

博士人材政策から見た米国 UMETRICS：
UMETRICS と博士人材データベース(JGRAD)の
国際比較研究

2017 年 7 月

文部科学省 科学技術・学術政策研究所
第1調査研究グループ
松澤 孝明

【調査研究体制】

松澤 孝明 文部科学省 科学技術・学術政策研究所 第1 調査研究グループ
総括上席研究官

【Authors】

Takaaki MATSUZAWA Director 1st Policy-Oriented Research Group
National Institute of Science and Technology Policy (NISTEP),
Ministry of Education, Culture, Sports, Science and Technology
MEXT

報告書の引用を行う際には、以下を参考に出典を明記願います。

Please specify reference as the following example when citing this paper.

松澤孝明, 「博士人材政策から見た米国 UMETRICS; UMETRICS と JGRAD の国際比較」,
NISTEP RESEARCH MATERIAL, No.263, 文部科学省科学技術・学術政策研究所.

DOI: <http://doi.org/10.15108/rm263>

Takaaki MATSUZAWA, “UMETRICS in term of human resource policy for PhD holders;
International comparative study between UMETRICS in US and JGRAD in Japan” *NISTEP
RESEARCH MATERIAL*, No.263, National Institute of Science and Technology Policy, Tokyo.

DOI: <http://doi.org/10.15108/rm263>

**博士人材政策から見た米国 UMETRICS:
UMETRICS と博士人材データベース (JGRAD) の国際比較研究**

文部科学省 科学技術・学術政策研究所 第1調査研究グループ

松澤 孝明

要旨

本報告書録は、2017年3月3日に文部科学省科学技術・学術政策研究所 (NISTEP) で行われた、オハイオ州立大学大学院経済学研究科教授 Bruce Weinberg 氏の講演会の内容を、講演者の了承のもとに当研究所においてとりまとめ、また、同氏との意見交換を通じて得られた情報をもとに、NISTEP が進める博士人材データベース (JGRAD) と UMETRICS の比較を行い、両者の類似性や相違点について考察を行ったものである。

本報告書のうち、講演録の内容は、講演の記録として講演者の見解を掲載したものであり、また JGRAD と UMETRICS の比較研究については、筆者の見解をまとめたものである。当研究所の公式の見解を示すものではないことに留意されたい。

**UMETRICS in term of human resource policy for PhD holders;
International comparative study between UMETRICS in US and JGRAD in Japan**

1st Policy-Oriented Research Group, National Institute of Science and Technology Policy (NISTEP), Ministry of Education, Culture, Sports, Science and Technology (MEXT)
Takaaki MATSUZAWA

ABSTRACT

This report elaborates the seminar presentation by Dr. Bruce Weinberg of Department of Economics, Ohio State University, taken placed in March 3, 2017 at the National Institute of Science and Technology Policy (NISTEP). Speaking minutes are acknowledged by the presenter. In addition, we analyze the similarities and differences between the Doctoral Human Resources Database in Japan (JGRAD) and UMETRICS (Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science), based primarily upon the presentation of this seminar. It should be noted that the opinions in this report are the sole responsibility of the contributor(s) and do not necessarily reflect the official views of NISTEP.

目次

概 要	i
1. 目的	i
2. 調査概要	i
3. UMETRICS と JGRAD の比較研究	iv
本 編	1
第 1 章 UMETRICS と博士人材データベース (JGRAD) の比較研究	1
1. はじめにー我が国の人材政策から見た UMETRICS の意義	1
2. 調査結果 1 : UMETRICS の概要について	1
(1) UMETRICS の特徴とデータ構造	1
(2) UMETRICS プロジェクトの運営	4
(3) UMETRICS の目的	4
(4) UMETRICS のミッション	5
(5) UMETRICS の誕生の背景ーSTAR METRICS の進展と UMETRICS からの独立	6
3. 調査結果 2 : 博士人材データベース (JGRAD) の現状と課題	8
(1) JGRAD の構築の経緯と大学参加状況	8
(2) JGRAD の登録状況	9
(3) JGRAD による情報提供等	12
4. 分析及び考察 : UMETRICS と JGRAD の比較分析	14
(1) プロジェクトを取り巻く政策的環境	14
(2) プロジェクトの形態と大学参加の規模	16
(3) プロジェクトの社会背景	17
(4) データ利用の原則	18
(5) データベースのユーザー	18
(6) データベースの構築目的と登録対象者の範囲	18
(7) コアデータ、データ連結の範囲、データ入力方式	20
(8) データベースとしての機能と登録インセンティブ	21
(9) まとめ	22
5. 結論 : UMETRICS に学ぶ JGRAD の将来展望	23
(1) データベース構築とサーベイ調査の連携	23
(2) データ連携の促進とメトリクス手法の導入	23
(3) データベースの機能の整理	23
(4) 大学との共同ベンチマーク	24
第 2 章 講演会	27
講演会概要	27
講演録 (日本語)	28
謝辞	80
調査体制	80
参考資料	81

概 要

1. 目的

米国における人材データベースプロジェクト UMETRICS (Universities: Measuring the ImpacTs of Research on Innovation, Competitiveness, and Science) についての情報収集を行なうため、オハイオ州立大学経済学研究科教授 Bruce Weinberg 氏をお招きし、講演会を開催した (2017 年 3 月 3 日)。また、講演会で得られた情報や文献調査を元に、UMETRICS と NISTEP が推進する博士人材データベース (JGRAD : Japan's Doctoral Human Resource Database) の類似点や相違点について比較分析を行なった。

2. 調査概要

(1) UMETRICS について

UMETRICS [1]とは、2015 年 1 月に、米国オバマ政権が進める STAR METRICS (Science and Technology for America's Reinvestment: Measuring the EffectTs of Research on Innovation, Competitiveness and Science) から分離・独立した人材データベースプロジェクトである [2]。「人材が研究の最も重要な成果物の一つである」という考え方に立って、研究の価値を測定するために「人材」に着目している点の特徴である。

UMETRICS の基盤となった STAR METRICS とは、2009 年、オバマ政権下で米国復興・再生法 (the American Recovery and Reinvestment Act of 2009; ARRA) の要請により、米国経済を刺激するために誕生した連邦政府のプログラムで、研究事業体 (research enterprise) に関する大規模な自動化されたデータプラットフォームである [1]。UMETRICS は多くの大学から得られた STAR METRICS のデータを元にしており、連邦政府 (及びいくつかの連邦政府以外) の研究費で雇用された人のファイルが、行政データ (税のデータ、給与記録、購入記録等) や、自然発生データ (学位論文、出版、引用、研究費 (funded grant)、特許等)、調査データ (統計局 (Census Bureau) が保有するサーベイ調査等) など、広範なデータソースにリンクされた構造になっている [3]。

UMETRICS プロジェクトは、米国中西部の主要大学 (Big Ten) が設立した大学間コンソーシアム CIC (Committee of International Cooperation) (1) の連携協力がベースとなって進められている (2)。その中核機関は、ミシガン大学科学イノベーション研究所 (IRIS) である。IRIS は加盟大学間のコンソーシアムで、IRIS の運営は各大学が年会費を払い担保している。文献によれば、IRIS は (総額で) 研究開発資金提供の 150 億ドルを意味する 24 大学と既にパートナーシップを組んでいる [4]。また、Weinberg 氏によれば、UMETRICS には、現在、一流大学を含む 50 大学がプロジェクトにコミットし、実際に 19 大学がデータを登録している状況にある。ミシガン大学をハブとして、そのノードがオハイオ州立大学やニューヨーク大学等に存在する。

(2) 博士人材データベース (JGRAD) について

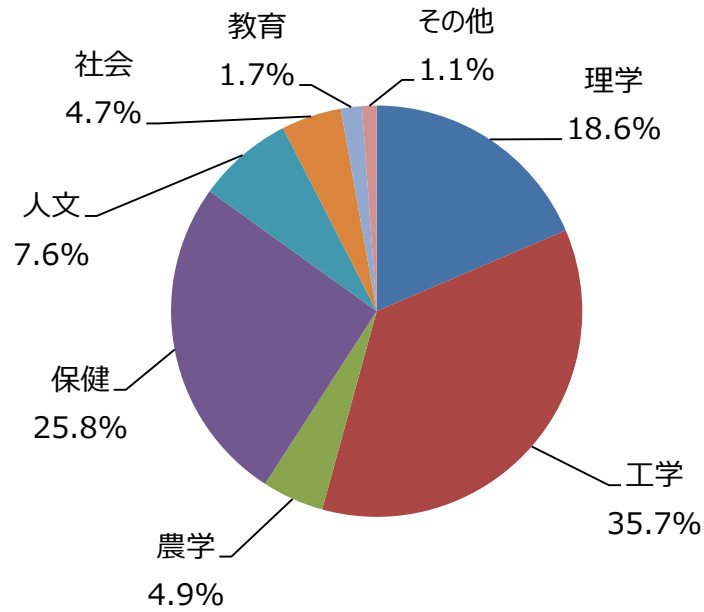
博士人材データベース (JGRAD) は文部科学省科学技術・学術政策研究所(NISTEP) が 2011 年度より進める人材データベースプロジェクトである。その目的は、博士課程修了者のキャリアパスを追跡し、博士人材の社会における活動状況を把握することにより、人材政策の基礎となるエビデンスを得ることである。

JGRAD は、2017 年 3 月末現在 (2017 年 4 月 3 日集計)、国公私立大学 42 大学、165 研究科が参加し、プロジェクトを進めている。キャリアパス把握の対象となるアカウント発行数は 24,525 人、実際の登録者 (JGRAD へのログイン数) は 7,072 人に達している。なお、2017 年 3 月末現在の状況を分析すると以下の通りである (概要図表 1)。

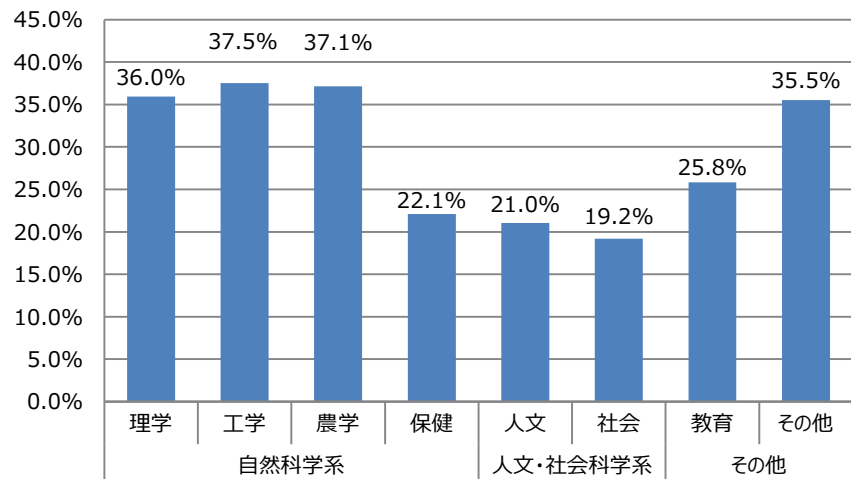
項目	登録状況
1. 登録者のジェンダー比	・ 「男性」が 59.9%、「女性」が約 24.0% (性別未回答者が 16.1%)
2. 登録者の分野別比率	・ 「理学」(18.6%)、「工学」(35.7%)、「農学」(4.9%)、「保健」(25.8%) ・ これらを合計した「自然科学系」が登録者の約 85%を占める。
3. 分野別登録率	分野別アカウント発行数に対する登録者数 (ログイン数) の割合を分野別の「登録率」として定義すると、 ・ 「工学 (37.5%)」「農学 (37.1%)」「理学 (36.0%)」の順に高い。 ・ 「保健 (22.1%)」や「人文 (21.0%)」「社会 (19.2%)」は低い。
4. 留学生	登録者 (7,072 人) のうち、1,455 人 (全登録者の 20.6%) が留学生。 ・ 留学生のうち 75.3%が「東アジア・東南アジア」出身。 ・ 上記に「南アジア (8.4%)」「西アジア (3.9%)」「中央アジア (0.8%)」を加えたアジアからの留学生は約 9 割 (88.4%)

概要図表 1 JGRAD の登録状況 (2017 年 3 月末現在) (n=7,072)

JGRAD は国立研究開発法人 科学技術振興機構(JST)の保有するデータベースとの連携を推進している。具体的には、2016 年 9 月 1 日より、JST の保有する JREC-IN Portal との連携を開始し、JGRAD に登録した情報をもとに、JREC-IN Portal で検索された求人情報を JGRAD 上で自動表示する「オーダーメイド型」の求人情報提供サービスを開始した[5]。また、登録者がキャリアパスを考える上で参考になるようなロールモデルの情報発信や、登録者のデータの再入力負荷を低減するため JST の研究者総覧データベース (researchmap) 等とのデータ連携についても検討を進めている。



概要図表 2 JGRAD の登録者の分野別比率 (n=7,072) (2017年3月末現在)
出典：博士人材データベース (JGRAD) 事務局 調べ



	理学	工学	農学	保健	人文	社会	教育	その他
アカウント発行数 (24,525)	3,652	6,730	926	8,254	2,543	1,739	453	228
登録者数 (7,072)	1,313	2,525	344	1,823	535	334	117	81
分野別アカウント発行数に対する登録率	36.0%	37.5%	37.1%	22.1%	21.0%	19.2%	25.8%	35.5%

概要図表 3 分野別アカウント発行数に対する登録率 (2017年3月末現在)
出典：博士人材データベース (JGRAD) 事務局 調べ

3. UMETRICS と JGRAD の比較研究

UMETRICS と JGRAD の類似点・相違点を比較すると概要図表 4 のようになる。

概要図表 4 UMETRICS と JGRAD の比較

比較項目		類似性	UMETRICS (米国)	博士人材データベース(JGRAD) (日本)
1. プロジェクトの概要	1 開始年	○	・2009年、STAR METRICS開始 ・2015年1月、STAR METRICSから独立	・2011年、JGRAD構想開始 ・2014年4月、試験運用開始
	2 データ収集の目的	×	・研究開発投資の社会における成果を多角的に捉える	・博士人材のキャリアパスの把握・追跡
	3 中核機関	○	・ミシガン大学科学イノベーション研究所 (Institute for Research on Innovation and Science; IRIS)	・文部科学省科学技術・学術政策研究所 (NISTEP) 第1調査研究グループ
	4 形態	○	・大学参加型(コンソーシアム型)	・大学参加型(コンソーシアム型)
	5 大学参加状況	○	・19大学 (Big Tenが中心) ・一流大学を含む、50大学がコミット	・42大学 (博士課程教育リーディングプログラム参加大学を含む) ・博士号授与件数トップ50大学が目標
	6 政策的位置づけ	×	・2009年 米国復興・再生法 (ARRA) の元で始まった STAR METRICSを技術的基盤とする ・2015年1月、STAR METRICSから独立	・2011年 SciREXのデータ・基盤事業 (NISTEPの研究プロジェクト) が開始 ・2016 第5期科学技術基本計画 (閣議決定) で位置づけられる
	7 政策環境	○	・サーベイ調査 (Survey of Earned Doctorates; SED) が実施される中でUMETRICSを後発的に開始	・サーベイ調査 (博士人材追跡調査: JD-Pro) が実施される中で平行してJGRADを構築
2. データ	8 コアデータ	×	・行政データ (自然データ、サーベイデータとリンク)	・サーベイデータ
	9 登録者	×	・研究者、ポスドク、事務員等、多岐に渡る ・米国内	・参加大学の博士課程在籍者・修了者 ・帰国留学生、海外転出者も追跡対象
	10 データ入力	×	・機関入力	・登録者本人が入力 (機関からのデータインポートも検討中)
	11 データ利用	○	・コンソーシアム内: IRIS、参加大学	・コンソーシアム内: NISTEP、参加大学 (自大学分のみ)
	12 サーベイ調査との関係	○	・SEDとの連携に向けて作業中	・将来的に、博士人材追跡調査 (JD-Pro) との連携を構想中
	13 データ連携の範囲	×	・広範なデータ連携 ・米国内勢調査局の持つ雇用と家計データや職歴データや、企業データ等の社会的経済的データ、出版・引用データなどを連結	・現時点では限定的なデータ連携、インポート促進 ・JSTのJREC-IN Portalとの連携開始 (2016年9月) ・JSTのresearchmapとの連携検討 (項目精査、新機能追加等) 開始 (2016年10月)
3. 機能	14 機能	×	・調査研究	・調査研究 ・インセンティブとしての情報提供 (公募情報、ロールモデル)

(注) ○ : 類似点、× : 相違点

出典 : 筆者作成

概要参考文献

[1] The UMETRICS Initiative, “Universities: Measuring the ImpacTs of Research on Innovation, Competitiveness, and Science”.

(<https://www.btaa.org/docs/default-source/umetrics/umetrics-synthesis-document.pdf?sfvrsn=4>) [アクセス日: 2017年7月17日]

[2] 国立研究開発法人 科学技術振興機構研究開発戦略センター (CRDS). 調査報告書 米国「科学イノベーション政策のための科学」の動向と分析. CRDS-FY2015-RR-04.

