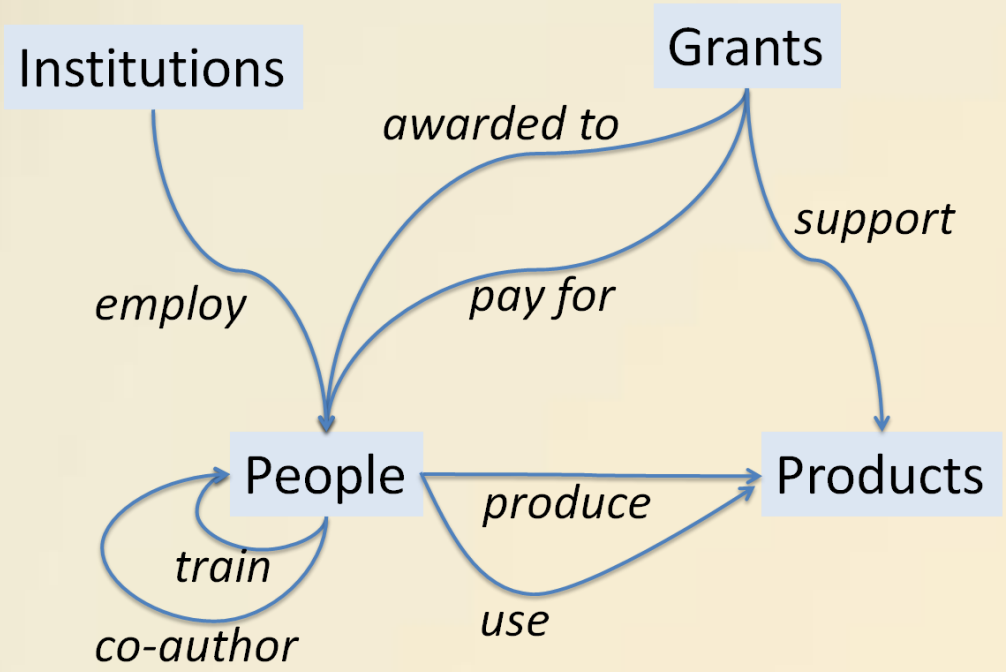
Common Data and Open Platforms for Transparent Science

John L. King
USDA Economic Research Service
USDA Office of the Chief Scientist

National Academy of Sciences
Committee on National Statistics
"Future Directions for the *NSF National Patterns of Research and Development Resources*"
September 6-7, 2012

How do you make science transparent for empirical analysis?

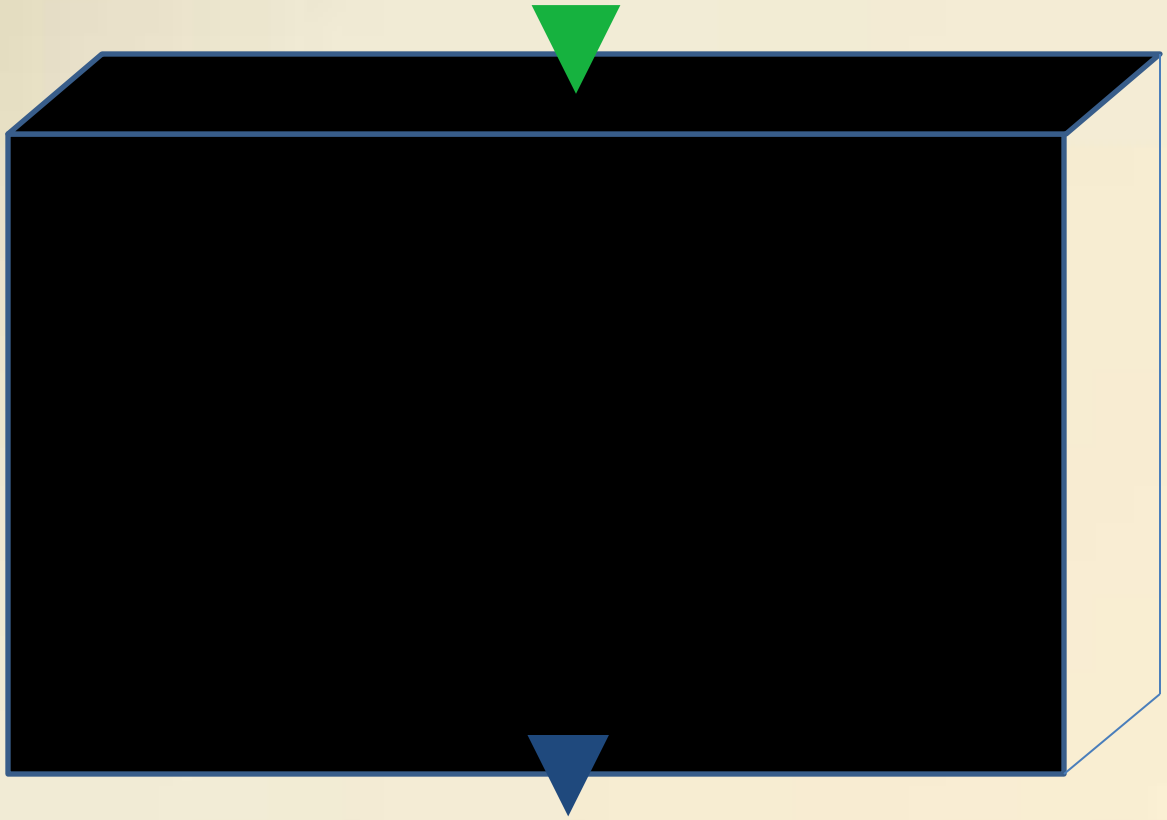
Approach #1: Conceptual framework



Source: Ian Foster

Approach #2:
Less complex framework

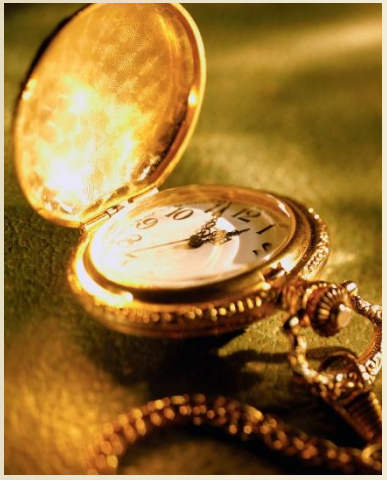
$$$



Research

Common data standards and open platforms to improve understanding of science

- **Star Metrics** establishes data standards to enable empirical studies of science impacts
- **VIVO** provides an open platform that helps different science institutions meet that standard



Talk today will describe these two approaches, and how they complement each other

What is STAR METRICS?

Science and
Technology for
America's
Reinvestment:

Measuring the
Effec**T**s of
Research on
Innovation,
Competitiveness
and
Science

- A partnership among Federal agencies and U.S. research institutions
- A platform for data collection and analysis of R&D investments
- An approach to data collection that relies on automated harvesting from systems of record
- New applications and tools to meet research needs and policy requirements

What is STAR METRICS?

Science and
Technology for
America's
Reinvestment:

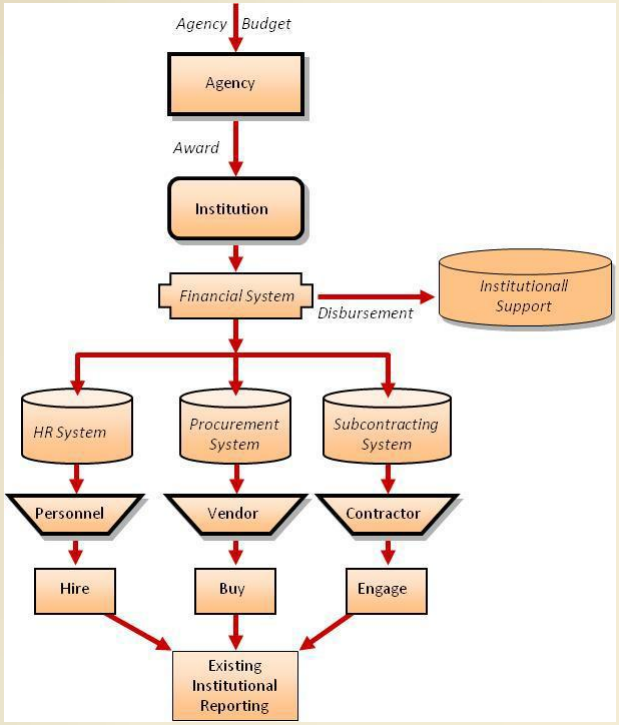
Measuring the
Effec**T**s of
Research on
Innovation,
Competitiveness
and
Science

- A partnership among Federal agencies and U.S. research institutions
- A platform for data collection and analysis of R&D investments
- An approach to data collection that relies on automated harvesting from systems of record
- New applications and tools to meet research needs and policy requirements

Star Metrics provides a common format for administrative data from different sources

Goal 1

Systematic collection of important data from authoritative sources



Goal 2

Minimizing adminstrative burden

Star Metrics provides a common format for administrative data from different sources

Federal Award

- **Principal Investigator**
- **Program Information**
- **Abstract/Proposal**
- **Obligated Funds**

Star Metrics provides a common format for administrative data from different sources

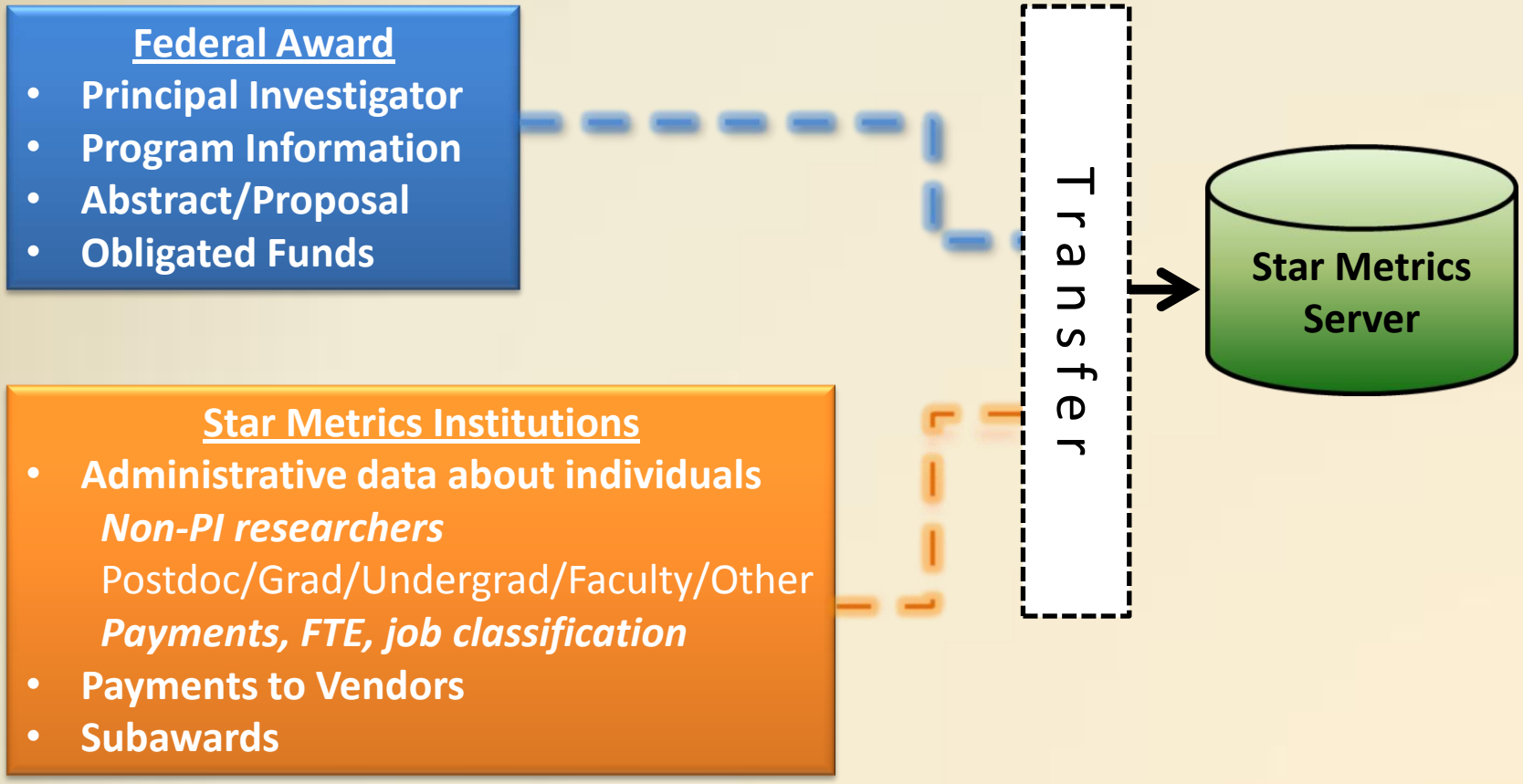
Federal Award

- **Principal Investigator**
- **Program Information**
- **Abstract/Proposal**
- **Obligated Funds**

