Open access as a path to increased scientific productivity

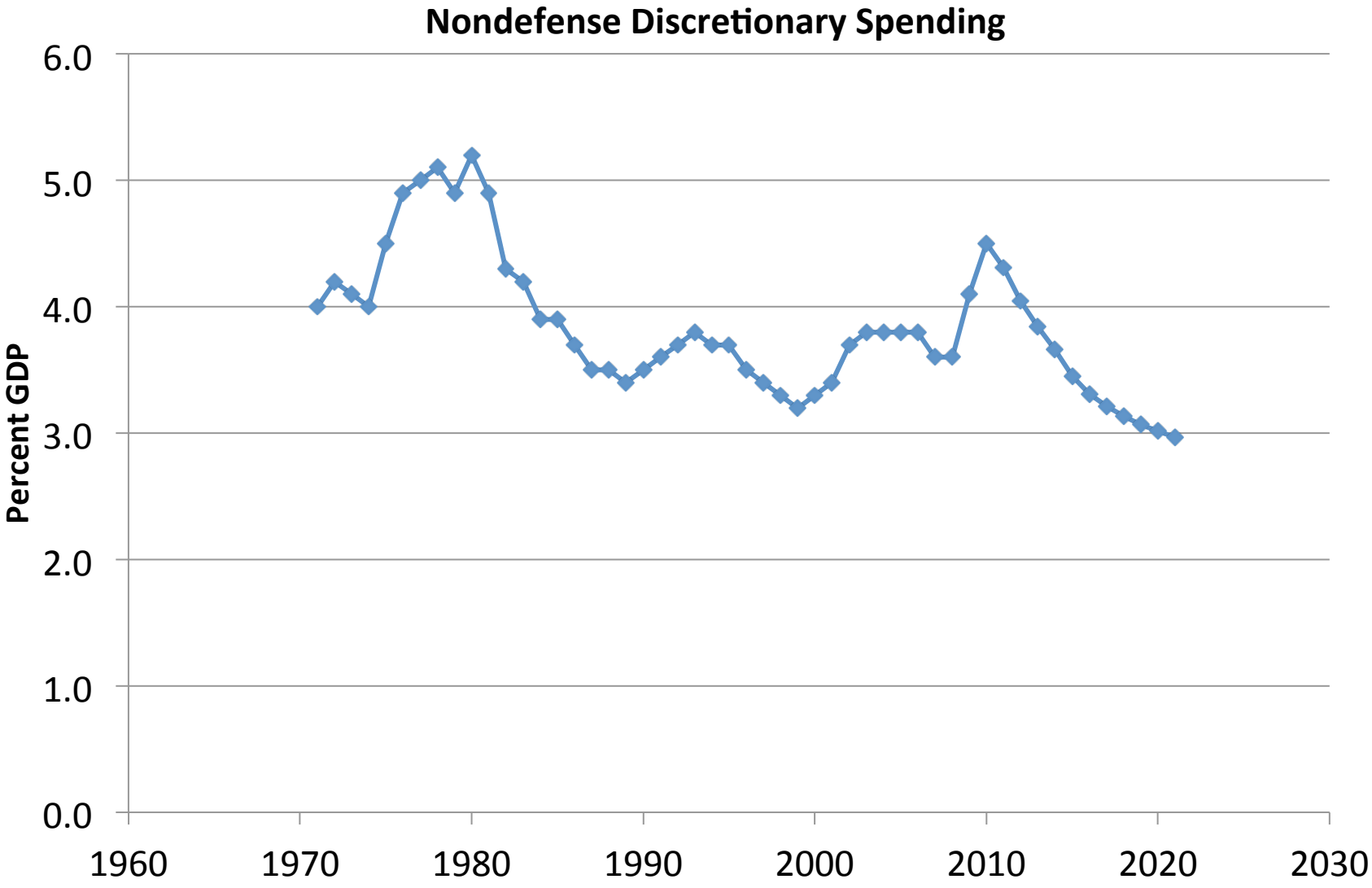
Neil M. Thakur PhD
NIH Office of Extramural Research

Berlin 9

Nov 8, 2011

This presentation does not necessarily reflect the views or policies of NIH

Big Picture- Threats to science funding



Solutions- how to do more with less

- Work Longer
- Work Cheaper
- True Productivity: Create value more efficiently

Literature and Scientific Productivity

Electronic distribution of scientific literature has lead to significant productivity gains

- Find articles faster and more accurately
- Retrieve articles from desktop
- Easier to store/organize
- Easier to cite

Literature and Scientific Productivity

Countervailing trend: As digital literature has made us more productive…

- More journals
- More papers
- More specialties and sub-specialties

Productivity Challenge

- Problem for most NIH funded scientists:
  - Information overload and time scarcity: *knowledge is costly*

- Partially resolved through
  - More focused science (less impactful?)
  - Team science and collaboration

Can scientific publishing help increase productivity?

A vision

Ensure access to the peer-reviewed scientific literature allows scientists incorporate new findings into their thinking faster, increasing the productivity and innovation of the scientific enterprise.

Using the literature to enhance productivity and innovation of the scientific enterprise

- **Full implementation**
  - **Increased productivity**: Instead of absorbing 40 papers a month, scientists absorb 4000 papers a month
  - **Increased funder ROI**: Proliferation of decision tools to apply the scientific literature to real world problems
    - How much fertilizer given my soil and weather conditions?
    - What is the best course and dose of chemotherapy for a patient like me?