[3] Catherine Buffington, Benjamin Cerf, Christina Jones, and Bruce A. Weinberg, “STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to 2010 Census,” American Economic Review: Papers & Proceedings, 2016, 106(5): 333–338 <http://dx.doi.org/10.1257/aer.p20161124>

(<http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.p20161124>)

[アクセス日: 2017年7月17日].

- [4] Michael Eisenstein, "Academic return: A broader understanding of 'impact' could help governments to measure the diverse benefits of their investment in research," *Nature* 533, S20–S21 (05 May 2016) doi: 10.1038/533S20a Published online 04 May 2016 (http://www.nature.com/nature/journal/v533/n7601_supp/full/533S20a.html) [アクセス日: 2017年7月17日].

- [5] 松澤孝明、篠田裕美, 「博士人材データベースのパイロット運用-政策・制度・運用の現状と改善に関する検討報告書-」. 文部科学省科学技術・学術政策研究所 調査資料-255, 2016年11月.

本 編

第1章 UMETRICS と博士人材データベース(JGRAD)の比較研究

1. はじめに—我が国の人材政策から見た UMETRICS の意義

2017年3月3日、文部科学省科学技術・学術政策研究所(NISTEP)において、米国よりオハイオ州立大学経済学研究科教授 Bruce Weinberg 氏をお招きし、UMETRICS (Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science) [1]に関する講演会を開催した。初めに、その背景や意義について説明したい。

NISTEP では、我が国における博士課程修了者のキャリアパスの把握を目的として、2011年度より「博士人材データベース (JGRAD)」の構築を進めている。第5期科学技術基本計画(2016年1月22日、閣議決定) [2]では「博士人材のデータベースの活用」が位置付けられており、JGRAD は2017年3月末現在、国公私立大学42大学が参加し、試験運用を行っている¹。

JGRAD の試験運用では、NISTEP 及び参加大学の意見交換の場として、年2回、「博士人材データベースのパイロット運用に関する連絡協議会」(以下、「連絡協議会」)が開催されている [3]。2016年度第1回連絡協議会(2016年9月15日)において、参加大学から海外における人材のデータベースプロジェクトについての情報提供を求める意見があった。そこで、2015年12月、Science誌に論文 [4]が掲載され、国際的にも話題となった UMETRICS による人材研究について、同論文の責任著者である Weinberg 氏をお招きして講演会を開催することとなった。

UMETRICS は、ミシガン大学科学・イノベーション研究所 (Institute for Research on Innovation and Science; IRIS) を中心に、ビッグ・テン (Big Ten) と呼ばれる中西部の主要大学が中心となってコンソーシアムを組んで進める人材データベースである [5]。連邦政府(及びいくつかの連邦政府以外)の研究費で雇用されたすべての人のファイルが、センサスなど、多様なデータソースにリンクされている [6]。「研究の価値」を測定する上で、「人材が研究の最も重要な成果物の一つである」(Weinberg 氏)²という視点から、博士人材を含む多くの人材研究が行われている。

米国では NSF 等 が中心となり、国による全国的な博士号取得者に対するサーベイ調査「Survey of Earned Doctorates (SED)」 [7] が行われている。UMETRICS はこのような環境の下で、大学のコンソーシアムを主体とする人材データベースプロジェクトとして新たに立ち上がり、SED とのデータ連携も目指している [5]。我が国においても、2014年度より、博士人材に対するサーベイ調査として「博士人材追跡調査 (JD-Pro)」 [8] が開始されたが、それと並行して JGRAD の構築を大学参加型で進めている状況は米国の状況と酷似している。そこで、本稿は JGRAD の今後の方向性や、JGRAD と JD-Pro の将来的な連携等について検討するための参考として、UMETRICS (米国) と JGRAD (日本) の類似点や相違点について、文献調査や Weinberg 教授と意見交換等で得られた情報をもとに比較分析を行ったものである。

2. 調査結果 1: UMETRICS の概要について

(1) UMETRICS の特徴とデータ構造

UMETRICS [1]とは、2015年1月に、米国オバマ政権が進める STAR METRICS (Science and Technology for America's Reinvestment: Measuring the Effects of Research on Innovation, Competitiveness and Science) から分離・独立した人材データベースプロジェクトである [5]。「人材が研究の最も重要な成果物の一つである」という考え方から、研究の価値を測定するために「人

¹ 文部科学省の要請を踏まえ、「博士課程教育リーディングプログラム」の参加大学が含まれている。

² Weinberg 氏の講演より。第2部参照。

材」に着目して考える点³が特徴である。

UMETRICS のデータは、連邦政府（及びいくつかの連邦政府以外）の研究費で雇用された人のファイルが、広範なデータソースにリンクされた構造になっている。例えば以下のようなデータを特定できる [6]。

1) 給与記録・購入記録

UMETRICS のデータは、参加研究機関における連邦政府が支出した研究費（federally funded grants）すべてについて、ベンダーからの購入やサブ・コントラクターから給与が支払われた人がすべて捕捉されている [9]。データには、大学の給与記録が含まれ、連邦政府（及びいくつかの連邦政府以外）の研究費が、全支出期間にわたり、支払いを受けた個人に対して、所得としてどのように分配されたかの比率がわかる [6]。

2) ファンディング・ソース/職位の記録

UMETRICS のデータには、「ファンディング・ソース」（連邦政府の研究費配分機関等）や研究プロジェクトに雇用されたすべての人の「職位（job title）」が含まれている [9]。「職位」は「学部（faculty）」や「大学院生」を含めて、6つのカテゴリーに分類されており、研究プロジェクトにより雇用されたすべての人の役職（position）を特定することができる [6]。

3) 博士人材の特定

博士号取得者は、UMETRICS データと ProQuest（学位論文データベース）とマッチングさせることで特定することができる⁴。これらのデータには、学位論文の著者名、分野、学位授与機関、及び授与された学位が含まれる [6]。

4) 広範なデータリンク

UMETRICS のデータは、以下のような広範なデータとリンクしている [9]。

- ① 自然発生データ：学位論文、出版、引用、研究費（funded grant）、特許等
- ② 統計局（Census Bureau）が保有するサーベイ調査や行政データ：米国内における仕事の試用期間（job placements）、スタートアップ企業に関する情報

Weinberg 氏は、講演の中で、これら行政データ、自然発生データ、サーベイ調査データの特徴を図表 2-1 のように整理し、UMETRICS の有力なデータソースと、それから得られる情報の関係性を図表 2-2 のように説明している。また、文献によれば UMETRICS のデータソースとして、図表 2-3 が挙げられている [1]。

³ Weinberg 氏の講演より。第 2 部参照。

⁴ UMETRICS の「大学院生（Graduate students）」は「修士課程学生（Masters students）課程学生」を含むため、「博士号取得者（PhD recipients）」を特定するためには、「ProQuest」学位論文データベースとのデータマッチングが必要となる [6]。

図表 2-1 UMETRICS のデータオプション

	行政	自然発生	サーベイ (質問表)
例	UMETRICS : 人材登録、税のデータ	出版、特許、引用	SED (博士号取得者調査)
利点	全数をカバー、正確		トピックに自由度が大きい
負担	初期のクリーニングとリンク、限られたトピックについて可能	初期のクリーニングとリンク	実施コスト、正確性、回収率

出典 : Weinberg 氏講演資料「Data Option」より筆者作成

図表 2-2 UMETRICS の有力なデータソースと得られる情報

データソース	得られる情報
・ チェックブック、小切手帳、クレジットカードの口座	1) 課金先 : 支出先・支払額、支出の目的 2) 購買 : 装置、備品、人
・ 出版データ・論文データ・引用データ・特許	1) (研究の) 新規性、革新性 2) (研究の) インパクト
・ NIH (バイオメディカルリサーチ) のデータ	1) 医薬品の承認 2) 学術アウトカム
・ 国勢調査データ	(10年ごとの) 膨大なデータ
・ 雇用者側の所得・税務申告データ	1) 就業動向 (教育機関を出てどのような業界に就職するのか) 2) 企業規模、設立年数 3) 年間所得

出典 : Weinberg 氏の講演 (第2章) をもとに筆者作成

図表 2-3 UMETRICS の主なデータソース (文献調査)

<ul style="list-style-type: none"> ・ CIC 機関の STAR METRICS データセット (後述) ・ Survey of Earned Doctorates (SED) ・ 統計局 (Census Bureau) の LEHD (Longitudinal Employer-Household Dynamics⁵) 及び失業データ ・ ヘルスケア (Medicare) ・ 米国特許庁 (USPTO) の保有するイノベーション関連データ ・ ファイナンス (VentureXpert、CRISP、IPO データベース) ・ 論文データベース ・ 産業公告 (Industry Announcement) ・ 履歴書 (Curricula vitae)
--

出典 : The UMETRICS Initiative, "Universities: Measuring the ImpacTs of Research on Innovation, Competitiveness, and Science" [1] をもとに筆者作成

⁵ 直訳すると「長期的な雇用者/世帯のダイナミクス」。統計局の統計の一種。

(2) UMETRICS プロジェクトの運営

UMETRICS プロジェクトは、米国中西部の主要大学(Big Ten)が設立した大学間コンソーシアム CIC (Committee of International Cooperation) (図表 2-4) [1]の連携協力がベースとなって進められている [5]。その中核機関は、ミシガン大学科学イノベーション研究所 (IRIS) であり、IRIS は加盟大学間のコンソーシアムである。

IRIS は、“2009 年にバラク・オバマ大統領が実施した研究開発に対する 520 億ドルの支出を含む経済刺激策により創出された研究業務を追跡するために、ジュリアン・レーン氏らが NIH と NSF において創設した連邦政府の IRIS はプログラムの「末裔」である” (Nature,2016) と言われている [10]。IRIS への参画は任意であり、その運営は各大学が年会費を払い担保している。文献では、IRIS の所長のジェイソン・オーエンスミス氏によれば、IRIS は (総額で) 研究開発資金提供の 150 億ドルを意味する 24 大学と既にパートナーシップを組み、目標としては、代表的な州立及びランド・グラント大学 (flagship state and land-grant universities [11]⁶)、並びに連邦研究開発費を最低 1 億ドル以上得ているすべての機関を包含することであると記載されている [10]。このスコープは、すべての連邦政府が支出する研究開発の 90%以上のデータを含むことになると解説されている [10]。

なお、Weinberg 氏によれば、UMETRICS は、現在、一流大学を含む 50 大学がプロジェクトにコミットし、実際に 19 大学がデータを登録している状況にあり、ミシガン大学をハブとして、そのノードがオハイオ州立大学やニューヨーク大学等に存在するとのことである⁷。

図表 2-4 CIC メンバー大学

シカゴ大学、イリノイ大学、インディアナ大学、アイオワ大学、ミシガン大学、 ミシガン州立大学、ミネソタ大学、ネブラスカ大学リンカーン校、ノースウエスタン大学、 オハイオ州立大学、パデュー大学、ウィスコンシン大学マジソン校

出典：The UMETRICS Initiative, “Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science” [1]をもとに筆者作成

(3) UMETRICS の目的

UMETRICS の目的は、「人材」に着目して研究の価値 (Value) を測定することである⁸。大学が UMETRICS に参加するメリットは、UMETRICS を用いた研究から得られる報告書を受け取ることにより、どのような形で (大学の) プログラムを改善したらよいのか、(大学として) どのような活動を行えばよいのか等、有用な情報を得ることができ、また、大学はこれら (UMETRICS) のデータにアクセスすることができることである⁹。すなわち、UMETRICS は「エビデンスベースの意思決定」や「信頼性のある論拠 (credible advocacy) をもとにした研究マネジメント」を報告するための「科学的なフレームワーク」を構築するものである [1]。

文献では、UMETRICS の目的を「研究政策コミュニティがアクセスできる結果を提示」し [1]、

⁶ 「land-grant universities」とは、1862 年及び 1890 年モリー法 (Morrill Acts of 1862 and 1890) により連邦の土地を与えられた米国の高等教育機関のこと (出典：best value schools, What is a Land-Grant University? <http://www.bestvalueschools.com/faq/what-is-a-land-grant-university/>) [11]

⁷ 2017 年 3 月 3 日の Weinberg 氏インタビュー時点。

⁸ Weinberg 氏の講演より。第 2 部参照。

⁹ Weinberg 氏の講演より。第 2 部参照。

「大学が支援者 (donors) や政策決定者 (policy maker)、あるいは鍵となるステークホルダーに対して、政策的な議論や研究の成果を効果的に伝達する能力に寄与」することであると説明されている [1]。この一例として、Weinberg 氏は講演の中で、UMETRICS を用いたオハイオ州やインディアナ州における研究投資の地域分布に関する研究事例をあげ、州議会や知事、上院・下院議員への説明のための情報として有効であることを説明している。

また、UMETRICS のもう一つの目的は、「研究行政官に直接役に立つツールを開発すること」である [1]。このため、個々のプロジェクトは、研究のインパクトを定量化するためのアルゴリズムや、パフォーマンスの高いチームを強化し、(パフォーマンスの高いチームを) デザインするためのツールを含んでいる [1]。

図表 2-5 UMETRICS の主な目的

- | |
|--|
| <ul style="list-style-type: none">・ 科学政策研究者によるエビデンスベースの意思決定と信頼性のある論拠が可能な研究管理報告のためのフレームワーク構築・ 大学が、政策的な議論や研究成果を、政策決定者やステークホルダー等に効果的に伝達するための能力構築・ 研究行政官に役立つツールセットの開発 (研究のインパクトの定量化のためのアルゴリズム開発、パフォーマンスの高いチームデザイン等) |
|--|

出典：The UMETRICS Initiative, “Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science” [1]をもとに筆者作成

(4) UMETRICS のミッション

図表 2-6 は、UMETRICS が取り組む「研究ワークフォースの構造」、「協力の本質と進化」、及び「研究成果」の3つのクリティカルな課題についてまとめたものである [1]。

図表 2-6 UMETRICS の3つのミッション

- | |
|---|
| <ol style="list-style-type: none">1. 研究ワークフォース：“近年、科学のワークフォースの構造が急速に変化していることは明らかであるが、データは不備であり、この変化が科学の生産性にどのような変化を与えているのかについてはほとんどわかっていない。”このため、“STAR METRICS データをもとに、これらの変化について、事実のセットを確立する” [1]。2. 研究協力：“複雑な科学的疑問に答えるためには、科学者は、複雑な学際的チームの中で協力することが、ますます必要となっている。”そこで、“効率の高い研究チームを作る科学的なベースを発展させ、効率の高いチームを構築するためのツールを開発する” [1]。3. 研究成果：“研究成果は、歴史的には研究資金活動の手製の報告書を機械的に文書化することで捕捉してきた”。しかし、“UMETRICS のアプローチは、人材中心で、科学者や彼らの協力を分析フレームワークの中心としている” [1]。 |
|---|

出典：The UMETRICS Initiative, “Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science” [1]をもとに筆者作成

(5) UMETRICS の誕生の背景—STAR METRICS の進展と UMETRICS からの独立

STAR METRICS—邦訳すれば「米国再投資のための科学技術：イノベーション、競争力、科学に関する研究の効果測定」—とは、2009年、米国復興・再生法（the American Recovery and Reinvestment Act of 2009 ; ARRA）の要請により、米国経済を刺激する努力の中で誕生した連邦政府のプログラムであり [1]、公的研究開発投資が雇用に与える影響を測定する目的で開始されたプログラムである [5]。UMETRICS は、この STAR METRICS プロジェクトにより可能となった大規模で、自動化された研究事業体（research enterprise）に関するデータプラットフォームを基盤としている [1]。

STAR METRICS は、連邦の科学への投資と、その結果について、詳細な情報を提供するための情報プラットフォームを構築するために始められたボランタリーな協力であり、ボトムアップなアプローチを取っている [1]。STAR METRICS には、約 100 の機関と 6 つの省庁（Agency）がプログラムに参加し、米国大学協会（AAU） [12]¹⁰、公立大学協会（APLU） [13]¹¹、米国医学カレッジ協会（AAMC） [14]¹²及び連邦デモンストレーションパートナーシップ（The Federal Demonstration Project; FDP） [15]¹³に支援されている [1]。このような連邦政府のプログラムへの多くの大学参加による膨大なデータセットの蓄積が、IRIS による UMETRICS 誕生の基盤となっている。

STAR METRICS は、2015 年、ツールとしての機能に特化した STAR METRICS レベル II /Federal RePORTER と、具体的な研究開発投資の社会的影響の測定を目指す UMETRICS とに分離した [5]。UMETRICS は、STAR METRICS プロジェクトにより可能となった大規模で、自動化された研究事業体（research enterprise）に関するデータプラットフォームを基盤とし [1]、多くの大学から得られた STAR METRICS データを含んでいる [1]。以上をまとめると、図表 2-7 のようになる。

文献によれば UMETRICS は、研究に関するビッグ・データを結合し、発掘し、分析する方法とツールにおいて、STAR METRICS よりも、さらに進んでいる [1]。例えば、データ収集は、UMETRICS チームのメンバーが開発した自動化ツールを導入しており、異なる情報源（ソース）から、データをシステムティックに収集し、取り入れている [1]。上述の図表 2-3 のような各種情報源から情報を抽出・構造化し、曖昧さをなくして個々のデータ及び CIC 機関から得られた STAR METRICS データにリンクしている [1]。それにより、UMETRICS は、大規模で、構造化され、結合され、更新されたデータセットを構成しており、様々なレベルで、研究組織を高質かつ大規模で分析することを可能にしている [1]。