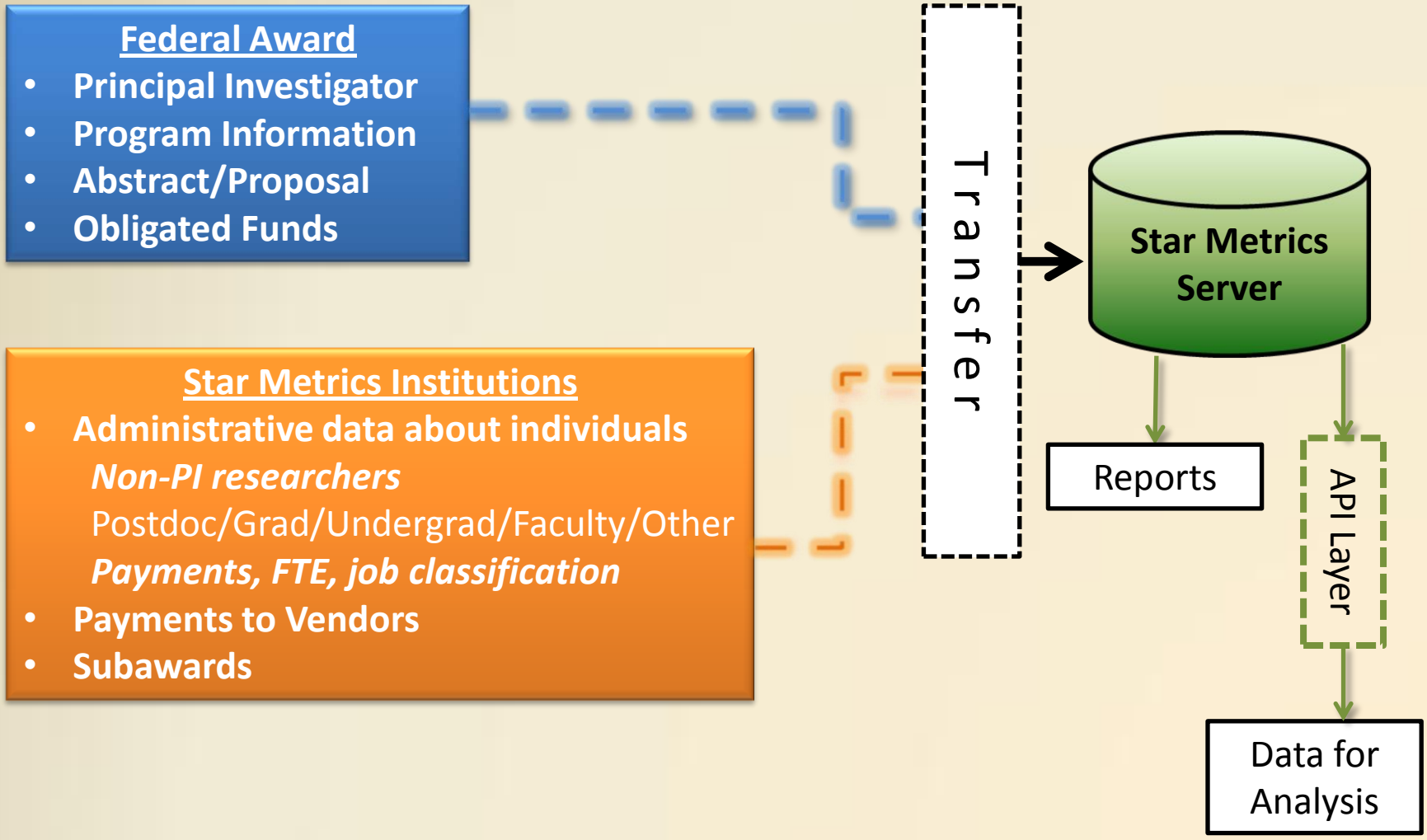
Star Metrics Institutions

- **Administrative data about individuals**
  - ***Non-PI researchers***
    Postdoc/Grad/Undergrad/Faculty/Other
  - ***Payments, FTE, job classification***
- **Payments to Vendors**
- **Subawards**

Star Metrics provides a common format for administrative data from different sources



Star Metrics provides a common format for administrative data from different sources



Workforce Reports

Number of individuals supported by funding sources

2011_Q2	Number of Individuals
All Funding (Federal & Non-Federal)	8,325
All Non-Federal	3,400
All Federal Funding	5,812
All Federal Science Funding	4,250

Rolling 12 months thru 6/2011	Number of Individuals
All Funding (Federal & Non-Federal)	11,668
All Non-Federal	5,563
All Federal Funding	8,043
All Federal Science Funding	6,053

Number of vendor jobs supported by funding sources

2011_Q2	Vendor Jobs
Non-Federal	47.3
Non-Federal Funding	47.3
Federal	124.0
Federal Science Funding	90.8
Federal Non-Science Funding	33.1
Grand Total	171.3

Rolling 12 months thru 6/2011	Vendor Jobs
Non-Federal	50.8
Non-Federal Funding	50.8
	102.5
	81.6
	20.9
	153.3

Federal Funding for Science and Jobs
STAR METRICS Report
for
Star State University
2011_Q2

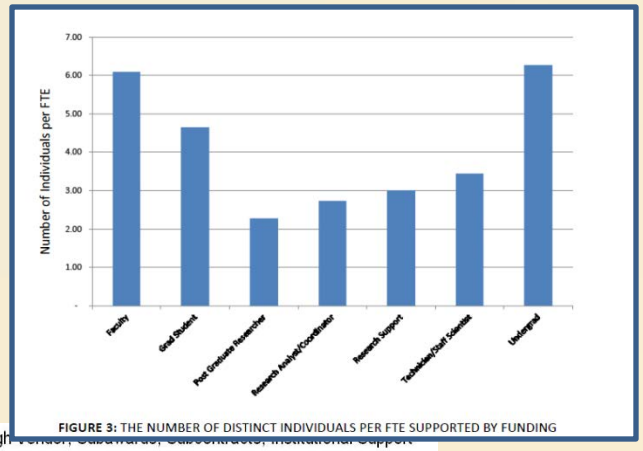


FIGURE 3: THE NUMBER OF DISTINCT INDIVIDUALS PER FTE SUPPORTED BY FUNDING

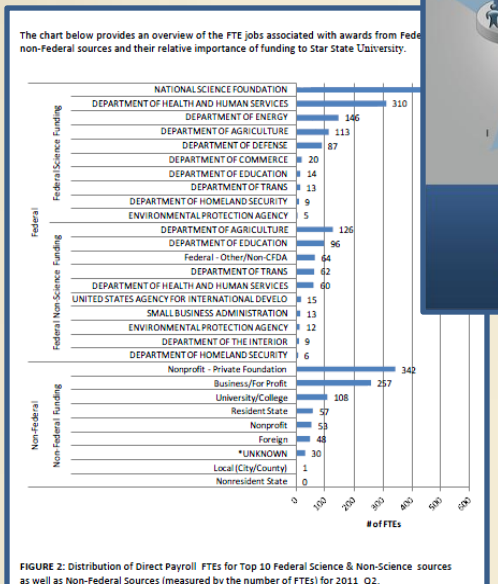
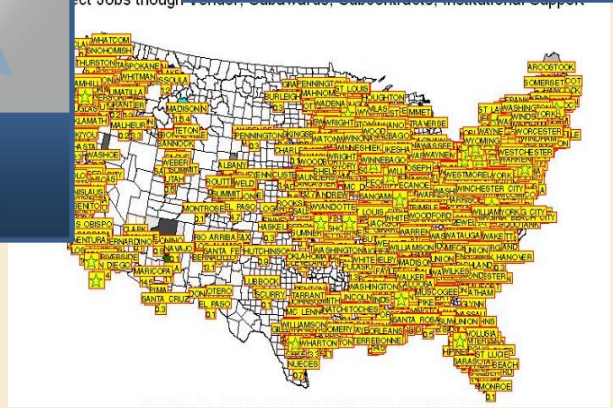


FIGURE 2: Distribution of Direct Payroll FTEs for Top 10 Federal Science & Non-Science sources as well as Non-Federal Sources (measured by the number of FTEs) for 2011_Q2.



Level One Core Data Elements

Information on Awards	• Unique Award Number • Recipient Account Number • Overhead Charged
Information on Individuals	• Unique Award Number • Recipient Account Number • Deidentified Employee ID Number • Occupational Classification • FTE Status • Proportion of Earnings Allocated to Award
Information on Indirect Costs	• Overhead Salary+Fringe to Total Ratio
Payments to Vendors Payments to Subawards	• Unique Award Number • Recipient Account Number • Vendor/Subaward DUNS Number • Vendor/Subaward Amount

Matching/mapping process captures quarterly updates of Level One data

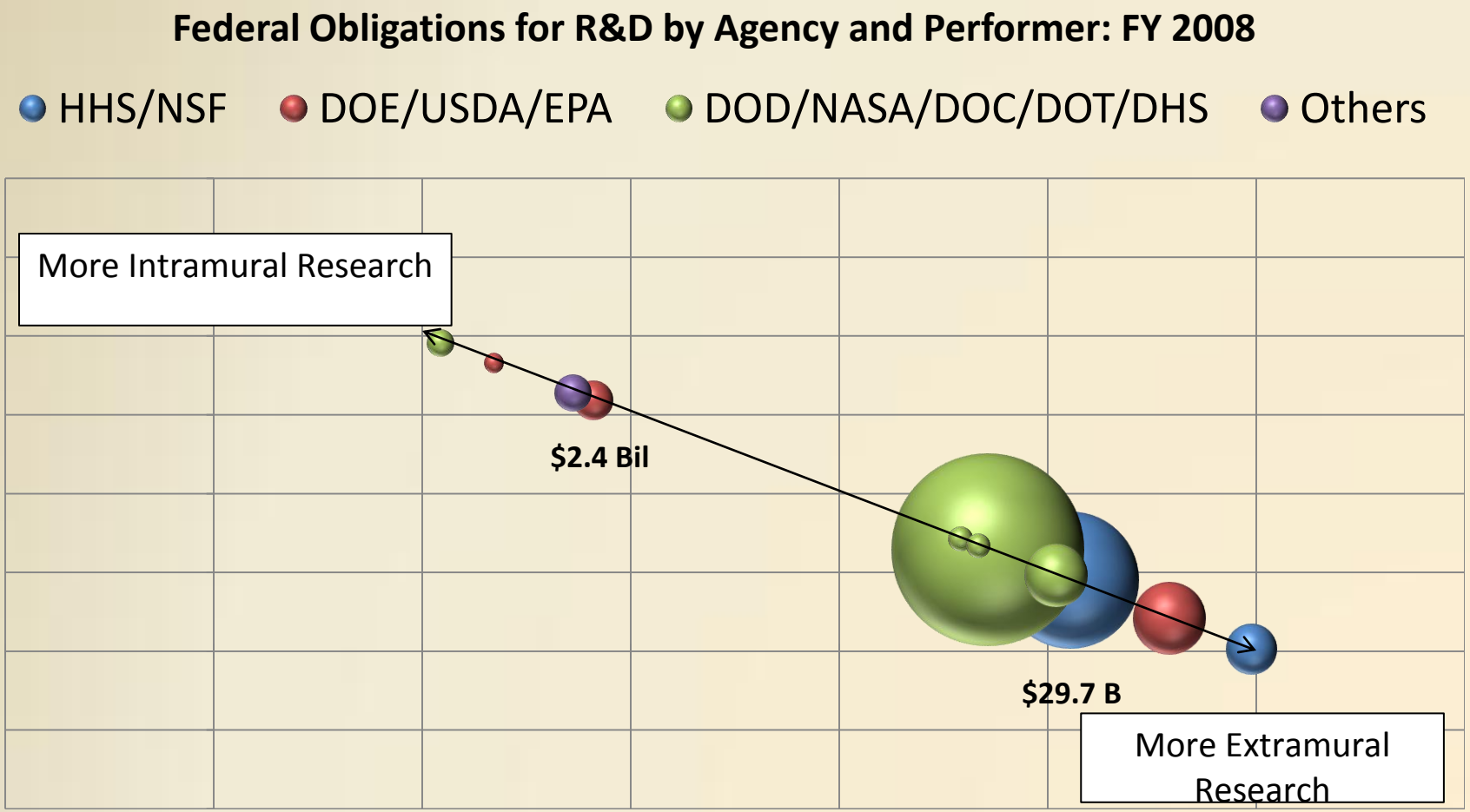
- Alphanumeric matching of grant numbers
- Institutional job series map to SM classes
  - Faculty, Postgrad, Graduate, Undergraduate, Research Analyst, Research Support, Technician
- CFDA codes map awards to
  - Funding agency
  - Science/non-science funding
- Same accounting framework possible for non-Federal research support
  - Corporate, endowment, foundation