- **How?** Development and widespread use of quantitative analysis of the literature

We are just starting to learn how to compute the literature

- **Literature based discovery**: Natural language processing to suggest new hypotheses
  - Christopher M. Miller, M.D; Thomas C. Rindflesch, Ph.D; Marcelo Fiszman, M.D., Ph.D; Dimitar Hristovski, Ph.D.; Dongwook Shin, Ph.D; Graciela Rosemblat, Ph.D; Han Zhang, M.S; Kingman P. Strohl, M.D. A closed literature-based discovery technique finds a mechanistic link between hypogonadism and diminished sleep quality in aging men. Accepted to *Sleep*, 5/24/2011. http://www.journalsleep.org/AcceptedPapers/SP-176-11.pdf .

- Developing analytic tools for all disciplines may require broad investment, development, and ideas from other fields
  - **Visualizations** of author and citation networks- we need to do this for research results, not just metadata
  - **Innovative search and screening of the literature**: Amazon's search by photograph- why can't we do that for tumors, plants, bacteria, and other images in papers? http://a9.com/-/company/visualsearch.jsp

A Realistic Goal?

In 10 years, a scientist will be able to incorporate 30% more papers into their thinking than they can today in the same amount of time

How do we get there?

- **A vision** of what we are trying to do
- **A measurable goal**
- **A shift in thinking-** What is the role of scientific publication and all the people involved (authors, reviewers, publishers, funders, libraries)?
- **Full Access** to scientific publication
  - Even though the technology is unclear, full text access will be required
  - Fragmented or siloed access limits impact of any application, and may introduce bias
  - Are current Open Access practices (licenses, machine friendliness) sufficient?

Shift in Thinking: A new role for scientific publishing

Goal	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation

Complementary roles for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity

Complementary roles for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	**Libraries**	**Scientists, Funders**

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	**Libraries**	**Science producers (Scientists, Funders)**

To the extent these groups want different things, we have a cross over:

Open access journals ask scientists and funders to solve a library problem

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	Libraries	Scientists, Funders
Solution	**Papers as one to many communication**	**Papers as infrastructure (all to all communication)**

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	Libraries	Scientists, Funders
Solution	Papers as one to many communication	Papers as infrastructure (all to all communication)
Primary Usage	**People reading PDFs and print copies**	**People using machines to pre-screen and analyze papers**

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	Libraries	Scientists, Funders
Solution	Papers as one to many communication	Papers as infrastructure (all to all communication)
Primary Usage	**People reading PDFs and print copies**	**People using machines to pre-screen and analyze papers**

Using machines to find what to read

versus

Using machines to help us understand the literature

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	Libraries	Scientists, Funders
Solution	Papers as one to many communication	Papers as infrastructure (all to all communication)
Primary Usage	People reading PDFs and print copies	People using machines to pre-screen and analyze papers
Necessary Access/ License	**Fair Use (read and download personal copies)**	**Analysis by machine, creation of decision tools and other 'derivative works'**

Uncertain access undermines infrastructure

- A comprehensive analysis requires one comprehensive analytic dataset of publications- often not feasible

- Barrier: fragmented literature
  - Technical barriers to access: Multiple data formats, download caps…
  - Multiple archives with different access protocols and usage rights:
    - Usage often cannot be determined or negotiated electronically on the fly
    - Contracts, not copyright law, limit access: the ability of machines to crawl, search and download publications for analysis
    - License and contract may also limit use: the creation of decision tools and derivative works

Uncertain access undermines infrastructure

Literature analytics highlight impracticalities and risks of a fragmented system with uncertain access…

- Additional cost, complexity and time
- Limited scope and/or validity of enquiry
  - Missing data, or biased sample?
- Disincentive to develop new search and analysis technology

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	Libraries	Scientists, Funders
Solution	Papers as one to many communication	Papers as infrastructure (all to all communication)
Primary Usage	People reading PDFs and print copies	People using machines to pre-screen and analyze papers
Necessary Access/ License	Fair Use (read and download personal copies)	Analysis by machine, creation of decision tools and other derivative works
Outcome Measures	**Citations, Downloads, Academic Impact**	**User Productivity, Public Impact**

How research achieves impact

- **Training and implementation**
  - We want clinicians to be trained on the latest NIH supported research, and adopt evidence based practices. Reprints help convince research users to change their practice.

- **Commercialization**
  - Bayh-Dole Act, to encourage commercialization of federally supported research
  - SBIR program: Federal set-aside for research by small businesses

- Do OA commercial restrictions facilitate the application of research findings?

A new role for scientific publishing

Domain	Disseminate Scientific Findings	Increase Scientific Productivity and Innovation
Problem	Information Scarcity	Time Scarcity
Primary Constituents	Libraries	Scientists, Funders
Solution	Papers as one to many communication	Papers as infrastructure (all to all communication)
Primary Usage	People reading PDFs and print copies	People using machines to pre-screen and analyze papers
Necessary Access/ License	Fair Use (read and download personal copies)	Analysis by machine, creation of decision tools and other derivative works
Outcome Measures	Citations, Downloads, Academic Impact	User Productivity, Public Impact

A vision

Ensure access to the peer-reviewed scientific literature allows scientists incorporate new findings into their thinking faster, increasing the productivity and innovation of the scientific enterprise.