¹⁰ The Association of American Universities (米国大学協会) (ホームページ <https://www.aau.edu/>) [12]

¹¹ The Association of Public and Land-grant Universities (APLU) (公立/ランド・グラント大学協会) (ホームページ <http://www.aplu.org/>) [13]

¹² The Association of American Medical Colleges (米国医学カレッジ協会) (ホームページ <https://www.aamc.org/>) [14]

¹³ The Federal Demonstration Project (ホームページ <http://sites.nationalacademies.org/PGA/fdp/index.htm>) [15]

図表 2-7 STAR METRICS の特徴（文献調査）

項目	概要
1. 背景	・ 2009 年米国復興・再生法（ARRA）の要請により、米国経済を刺激する努力の中で誕生 [1]
2. 目的	・ 連邦の科学投資とその結果について詳細な情報を提供する情報プラットフォームを構築 [1] ・ 公的研究開発投資が雇用に与える影響を測定する目的で開始 [5]
3. アプローチ	・ プロジェクトは、ボランタリーな協力、ボトムアップなアプローチがベース [1]
4. 参加機関	・ 約 100 の機関と 6 つの省庁がプログラムに参加 [1] ・ AAU、APLU、AAMC 及び連邦デモンストレーションパートナーシップによる支援 [1]
5. データ	・ 大学および研究機関の財務や人事の運用データを比較・分析できるようにデータの単位等を揃えて統合したものが基礎 [5] ・ 米国国立科学財団（NSF）や米国国立衛生研究所（NIH）、米国エネルギー省（DOE）等省庁の採択研究費のデータが連結され、研究者、ポスドク、事務員等個々のデータとリンクされている [5]
6. データの利用	・ レベル I に参加している大学および研究機関の研究者は STAR METRICS のデータを使用して研究することができる [5] ・ 目的や研究範囲によりデータは加工・制限される [5]
7. プロジェクト構成	(1) レベル I（2009） ・ 研究開発費によって生み出される大学・研究機関における雇用と関連企業への資金波及を測定することから始まった [5] ・ 特に公的研究開発費が「研究機関の雇用に与える影響」と「波及効果として企業の支払いに与える影響」について着目 [5] (2) レベル II（2012） ・ 広範な研究開発投資の効果を測るべく、特許、論文引用数、研究上の成果等より多くの成果のデータと研究開発投資のデータを結合 [5]
8. プロジェクトの継承	2015 年に以下の 2 つのプロジェクトに分離。 (1) STAR METRICS レベル II / Federal RePORTER：ツールとしての機能に特化 [5] (2) UMETRICS：具体的な研究開発投資の社会的影響の測定を目指す [5]

出典： 1) The UMETRICS Initiative, “Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science” [1]、2) 国立研究開発法人 科学技術振興機構 研究開発戦略センター（CRDS）. 調査報告書 米国「科学イノベーション政策のための科学」の動向と分析 [5]をもとに筆者作成

3. 調査結果 2: 博士人材データベース(JGRAD)の現状と課題

米国の UMETRICS 同様、わが国においても大学参加型のコンソーシアムをベースとして博士人材データベース (JGRAD) を構築している。これまでの試験運用の知見を踏まえ、2016 年度よりプロジェクトの見直しと、本格運用に向けて登録項目の精査等、機関間の連携協力等が推進されている [3]。ここでは UMETRICS と JGRAD の比較を行うための基礎調査として、JGRAD のこれまでの経緯や現状について概観する¹⁴。

(1) JGRAD の構築の経緯と大学参加状況

JGRAD は、2011 年、文部科学省の「政策のための科学 (SciREX)」事業の一部として、科学技術・学術政策研究所 (NISTEP) の研究開発プロジェクトとして構想が開始された [3]。その目的は、博士課程修了者のキャリアパスを追跡・把握し、社会での活動状況を可視化することにより、人材政策の基礎となるエビデンスを得ることである。

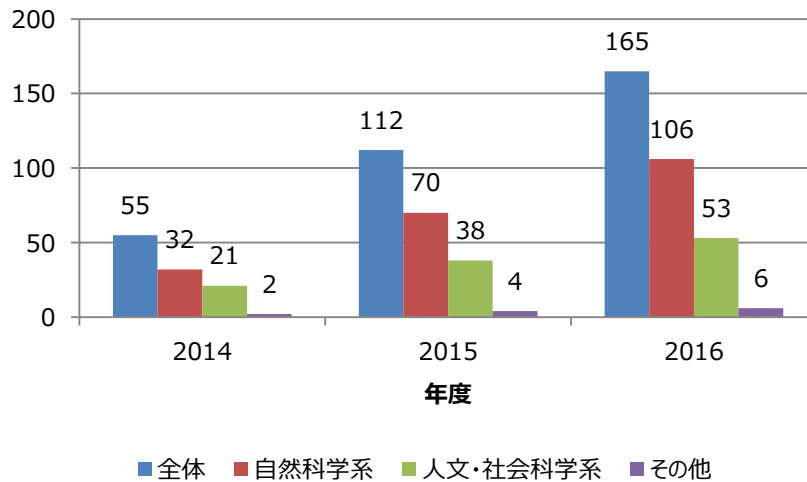
JGRAD は、博士課程在籍中から登録を開始し、博士課程修了後も定期的にデータ更新を登録者自身にお願いしている。2014 年 4 月より、当初 12 大学で試験運用を開始したが、その後、「第 5 期科学技術基本計画 (2016 年 1 月 22 日、閣議決定)」において政策的位置づけが強化された [3]。また、2017 年 2 月より、文部科学省の要請を受け、「博士課程教育リーディングプログラム」(以下、「リーディングプログラム」) の参加大学にも JGRAD への登録をお願いすることとなった。これにより、博士人材プログラムの PDCA サイクルにおいて、JGRAD はエビデンスに基づきプログラムをチェックし (C: Check)、改善につなげる (A: Action) 上で、一定の役割を担うようになったものと考えられる。その結果、2017 年 3 月末現在、参加大学数は 42 大学に拡大している¹⁵。

大学数だけでなく、JGRAD に参加する研究科数も着実に増加しており、2017 年 3 月末現在、165 研究科 (自然科学系 106、人文・社会科学系 53、その他 6) が参加している¹⁶ (図表 3-1)。ただし、研究科数を見る限り、自然科学系が人文・社会科学系の約 2 倍と多く、今後は人文・社会科学系の研究科に対しても、協力を呼びかけていく必要がある。

¹⁴ JGRAD のデータ取得は毎月、月当初の営業日のデータを前月末のデータとして集計し、登録状況の確認を行っている。本報告書では、原則、2017 年 3 月末 (データ取得日 2017 年 4 月 3 日) のデータをもとに分析している。

¹⁵ 文部科学省の要請を踏まえ、「博士課程教育リーディングプログラム」の参加大学が含まれている。

¹⁶ 改廃された研究科は含まない。また、参加大学において、参加研究科の確定が遅れている場合や、プログラム単位での参加を行っている場合、数字に含まれない。



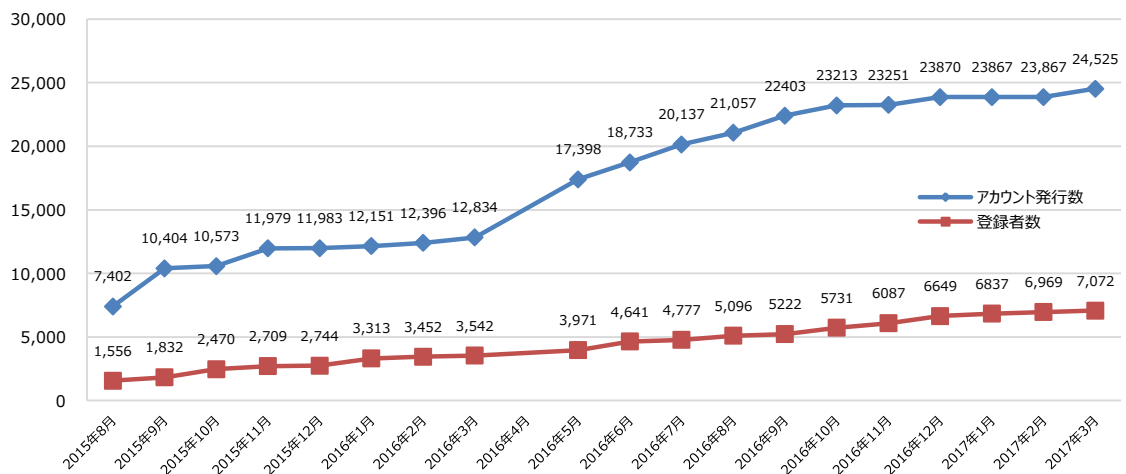
図表 3-1 年度ごとの参加研究科累積数 (2017年3月末現在¹⁷⁾
出典：博士人材データベース (JGRAD) 事務局調べ¹⁸

(2) JGRAD の登録状況

1) アカウント発行数と登録者数 (ログイン数)

参加大学の拡大の結果、登録者も着実に増加している。潜在的なキャリアパス把握の対象であるアカウント発行数は、2017年3月末現在、24,525人、実際にJGRADにログインしたことのある登録者 (ログイン数) は7,072人である (図表 3-2)。JGRAD 参加大学及び参加研究科の増加に伴い、アカウント発行数が順調に増加するとともに、登録者数もリニアに増加している。

なお、事業管理上、アカウント発行から実際に登録が行なわれるまでの期間 (タイムラグ) を約4ヶ月として、登録者数 (ログイン数) を4ヶ月前のアカウント発行数で割った値を「登録率」と便宜的に定義している。この登録率は、2016年10月以降、約30%前後で推移している。



図表 3-2 博士人材データベースのアカウント発行数と登録者数 (ログイン数)
出典：博士人材データベース (JGRAD) 事務局調べ (2017年3月末現在¹⁹⁾

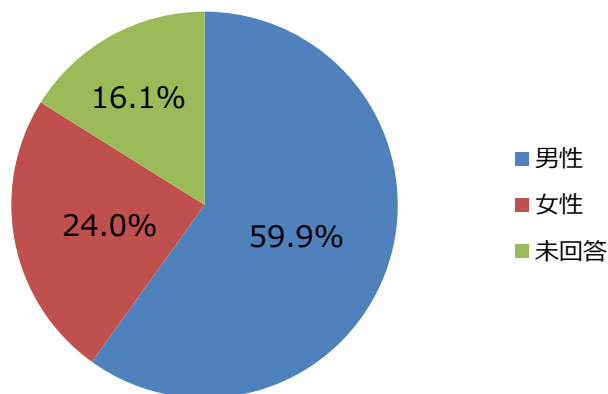
¹⁷ 2017年4月3日集計

¹⁸ 2016年度の博士人材データベース事務局は、株式会社 日立コンサルティングに委託した。

¹⁹ 2017年4月3日集計

2) 登録者のジェンダー分布

2017年3月末現在（2017年4月3日集計）、7,072人の登録者のジェンダー構成比を見ると、「男性」が59.9%、「女性」が約24.0%（性別未回答者²⁰が16.1%）である（図表3-3）。



図表 3-3 JGRAD の登録者のジェンダー比（n=7,072）（2017年3月末現在）

出典：博士人材データベース（JGRAD）事務局 調べ

（注）図中の%は、男性59.91%、女性23.98%、未回答16.11%を小数点第2位で四捨五入している

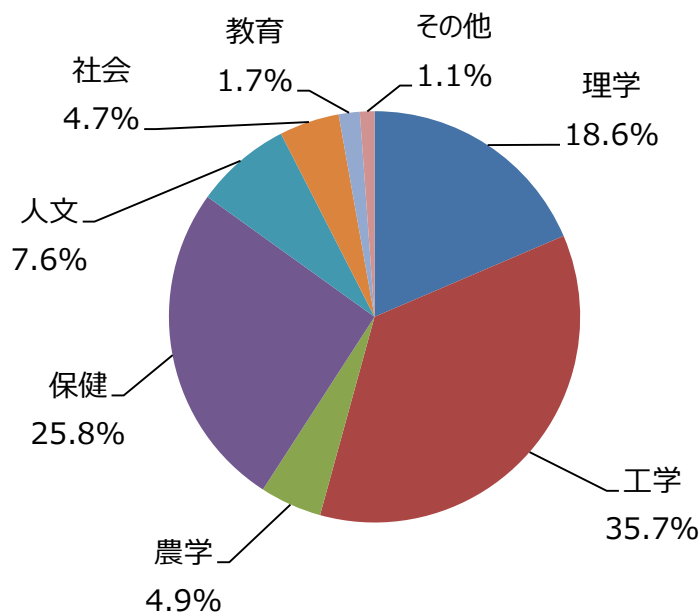
3) 登録者の分野分布

JGRADの登録者を分野別²¹に見ると、「理学」が18.6%、「工学」が35.7%、「農学」4.9%で、これに「保健」が25.8%を加えた「自然科学系」の博士人材が、登録者の85.0%を占める（図表3-4）。

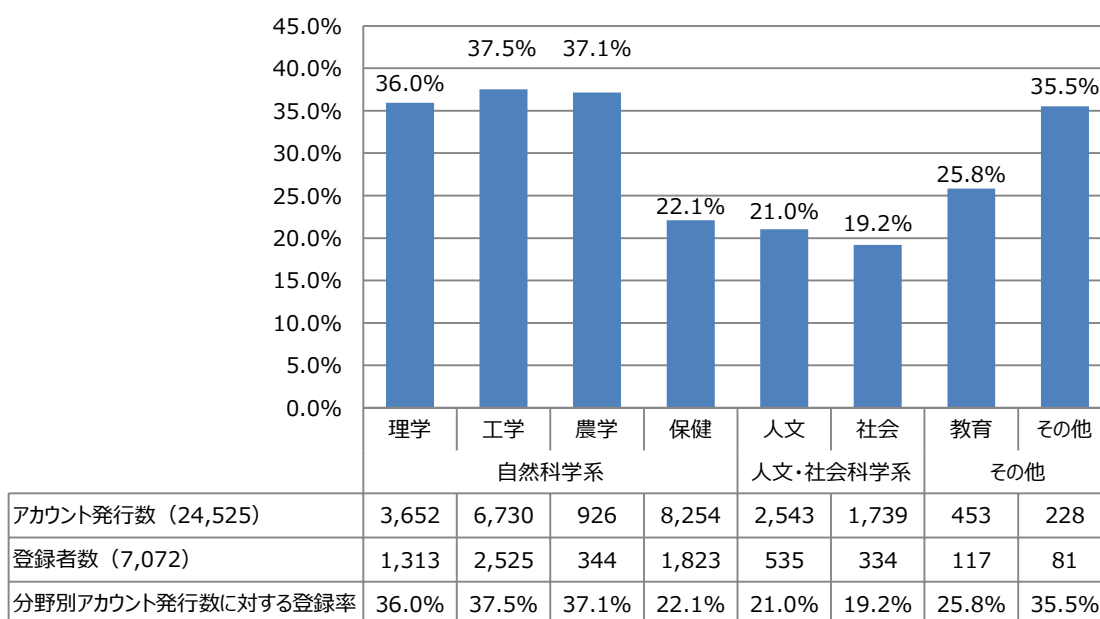
一方、アカウント発行数に対する登録者数（ログイン数）の割合を「登録率」として定義し、分野別に見ると、「工学（37.5%）」「農学（37.1%）」「理学（36.0%）」の順に高く、「保健（22.1%）」や「人文（21.0%）」「社会（19.2%）」では低い。（図表3-5）すなわち、専門分野により登録状況には差異があることがわかる（図表3-5）。ただし、ここで言う「登録率」は1）のように4ヶ月のタイムラグを考慮しておらず、同月の登録者数（ログイン数）を同月のアカウント発行数で割った値である。

²⁰ 人材関係の調査では、性別に関して回答を求めるものが多いが、JGRADでは、性別未回答者がいる。また、2016年度に実施された博士人材データベースの本格運用に向けた登録項目の精査においても、性別の回答に関してはLGBT対応を考慮することとなった。

²¹ 本報告書で利用する分野分類は、登録者が所属する研究科の分野に基づいている。なお、理工学研究科等、複数の専門分野が統合された研究科の場合は、所属する登録者が登録した専門分野のうち一番多い分野を研究科の分野として選択している。