Star Metrics Partner Institutions

Star Metrics Partner

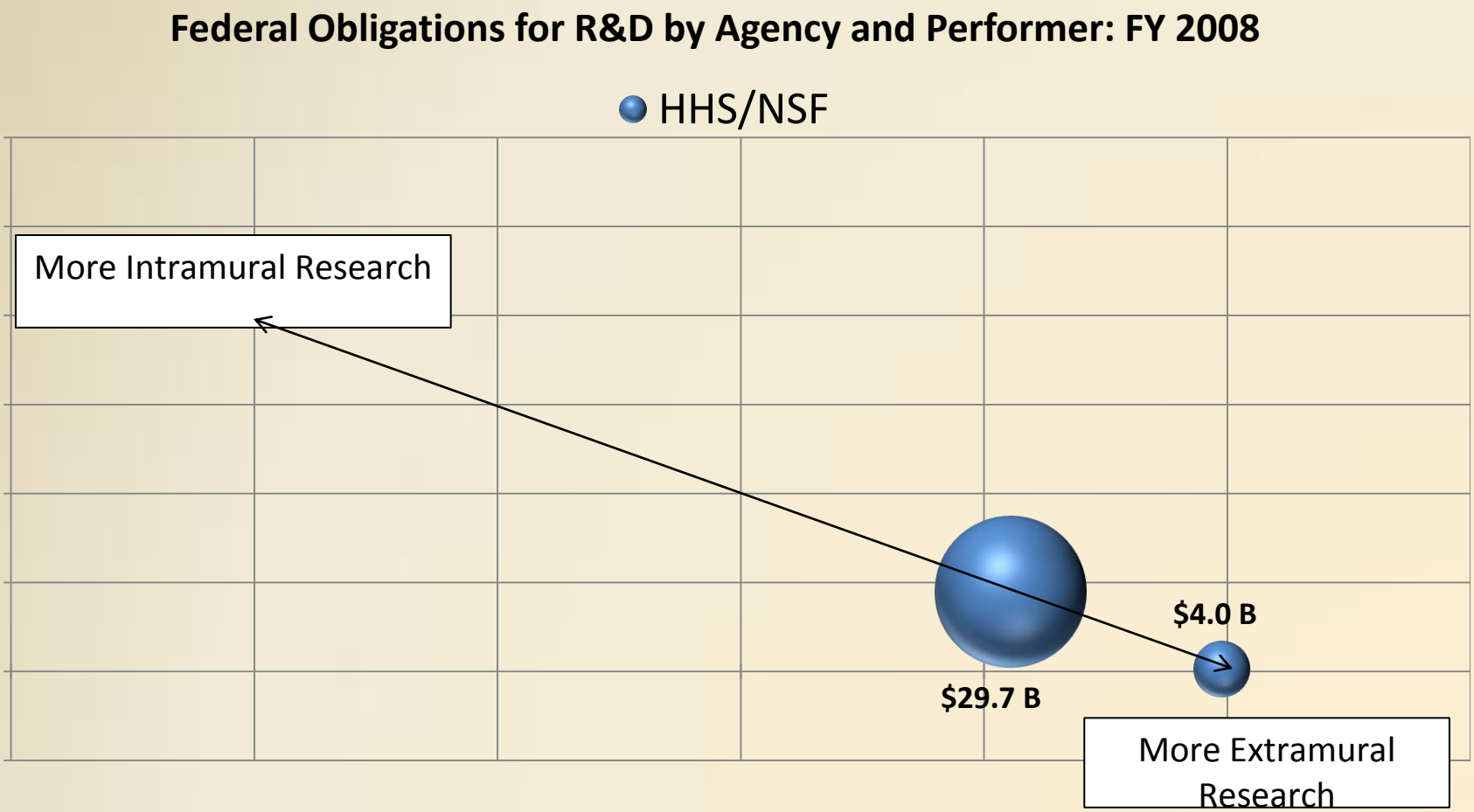
Star Metrics Partner/Land Grant University

The Intramural – Extramural Continuum of Federal R&D



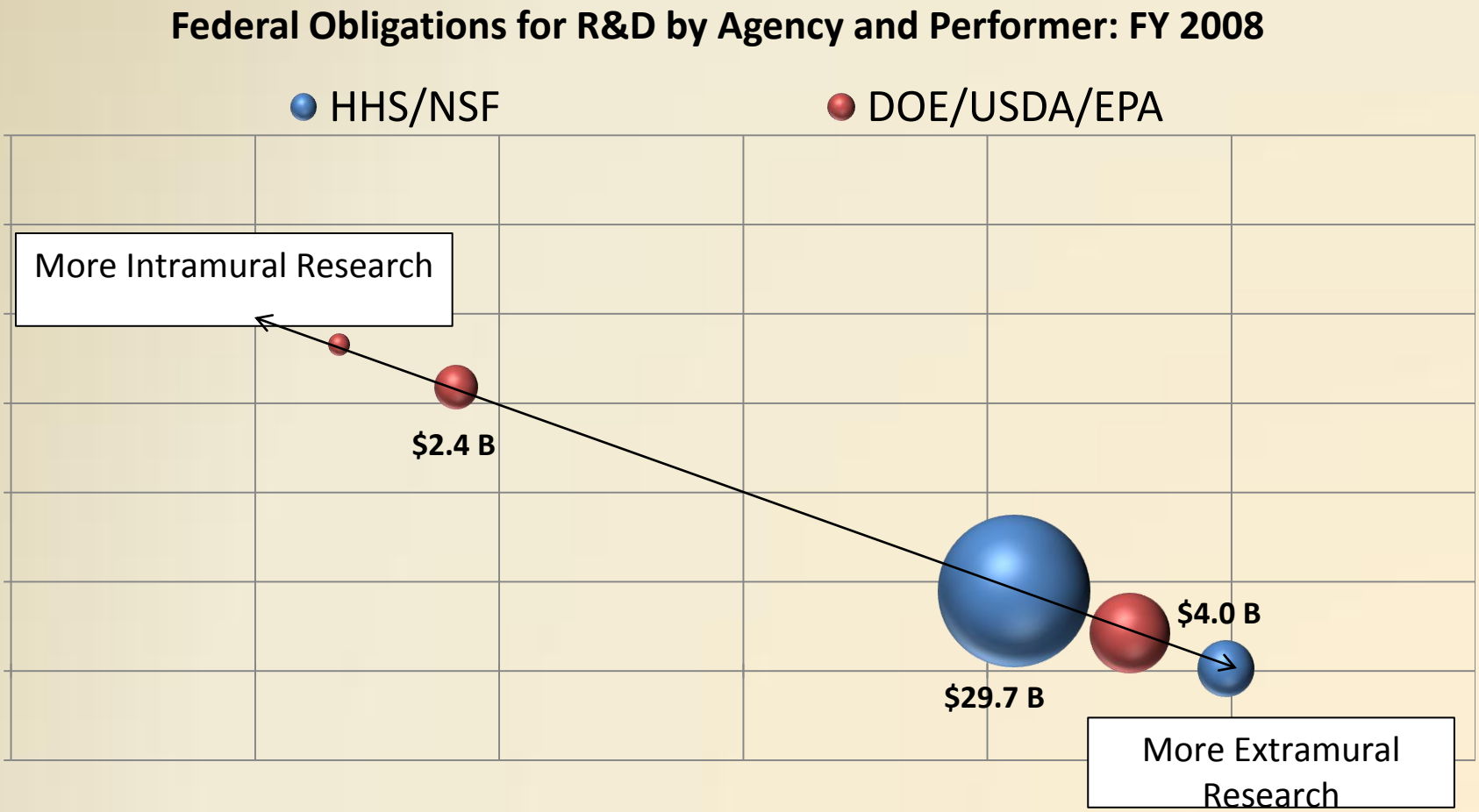
NSF Science and Engineering Indicators 2010, Table 4-9. (Total: $115 Billion)

New Star Metrics Partners have Different Research Portfolios



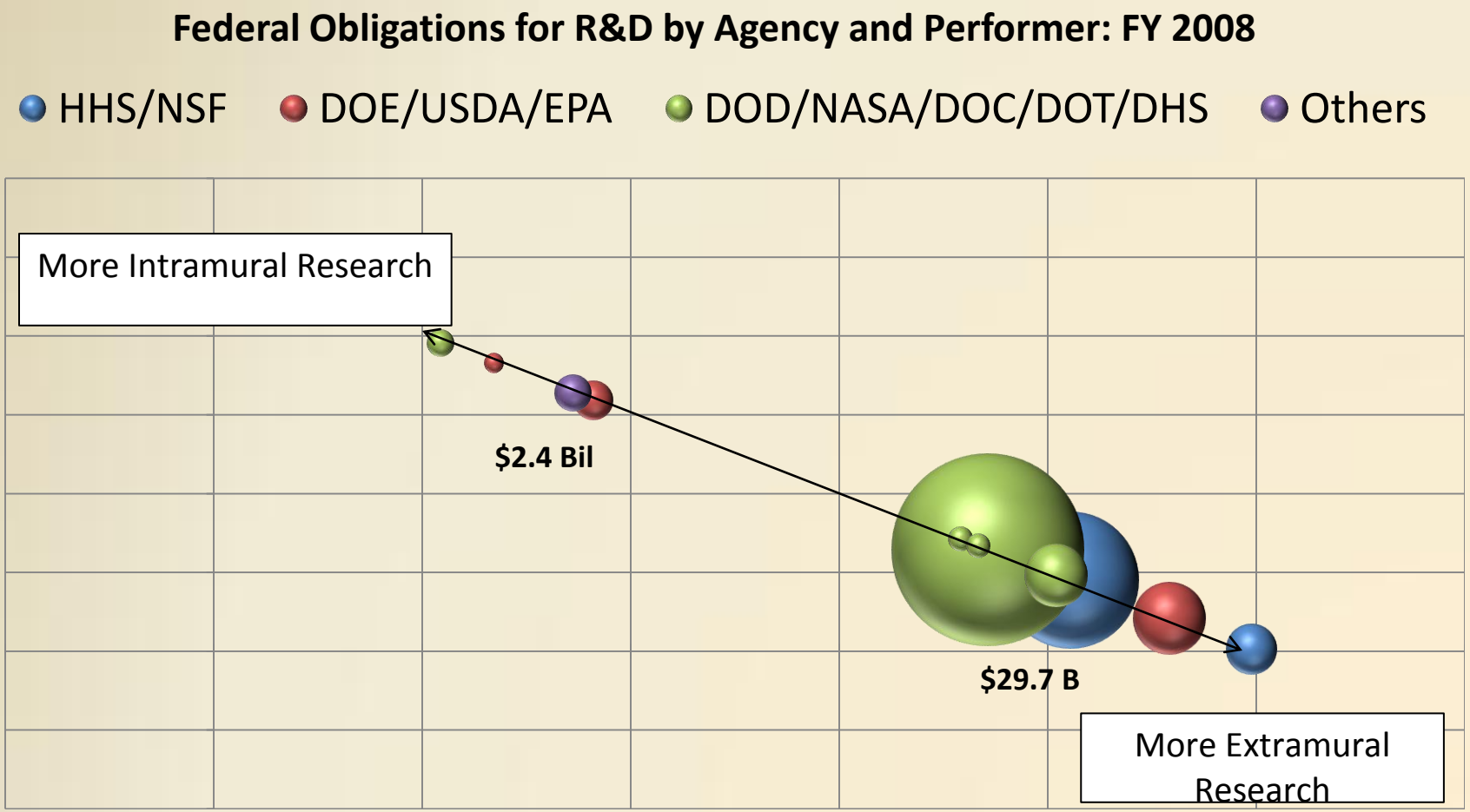
NSF Science and Engineering Indicators 2010, Table 4-9. (Total: $115 Billion)

New Star Metrics Partners have Different Research Portfolios



NSF Science and Engineering Indicators 2010, Table 4-9. (Total: $115 Billion)

New Star Metrics Partners have Different Research Portfolios



NSF Science and Engineering Indicators 2010, Table 4-9. (Total: $115 Billion)

Challenge: Make different research models transparent and comparable

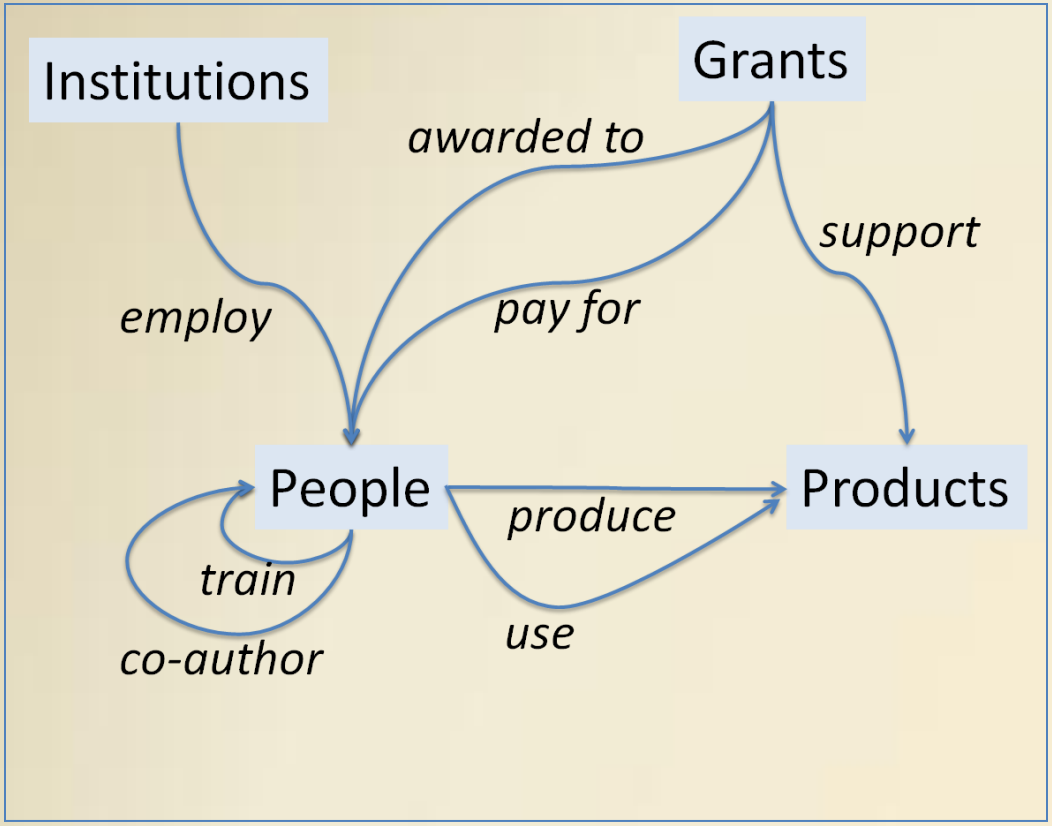
- Extramural and intramural research report different information (e.g. no intramural RPPR)
- Success will help us address our tough questions.
  - Does intramural research engage different topics of inquiry? (e.g. natural resources)
  - How to compare R&D portfolios across programs, agencies, or departments?
  - How does scientific discovery differ across settings?
  - What incentives and rewards do scientists encounter?

USDA adoption of the VIVO platform designed to meet this challenge, with additional benefits

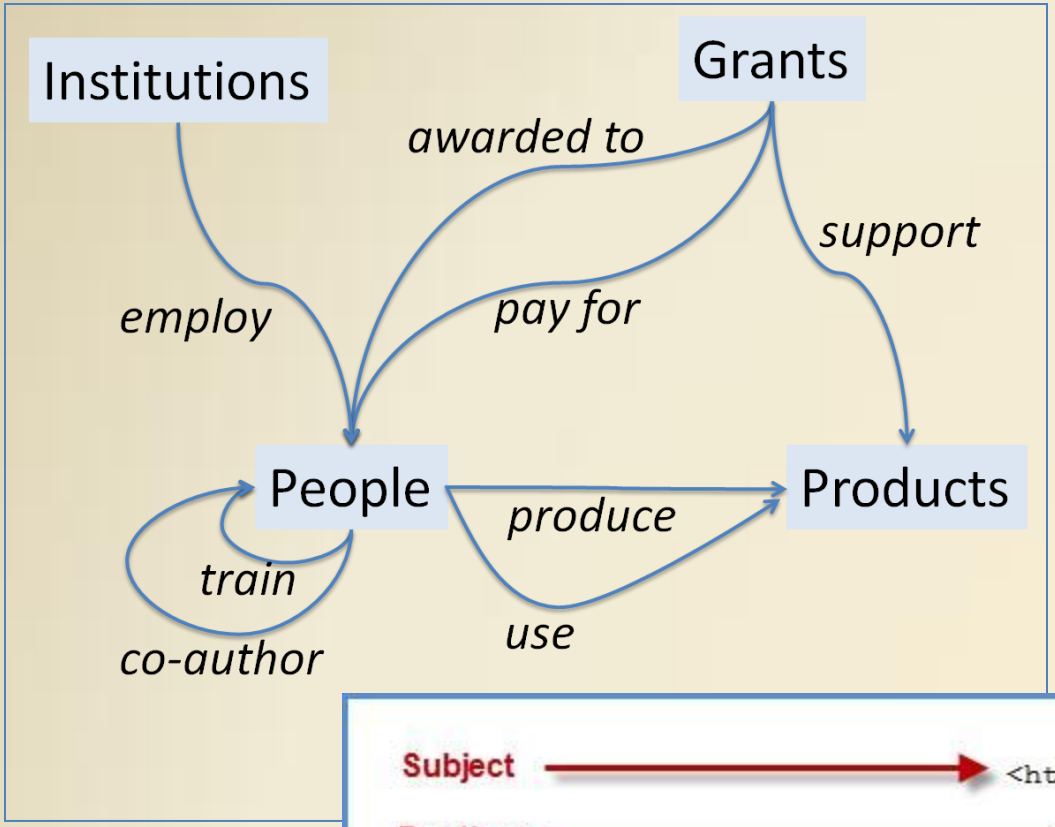
USDA/VIVO provides tools for research discovery

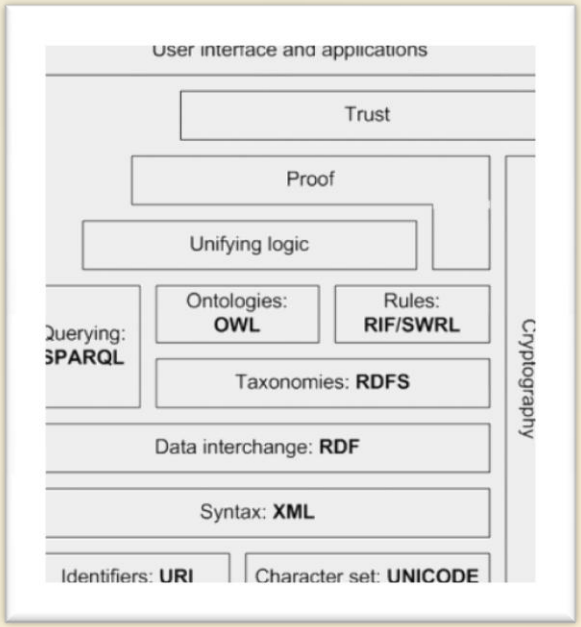
VIVO is a *semantic web* approach to describing complex, related data in a simple, flexible format

Source: Ian Foster



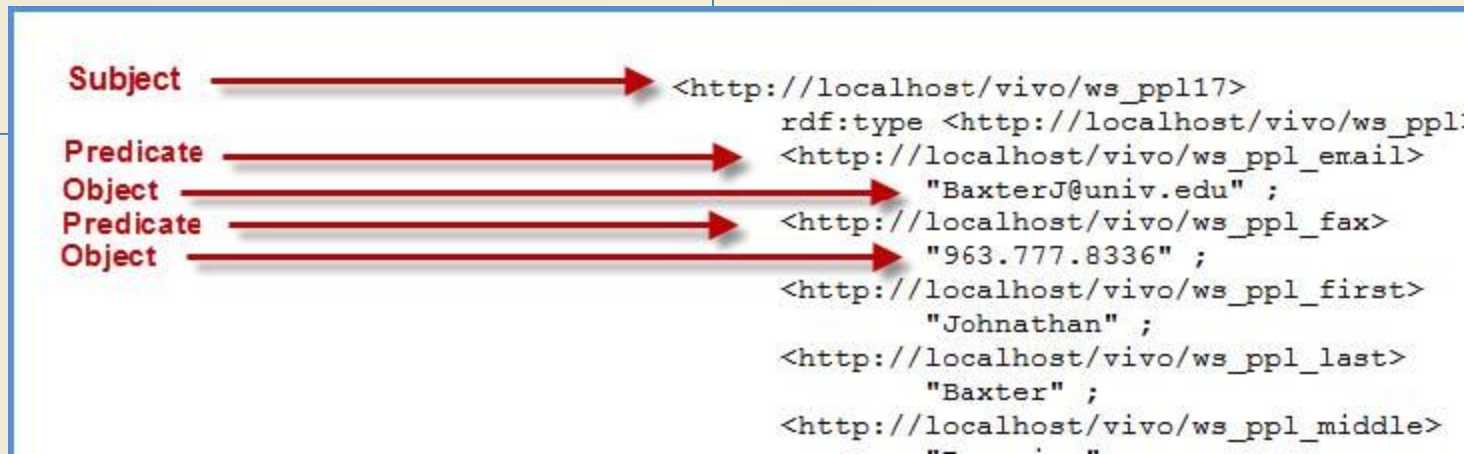
Source: Ian Foster





Source: W3C

Source: sourceforge.net



USDA VIVO implementation has "ordinary" and "extraordinary" benefits

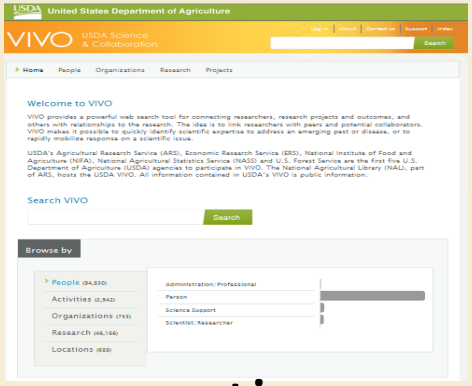
"Ordinary" benefits of VIVO

- Connecting USDA researchers for collaboration
- Public-facing expertise locator for the full scope of USDA research
- Connection to other VIVO institutions via ontology

"Extraordinary" benefits of VIVO

- Uniform data structure across USDA science agencies
- Source of clean data to document outcomes of intramural science
- Sharing USDA/VIVO with other Federal R&D agencies

USDA/VIVO launch planned for 2012



Intramural Research

– 5,000+ individuals

– Publications, projects, and patents

– 2008 to present, deepening coverage over time

Extramural Research

– Ingest complete

– 90,000+ individuals, pending disambiguation

Interface with other Federal Star Metrics agencies

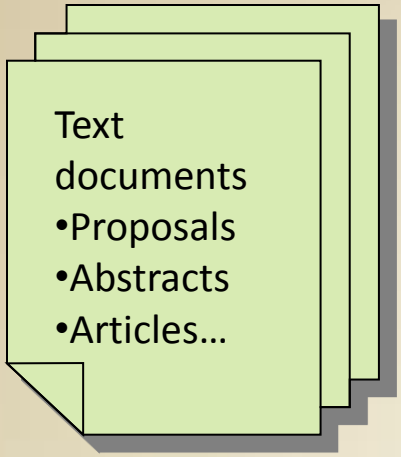
–Through VIVO directly (EPA), or VIVO-enabled data

Applications for Science Agencies

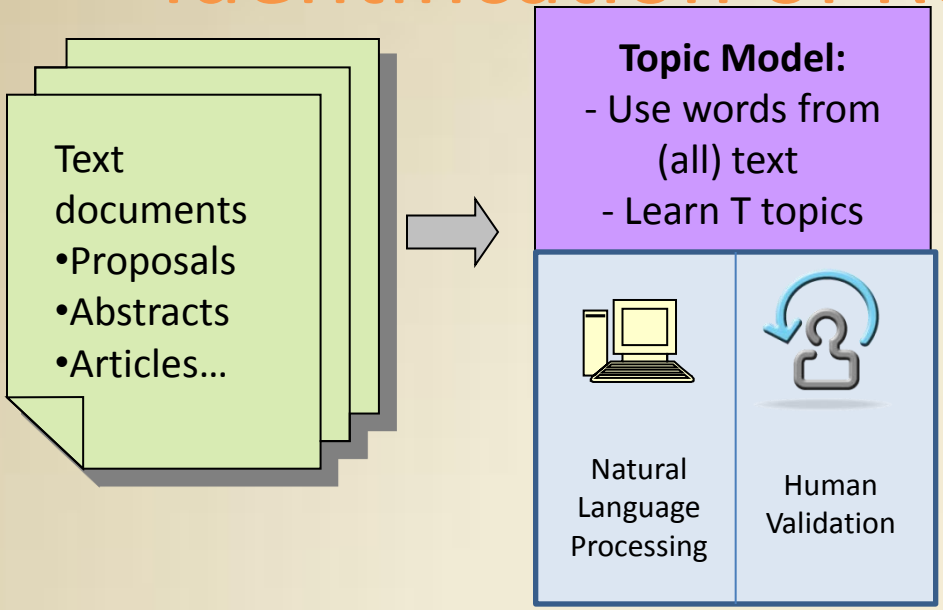
- Data to support management analysis and decision-making
  - Priority setting and budgeting
- Integrated performance measures
- Hypothesis-driven analysis

Topic Modeling: Automatic Identification of Related Research

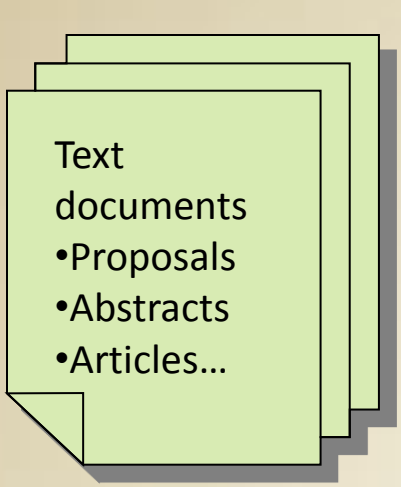
Topic Modeling: Automatic Identification of Related Research

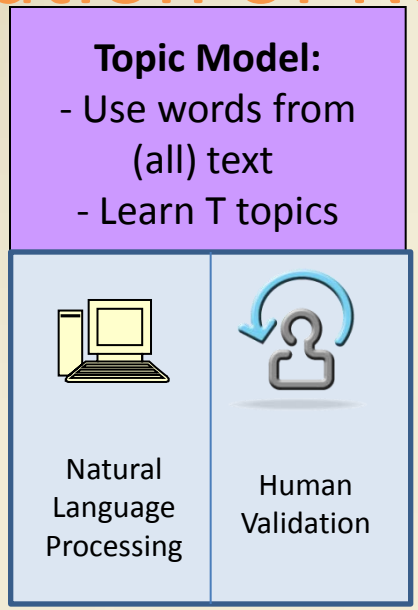


Topic Modeling: Automatic Identification of Related Research



Topic Modeling: Automatic Identification of Related Research





Automatically learned topics (e.g.):