図表 3-4 JGRAD の登録者の分野別比率 (n=7,072) (2017年3月末現在)
出典：博士人材データベース (JGRAD) 事務局 調べ



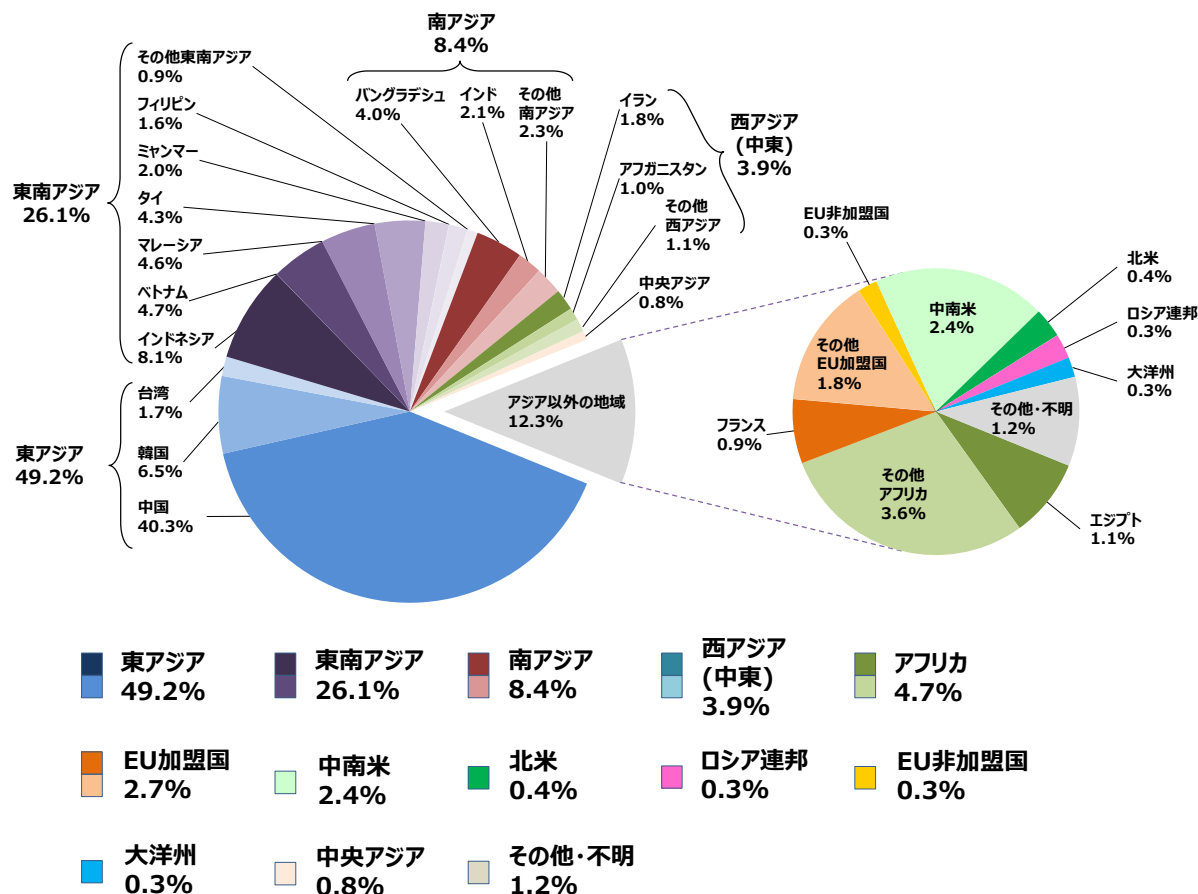
図表 3-5 分野別アカウント発行数に対する登録率 (2017年3月末現在)
出典：博士人材データベース (JGRAD) 事務局 調べ

4) 留学生の登録状況

JGRAD の特徴の一つは、我が国の博士課程に在籍している留学生や、我が国の博士課程を修了し、海外で活躍する研究者も追跡対象としていることである。このため、日本語・英語の2か国語での入力が可能であり、また JGRAD を用いたアンケート調査や連絡・情報提供も、原則、日本語・英語で行われている。

JGRAD では、このような取組の結果、2017年3末日現在、1,455人の留学生が登録されてお

り、これは全登録者の20.6%（日本人61.2%、未回答18.2%）にあたる。これを出身地域別にみると、留学生のうち「東アジア（49.2%）」及び「東南アジア（26.1%）」で75.3%を占める。これに「南アジア（8.4%）」「西アジア（3.9%）」「中央アジア（0.8%）」を加えたアジアからの留学生は約9割（88.4%）である（図表3-6）。



図表 3-6 留学生の地域別割合²² (n=1,455) (2017年3月末現在)

出典：博士人材データベース (JGRAD) 事務局 調べ

(3) JGRADによる情報提供等

JGRADのもう一つの特徴は、博士人材のキャリアパスを「調査」するための機能だけでなく、構築の当初から、登録者のインセンティブ向上のための施策として、登録者への情報提供事業（サービス）等も視野に入れながら進めてきたことである。これはJGRADの収集する情報が個人情報であり、かつ収集方法が登録者自身による入力というサーベイ調査型の手法に依存していることによる。調査機能と情報提供機能を一つのデータベースで達成しようとしたところにJGRADの難しさが内在している。

現在検討されている主なサービス機能としては、登録者に役に立つ「情報提供」と登録価値の向上のための「新機能」に分かれる。具体的な検討状況は以下のとおりである。

²² グラフに利用した国名及び地域名は、基本として、国連による世界地理区分“United Nations, Standard country or area codes for statistical use (M49) (<https://unstats.un.org/unsd/methodology/m49/>)”に基づき作成している。なお、欧州に分類される国々については、欧州連合の加盟、非加盟、及びロシア連邦で分類している。

1) 求人情報の提供

求人情報の提供については、登録者や大学からの要望の高い情報の一つである。そこで、2016年9月より、国立研究開発法人 科学技術振興機構 (JST) の保有する JREC-IN Portal とのデータ連携を開始し、JGRAD に登録した情報をもとに、JREC-IN Portal で検索された求人情報を JGRAD 上で自動表示する「オーダーメイド型」の求人情報サービスを開始した [3]。

JGRAD を用いた先行研究によれば、博士課程在籍者が課程修了後、アカデミアに進むか、非アカデミアに進むかを選択する場合、アカデミアと非アカデミアでは選択の時期に違いがあることが示唆されている [16]。したがって、博士課程在籍時の早期から、自らの専門分野に関する求人情報に触れておくことは、博士課程在籍者のキャリア形成に対する意識を喚起する上でも意味があると考えられる。

また、これまで研究者を中心に求人情報を提供してきた JREC-IN Portal の利用者層が博士課程在籍者まで拡大する効果も期待される。

2) ロールモデルの配信の検討

登録者がキャリアパスを考える上で参考となるようなロールモデルの情報発信については、登録者の関心の高い事項の一つである。各大学においてもキャリア・センター等が主催する講演会等が行われているが、忙しい研究の合間を縫って、なかなか参加しにくいという声も聞かれ、JGRAD による配信の要望も大きい²³。

ロールモデルのコンテンツは、雑誌の記事や大学・関連機関のホームページなど、様々なところで紹介されており、また各大学でも講演会等の実施や、コンテンツの開発等が進められている。問題は、博士人材のニーズに合わせてコンテンツが整理され、ワンストップで検索・閲覧できるような仕組みになっていないため、多くのコンテンツが利用しにくい状態で存在することであると考えられる。また、コンテンツを開発する側の意図と、それを利用する博士人材のニーズが、必ずしも一致しているとは限らないことや、コンテンツがキャリアパス形成にどのように役立ったかを測定することが難しく、プロジェクトの政策効果を測定しにくいことである。

そこで大学や関係機関との連携により、開発されたロールモデル・コンテンツを収集し、分類・整理して JGRAD 上で閲覧できるようになれば、登録者の利便性も向上し、キャリアパス形成の参考になるのではないかと考えている。また、将来的には、コンテンツに対するアクセス記録等を分析することで、どのような属性を持つ登録者が、どのようなコンテンツにニーズがあるのか等、知見を蓄積し、登録者のニーズにかなったコンテンツの開発や配信を行うことができるようになることが期待されている。

3) JST の研究者総覧データベース (researchmap) 等とのデータ連携

先行研究によれば、博士課程在籍者のうち、約6割は課程修了後アカデミアに進む [8]。このことから、博士課程在籍時に JGRAD に登録したデータが、本人の希望により、JST の研究者総覧データベース (researchmap) 等、他の研究者データベースに引き継ぐことができれば、登録者の再入力負荷の低減や利便性の向上につながるものと期待される。

このような観点から、JGRAD と researchmap のデータ連携を図ることとし、2016年度に「博士人材データベースの今後の運用に向けた検討委員会 (座長：斎藤貴浩大阪大学教授)」を設置し、登録項目の精査など、検討が行われた [3]。その結果をもとに、2017年度より、具体的な項目整備

²³ 2015年に筆者らが行った登録者へのインタビューによる。

に着手する予定である。

項目整備の中で、例えば、論文成果等データベースに登録された研究業績等を指定のフォーマットで出力する機能（履歴書機能）等、登録者からの要望の高い新機能についても検討する予定である。また、将来的には、JST以外の機関や大学の保有するデータベースともデータ連携を進め、データインポート等、登録者の入力負荷が少ない登録方式等についても検討している。

4. 分析及び考察：UMETRICSとJGRADの比較分析

UMETRICSとJGRADは、いくつかの点でプロジェクトとして類似した特徴を有している。ここでは、両者の類似点・相違点を比較分析してみたい。

(1) プロジェクトを取り巻く政策的環境

1) サーベイ調査実施下での人材データベースプロジェクトの発展

UMETRICSおよびJGRADは、ともに、国レベルでのサーベイ調査が存在する中で誕生した人材データベースプロジェクトであり、両プロジェクトを取り巻く政策的環境は類似していると考えられる。例えば、米国では、毎年、NSF等による博士課程修了時の調査(Survey of Earned Doctorates; SED)が、また、2年に一度、博士課程修了後の追跡調査 (Survey of Doctorate Recipients; SDR) がサーベイ調査として実施されている (図表 4-1) [7]。UMETRICSは、このような環境の中で、大学コンソーシアムによるデータベース型の人材プロジェクトとして後発的に (2015年) 開始された。このようなUMETRICSとSED等サーベイ調査の関係は、2014年より、サーベイ型の「博士人材追跡調査 (JD-Pro)」 [8]が行われる中で、大学コンソーシアム型でJGRADの試験運用が行われている我が国の状況と酷似している。

図表 4-1 米国における博士課程修了者調査の概要

		博士課程修了時の調査	博士課程修了後の追跡調査
調査名		SED(Survey of Earned Doctorates)	SDR(Survey of Doctorate Recipients)
調査基盤	調査の統括・後援	NSF、NIH、USED、USDA、NEH、NASA	NSF、NIH
	調査の実務担当	シカゴ大学のNORCおよび博士号を授与する大学	シカゴ大学のNORC(出身大学は原則として関与しない)
	調査開始年	1957年[1957年7月～1958年6月修了者]	1973年
	調査周期/時期	毎年調査/博士課程修了時点	2年に一度の隔年調査/調査時点は10月
調査概要	調査対象者	研究関連の博士課程修了者全員(M.D.等の実務に関する学位は除くが、人文関連も含む全分野)	米国の研究関連の博士課程修了者(人文関連を除く分野)で75歳以下の米国内在住者
	母集団の人数	48,069人【2010年度調査:418機関】	752,000人【2008年SDR調査による推定値】
	標本の抽出方法	なし(悉皆調査)	前回のSDRの対象者約9割と直近2年のSEDの対象者約1割を層化抽出(母集団の推定値の5.3%に相当する40,093人に調査実施)
	回収率	92.9%【2010年度調査:418機関】	母集団復元用のウェイトをつけない場合は80.7%(ウェイトをつけると80.5%)【2008年SDR調査】
データ間連携 情報伝達・ 非回答者の 属性把握		性別、生年月日といった博士課程修了者の個人データはSEDで始めて取得(SED費回答でも大学からの基本情報の提出によって情報を補完)	SEDや過去のSDRの調査結果と個人単位でマッチングを行い、組み合わせて分析可能(SDRではSEDで調査済のことは調査しないが、マッチングのミス防止のため、生年月日は再度調査)
近年の調査・データ 利用の変化		2009年の調査結果から図を中心にした概要や対話式ウェブサイトを作成し、一般向けに簡潔な情報提供を開始	2003年から米国で博士号を取得後、国外に移動した者を調査するISDR(International Survey of Doctorate Recipients)を実施

出典：斉藤経史、伊藤裕子、富沢宏之。「博士課程修了者の状況把握のシステム設計—博士人材データベースの構築背景および海外の博士課程修了者調査—」,文部科学省科学技術政策研究所、調査資料-216, 2012. [7]

2) メトリックス型データベースとサーベイ調査の有機的な連携

先行研究では「UMETRICS は、米国の大学で博士号取得者の動向を毎年追跡調査している (SED)との連携に向けて作業中であり、政府が博士号取得者を支援することの長期的な社会影響を測定することを目標にしている」ことが報告されている [5]。そこで、大学間コンソーシアムによるデータベース構築 (メトリックス型) とアンケートによるサーベイ調査の将来的な関係性について、Weinberg 氏に問題提起した。Weinberg 氏は、UMETRICS とサーベイ調査の関係は、「独立」より「連携」であることを強調し、「相互に補完しあうもの」としてデータの連携・統合が必要であると説明した²⁴。また、講演においても、UMETRICS のデータ連結には、サーベイ調査データが含まれることを説明している。なお、米国においても、データ連結には慎重な機関もあるらしいが、それでも UMETRICS は米国統計局のデータベースや SED との連携等、行政機関の垣根を越えた検討を行っている。

この状況は我が国にも重要な示唆を与えるものである。わが国の場合、現在、博士人材に対するサーベイ調査が多数存在し、大学や回答者の大きな負担になっていることが問題となっている。この原因は、調査主体が多く、異なる主体が似たようなデータを収集し、個別に保持されているからであると考えられる。各機関が行うサーベイ調査に伴う回答者の負担を軽減していくために、例えば JGRAD と JD-Pro の関係を整理し、将来的に、両者のデータ連携・統合等を実現していくことが一つの検討課題となっている。その意味で、米国における UMETRICS と SED 等のサーベイ調査の関係は注目に値するところである。我が国においても、UMETRICS を一つのモデルとし、JGRAD とサーベイ調査のデータ連携を促進することは、回答者の負荷低減につながるだけでなく、収集されたデータを有効活用し、より高度な分析へと発展させる意味でも重要であると考えられる。

2016 年度は「博士人材データベースの今後の運用に向けた検討委員会」を設置し、他のデータベースとの連携を図るため、JGRAD の本格運用に向けた登録項目の精査が行われた。この中で、上記の視点から、JGRAD がデータベースとして確保すべき登録項目 (プラットフォーム) と、JGRAD のもとで実施すればよいアンケート調査項目 (サーベイデータ) についても精査が行われた。これは Weinberg 氏の説明するメトリックス型データベース (UMETRICS) とサーベイ調査 (SED) の連携に向けた関係整理に他ならない。今後は、JGRAD と関係機関との協力の輪を広げ、データ連携を促進することで、サーベイ調査との有機的な連携による回答者の負担軽減を図り、より正確で利便性の高いデータセットを構築していく必要があると考えられる。

3) メトリックス型データベースとサーベイ調査の国際動向

人材政策における伝統的なサーベイ調査と、近年成長しつつあるメトリックス型のデータベースの関係については、米国や我が国のみならず、今後、より国際的な視点から検討する必要がある。米国に限らず、欧州諸国でも高等教育機関卒業直後の調査や卒業後の追跡調査が広く行われており、例えば、英国においては図表 4-2 のような HESA (Higher Education Statistics Agency) によるサーベイ調査が行われている [7]。一方で、Weinberg 氏は質疑の中で「オーストラリアもこの方向 (人材データベース) で一定額の投資を行っている」ことに言及している²⁵。

UMETRICS の事例でも明らかのように、メトリックス型の人材データベースの開発には、データ収集やデータリンクに関する技術の発展が不可欠であることは言うまでもない。また、Weinberg 氏は、「人または集団として、このようなこと (人材データベースを構築し研究するこ

²⁴ Weinberg 氏へのインタビューによる。

²⁵ Weinberg 氏の講演より。第 2 部参照。

と)に価値を見出すかどうか」、また「障壁を取り除き、コンソ(コンソーシアム)を組む、又は機関連携を行う能力による」とコメントしている²⁶。すなわち、データベース構築を進めるための政策的・社会的環境の整備が重要なポイントであり、各国の国情の違いもあるものと考えられる。

このような状況の中で、今後、博士人材に対するサーベイ調査とメトリックス型のデータベースプロジェクトが、国際的にどのように整理され、どのように展開していくのか。Weinberg氏は、講演の中で、人材データベースの構築が各国に拡大することで、将来的に国際的な連携協力が可能となれば、「研究の価値」を国際的な視点から捉えていくことが可能になると期待を込めて言及していた²⁷。したがって、我が国においても国際的な動向について引き続き情報収集しながら、JGRADとサーベイ調査の関係について検討していく必要がある。

図表 4-2 英国における高等教育機関の卒業後調査の概要

		高等教育機関卒業直後の調査	高等教育機関卒業後の追跡調査
調査名		Early DLHE (Destinations of Leavers from Higher Education)	Longitudinal DLHE (Destinations of Leavers from Higher Education)
調査の統括・後援		HESA (HEFCE、RCUK等の政府機関の後援とデータ共有)	
調査基盤	調査の実務担当	学位を授与する各大学 (大学が調査会社を利用することもある)	過去3回はIFF Research社 (出身大学は卒業生の連絡先を提供)
	調査開始年	FDSの後継として2002年 [2002年8月～2003年7月卒業生]	2002/03の卒業生に対して、約3年半後の2006年に第一回調査を実施
	調査周期/時期	毎年(年2回実施して1年分にまとめる)/卒業の約6ヶ月後	2年に一度の隔年調査/卒業の約3年半後
	調査対象者	高等教育の卒業生全員 (2010/11卒業生調査までは入学前の居住地がEU外の者および職業資格に関連する課程は除く)	Early DLHEの回答者全員(Early DLHE回答時に追跡調査への協力を拒否した者には調査を行わない)
調査概要	母集団の人数	2010/11卒業生全体:526,225人 うち博士課程修了者:10,395人 【2010/11卒業生調査:165機関】	2006/07のEarly DLHEに回答し、追跡調査を拒否しなかった者:332,110人 うち連絡先の判明者:220,841人
	標本の抽出方法	なし(悉皆調査)	3種の調査手段を段階的に用いるA群を層化抽出し、他はメールのみで調査(Early DLHE回答時に追跡調査を拒否しなかった者は悉皆調査)
	回収率	全体の回収率:78.8% フルタイム学生回収率:79.9% パートタイム学生回収率:74.7% 【2010/11卒業生調査:165機関】	調査依頼を行った者全体に対する回収率:22.2% A群:回収率43.7% A群以外:回収率12.8% 【2006/07卒業生に対する調査】
データ間連携 情報伝達・ 非回答者の 属性把握		HESAは学生時から同一人物を識別したIDを全学生に付与し、在学時のStudent recordから個人IDでマッチングし情報を補完 (Early DLHE内で性別や生年を調査しない)	在学時のStudent recordおよびEarly DLHEと個人単位でマッチングし情報を補完(Longitudinal DLHEでは、性別、生年、卒業直後の状況を調査しない)
近年の調査・データ 利用の変化		2011/12卒業生調査から入学前の居住地がEU外の者および職業資格に関連する課程が調査対象に追加	第1回調査は標本調査であったが、2008年11月を調査時点とした第2回調査から、追跡調査を拒否しなかった者全員に調査依頼

出典：斉藤経史、伊藤裕子、富沢宏之、「博士課程修了者の状況把握のシステム設計—博士人材データベースの構築背景および海外の博士課程修了者調査—」,文部科学省科学技術・学術政策研究所、調査資料-216,2012.