…

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 …

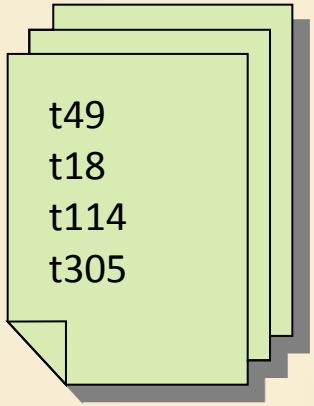
t11. speech sound acoustic recognition human …

t12. museum public exhibit center informal outreach

t13. particles particle colloidal granular material …

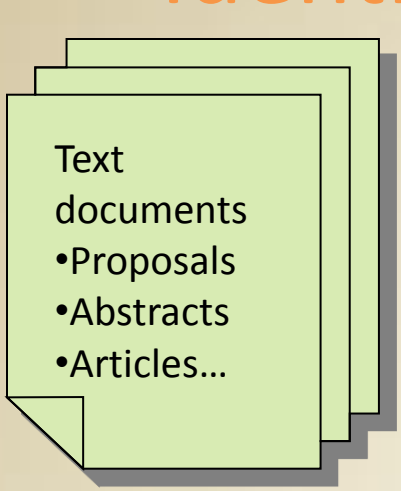
t14. ocean marine scientist oceanography …

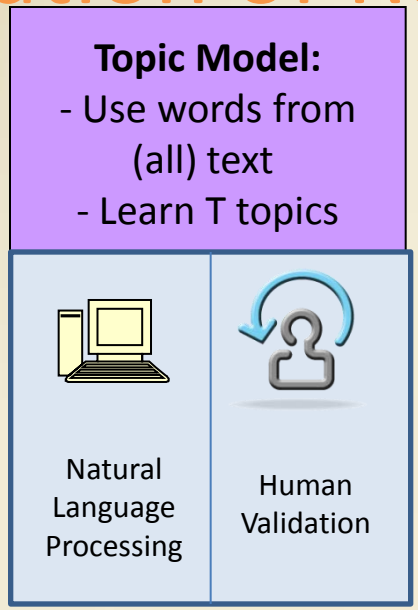
…



Topic tags for each and every document

Topic Modeling: Automatic Identification of Related Research





Automatically learned topics (e.g.):

…

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

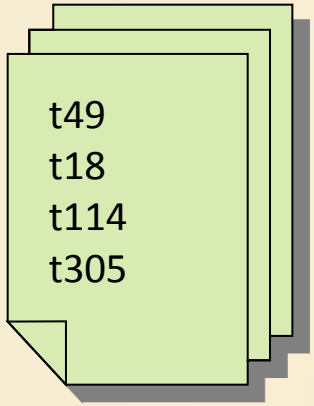
t13. particles particle colloidal granular material …

t14. ocean marine scientist oceanography …

…

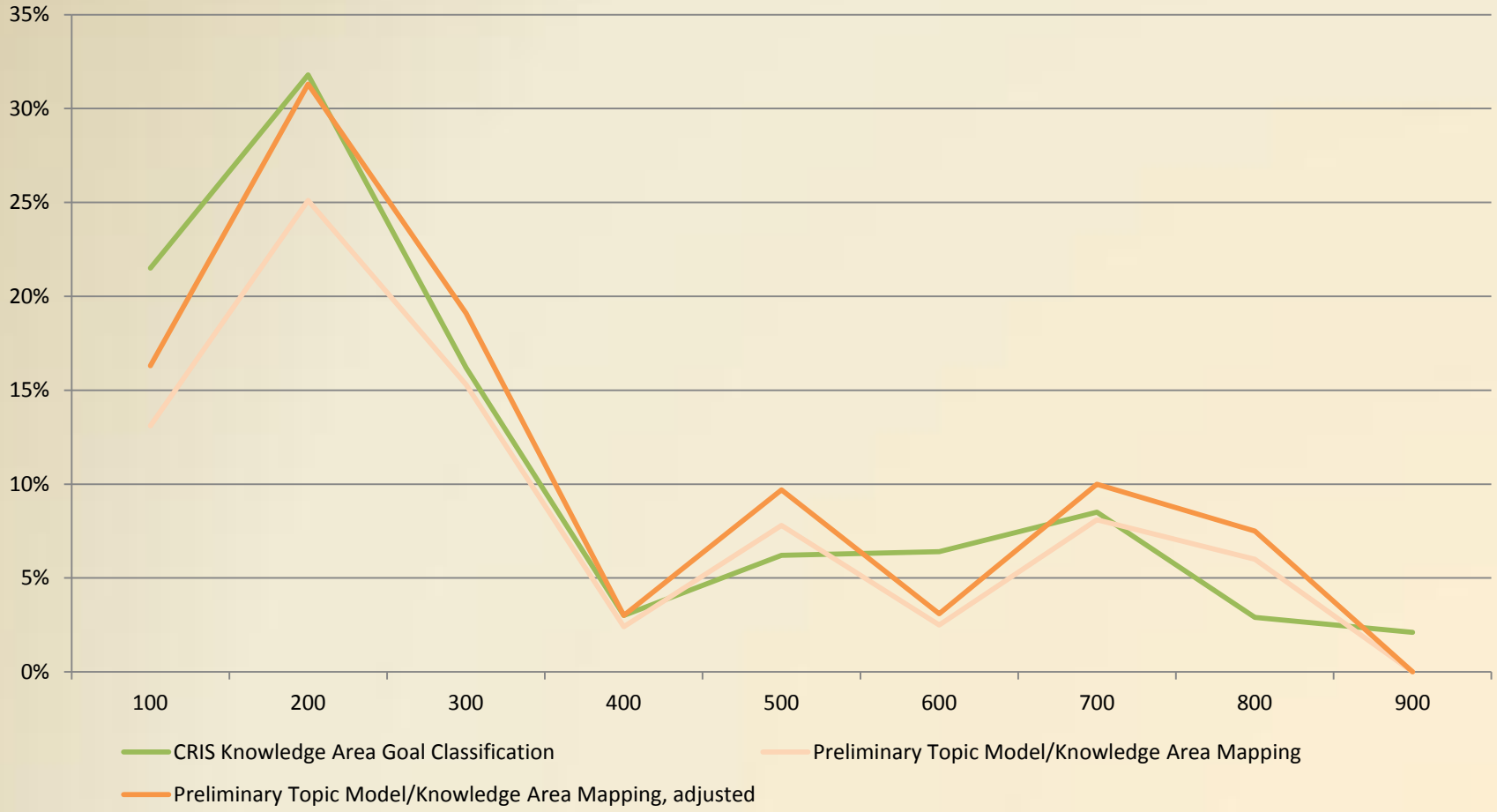
USDA's approach:

1. **"Learn" from projects**
2. **Apply to publications**

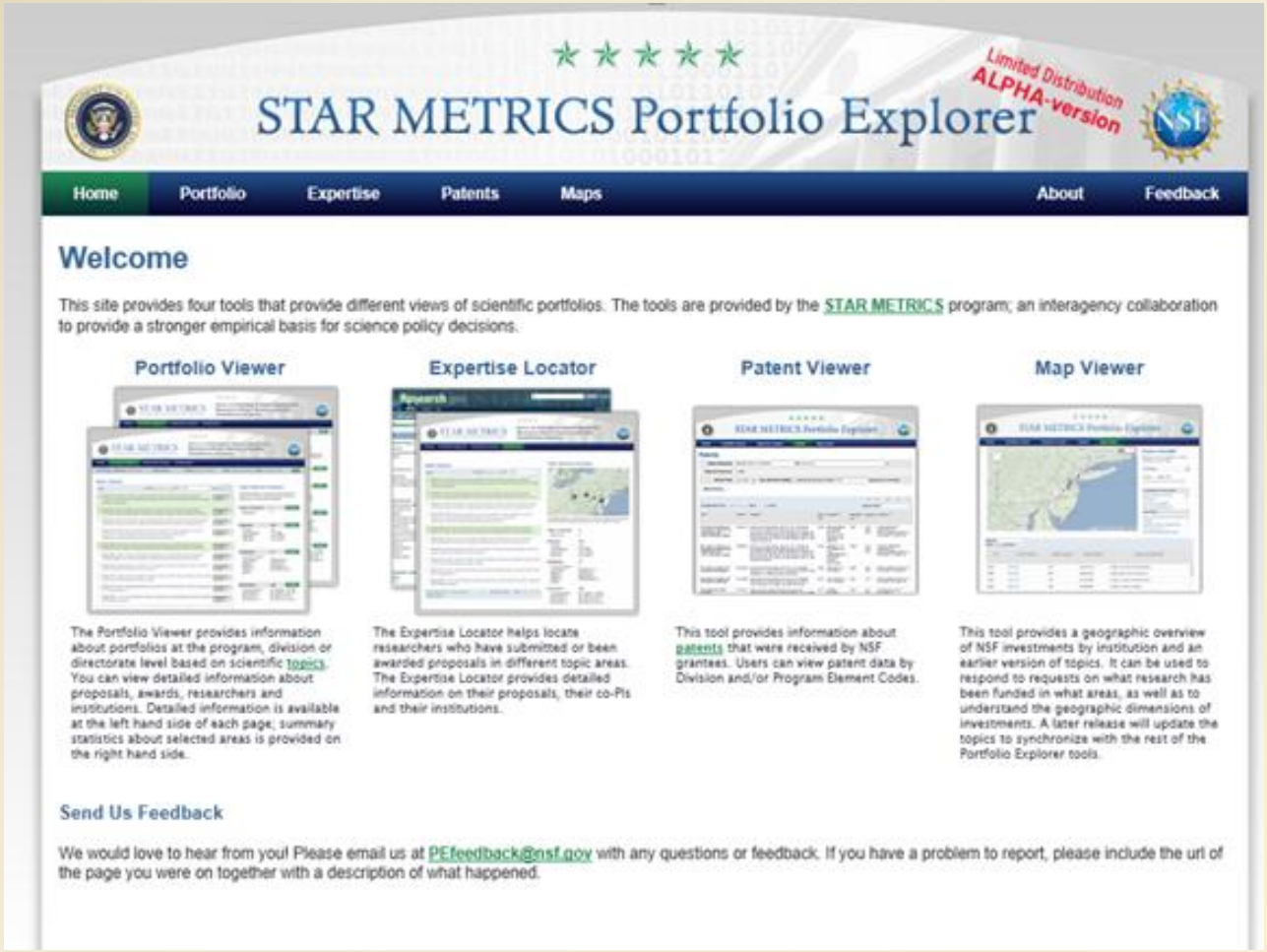


Topic tags for each and every document

Preliminary topic model results are promising…



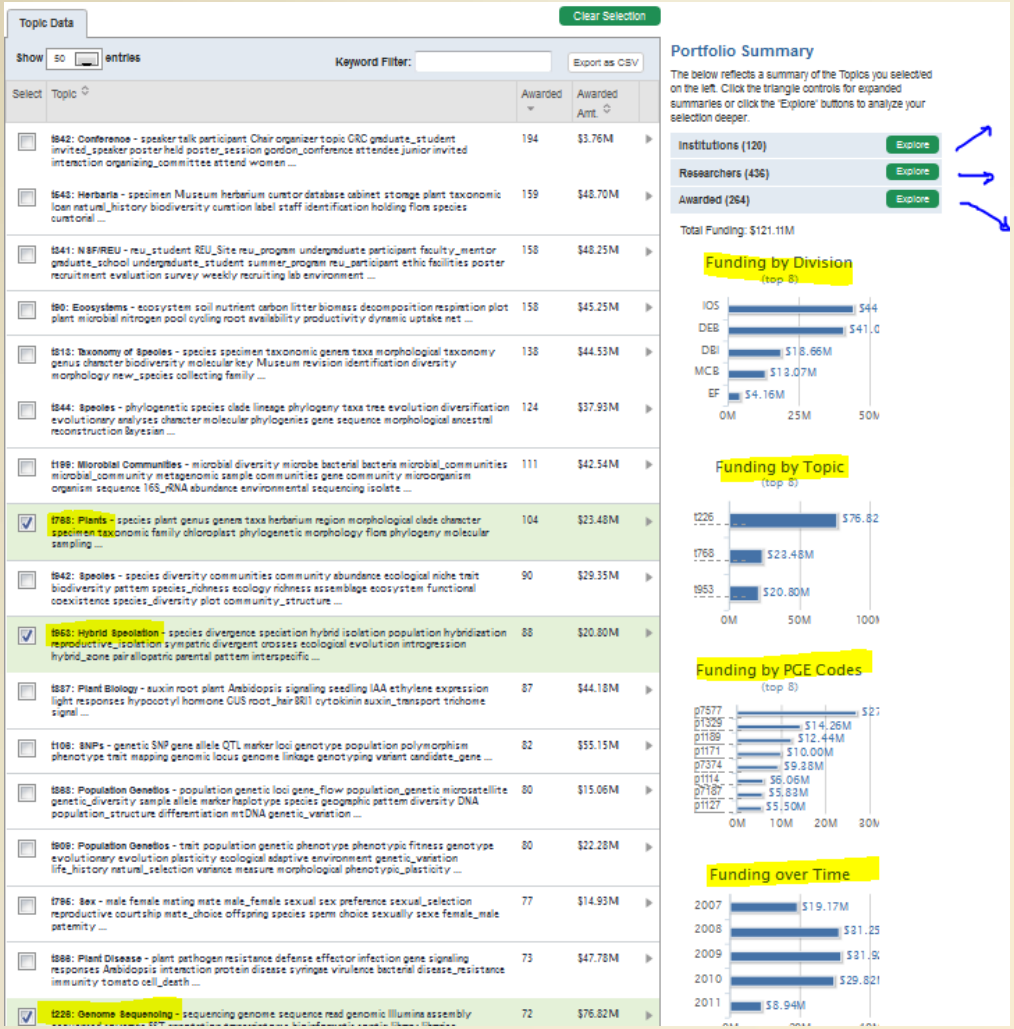
Application of Topic Modeling: Portfolio Explorer



Using Portfolio Explorer to find research projects

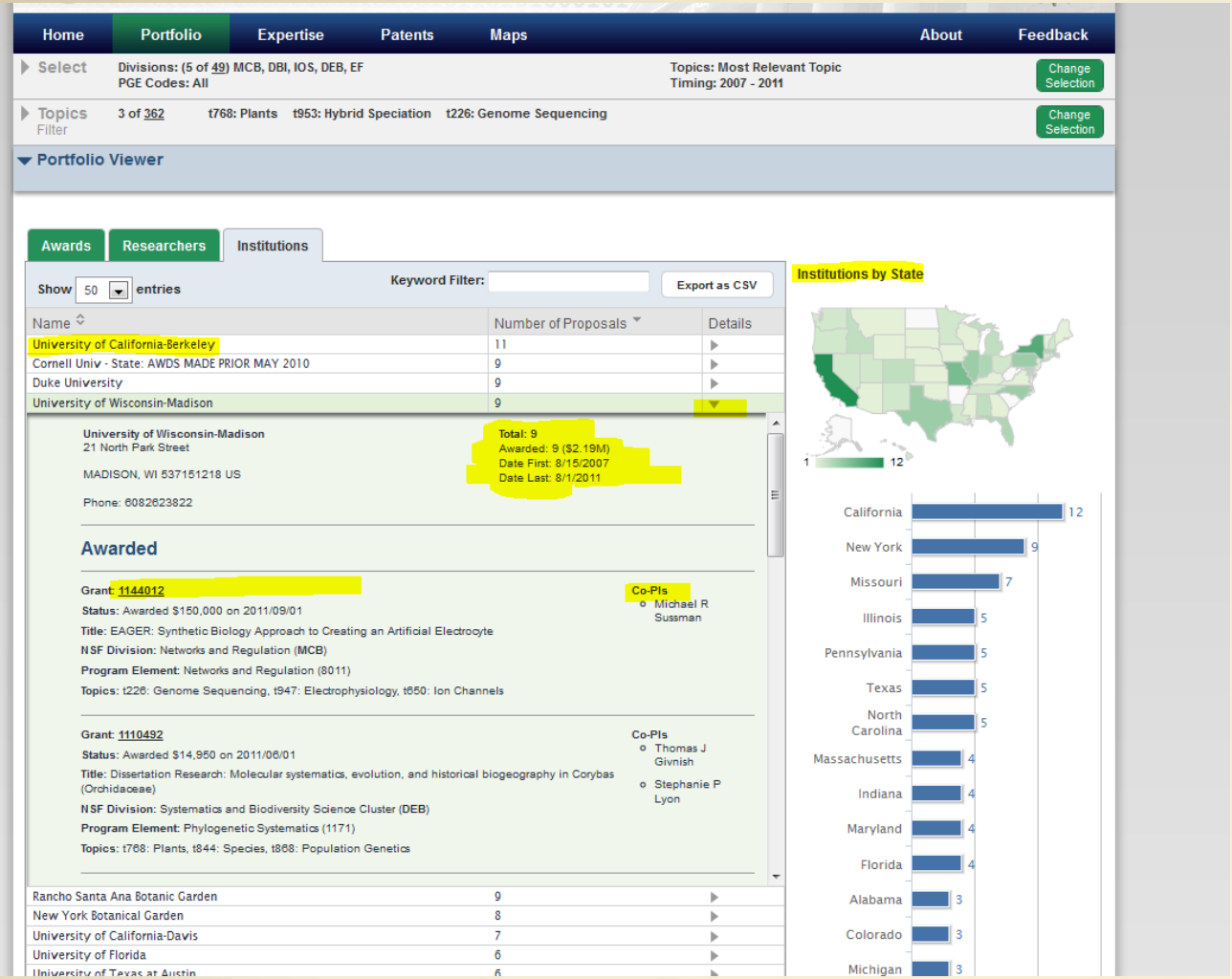
1. Check boxes for program areas of interests (e.g. NSF Directorate for Biological Sciences
2. Drop-down menus for division descriptions (shows common topics and project codes)
3. "View Topics" button pulls topics for these 7,500 projects

‘Topic filter’ selects topics (e.g. Plants, Hybrid Speciation, Genome sequencing)



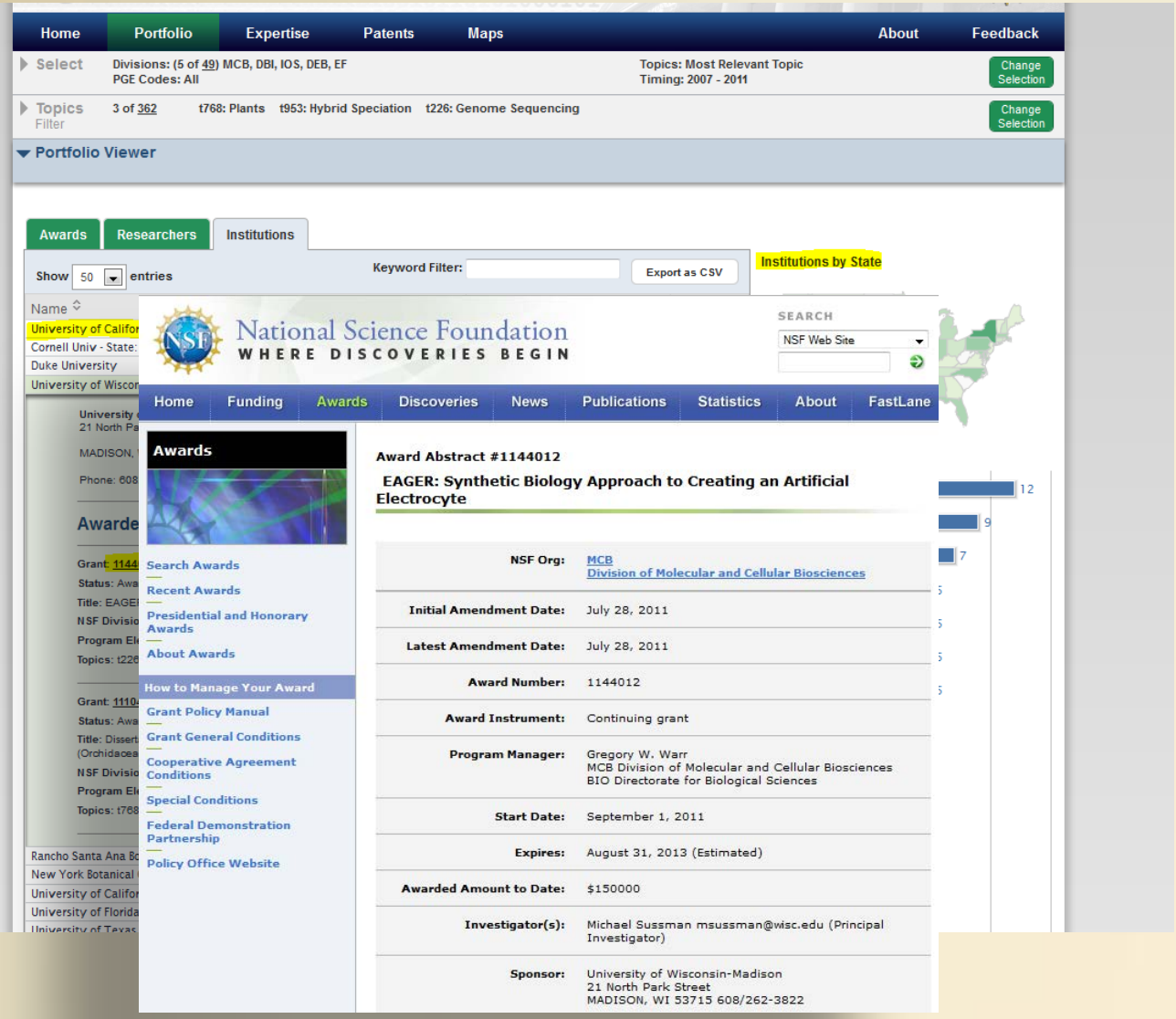
- Graphs by:
  - NSF division
  - Topic
  - NSF proj codes
- Portfolio Summary views:
  - Institutions
  - Researchers
  - Awarded

‘Institutions’ view



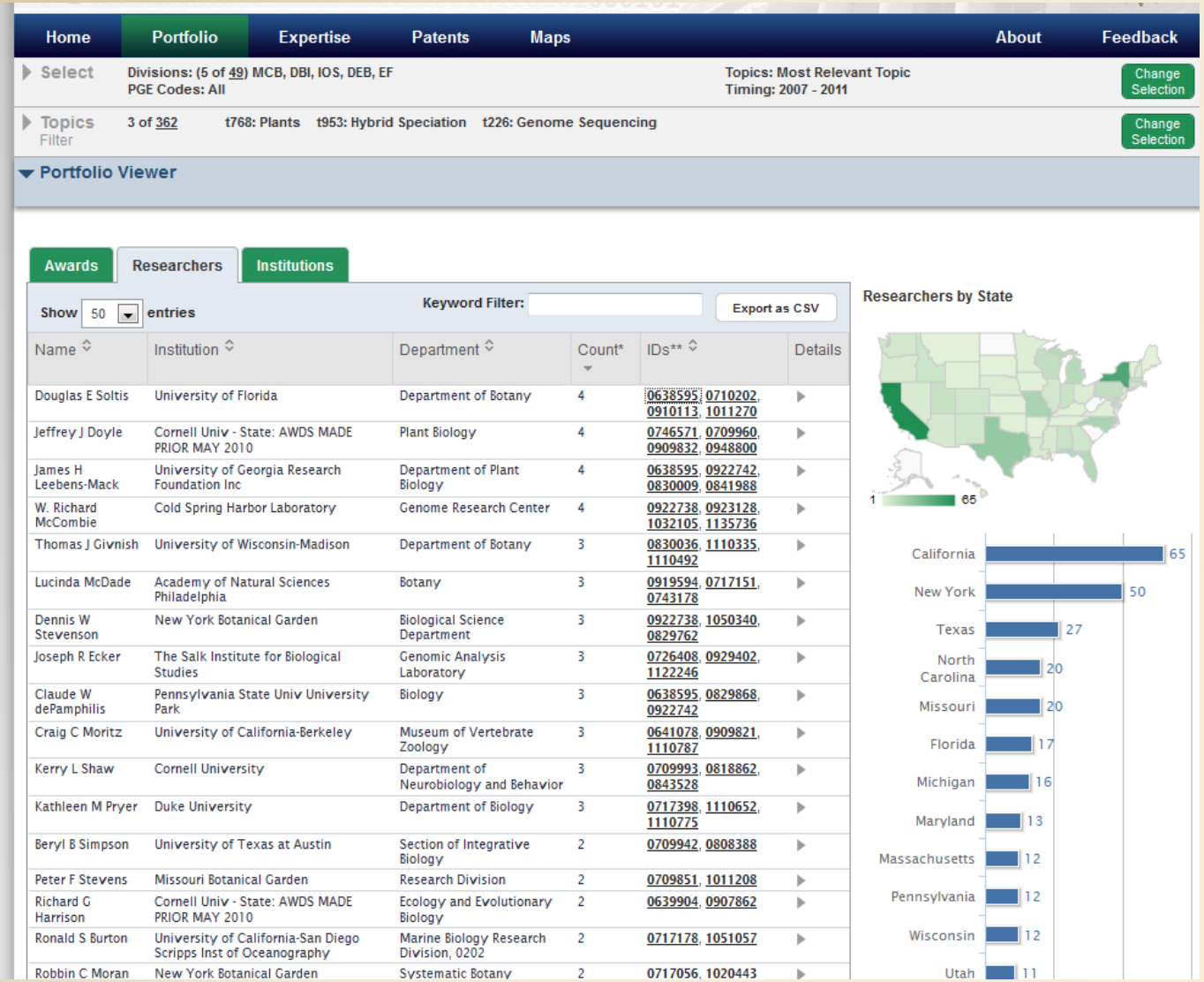
- Long list of institutions, with map by state
- Dropdown box with summary stats, projects
  - Linked grant records