(2) プロジェクトの形態と大学参加の規模

UMETRICSとJGRADは、共に研究所(IRISとNISTEP)をハブとした、参加大学によるコンソーシアム型のプロジェクト構成をとっている。したがって、サーベイ調査(SEDやJD-Pro)のような全数調査ではなく、データベースに登録された有力大学を調査対象(したがって部分調査)としている。また、プロジェクトの開始時期もUMETRICSは2015年1月(STAR METRICSから

²⁶ Weinberg氏の講演より。第2部参照。

²⁷ Weinberg氏の講演より。第2部参照。

の独立)、JGRADは2014年4月(試験運用開始)と類似している。

両者は、大学参加の規模感も非常に近い。例えば、Weinberg氏によると、UMETRICSはBig Tenを中心に、米国の主要50大学がコミットし、19大学がデータ登録を行っている²⁸。また、UMETRICSを進めるIRISは、連邦研究開発費を最低1億ドル以上得ているすべての機関を包含することをスコープに入れ、すべての連邦政府が支出する研究開発の90%以上のデータを包含することを目指している [10]。

一方、JGRADはNISTEP(文部科学省の研究機関)がハブとなり、資金的な指標ではないが、博士号授与件数を事業管理の一つの指標として、博士号授与件数の多い上位50大学の参加を当面の政策目標としてプロジェクトを進めている。これは、わが国で年間に輩出される博士号取得者の約75%をカバーすることになるので、スコープの規模感にも相似性がある [3]。

現在、JGRADは、国公立大学42大学の参加を得るなど、参加大学数の上では、当初の目標に近づきつつある。しかし、JGRADの場合、研究科単位での参加や、一部プログラム単位での参加を認めていることや、データ入力を登録者の意思に任せているため、データの正確性や網羅性、充足率の点では、参加大学において連邦研究費が支払われたすべての人材を包含するUMETRICSに比べ改善の余地があるといえる。このため、引き続き参加大学に対し、全学参加に向けて呼びかけていくとともに、登録率の向上に努めていく必要がある。

(3) プロジェクトの社会背景

UMETRICSとJGRADの立ち上がった社会背景には、かなり違いがあると考えられる。UMETRICSは、上述のようにSTAR METRICSを、その技術的基礎としており、STAR METRICSへの参加はボランティアでボトムアップ・アプローチを取っている点では、確かにJGRADの大学参加の枠組みと類似している。

しかし、STAR METRICSの場合、2009年、米国復興・再生法(ARRA)の要請と、それを牽引する当時のオバマ政権の強い政治的リーダーシップが社会背景として存在し、また、経済刺激策により創出された研究業務を追跡しようというジュリア・レーン氏らによるリーダーシップと献身があったことは言うまでもない [10]。つまり、政府の研究開発投資を雇用効果という共通の指標で評価しようという目標と、それに対する政策的・社会的なコミットメントが存在したことが、機関間の壁を乗り越えて、政府や大学の保有するデータ間の膨大な連携・統合を可能にした背景ではないかと考えられる。

その意味では、JGRADの場合、このような政策的・社会的な共通目標やコミットメントがはるかに弱かった。当初、文部科学省の推進するSciREX事業の一部一すなわち、「NISTEPの調査研究プロジェクト」として立ち上がったが、STAR METRICSのような強力な社会的推進力があつたとはいえない。ナショナル・プロジェクトとしての政策的基盤が強化されたのは、「第5期科学技術基本計画」(2016年1月22日、閣議決定)からである²⁹。また、JGRADは構築当初は、「個人情報保護」と「登録者のインセンティブ」が障壁となり [3]、米国のように「研究の価値を測定すること」に参加大学も含めた共通の価値を醸成するレベルにまでは至っていなかったと考えられる。こうした状況から、本来調査を目的としたデータベースの構築が、プロジェクトの過程で登録者へのインセンティブに議論がシフトし、登録者向けのサービス提供を含む構想へと変化したこ

²⁸ Weinberg氏へのインタビューによる。

²⁹ ただし、「博士人材のデータベース」という表現が用いられており、JGRADへの直接的な言及は行われていない [2]。

とも、米国とは状況を異にしている。

(4) データ利用の原則

UMETRICS と JGRAD は、一定のルールの下で、ハブ機関と参加大学がデータを利用できる「クローズドなデータベース」である。UMETRICS の利用条件等の詳細については、引き続き調査する必要があるが、少なくともデータ利用が参加大学に限られる点では、両者は共通性がある。

ただし、JGRAD の場合、現在、参加大学は「自大学のデータの利用」が認められているが、データ全体の分析は、国(NISTEP)主導で行うことを想定している。しかし、国の公的な立場で大学の協力により収集した個人情報などを NISTEP の研究者が独占的に利用するスキームについては、大学にも協力が得られにくい原因の一つでもあり、今日、大学とのパートナーシップを重視する方向でプロジェクトを進めている。例えば、NISTEP の客員制度等を活用すれば、参加大学の研究者が NISTEP の客員研究員として JGRAD を用いたデータ全体の分析に参画することは可能である。

データ利用の原則は、参加大学との「価値の共有」及び「機関間連携能力」に大きく影響する。将来的には、上述の客員制度等も活用しながら、UMETRICS のように、いくつかの研究ミッションをベンチマークとして立ち上げ、参加大学の研究者とともに、データの有効利用と人材研究の高度化を図る仕組みについて検討していく必要がある。

(5) データベースのユーザー

UMETRICS と JGRAD は、科学政策研究者による調査研究への利用に関する限り、原則、大きな違いがあるとはいえない。しかし、各大学で想定されるデータベースの利用者は、UMETRICS と JGRAD では若干違いがあるように思われる。

UMETRICS の場合、「研究の価値」を測定・評価し、ステークホルダー等に説明するために、あくまで「科学政策研究者 (Science Policy Researcher)」が利用することを中心に構築が行われているようである。それは大学の経営を支えるエビデンスであり、また科学行政に役立つツールでもある。

これに対し、JGRAD の場合、人材政策のエビデンスを得ることを目的に整備が進められているが、もともと「博士人材のキャリアパスの把握」という目的のためにデータベースが構築されたこともあり、科学政策研究者のみならず、参加大学のキャリア・センター等、学生の就職支援部門などの実務的な利用も視野に入れる必要がある。また、近年、大学の IR 活動への利用の観点から、企画評価部門を JGRAD の窓口とする大学も増えており、大学経営のためのエビデンスの提供という意味では JGRAD も UMETRICS の利用に近づきつつあると考えられる。いずれの場合も、JGRAD の現状は、科学政策研究者のみならず、就職支援部門や評価部門等、大学の事務部門等の活用もある程度想定しつつ、プロジェクトを推進する必要があると考えられる。

(6) データベースの構築目的と登録対象者の範囲

1) 構築目的と登録対象者の範囲の違い

データベースの「構築目的」と「登録対象者の範囲」において、UMETRICS と JGRAD には違いが見られる。

UMETRICS は、もともと、公的研究開発の雇用効果等の測定を行うための STAR METRICS を起源とし、「人材」を対象とするデータベースとして独立・発展してきた。このため、UMETRICS は確かに大学院生を対象に多くの研究が行われているが、連結されるデータの対象者の範囲は、そ

れに限らず、研究者、ポスドク、大学院生、事務職員等、研究費からの給与を受けた者や支出に携わった者などすべてに渡る。また、博士号取得者の特定は、上述のとおり UMETRICS データと ProQuest (学位論文データベース) とマッチングさせる必要がある [6]。

これに対して、JGRAD は、「博士人材のキャリアパス把握」という特化した目的のために整備を進めており、したがって、わが国の大学の博士課程在籍者及びその修了者に限定されている。国内に存在する「海外の博士課程を修了した博士人材」は、現在のところ登録対象とはなっていない³⁰。なお、近年、博士課程の一貫化が進み、博士前期過程(修士)に登録を拡大してもいいのではないかという声がある。実際、修士課程学生の中にも自ら登録を希望する者がいることや、リーディングプログラムの支援対象者の一部等では博士課程前期からの登録事例もあることなど、修士課程学生に対するモニター制度等も今後検討することとしているが、原則は、博士課程後期からの登録としている³¹。このように、JGRAD の登録対象者は UMETRICS の登録対象者に比べて限定的である。

2) 博士人材の国際的アウトフローの把握

JGRAD は登録対象者の範囲が UMETRICS に比べて限定的ではあるが、データ・カバレッジの点で優位性があると考えられる³²。すなわち、UMETRICS は米国内にいる人材しかカバーしていないのに対し、JGRAD は、わが国の博士課程を修了した者を把握の対象にしているため、例えば母国に帰国した留学生や海外に出国した博士人材—すなわち、博士人材の国際的アウトフロー—も把握の対象としている。捕捉の程度はともかくも、少なくとも年に1度、これら海外居住者・帰国留学生の状況も把握しようとしている点では、JGRAD の方が UMETRICS よりスコープがやや広い。この点は、Weinberg 氏が JGRAD に関心を寄せている理由の一つである。Weinberg 氏は、「研究の価値」については国際的な意味合いで考える必要があることを強調し、JGRAD と UMETRICS の将来的な連携協力にも関心を示した。例えば、仮に、JGRAD と UMETRICS が将来連携することが可能となれば、わが国の大学を修了し、米国へ転出した博士人材を JGRAD で補足し、その博士人材の米国での活動状況を UMETRICS で把握する、といった国際的な研究活動の価値をトラッキングすることが理論的には可能となるからである。

また、上述のとおり、現在、人材データベースの開発を進めている国は、日米でだけでなく、オーストラリア等にも広がっている³³。こうした動きが、今後、世界的に拡大していけば、データ利用等、制度的な問題はあっても、もし将来的に連携することができるようになれば、人材の国際流動性を把握することも可能となるだろう。そのような意味でも、引き続き海外動向については注意深く情報収集していくことが必要である。

³⁰ 将来的に海外で博士号を取得した者や社会人博士等にも登録を促す取り組みをしてはどうか、という議論はあるが、現時点では実現していない。

³¹ 博士課程教育リーディングプログラムの一部など、修士課程・博士課程を切り分けることが難しい場合、修士から登録している事例もある。また、近年、修士課程の学生からも登録の希望が寄せられている事例があるため、今後、これらの希望者に対する取り扱いを検討することとしている。

³² Weinberg 氏との意見交換より。

³³ Weinberg 氏の講演における質疑より。第2部参照

(7) コアデータ、データ連結の範囲、データ入力方式

UMETRICS と JGRAD の大きな違いは、データベースを構成するコアデータの種類とデータ連結の範囲である。

UMETRICS はデータ連結の範囲が非常に多岐にわたり、行政データ、自然データ、調査データを利用している。特に、行政データの利用は、登録者本人への負荷が少なく、データも正確で、カバー範囲が広いことから、UMETRICS のコアを形成している。しかし、行政データは、初期のクリーニングや情報とのリンクが難しく、またデータ収集のトピックに限られるという問題点があるため、出版や特許等の自然発生データやトピックの設定に自由度が大きい調査データ（サーベイ）と組み合わせることが必要となる³⁴。

先行研究では、このような多様なデータソースとリンクされた UMETRICS の特徴を、「米国国勢調査局の持つ雇用と家計データや職歴データ、企業データ等の社会経済的データ、出版・引用データなどを連結させることで、研究開発投資の社会における成果を多角的に捉えようとしている」（CRDS, 2015）と説明している。しかし、このような広範なデータ連結を可能としたのは、上述のとおり、米国における STAR METRICS の技術的成功と、それを推進した社会背景が存在したからである。

これに対して、JGRAD のデータの中心は、登録者自らが入力・更新するサーベイデータである。この背景としては、第1に、JGRAD の設計当初、諸外国の博士人材のキャリアパス追跡は、伝統的なサーベイ調査が主流であり [7]、UMETRICS のような比較的新しいメトリクス型のデータベースプロジェクトに対する知見が乏しかったことが挙げられる。第2に、JGRAD 設計当時、国内においてモデルとなるデータベースが、JST の研究者総覧データベース（researchmap）しかなかったことから、システムの構成や入力形式も、researchmap に準拠したと考えられる [3]。第3に、JGRAD に対する国としての政策的なコミットメントが弱く、NISTEP の研究プロジェクト（ボトムアップ型）として開始された結果、「個人情報保護」や「機関間の壁」などが原因で、「行政データ」の利用や、他機関の保有するデータとの連携が困難であり、登録者の任意の協力を依存するサーベイデータの形式をとらざるを得なかったことが挙げられる。

JGRAD は、当初、サーベイ調査（アンケート）を行うための、いわば「登録台帳」（登録型データベース）として、データベースシステムを構想したようであるが、その後、調査項目を増やす過程で、サーベイ調査をデータベース上で行う「調査型データベース」に変化したようである³⁵。しかし、このような変遷を経てもサーベイデータを収集しデータベースを構築するという基本的な発想に変わりはない³⁶。このため、JGRAD において、大学の保有する行政データは、登録者のアカウント/ID 発行に必要な一部データ（学籍番号、所属研究科等）を大学から入手する意外、使われていないし、また、すでに大学や機関が保有する基本情報についても、登録者本人が改めて JGRAD に入力・更新するという方式を伝統的に採用している。また、このような JGRAD の性格から、既存のネット・アンケートとの違いが議論となったり、JGRAD の登録項目について調査型データベースとして増やすべきか、それとも登録型データベースとして絞り込むべきかなど、構築の方向性についても議論されてきた [3]。

JGRAD は、JST をはじめとする他のデータベースとの連携を進めている点で UMETRICS と共

³⁴ Weinberg 氏は講演の中でも、行政データ、自然発生データ、調査データのそれぞれの長所短所を比較したうえで、これらのデータ連携の重要性を説明している。（図表 2-1 参照）

³⁵ 関係者へのヒアリングによる。

³⁶ その意味ではネット・アンケート調査の発想に近かったと考えられる。

通性がある。しかし、JGRAD のデータ連携は、現在、米国に比べて極めて限定的である。2016 年度より、JST の JREC-IN Portal や researchmap とのデータ連携を推進しているものの [3]、今後、更なる機関間の連携推進や、大学からのデータインポートを促進する等、登録者に負荷のかからない入力方式の検討を行う必要があるのではないかと考えられる。

(8) データベースとしての機能と登録インセンティブ

1) 登録者への情報提供機能の有無

UMETRICS と JGRAD のもうひとつの大きな違いは、データベースとしての機能である。これは、データベースを構成するコアデータの違いと、データ入力方式の違いによるところが大きい。

UMETRICS は、行政データをコアデータとしているため、データとしては全数を覆っており、サーベイデータのように回収率を上げることにコストをかける必要がない。データ入力も大学や機関が保有するデータを収集することが中心なので、登録者にデータ入力・更新の負荷を与えることはほとんどない。

これに対して、JGRAD のコアデータは登録者の入力データ（サーベイデータ）である。このため、構築当初から登録率を向上させるためのインセンティブが議論され、UMETRICS にはない「情報提供（サービス）」のためのデータベース機能の議論に多くの時間を割いてきた [3]。データベースを構築する側は、キャリアパス把握のための調査データの取得を一義的な目的とし、一方で、データを提供する大学や登録者の側は、情報提供サービスを要求するという、非対称な構造がこのプロジェクトには内在している。