‘Institutions’ view



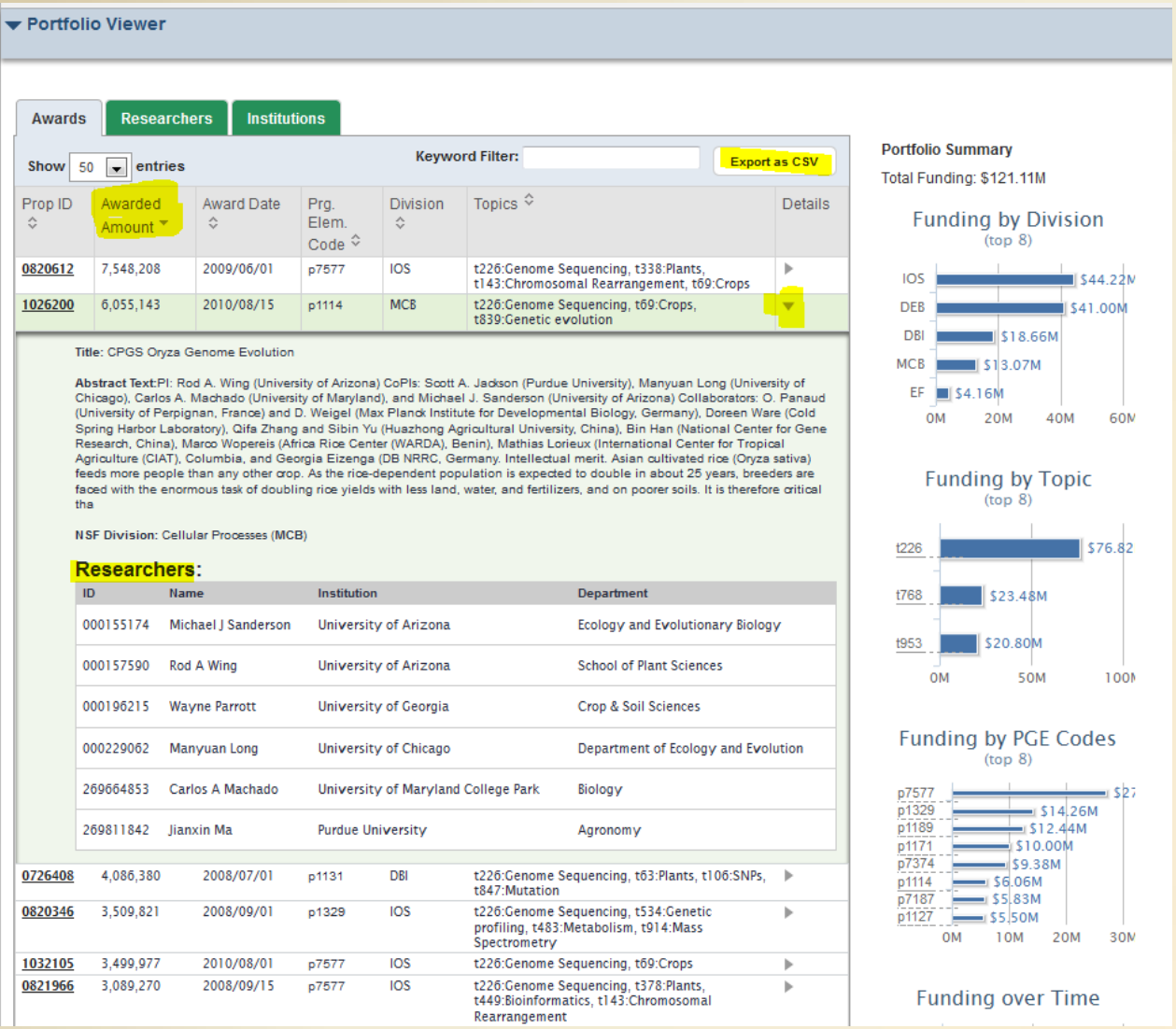
- Long list of institutions, with map by state
- Dropdown box with summary stats, projects
  - Linked grant records

‘Researchers’ view



- Researchers with linked grant records

‘Awards’ view



- Shows all awards in selected topics
  - dropdown for detail
  - CSV export

Other potential agency applications

- Complementarity analysis
  - Gaps and hotspots
- Compare research investments to outputs
- Compare funded/non-funded projects
- Expertise locators for review panels, funding announcements

Thank you

John.L.King@osec.usda.gov

Overview of Small Area Estimation Methods

Julie Gershunskaya,
U.S. Bureau of Labor Statistics

NAS Workshop
Washington DC, September 2012

BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

Any opinions expressed in this presentation are those of the author and do not constitute policy of the Bureau of Labor Statistics

Outline

I. Introduction (what is small area?)

II. Domain estimation
   - Direct design-based estimators (HT, Ratio, modified direct)
   - Indirect estimators (synthetic estimators, composite estimators)

III. Methods based on the mixed effects models
   - Two types of models (area- and unit-level models)
   - Fay-Herriot area-level model
   - Nested error regression model
   - Combining cross-sectional and time-series data
   - Generalized linear mixed model

IV. Discussion: possible general scenarios for modeling BRDIS R&D data

V. Summary

Introduction: What is Small Area?

Planning for domains of interest at the sampling design stage is very important. Yet there are always domains that lack sample

Small area is a domain of interest where the sample size is not large enough to make direct sample based estimates of "adequate precision"

These can be

- geographic entities
- industrial subdivisions
- socio-demographic groups
- intersections of geography, industry, or demography

Small Area Estimation (SAE) methods are statistical tools used to produce more stable estimates

Two types of estimators for domains

(1) Direct estimators

- use values of the variable of interest only from the sample units in the domain and time period of interest
- usually unbiased or nearly unbiased
- may be extremely inefficient

(2) Indirect estimators

- "borrow strength" outside the domain and/or time period of interest
- are based on implicit or explicit assumptions (model)
- have smaller variance but may be biased if assumptions fail

The objective is to strike the balance between bias and variance

Direct estimators for domains (HT)

Target quantity: total Y_d for domain d of the population.

Each sampled unit j has associated sampling weight:

w_j = inverse of a unit's probability of being selected

(rough interpretation: weight corresponds to the number of pop units represented by a sample unit)

Horvitz-Thompson (HT) estimator of domain total Y_d :

$$\hat{Y}_d^{(HT)} = \sum_{j \in s_d} w_j y_j$$

sum over sampled units in domain d

HT estimator may be very unstable, especially in small domains

Alternative direct estimators may be improvement over efficiency of HT

Direct estimators for domains (Ratio estimator)

y_j: measurement of interest for sample unit j

x_j: auxiliary data for sample unit j

X_d: known population total for domain d
(from administrative or census data)

R&D Example:

y_j is R&D expenditure for company j (broken down by Industry & State)

x_j is total payroll for company j (broken down by Industry & State)

X_d is true population total payroll in a State/Industry (domain d)

Compute the ratio using sample data: $\hat{B}_d = \dfrac{\hat{Y}_d^{(HT)}}{\hat{X}_d^{(HT)}}$

Apply this ratio to known population total X_d: $\hat{Y}_d^{(R)} = X_d \hat{B}_d$

Direct estimators for domains (Post-stratified estimator)

Post-stratified estimator is a particular case of the ratio estimator

N_d is the number of population units in the domain (suppose it is known)

$\hat{N}_d^{(HT)} = \sum_{j \in s_d} w_j$ is the sample estimate of N_d

The ratio estimator becomes

$$\hat{Y}_d^{(PS)} = N_d \frac{\hat{Y}_d^{(HT)}}{\hat{N}_d^{(HT)}},$$

This estimator usually is improvement over HT

Yet when domain variance is large and sample size is small, PS estimator still would not provide satisfactory precision

Direct estimators for domains (GREG)

The ratio estimator is a particular case of the

Generalized Regression (GREG) estimator

$$\hat{Y}_d^{(GREG)} = \hat{Y}_d^{(HT)} + \left(\mathbf{X}_d - \hat{\mathbf{X}}_d^{(HT)} \right)^T \hat{\mathbf{B}}_d ,$$

where

$\mathbf{X}_d$ is a vector of known population totals for domain d

$\hat{\mathbf{X}}_d^{(HT)}$ is a vector of HT estimates of $\mathbf{X}_d$

$\hat{Y}_d^{(HT)}$ is HT estimate of Y_d

$\hat{\mathbf{B}}_d$ is a vector of coefficients
(derived from the sample using a particular formula)

Modified direct estimator (a.k.a. Survey Regression)

Note that $\hat{B}_d$ in the ratio estimator is based on sample data.

When sample is small, $\hat{B}_d$ may be unstable.

Instead, we can use pooled sample: $\hat{B} = \dfrac{\sum_{j \in s} w_j y_j}{\sum_{j \in s} w_j x_j}$

sum over units in the pooled sample

Survey Regression (SR) estimator:

$$\hat{Y}_d^{(SR)} = \hat{Y}_d^{(HT)} + \left(X_d - \hat{X}_d^{(HT)} \right) \hat{B}$$

$\hat{B}$ is based on pooled sample

Modified direct estimator (Survey Regression, cont.)

R&D example

X_{im} is known population payroll in industry i and State m

$\hat{X}_{im}^{(HT)}$ is HT estimate of payroll in industry i and State m

$\hat{X}_{i}^{(HT)}$ is HT estimate of payroll, national total for industry i

$\hat{Y}_{i}^{(HT)}$ is HT estimate of R&D, national total for industry i

Use national data for industry i to find $\hat{B}_i = \dfrac{\hat{Y}_i^{(HT)}}{\hat{X}_i^{(HT)}}$

SR estimator for R&D in industry i and State m is

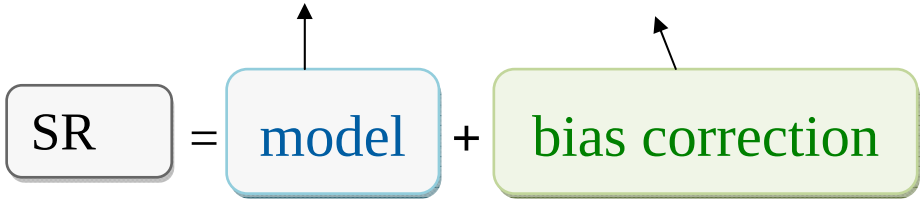
$$\hat{Y}_{im}^{(SR)} = \hat{Y}_{im}^{(HT)} + \left(X_{im} - \hat{X}_{im}^{(HT)} \right) \hat{B}_i$$

Modified direct estimator (Survey Regression, cont.)

Although $\hat{B}$ in SR estimator is based on a larger sample, effective sample size of the SR estimator is still equal to the domain size

To see this, re-write SR as

$$\hat{Y}_d^{(SR)} = X_d \hat{B} + \sum_{j \in s_d} w_j \left(y_j - x_j \hat{B} \right)$$



Bias correction is a sum of weighted residuals

Ultimately, the efficiency of SR still depends on variability of residuals and on the domain sample size

Indirect estimators (synthetic estimators)

Direct sample based estimators are unbiased (or nearly so) but they may have unacceptably large variance

To overcome this problem, certain assumptions about similarity or relationships between areas and/or time periods are made

Assumptions allow using larger sample, thus “borrowing strength”

Synthetic estimators are sample-based estimators from larger (or combined) areas and/or from other time periods when applied to a small area

The synthetic estimator has lower variance but it is biased if assumptions fail

Synthetic estimator example: Estimation of domain mean from a simple random sample

Direct domain sample average

$$\bar{y}_d = \frac{1}{n_d} \sum_{j \in s_d} y_j$$

sum over sampled units in domain d

is not reliable

Assumption: domain mean = overall population mean

This assumption allows us to use larger sample:

$$\bar{y}_d^{(Syn)} = \frac{1}{n} \sum_{j \in s} y_j$$

sum over all sampled units

The "synthetic" average is much more stable. However, it may be badly biased if the assumption about common mean does not hold

Synthetic estimator example (common slope)

Consider again Survey Regression estimator in the form

$$\hat{Y}_d^{(SR)} = X_d \hat{B} + \sum_{j \in s_d} w_j \left(y_j - x_j \hat{B} \right)$$

SR = model + bias correction

The model part of SR is a synthetic estimator $\hat{Y}_d^{(Syn)} = X_d \hat{B}$

R&D example:

A synthetic estimator of R&D expenditure in industry i and State m is

$$\hat{Y}_{im}^{(Syn)} = X_{im} \hat{B}_i, \qquad \text{where } \hat{B}_i = \frac{\hat{Y}_i^{(HT)}}{\hat{X}_i^{(HT)}}$$

Assumption: common ratio $\hat{B}_i$ of R&D to payroll holds for all States in industry i

BRDIS State Estimator (based on Slanta and Murlow, 2004)

R&D for State m is estimated as

$$\hat{Y}_m = Y_{m,s} + \hat{Y}_{m,c}$$

$Y_{m,s} = \sum_{j \in s} y_{m,j}$ is observed sampled total for R&D in State m

$\hat{Y}_{m,c}$ is prediction of non-sampled part of pop for R&D in State m,

computed as $$\hat{Y}_{m,c} = \sum_{i=1}^{I} R_{im} \hat{Y}_{i,c},$$

where

R_{im} is the ratio of payroll in State m to National total for industry i (expected value for sample-complement is used)

$\hat{Y}_{i,c} = \sum_{j \in s} \left(w_j - 1 \right) y_{i,j}$ is prediction for non-sampled part of R&D in i

Assumption: in each industry i, R&D is distributed among States proportionately to State's total payroll

BRDIS State estimator discussion

Let us compare the State estimator (SM 2004) with the synthetic estimator based on a common industry slope

For simplicity, consider estimation for the whole population (rather than non-sampled part only):

SM estimator:

$$\hat{Y}_m^{(SM)} = \sum_{i=1}^{I} \frac{X_{im}}{X_i} \hat{Y}_i^{(HT)}$$

Common industry slope (CIS):

$$\hat{Y}_m^{(CIS)} = \sum_{i=1}^{I} X_{im} \frac{\hat{Y}_i^{(HT)}}{\hat{X}_i^{(HT)}}$$

Both are synthetic estimators based on similar assumptions.

Notice in the denominators: population X_i vs. estimated $\hat{X}_i^{(HT)}$

It might be worth evaluating estimators using BRDIS data:

if indeed R&D is correlated with payroll, CIS may be more efficient

Synthetic estimators (cont.)

$\hat{Y}_{\bullet 1}$	$\hat{Y}_{\bullet 2}$	…	$\hat{Y}_{\bullet M}$	
C_{11}	C_{12}	…	C_{1M}	$\hat{Y}_{1\bullet}$
C_{21}	C_{22}	…	C_{2M}	$\hat{Y}_{2\bullet}$
…	…	…	…	…
C_{I1}	C_{I2}	…	C_{IM}	$\hat{Y}_{I\bullet}$

***S**tructure **Pre**serving **E**stimator* **(SPREE)**

The goal: estimate Y_{im} for areas (im) defined by intersection of industry i and geography m, say

Cell counts C_{im} for characteristic of interest are available from administrative data (e.g., most recent census)

The sample is large enough for reliable marginal estimates $\hat{Y}_{i\bullet}$ or $\hat{Y}_{\bullet m}$.

Iterative **P**roportional **F**itting (**IPF**, a.k.a. raking) algorithm adjusts old census counts C_{im} to fill out the table to conform to new margins.

Assumption: the structure remains unchanged since last census, e.g.,

$$\frac{C_{ij}}{C_{kj}} \Big/ \frac{C_{il}}{C_{kl}} = \frac{Y_{ij}}{Y_{kj}} \Big/ \frac{Y_{il}}{Y_{kl}}$$ for any combination i, j, k, l.

Indirect estimators (composite estimators)

Composite estimator is a convex combination of the direct and synthetic estimators, as a compromise between bias and variance:

$$\hat{Y}_d^{(C)} = \gamma_d \hat{Y}_d^{(Direct)} + \left(1 - \gamma_d\right) \hat{Y}_d^{(Model)}$$

The central question: **How to choose weights γ_d?**

Possible approaches:

- *Sample Size Dependent Method*

  Define weights based on the sample coverage in a given area, e.g.,

  γ_d is proportional to $\hat{N}_d / N_d$

  This method does not account for area variation

- Find weights to *minimize mean square error (MSE)* of the resulting estimator

  Depends on estimates of MSE of composite parts that are unstable

Methods based on explicit models

Explicitly stated modeling assumptions allow for application of standard statistical methods for

- model selection
- model evaluation
- estimation of model parameters
- producing measure of uncertainty (confidence intervals, mean squared error) under the assumed model

Methods based on explicit models constitute the core of modern Small Area Estimation methods

Most popular are methods based on the

- linear mixed model (for continuous variables) or
- generalized linear mixed model (for binary or count data)

Two general approaches

Area-level model (assumptions on aggregate, area-level data)

- applied when only area-level auxiliary data is available
- direct sample-based estimates play the role of data in the model; their variances are assumed known
- generally, easier to apply (than the unit-level models)
- usually, direct sample-based estimators already take into account sampling design. (However, it may happen that some areas are not sampled: if areas are selected into sample with unequal probabilities related to the true area means, bias may occur)

Unit-level model (assumptions on individual units)

- more efficient use of information
- need care accounting for sampling design

Fay-Herriot (FH) area-level model

Fay & Herriot (1979): estimation of per-capita income for small places

Auxiliary variables:

- county level PCI (census);
- value of owner-occupied housing (census),
- average adjusted gross income per exemption (IRS)

Assumtions for $d = 1, \ldots, D$:

Sampling model: $\hat{Y}_d^{(Direct)} = \theta_d + \varepsilon_d$

Linking model: $\theta_d = \mathbf{X}_d^T \boldsymbol{\beta} + \nu_d$

ν_d and ε_d are mutually independent random terms

The model can be written as a linear mixed model:

$$\hat{Y}_d^{(Direct)} = \mathbf{X}_d^T \boldsymbol{\beta} + \nu_d + \varepsilon_d$$

Fay-Herriot (FH) area-level model (cont.)

The "best" (in a certain sense) estimator under the model has the composite form $\boxed{\hat{\theta}_d = \gamma_d \hat{Y}_d^{(Direct)} + \left(1-\gamma_d\right)\mathbf{X}_d^T\hat{\boldsymbol{\beta}}}$

where $\gamma_d = \dfrac{A}{A+V_d^{(Direct)}}$.

- direct estimates are shrunk toward the synthetic part
- area-specific random term v_d in the "linking model" captures additional variability not explained by auxiliary variables. This is achieved by introduction of parameter A, the variance of v_d
- the smaller the value of A, the more weight goes to the synthetic part
- variance $V_d^{(Direct)}$ of the direct sample estimator is considered known. Areas with larger sample variance place relatively more weight on the synthetic part

BRDIS: Estimation for State by Industry

$\hat{Y}_{im}^{(Direct)}$ is a direct sample estimator for R&D in industry i and State m

Direct sample estimator provides unbiased measurement of unobserved truth θ_{im}, with some random error:

$$\hat{Y}_{im}^{(Direct)} = \theta_{im} + \varepsilon_{im} \qquad \text{(sampling model)}$$

The assumption that in industry i, State level R&D is proportional to State's total payroll can be expressed as

$$\theta_{im} = X_{im} B_i + v_{im} \qquad \text{(linking model)}$$

We allow deviations from our assumption by using the random term v_{im}.

The resulting composite estimator can be written as

$$\boxed{\hat{\theta}_{im} = X_{im}\hat{B}_i + \gamma_{im}\left(\hat{Y}_{im}^{(Direct)} - X_{im}\hat{B}_i\right),} \qquad \gamma_{im} = \frac{\hat{A}_i}{\hat{A}_i + V_{im}^{(Direct)}}$$

B_i and A_i are estimated from the data using a simple algorithm

Nested Error Regression (NER) unit-level model

Battese, Harter and Fuller (1988): estimation of areas planted with corn and soybeans for 12 counties in Iowa

Survey variables:

y_{dj} number of hectares of corn (or soybean) per segment j in county d

Auxiliary variables: satellite data

x_{1dj} number of pixels planted with corn per segment j in county d

x_{2dj} number of pixels planted with soybean per segment j in county d

For each sample unit $j = 1,\ldots,n_d$ in each areas $d = 1,\ldots,D$, assume

$$y_{dj} = \beta_0 + \beta_1 x_{1dj} + \beta_2 x_{2dj} + v_d + \varepsilon_{dj},$$

v_d and ε_{dj} are mutually independent random terms

Nested Error Regression (NER) unit-level model (cont.)

The "best" estimator under the model:

$$\hat{\theta}_d = \gamma_d \bar{y}_d + \left(1-\gamma_d\right)\left(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_{1d} + \hat{\beta}_2 \bar{x}_{2d}\right)$$

where $\gamma_d = \dfrac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2 / n_d}$

- regression coefficients $\beta_0, \beta_1, \beta_2$ are unknown model parameters estimated from the data
- σ_v^2 and σ_e^2 are variances of the random terms in the model. They are also unknown model parameters and are estimated from the data
- note that the larger the sample size n_d of an area the more relative weight is on the sample part

Both FH and NER models are particular cases of *the linear mixed model*

Combining cross-sectional and time-series data

In surveys repeated over time, we can use previous period's data

Rao and Yu (1994) introduced the following model for area-level data:

For areas $d = 1,\ldots,D$ and time periods $t = 1,\ldots,T$, assume

$$\hat{Y}_{dt}^{(Direct)} = \mathbf{X}_{dt}^{T}\boldsymbol{\beta} + v_d + u_{dt} + e_{dt},$$

and random terms u_{dt} follow AR(1) model

$$u_{dt} = \rho u_{d,t-1} + \varepsilon_{dt}, \qquad |\rho| < 1$$

The "best" estimator for the current period is a weighted sum of the synthetic estimator for the current period and model residuals from all time periods:

$$\hat{\theta}_{dT} = \gamma_{dT}\hat{Y}_{dT}^{(Direct)} + \left(1-\gamma_{dT}\right)\mathbf{X}_{dT}^{T}\hat{\boldsymbol{\beta}} + \sum_{t=1}^{T-1}\gamma_{dt}\left(\hat{Y}_{dt}^{(Direct)} - \mathbf{X}_{dt}^{T}\hat{\boldsymbol{\beta}}\right)$$

Generalized linear mixed models

y_{dj} is binary.

Target quantity: small area d proportion $P_d = N_d^{-1} \sum_{j \in d} y_{dj}$

- It is possible to formulate an area-level FH-type model using direct sample estimates of proportions
- One possible drawback of an area-level model: some areas may have no sample units reporting R&D and will be dropped from the model
- A unit-level generalized linear mixed model may be more efficient

y_{dj} is 1 with probability p_{dj} and is 0 with probability $1 - p_{dj}$

$$\log\left(\frac{p_{dj}}{1-p_{dj}}\right) = \mathbf{x}_{dj}^T \boldsymbol{\beta} + v_d$$

A couple scenarios for modeling R&D data

A unit-level model scenario:

- *wages*, *employment* (*other covariates?*) are obtained from administrative data for all units in the population of businesses
- using sample data, establish relationship between R&D of interest and auxiliary variables, estimate the model parameters. Apply results for prediction of R&D in the non-sampled part

  (There is no explicit question on State (or county) by Industry R&D in the BRDIS questionnaire. A proxy has to be derived)
- in the modeling, it is important to account for sampling design

An area-level model scenario:

- current "hybrid estimator" (or other area-level predictors) can be used in the model part of the composite estimator
- consider alternative direct estimators of R&D, for use in the area-level model
- what if some areas are not in the sample? Would this cause bias?

Direct sample estimators for State R&D

Let $\delta_{m,j} = \begin{cases} 1, & \text{if company } j \text{ reports R\&D in State } m \\ 0, & \textit{otherwise} \end{cases}$

A few examples of the model-based estimators that account for sampling design:

- No auxiliary info:

$$\hat{Y}_m = \sum_{j \in s} y_{m,j} + \left(N_m - n_m\right) \frac{\sum_{j \in s} \left(w_j - 1\right) y_{m,j}}{\sum_{j \in s} \delta_{m,j} \left(w_j - 1\right)},$$

- Analogue of the ratio estimator, using company's payroll x_j or some other auxiliary variable (maybe payroll per employee is better?):

$$\hat{Y}_m = \sum_{j \in s} y_{m,j} + X_{m,c} \frac{\sum_{j \in s} \left(w_j - 1\right) y_{m,j}}{\sum_{j \in s} \delta_{m,j} \left(w_j - 1\right) x_j},$$

where $X_{m,c}$ is total payroll in the non-sampled part of State m

Direct sample estimators for State R&D (cont.)

- Analogue of the modified direct estimator:

$$\hat{Y}_m = \sum_{j \in s} y_{m,j} + \sum_{i=1}^{I} X_{im,c} \hat{B}_i + \left(N_m - n_m \right) \hat{M}_{m,c},$$

where

$$\hat{B}_i = \frac{\sum_{j \in s} y_{i,j}}{\sum_{j \in s} \delta_{i,j} x_j}, \quad \delta_{i,j} = \begin{cases} 1, & \text{if company } j \text{ reports R\&D in industry } i \\ 0, & \textit{otherwise} \end{cases}$$

$$\hat{M}_{m,c} = \frac{\sum_{j \in s} \delta_{m,j} \left(w_j - 1 \right) \left(y_{m,j} - \sum_{i=1}^{I} x_{im,j} \hat{B}_i \right)}{\sum_{j \in s} \delta_{m,j} \left(w_j - 1 \right)}$$

This is intended only as illustration of alternative direct estimators. A better model can be devised, especially if there are more auxiliary variables.

Some important points

- It is important to plan for domains of interest at the sampling design stage
- Finding a set of good auxiliary variables is the formula for success, at every stage of a survey
- SAE methods are based on assumptions: evaluation of resulting estimates is of utmost importance
- Stating assumptions using a statistical model supports systematic approach to a problem:
  - explicitly stated assumptions
  - model selection and checking
  - availability of measures of uncertainty (MSE, CI)
- It is important to account for sampling design (unequal probabilities of selection) in the model formulation and fitting

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data, *Journal of the American Statistical Association*, 83, 28-36

Fay, R.E. and Herriot, (1979). Estimates of Income for Small Places: an Application of James-Stein Procedure to Census Data, *Journal of American Statistical Association*, 74, 269-277

Rao, J.N.K. (2003). Small Area Estimation, New-York, John Wiley & Sons, Inc.

Rao, J.N.K., and Yu, M. (1994). Small-area estimation by combining time series and cross-sectional data. *The Canadian Journal of Statistics*, 22, 511-528

Sverchkov, M., and Pfeffermann, D. (2004). Prediction of Finite Population Totals Based on the Sample Distribution. *Survey Methodology*, 30, 79–92

Contact information

Julie Gershunskaya
Mathematical Statistician
Bureau of Labor Statistics
SMS/OEUS
(202) 691-6360
Gershunskaya.Julie@bls.gov