2) 機関間連携による情報提供

JGRAD では、登録者に対する情報提供のコミットメントを達成するため、例えば博士人材の求人情報やロールモデルについての情報提供等について議論され、また、多機能化についての議論が進められてきた³⁷ [3]。しかし、これらのサービスを NISTEP の限られた資源の中で実現するのは自ずと限界もある。そこで、2016 年度より、国の機関の連携により対応することとした。具体的には、上述のとおり、2016 年 9 月より、JREC-IN Portal との連携により、JGRAD の登録情報を元に関連する求人情報を JGRAD 上で表示する情報提供を開始した [3]。また、ロールモデルの提供についても、引き続き、JST や大学との連携協力を検討することとしている。

3) 大学との連携とデータインポート

JGRAD は JST だけではなく、大学中心に進めているいくつかのデータベースとも、連携の可能性について議論している。JGRAD を介して、大学のデータベースの「入り口側」から、researchmap への「出口側」まで、一貫したデータ連携が可能となれば、登録者の負荷はさらに低減し、利便性の向上が期待される。このため、本人の同意の下で、大学の保有するデータを JGRAD にインポートする方式についても技術的・制度的な検討が進められている。

また、データのインポート方式が普及すれば、JGRAD のコアデータを、これまでのサーベイ型からメトリックス型への高度化することができ、UMETRICS 同様、登録者の負荷が少ない正確なデータベースへと、少しずつ進化するかもしれない。そのためには、個人情報保護の問題等、法務的な問題を解決するだけでなく、UMETRICS のような政策的・社会的コミットメントと、大学や関

³⁷ このため JGRAD には相互検索機能等を付加する改修も行われているが、個人所法の保護等、制度上の問題もあり、現在、使用を停止している [3]。

係機関による共通の価値の醸成が不可欠ではないかと考えられる。

(9) まとめ

上記の分析から、UMETRICS と JGRAD の類似性と相違点をまとめると、図表 4-3 のようになる。

図表 4-3 UMETRICS と JGRAD の比較

比較項目		類似性	UMETRICS (米 国)	博士人材データベース (JGRAD) (日 本)
1. プロジェクトの概要	1 開始年	○	・2009年、STAR METRICS開始 ・2015年1月、STAR METRICSから独立	・2011年、JGRAD構想開始 ・2014年4月、試験運用開始
	2 データ収集の目的	×	・研究開発投資の社会における成果を多角的に捉える	・博士人材のキャリアパスの把握・追跡
	3 中核機関	○	・ミシガン大学科学イノベーション研究所 (Institute for Research on Innovation and Science; IRIS)	・文部科学省科学技術・学術政策研究所 (NISTEP) 第1調査研究グループ
	4 形態	○	・大学参加型 (コンソーシアム型)	・大学参加型 (コンソーシアム型)
	5 大学参加状況	○	・19大学 (Big Tenが中心) ・一流大学を含む、50大学がコミット	・42大学 (博士課程教育リーディングプログラム参加大学を含む) ・博士号授与件数トップ50大学が目標
	6 政策的位置づけ	×	・2009年 米国復興・再生法 (ARRA) の元で始まった STAR METRICSを技術的基盤とする ・2015年1月、STAR METRICSから独立	・2011年 SciREXのデータ・基盤事業 (NISTEPの研究プロジェクト) が開始 ・2016 第5期科学技術基本計画 (閣議決定) で位置づけられる
	7 政策環境	○	・サーベイ調査 (Survey of Earned Doctorates; SED) が実施される中でUMETRICSを後発的に開始	・サーベイ調査 (博士人材追跡調査: JD-Pro) が実施される中で平行してJGRADを構築
2. データ	8 コアデータ	×	・行政データ (自然データ、サーベイデータとリンク)	・サーベイデータ
	9 登録者	×	・研究者、ポスドク、事務員等、多岐に渡る ・米国内	・参加大学の博士課程在籍者・修了者 ・帰国留学生、海外転出者も追跡対象
	10 データ入力	×	・機関入力	・登録者本人が入力 (機関からのデータインポートも検討中)
	11 データ利用	○	・コンソーシアム内: IRIS、参加大学	・コンソーシアム内: NISTEP、参加大学 (自大学分のみ)
	12 サーベイ調査との関係	○	・SEDとの連携に向けて作業中	・将来的に、博士人材追跡調査 (JD-Pro) との連携を構想中
	13 データ連携の範囲	×	・広範なデータ連携 ・米国内勢調査局の持つ雇用と家計データや職業データや、企業データ等の社会経済的データ、出版・引用データなどを連結	・現時点では限定的なデータ連携、インポート促進 ・JSTのJREC-IN Portalとの連携開始 (2016年9月) ・JSTのresearchmapとの連携検討 (項目精査、新機能追加等) 開始 (2016年10月)
3. 機能	14 機能	×	・調査研究	・調査研究 ・インセンティブとしての情報提供 (公募情報、ロールモデル)

(注) ○ : 類似点、× : 相違点

出典 : 筆者作成

5. 結論: UMETRICS に学ぶ JGRAD の将来展望

JGRAD による博士人材のキャリアパスの把握は、社会における博士人材の地位向上と活動の場の拡大、そのための政策的支援を行うため、エビデンスを得ることが目的である。JGRAD の将来展望を UMETRICS と比較しつつ考えたとき、いくつかの論点が見えてくる。

(1) データベース構築とサーベイ調査の連携

JGRAD が UMETRICS に学ぶべき第一の点は、データベースとサーベイ調査とのデータ連携である。欧米諸国では 1970 年代～2000 年代に、サーベイ型の博士人材追跡調査が普及したが、わが国は 2014 年に 2012 年度の博士課程修了者を対象に JD-Pro が開始されたばかりである。

このような環境下で、大学コンソーシアムを中心とする人材データベースの取り組みは、比較的新しい動向として、米国の UMETRICS をはじめ、他国（オーストラリア等）にも拡大しつつある。その意味では、JGRAD も、国際的な潮流の中にあり、成長過程であるといえる。一方で、JGRAD と JD-Pro の関係は、整理されないまま現在に至っており、将来的には米国の UMETRICS に見られるように JGRAD と JD-Pro が連携・統合を深めていくことが鍵である。特に JGRAD の場合、データ収集方式がサーベイ調査型で重なることから、登録者の負担軽減の観点からも、JGRAD の登録項目と JD-Pro の情報収集項目を適切に仕分けするなど、両者が連携することでキャリアパスの把握を行うことが必要である。

(2) データ連携の促進とメトリックス手法の導入

第二に、関係機関とのデータ連携の範囲を拡大していく必要がある。JGRAD の情報プラットフォームとしての役割を確かなものにしていくためにも、将来的には、JST だけでなく、大学や他の機関との連携協力を進め、登録者の利便性やデータベースとしての機能向上を図ることが必要である。

また、機関間連携によりデータインポートの普及を図り、登録者にとって負荷の少ないメトリックス型データベースの要素を取り入れていくことも必要ではないかと考えられる。JGRAD は、UMETRICS に比べ、目的がフォーカスされており、その意味ではデータ収集のスコープも設定しやすい。関係機関との協力の輪を広げ、データ連携を促進することで、より正確で利便性の高いデータセットを構築していく必要がある。

そのためには米国オバマ政権下で、米国復興・再生法の要請のもと STAR METRICS が発展し、UMETRICS の基礎を築いたように、政策的リーダーシップや社会的コンセンサスが不可欠であり、機関の壁を乗り越えるための共通の価値を醸成することが必要である。この点は、大学のご理解と政策当局のご支援に期待したい。

(3) データベースの機能の整理

JGRAD は UMETRICS と異なり、これまでの構築の経緯から、登録者に対して情報提供等のサービスを提供することをミッションの一つとしている。すなわち、1つのデータベースで、博士人材のキャリアパスに対する「調査」と博士人材への情報支援という「対策」を同時に実施しようとしている点が、プロジェクトを複雑なものとしている。両者の関係をどのように整理していくのが第三の論点である。

情報提供等のサービスは、ユーザーである登録者が常時、安定的に享受できることが必要であり、そのための体制整備が必要である。しかし、調査研究機関である NISTEP の本来のミッションか

らは、大きく一步踏み出しており、NISTEP の能力の範囲では自ずと限界があることも事実である。したがって関係機関と連携しつつ、サービスの向上を図ることが不可欠であり、将来的には情報提供を定型的に実施する体制整備し、調査分析チームから独立させることも視野に入れて検討する必要があると考えられる。

また、JGRAD による情報提供は、あくまで登録に対するインセンティブを確保することが狙いで、それ自体が目的ではない。したがって、データインポートを推進することで、インセンティブによらないデータ収集方法についても検討していく必要がある。

(4) 大学との共同ベンチマーク

最後に、共通の価値を醸成するための先駆的な実証研究（ベンチマーク）を実施する必要がある。UMETRICS は、コンソーシアムを形成し、参加大学から選ばれた科学政策研究者が、上述の「3つのミッション」について研究チームを構成し、研究を行っている。このようなやり方は今後の JGRAD のデータ分析においても一つの参考となる。

ただし、JGRAD の場合、研究者のみならず、就職支援部門や IR 部門の事務方等の利用も想定されるため、JGRAD によって何ができるのか、先駆的なベンチマークを通じて提示していく必要がある。NISTEP の調査研究だけでなく、国のプログラムのフォローアップや、大学における IR 活動等への利用等も視野に入れながら、その活用を具体的に検討していくフェーズに移りつつあると考えられる。

こうした視点から、JGRAD では、2015 年より、大学や登録者への情報のフィードバックを強化することを目的に、試行的にアンケート機能を活用した登録者の意識調査を実施したが [16]、今後は参加大学のプログラム評価や、登録者のキャリアパス形成など、現場のニーズを踏まえた調査や情報提供を実施していくこととしている。

JGRAD の登録者も、すでに 7,000 人あまりに達した今日、UMETRICS のような参加大学が享受できる価値を生み出すまで、もう一步のところまで近づきつつあると思われる。博士人材及び参加大学の皆様には、引き続き、JGRAD へのご協力を是非、お願いしたい。

引用文献

- [1] The UMETRICS Initiative, "Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science".
(<https://www.btaa.org/docs/default-source/umetrics/umetrics-synthesis-document.pdf?sfvrsn=4>) [アクセス日: 2017年7月17日]
- [2] 総合科学技術・イノベーション会議「第5期科学技術基本計画」(2016年1月22日閣議決定)
- [3] 松澤孝明、篠田裕美, 「博士人材データベースのパイロット運用-政策・制度・運用の現状と改善に関する検討報告書-」. 文部科学省科学技術・学術政策研究所 調査資料-255, 2016年11月.
- [4] Nikolas Zolas, Nathan Goldschlag¹, Ron Jarmin, Paula Stephan, Jason Owen-Smith, Rebecca F. Rosen, Barbara McFadden Allen, Bruce A. Weinberg, Julia I. Lane, L. Nikolas Zolas, "Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D.recipients," Science 11 Dec 2015: Vol. 350, Issue 6266, pp. 1367-1371, DOI: 10.1126/science.aac5949
(<http://science.sciencemag.org/content/350/6266/1367.full>) [アクセス日: 2017年7月17日].
- [5] 国立研究開発法人 科学技術振興機構研究開発戦略センター (CRDS). 調査報告書 米国「科学イノベーション政策のための科学」の動向と分析. CRDS-FY2015-RR-04.
- [6] Catherine Buffington, Benjamin Cerf, Christina Jones, and Bruce A. Weinberg, STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to 2010 Census, American Economic Review: Papers & Proceedings, 2016, 106(5): 333-338 <http://dx.doi.org/10.1257/aer.p20161124>
(<http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.p20161124>) [アクセス日: 2017年7月17日].
- [7] 斉藤経史、伊藤裕子、富沢宏之, 「博士課程修了者の状況把握のシステム設計ー博士人材データベースの構築背景および海外の博士課程修了者調査ー」, 文部科学省科学技術・学術政策研究所、調査資料-216, 2012.
- [8] 文部科学省科学技術・学術政策研究所, 「博士人材追跡調査第一次報告書-2012年度博士課程修了者コホート」, NISTEP REPORT No.165, 2015.
- [9] Barbara McFadden Allen, Julia I. Lane, Rebecca Rosen, Jason Owen Smith, Bruce A. Weinberg "UMETRICS as a Tool for Quantifying the value of Research and Assessing Underrepresentation," feature, association for women in science, summer 2015, pp. 36-39.
(<http://www.btaa.org/docs/default-source/umetrics/awis-umetrics-august-2015.pdf>) [アクセス日: 2017年7月17日].

- [10] Michael Eisenstein, "Academic return: A broader understanding of 'impact' could help governments to measure the diverse benefits of their investment in research," *Nature* 533, S20–S21 (05 May 2016) doi: 10.1038/533S20a Published online 04 May 2016 (http://www.nature.com/nature/journal/v533/n7601_supp/full/533S20a.html) [アクセス日: 2017年7月17日].
- [11] "best value schools," (<http://www.bestvalueschools.com/faq/what-is-a-land-grant-university/>) [アクセス日: 2017年4月30日].
- [12] "The Association of American Universities," (<https://www.aau.edu/>) [アクセス日: 2017年4月30日].
- [13] "The Association of Public and Land-grant Universities (APLU)," (<http://www.aplu.org/>) [アクセス日: 2017年4月30日].
- [14] "The Association of American Medical Colleges," (<https://www.aamc.org/>) [アクセス日: 2017年4月30日].
- [15] "The Federal Demonstration Project," (<http://sites.nationalacademies.org/PGA/fdp/index.htm>) [アクセス日: 2017年4月30日].
- [16] 篠田裕美 松澤孝明, 博士人材データベース (JGRAD)を用いた博士課程在籍者・修了者の所属確認とキャリアパス等に関する意識調査, 文部科学省科学技術・学術政策研究所, 2016.
- [17] BIG Academic Alliance ホームページ, "UMETRICS Universities: Measuring the Impacts of Research on Innovation, Competitiveness, and Science" (<https://www.btaa.org/projects/umetrics>) [アクセス日: 2017年3月30日 12:50].

第2章 講演会

講演会概要

演題:「博士取得者の初期キャリアの設定:米国 UMETRICS プロジェクトからのエビデンス(原題: Early Career Outcomes of Doctoral Recipients: Evidence from the UMETRIC)」³⁸

講師: Bruce Weinberg 氏

オハイオ州立大学大学院経済学研究科教授、全米経済研究所(NBER) 上席研究官

日時: 2017年3月3日(金) 15:00~17:30

場所: 科学技術・学術政策研究所会議室³⁹

概要: 「研究の価値」を測定することは、研究資金の獲得、政策立案、あるいは研究者がよりよい仕事をするために必要である。UMETRICS は、ミシガン大学科学イノベーション研究所(IRIS)を中心に、全米大手主要10大学(big Ten)、全米大学協会(AAU)、国勢調査局等のパートナーシップにより実施されるプロジェクトであり、米国における「研究の価値」について測定するために2つの取り組みが行なわれている。一つは、研究組織の一義的な成果物として「人間」について考えることであり、もう一つは研究の新規性を測るために「メトリクス(METRICS)」を策定することである。

研究プロジェクトの資金情報(課金先等)から、研究にかかわる「人」や装置、備品などの情報が特定でき、これらのデータを「行政データ(所得、税務申告等)」や「自然発生データ(論文データ、特許等)」、「調査データ(国勢調査等)」等とリンクさせる手法により、博士人材などの人材研究に活用している。

UMETRICS の参加機関は、どのような形でプログラムを改善したらよいか、どのような活動を行えばよいか等の情報を得ることができ、またデータにアクセスすることができる。

科学技術・イノベーションは国際的なものであり、「研究の価値」を理解するためには、国際的なフローも理解しなければならない。その第一段階が各国による取り組みであり、米国ではかなりの作業が行なわれている。しかし、究極のゴールは「データ統合」であり、国際連携が必要である。したがって、データプラットフォームの相互連携等が今後、ますます重要性を増すと考えている。

講師略歴:

1996年博士号取得(経済学、シカゴ大学)。イノベーション創出人材の流動性や国家間のイノベーション競争力等についての実証研究を行い、合理的な科学イノベーション政策立案のために必要な科学的根拠となる活動を推進。また、研究人材の賃金差異などの労働経済学にも携わっている。UMETRICS では主要研究者の一人として、対象大学における博士号取得者の動向の追跡調査などを通じて、研究の社会的影響を測るための定量的アプローチに携わっている。また、クリーブランド連邦準備銀行やフーヴァー研究所等にも招聘され共同研究を行ってきた。

※司会: 松澤孝明 科学技術・学術政策研究所第一調査研究グループ総括上席研究官

³⁸ スライドタイトル: UMETRICS: The Database of Doctoral Holders and their Impacts

³⁹ 本講演会は、同日開催された「2016年度博士人材データベース(JGRAD)パイロット運用連絡協議会」終了後、休息を挟んで、引き続き開催された。

講演録(日本語)

※本講演録は、Weinberg 教授の講演（英語）に対する日本語同時通訳速記録をもとに作成されたものである。

1. オープニング・講演者紹介

○司会者（松澤総括上席研究官、NISTEP）： それでは、時間になりましたので、後半のUMETRICS⁴⁰についての講演会を始めさせていただきます⁴¹。

本日は、多数講演会に御参加いただきまして、ありがとうございました。私は、今回の講演会を主催させていただいております科学技術・学術政策研究所第1調査研究グループの総括上席研究官であります松澤でございます。

今回、我々がUMETRICSに関しまして、アメリカのオハイオ州立大学から、Bruce Weinberg 博士をお呼びして、ご講演いただくように手配した経緯について最初にご説明できればと思います。

UMETRICSというプロジェクトは、8年前にオバマ政権（米国）が研究開発の公的投資についての経済効果、特に雇用効果について調べるためのSTAR METRICSというナショナル・プロジェクトを開始したことに端を発していると聞いております。その後、それを中西部の大学が中心となりまして、現在、ミシガン大学の研究所でございますIRIS（The Institute for Research on Innovation & Science）が大学とコンソーシアムを組む形で進めているプロジェクトで、特にMETRICSの手法を用いまして、人材政策などについて幅広い研究が行われていると聞いております。

昨年の2015年12月に、「サイエンス」誌にこの結果が公表されて世界的に反響を呼んだプロジェクトでございまして、我々のほうにも多数、プレスやジャーナリズムから、このプログラムについての評価や今後の展望について、また我が国のプロジェクトとの相似性について問い合わせが寄せられたという経緯があります。そこで、まずは博士人材データベース（JGRAD）の関係大学にぜひ御紹介したいという趣旨で、このたび、我が国で初めてこのプロジェクトについて講演会を企画しました。実はWeinberg先生は、昨日、日本に入られて、明日米国に帰ってしまうという非常に忙しいスケジュールの中、今回、特別にご講演をいただく機会をつくりました。

Weinberg博士についてご紹介させていただきます。1996年にシカゴ大学で経済学の博士号を取得されまして、イノベーションの創出人材の流動性や国家間のイノベーション競争等についての実証的な研究を行っている第一人者でございます。UMETRICSでは主要研究者の1人として、対象大学における博士取得者の動向の追跡などを行っております。また、クリーブランド連邦準備銀行やフーヴァー研究所、国家経済研究局（National Bureau of Economic Research）等にも招聘され、共同研究を行っているという方でございます。

私どもがこのプロジェクトに非常に興味をもっておりますのは、皆様もご存じのとおり、アメリカではNSF⁴²が中心となって、全国的なSED（Survey of Earned Doctorates）という博士号取得者に対するサーベイ調査が行われていますが、そのような環境の中で大学のコンソーシアムを主体と

⁴⁰ UMETRICS : Measuring the Impacts of Research on Innovation, Competitiveness, and Science

⁴¹ 本講演会は、「博士人材データベースのパイロット運用に関する連絡協議会」の終了後、休息を挟んで、引き続き開催された。

⁴² NSF : National Science Foundation

するUMETRICSプロジェクトが立ち上がり、また新たな展開をみせているということです。我が国でも「博士人材追跡調査(JD-Pro)」というサーベイ調査を我々のグループは定期的に行っているのですが、その中で、今後の博士人材データベース (JGRAD) のあり方や将来展望を考える上で、非常に大きな示唆をいただけるのではないかと期待しております。

それでは、先生、よろしく願いいたします。

2. 講演

○Weinberg教授： 本日はお招きいただき、まことにありがとうございます。

既に御承知かと思えますけれども、イノベーション、サイエンス、リサーチといったものは国際的な取り組みであることは明らかです。そして、米国における研究 (research) やその価値について理解する上で、我々は、大きな進歩を目指していきたくて考えています。当然それにはデータも必要になります。データなくして国際的な意味での研究の価値は理解できないと思います。

米国において生まれたアイディアは日本や他の国へ国境を越えた動き、また、研究者は国から国へ移動いたします。米国で研究から得る価値をみる場合には、ほかの国で生まれた研究からの価値も見なければ重要な部分を見失ってしまいます。いずれにしても、申し上げたいのは、このような研究を行うためには、国際的なコネクションやリンケージに非常に価値があり、非常に重要であるということです。私が最初申し上げましたけれども、本日こうやってお招きいただき非常にうれしく思っております。本日は幾つか我々が行っていることをご紹介させていただきますけれども、これがJGRADのイニシアティブにもお役に立てばと思います。

まず、我々がやっていることでありますけれども、“どうやって研究の価値を測定するか？”ということでもあります。研究の価値を理解するための多くの活動は、例えば出版物の数を数える、あるいは引用数を数える、研究が発表されたジャーナルのインパクト・ファクターを見るなど、ビブリオ・メトリックスに焦点を当てています。そして、よりよい測定を得ようと多くの異なる取り組みが行われています。けれども、我々がやろうとしていることは、研究の価値を2つの方法で、もう少し掘り下げて考えてみたいということです。

一つは、まずは研究組織の一義的な成果物の一つとして、「人間」について考えるということです。リサーチプロジェクトに雇用される学部生、大学院生、ポスドクは、彼らが行っている研究の最も重要な成果物の1つのだと考えているからなのです。よって、そういった経験から取得する価値を見てみたい、彼らが卒業した後どういった価値を生み出してくれているか見ることができずし、彼らを雇用した産業界やビジネスをみることもできます。また、彼等に起こっていることを見ることで彼らのスキルの価値を理解することができます。

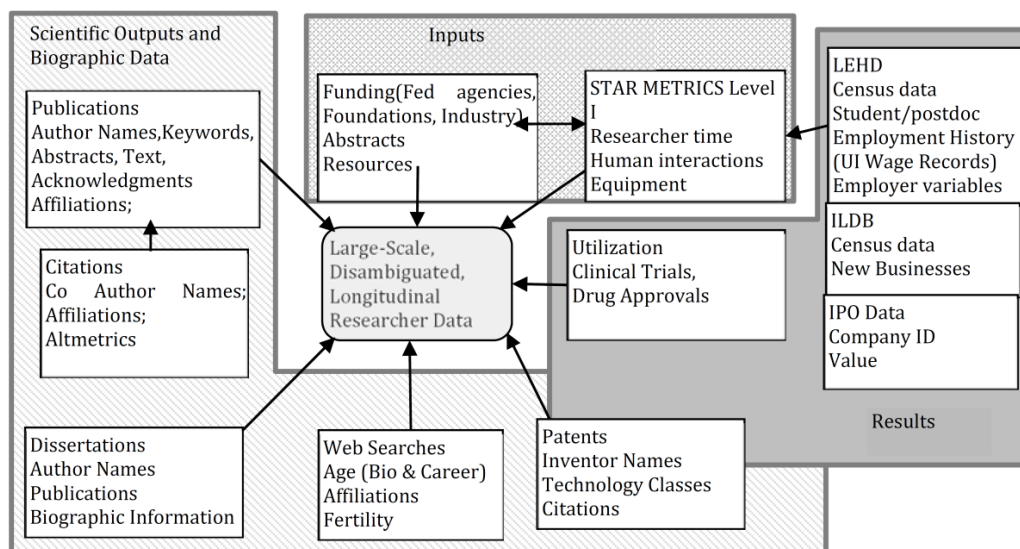
もう1つ、我々が行っていることは、UMETRICSについて我々が行っていることと、ここNISTEPで起こりつつあることとの関連について、今朝、お話がありましたことです。スライドを幾つかご紹介しますが、ひょっとしたら皆様には余り関係ないものもあるかもしれませんし、お手元のスライドの写しより、皆様のところに関連しているかもしれません。スライドにはちょっと手を加えさせていただきまして、皆様に関連した形でおみせできればと思っているのです。もう1つ我々がやろうとしていることは、研究の新しさや新規性のためのMETRICSの策定です。これらは、我々が多くのケースについて行った研究インパクトに明らかに関係しており、研究の新規性や、イノベーション性や、革命的な要素を捉えようとしている点で、インパクトとはちょっと別のものです。

私、現在、2つの取り組みにかかわっております。1つが、ミシガン大学の「科学イノベーション研究所 (IRIS)」であります。IRISの取り組みがUMETRICSの中心体となっているわけです。UMETRICSですけれども、大学間のパートナーシップによって成り立っております。また、米国の国勢調査局も関与しております。それからミシガン大学です。また、連携先としてオハイオ州立大、ニューヨーク州立大などもあります。これから大学もさらに増えると思います。そして、非常に多くの方面からご支援を受けることができっております。

それから、高齢化社会におけるイノベーションでありまして、余り時間はないかもしれませんが、若干触れたいと思っています。研究者はキャリアの早期で非常にイノベーションが高いということがよくいわれております。もちろんさまざまな英知を提供するわけですけれども、今何が起きているかといいますと、そういった研究者が年をとってきているということでもあります。つまり研究者のコミュニティが高齢化するということですので、生産性が影響を受けるかもしれない。米国もそうですし、日本でも高齢化が進んでいるわけです。

こういった2つの取り組みでありますけれども、非常に大きな規模のデータをつくっています。こういったデータを活用する、そしてほかの人からもアクセスできるようなデータづくりを目指しています。

こちらの概念図がデータのインフラストラクチャーになります。



幸いなことに、私から詳細まで入って細かいところは御説明いたしません。ただ、ここでご指摘しておきたい点なのですが、幾つかの主要なデータセット、また主要なデータのつながりについて触れておきたいと思います。

まず、データの中心となりますのがUMETRICSのデータです。UMETRICSのデータは、ご存知ない方もいらっしゃるかもしれませんが、STAR METRICSというイニシアティブがあって、そこから派生したものでありまして、基本的にスポンサーがついているリサーチプロジェクトがあるわけです。

例えば、チェックブック、小切手帳ですとかクレジットカードの口座を思い出していただきたい

のですけれども、基本的には、誰に支払いが行われているのか、どのくらいの金額なのか、何のためにそれが支出されているのか、どの口座なのかなどを知る必要があるわけです。つまり課金先がどうなのかを知りたいわけです。

これは非常に強力です。といいますのは、リサーチのプロジェクトについて考えていただいたときに何をするか。これは、そのプロジェクトにかかわる人に支払いをしなくてはいけないわけなので、リサーチにかかわる研究者を一人一人知らなくてははいけませんし、これが論文の執筆者なのか、あるいは教職員なのか、それともスタッフなのか、学部生なのか、結局論文には名前が挙がってこない人も分かるわけです。

それから、購買についてもわかります。このプロジェクトで例えばこういった装置、備品を買いましたといったものも特定化ができるということ。全てのインプットがあるわけです。これが研究にかかわってくるわけです。ですから、人、それから装置、備品など全てがわかる。

それでもって、そういったデータを幾つかのものと連携させるリンク先としましては、例えば出版データとつなげるということです。それによって、研究室（ラボ）に入ったものをみるだけではなくて、そこから出てきた論文データ、引用データですとか、新規性とか革新性といった研究のインパクトもわかる、特許などもみることができるということです。

いろいろNIHとも連携しておりまして、彼らはバイオメディカルリサーチを中心に興味あるところでありましてけれども、例えば医薬品の承認とかといったデータとも連携する、学術アウトカムのリンクができるということです。

また、関心が高まっているものとしたしましては、経済的価値ですとか社会的価値の研究です。我々は、自分もっているデータを例えば国勢調査のデータと連携させるということをやっています。

連邦政府によって10年ごとに国勢調査が行われているわけですがけれども、非常に膨大なデータがあるわけです。あるいは雇用者側からの、例えば所得、税務申告のデータなどもあります。こうした税務データを各人と合法的にマッチングさせる。リサーチプロジェクトにかかわった人が教育機関を出て、どういった業界に就職するのか、企業の規模、設立年数、どのくらいの年間所得を得ているかといった情報も得ることができるわけです。後ほど幾つか御紹介したいと思っております。

[挿入]

○Weinberg教授： もし御質問があれば後ほど質疑の時間を設けたいと思います。私は経済学者です。それぞれが介入し合うというのは経済学者の常でありますので、説明をしている内容で明確でないところがありましたら、いつでもとめて確認をしていただければと思います。どれぐらいでもずっと話し続けることもできますけれども、どちらでも結構です。

では、IRIS、またUMETRICSについて少しお話をさせていただきたいと思っております。こちらのスライドは、(配布物に)入っているもの、それから本日加えられたものもございまして、内容に少し違いがございます。

IRIS、それからUMETRICSは、全米大手主要10大学（ビッグ・テン）、AAU（全米大学協会）、国勢調査局、それ以外の主要大学も加わっています。ハブとして中心になっておりますのはミシガン大学です。それ以外のオハイオ州立大学、ニューヨーク大学などが参画しております。安全裏にデータを保有するのを行っております。Sloan、それからKauffmanなどがインフラストラクチャーに関して支援しております。また、それ以外に研究を支援しておりますのは、NIH、NSF、

SBA（中小企業庁）、USDA（農務省）、それから特許商標局などであります。

IRISは加盟している大学間におけるコンソーシアムであります。加盟大学から情報、データが提供され、安全裏に保管されまして、年次会費をそれぞれの大学が支払いまして運営を担保しています。そして、その引きかえとして大学が報告書を受け取ります。実際にIRISのガバナンスにも参画します。これは役員の選任にもかかわるわけです。そしてIRISの運営を行っております。

IRISへの参画については完全に任意です。ですから、重要なのは全ての利害関係者、大学、そして国勢調査局、IRISに参画している全ての機関が参画から利益を受けるということでもあります。このような取り組みに参画をすることによって、何がしかのメリットがなければならないということです。

では、どのようなメリットが得られるのでしょうか。研究に関する情報が得られます。どのような形でプログラムを改善したらよいか、どのような活動を行えばいいか。例を後ほどおみせいたします。それから、一時的にデータにアクセスをすることができます。国勢調査局が提供する情報は政策がありますので、情報の提供に関しては一部課金されるものもありますけれども、大半のものは無料で提供されています。そして、それぞれの加盟大学にとって有用であろうと我々が願っております報告書が提供されます。

こちらはオハイオ州の地図となっております。コロンバスが私のいるところです。オハイオ州立大学がオハイオ州内のそれぞれの郡にどれぐらいの研究資金の提供を行っているかを示している地図です。そして、写真に写っておりますこの女性はCaroline Whitacreです。彼女はオハイオ州の研究に関する部門を統括しているバイスプレジデントです。州議会、また下院にも上院にもアプローチをすることで、オハイオ州立大学に対する研究の支援を募っているわけです。この地図は彼女が作成したものです。オハイオ州内において、どのような形で研究資金が分配をされているかということです。同様の地図が全米のものでもあります。皆さんにとって、実際に書かれている内容をお読みになるには、ちょっと字が細か過ぎると思うのですが、これがまさに下院の議員、そして上院の議員の知りたいことなのです。お金を渡すのだから、これが私の選挙区においてどのように使われているのかを議員は知りたいわけです。より詳細な内容、また非常に精度も高いものでありまして、同様のデータは、以前はありませんでした。

[質疑]

○斎藤（総務研究官、NISTEP）：済みません、この地図の色分け、白になっているところはどのような地域ですか。

○Weinberg教授：それぞれの郡ごとに割り当てた拠出の金額ごとに濃淡がついています。オハイオ市内の地理的なことに皆さんご関心は余りないと思うのですが、こちらはオハイオのクリーブランドです。こちらはシンシナティ、そしてこちらがオハイオ大学のある場所です。それからトレド。ですので、金額に応じた形で色分けがされています。15ドル以下のお金が使われているところが白の表示となっております。ほかに何かありますか。

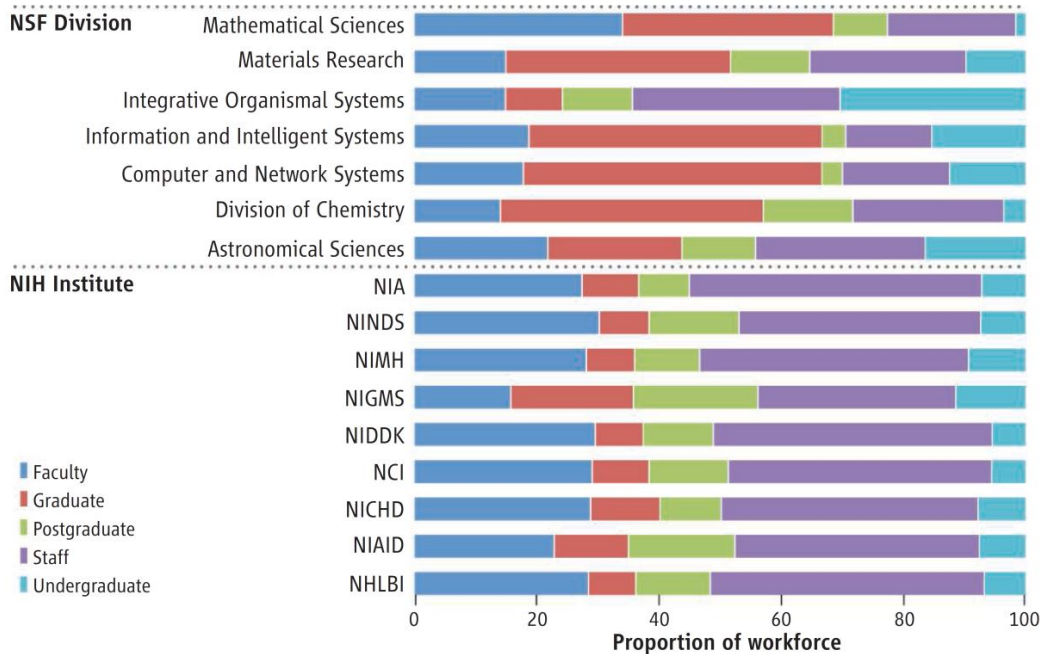
○Weinberg教授： こちらは、既に申し上げましたけれども、米国主要10大学が我々に最初の支援をしてくれているということでありまして、初期から参画しております。それぞれの大学ごとに もっと詳しい内容についてこれから申し上げますけれども、これは一例です。

パデュー大学に対して提出された報告書です。インディアナに非常に強力な支援をしているということがわかります。インディアナ州でも、先ほどオハイオでおみせしたものと似通ったマップが参照されています。これは黒と金で表示されています。ですから、このようなことが、大学が情報として得るものであります。これは州議会の議員、また知事などに提供されるものであります。

加えて、データも生成しています。加盟大学がデータを提供します。これは生データです。そして、処理します。一部は国勢調査局に送られます。研究者もこのデータに安全裏にアクセスすることができます。

こちらは既に申し上げましたが、研究に参画するということで、加盟大学がこのような形で一連の取り組みにも参画しているということになるわけです。

実際に出てきた成果物を幾つか御紹介させていただきたいと思います。こちらは2014年に発行された「サイエンス」であります。ポリシーフォーラムの中で発表されたものですが、こちらの表では、それぞれの労働力が研究プロジェクトごとにどのような形で雇用されているかの内訳を示しています。



Differences in workforce composition in projects funded by NSF divisions and NIH institutes. NIA, National Institute on Aging; NINDS, National Institute of Neurological Disorders and Stroke; NIMH, National Institute of Mental Health; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; NICHHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NIAID, National Institute of Allergy and Infectious Diseases; NHLBI, National Heart, Lung, and Blood Institute. (See SM.)

上段の部分はNSFのそれぞれの部門ごとの内訳です。それから、NIHのそれぞれの下部機関の内訳です。米国における研究開発は、多くがファカルティー側で使われていると思われておりますけれども、これはブルーで記しているところであります。左側です。

ここで申し上げたい点は、非常に限られた人たちがリサーチプロジェクトに参画している、またファカルティーにかかわっているということです。リサーチプロジェクトで資金援助を受けている人たちの大半は、何がしかの形で研修のパイプラインの中に入っている人たちです。学部生、そして大学院生は赤です。とりわけ全米科学財団に関係するような人たち、それからポスドクがグリーンです。そしてスタッフが紫となっています。恐らく2割ぐらいの人たちがファカルティーです。

大学院生の参画が根本的であると認識しておりますので、多くの研究を大学院生、博士号取得者に焦点を絞っています。こちらの研究に関してはまだ出版されておませんが、メディアマトリックスのデータとSED——これは全米科学財団(NSF)のほうで提供しているものですが——の2つをマッチングした研究を行っております。

ここで興味深い内容もみることができるのですが、もう1つは、異なる2つの種類のデータを組み合わせることによって——これは実際に行うことができるわけですが——そのようなアプローチを考える場合、このようなものを組み合わせたデータインフラをつくるに当たって、

どのようにするかということなのですが、皆さんは4番目、5番目の分類についてもお考えになれるかもしれませんが、私は3つ考えました。1つ目の分類は行政のデータです。そして、第2の分類は自然発生的なデータです。例えば発行物、引用、それから特許などです。そして、第3の種類のデータは調査データです。

それぞれのアプローチについて幾つか例を挙げて、費用と便益に関して御紹介させていただきたいと思います。ここからみせる全ての情報は博士号取得者に関するデータです。

例えば行政データの利点は、かかわる人の負担が低いということです。誰のところにも行きません。そして、このようなことを教えてくださるような労力をかけることはありません。何もする必要はありません。これが現行のものだということについても必ずしも確認する必要はないわけです。これによって完全なカバーを行うことができます。例えば調査を行って回答してくれないということになりますと、情報が欠けますけれども、行政データは完全です。ほとんどの研究者が税務申告を行っていますから、そういうデータもあるわけです。そして正確です。非常に豊富なデータを得ることができるわけです。よって、完全な研究者のネットワークの像を得ることができます。どのような機器類が買われているのか、どのような道具が使われているのか、あらゆることが得られるわけですから、非常に強力です。

では、行政データに依存することについての問題は何かでしょうか。これは実際にデータをクリーンアップするためのアルゴリズムを書くのにコストがかかるということです。そして、行政からデータセットを取得すること自体が難しいです。

ですから、人が知りたいことのうちの1つですけれども、例えば博士号取得者で実際に経済界、産業界に進んだ人たちが本当にそれを望んでいたのか、何をやりたかったのか、大学院に進んだ時点では当事者はわからなかったかもしれません。または、大学院に進んだけれども、親から何か介入されたということでキャリアを決めたかもしれません。そのようなことは行政のデータセットからはわかりません。そこら辺はやはり調査を行わないとわからないことです。

そして、自然発生的なデータです。これも負担が低いです。カバレッジも広いです。そして正確です。また、ある程度のデータ同士の連携、そしてクリーニングも行う必要があります。

3つ目のアプローチですけれども、ある種の調査を行うということです。つまり回答者に質問を投げかけて、なぜ大学院に進学したのでしょうか、そしてポストドク、その後、職について満足していらっしゃいますかといった調査を行うわけです。そういった調査を行うことによって回答を引き出す、その問題となる点ですけれども、調査を行うために費用が発生してしまうということです。回答率が高いものを目指せば目指すほどコストも上がります。高い回答率を得るとしても、これは完璧なものになるということはずなないです。回答者は忙しいでしょうし、調査に回答してもらえないこともなかなか難しいかもしれません。この調査には正確性にも問題があります。

といいますのは、例えば、昨年の年間所得はどのくらいでしたかと質問する場合、多くの国々、文化では、やはりこういったものは答えたくないという人が多いです。IRS（内国歳入庁）に答えなければもちろん牢屋に入れられてしまいますから、答えなくてはいけないかもしれませんが、一般的には正確に回答を引き出すのは難しいということです。したがって、行政データとは全く違います。

それから、研究者同士の連携はありますけれども、それを全て把握できているわけではないです。私自身もそういったネットワークを全部覚えているわけではありません。ただ、パブリケーションで例えばアルゴリズムがあれば捉えられます。

今申し上げた3つのアプローチは、それぞれ強み弱みがあるわけなのですが、例えば行政

データと自然発生データを必要なとき、使いたいときに使えるということはありません。

それから、こういった手法では得られないデータもあるわけです。そういった場合は別の手法を使うことになるわけでありまして、博士人材の研究に関して、こういった手法を活用したという成果物を幾つか御紹介したいと思っています。

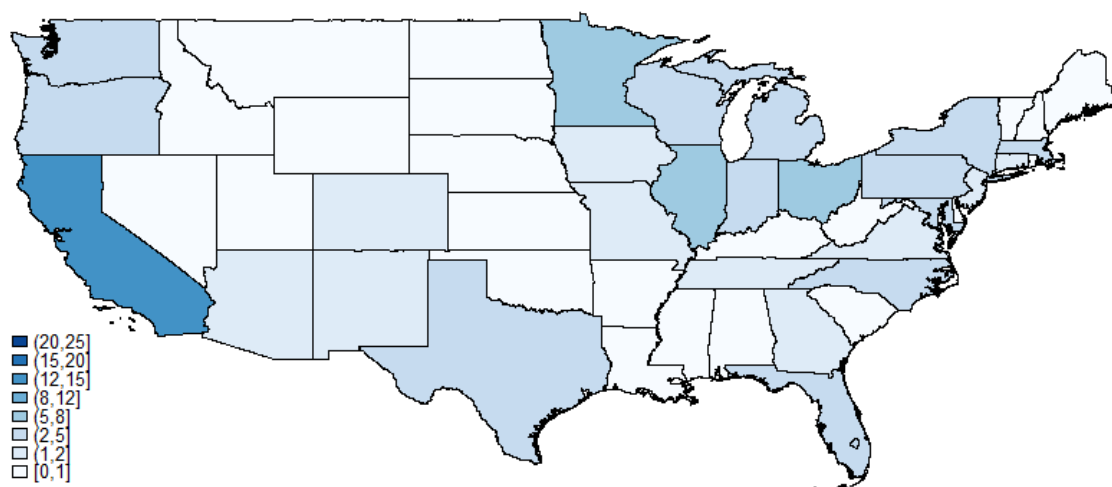
まず、研究領域別にどういった資金提供を受けているかということを見たいと思いました。このようにほとんどの領域が連邦資金を受けているわけですが、60、70、80%ぐらいの比率です。ただ、右から2番目、サイコロジー（心理学）は比較的低い。それから、真ん中の人文とか社会科学は比率が低くなっています。グラフの半ばです。これはそれほど驚くべき結果ではないと思います。

それから、これは非常に強力なメッセージが伝わるかと思うのですが、ちょっと複雑でもあるので、1つずつ数字を御説明することはいたしませんけれども、何をやったかといいますと、まずは、こういったいろいろな機関をみまして、3,329の学生を2つの大学でどのように連邦予算で支援を行ったかということで、このコラムに御注目ください。

院生の7%は国防省からの支援を受けています。資金提供元は国防省だけあります。国防省と科学財団から受けているのがこれだけの人数。そして、大体5%の学生は、国防省を含む3つ以上の機関から援助を受けていたということがわかります。国防省で正確な把握はできていないと思うのですが、もし国防省の記録があったとしても、国防省だけを見ても、自分たちの研究者は、例えばNIHとかNFS、ほかの連邦政府機関があるわけです。そこも学生と接触があるわけですから、そこまで情報は把握できないわけです。

ところで、こういった問題に答えようとするときに、ポストドクですとか院生、彼ら自身、実際どういった機関からサポートをもらっているかというのは答えられませんでした。プロジェクトからプロジェクトへと渡り歩いて、例えばこの機関はこのプロジェクト、そこまで知っているわけではないです。つまり、NIHから例えばトレーニングなどの要求がある。ただ、基本的にどういった機関からということは、なかなか本人たちはわからないわけです。

こちらでござんいただいていますのは、1年ほど前に出たものでありますけれども、執筆者はZolas氏。UMETRICSデータをもとにしています。



このピンの赤いマークはこのプロジェクトの加盟大学の位置を示しております。博士人材が米国内のどの州に行くかという行き先を示しております、博士号取得者は、ほかの州に比べてはるかに多くの人材がカリフォルニア州に向かっております。

ただ、こちらの8加盟大学のうちの6つの大学は、より多くの人々が、自分たちが卒業したその州にとどまっているわけです。ほかの州ではなくて、学んだ州にとどまるということです。米国の地理状況はこのようになっておりますけれども、こちらがアイオワです。これがインディアナ州。これらは皆様それほど魅力的な州とお考えになっていらっしゃるかもしれないですけれども、8大学のうち6つの州をみましても、多くがとどまるということです。

これは信じられないぐらい非常に重要なことなのです。オハイオ州立大の研究部長もいっていただきましたけれども、なぜオハイオをサポートすることが重要なのか。子供たちがオハイオ州に行き、そして彼らが卒業後オハイオ州にとどまってくれるのであればということを理解してもらうということです。オハイオ州の博士人材の支援がなぜ重要なのか、あるいは、ほかのところからなぜオハイオに行くのかということの説明の上で、こういった事実がわかることで、なぜこういった研究が重要なのかという意味をもってくるわけです。

こちらの表も時間の関係で詳細は割愛いたしますけれども、幾つか結果を紹介させていただきたいのですが、ご注目いただきたいのが上のほうです。

Table 1. Postgraduation employment of UMETRICS doctoral recipients who were paid by research grants and left the university between 2010 and 2012. The national workforce distribution is calculated from all employment in all establishments covered by the Census's LBD between 2010 and 2012.

Locale and small	Doctoral recipients placed in sector (%)				All
	Industry		Academia	Government	
	R&D firms	Non-R&D firms			
Placed within sector	17.0	21.7	57.1	4.1	100.0
National sample (M)	10.8	75.0	10.7	3.5	100.0
Of those in sector, percent placed:					
Within 50 miles	10.1	23.5	8.9	18.2	12.7
Within state	16.6	36.0	18.0	25.8	22.0

こちらで示しておりますのは、どこで就職先を見つけるかということであり、つまり学会から産業界にどのように行くのかということです。逆方向はまだみていないのですけれども。

これは博士号取得者がどういった業界、産業界、学会、また政府に——このように国関係、官で仕事を得るのは非常に少ないということであり、大学は驚くに値しないと思うのですけれども、57%ほどが学術界に行くということです。恐らくポストドクに行く人も多いと思いますけれども、そして4割ほどの学生が産業界、つまり業界に職を求めます。業界の中で、かなり多くのところがR&D、研究開発を行っている企業、すなわち、こういった人たちを雇うことによって恐らく利益を得られるであろう企業、彼らのスキルを使ってくれそうな企業に就職するわけです。

2番目の行は、米国の労働力全体のうち、R&D企業に就職する労働者は非常に少ないわけです。

そういった意味で、この人たちは特殊な存在なわけですが、これはハイテク企業ですとか、ほかの種類にまた細かく分けることもできますが、実態はほぼ似たようなトレンドであります。

こちらはまた非常に示唆に富んだスライドなのですが、

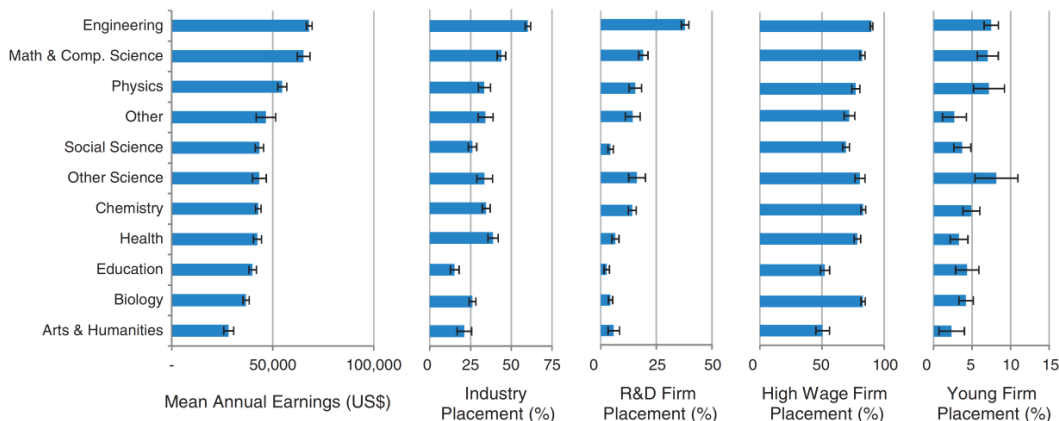


Fig. 3. The annual earnings and placement of doctoral recipients supported by grants vary by field. Young firm are defined to be those <5 years old. High-payroll per worker establishments are defined as those with a payroll per worker above the median for the establishments within their six-digit industry. Mean annual earnings are stated as U.S.\$1 ×1000. Means and standard errors for each variable.

別にこの後にテストをするということではないのですが、ここではProQuestという論文に関するデータベースをもっているところからデータを提供してもらっています。ProQuestの博士号取得者の論文のデータベースとマッチングするということで、アブストラクト、自然言語のプロセッシングなどのデータベースがあるわけですが、この学生が論文でどういったテーマ、サブジェクト、主題をみていたということがわかります。

マッチングしたところ、左側に修了後1年後の年次所得を示しておりますが、工学、エンジニアリング、数学、コンピュータサイエンスが最も所得レベルが高いということがわかります。これは驚くに値しないのですが、下に行くときとどんどん所得は下がっていきまして、人文系が、一番所得が少ない、2番目に低いのが生物学です。それから、化学もちょっと低目です。ポスドクに行く人が多いと思います。バイオロジーとか化学はポスドクが少ないのではないかと思います。したがって、その結果所得が低くなっていると思います。そういったポスドクの状況があると思います。

それから、アカデミアを出て産業界に職を求める比率は、例えば化学などは高いと思います。所得とは対照的に、R&D企業への就職率、それから賃金の高い企業に就職する比率、起業家精神ということで、設立5年以内の若い企業に就職する比率などを示しております。

こちらはさらに、資金提供状況、リサーチプロジェクトの支出先がどこか。またこちらに地図の御紹介ですが、米国のどの部分の大学に支出されているかを示しております。多くの資金は中西部に集中している。これは大学の所在地でもありますので当然ですが、東海岸、そして西海岸にも行っています。フロリダですとかテキサスなど、こういった諸州でもかなりの支出額があるということになります。

似たような調査をフランスの大学から、またスペインの大学からも、もらっています。米国が特別ということではありません。共通していると思います。基本的には、どこにお金が行っているのかというのを皆確認したいわけです。ですから、このようなものを日本でも追ったらどうだろうかということなのです。

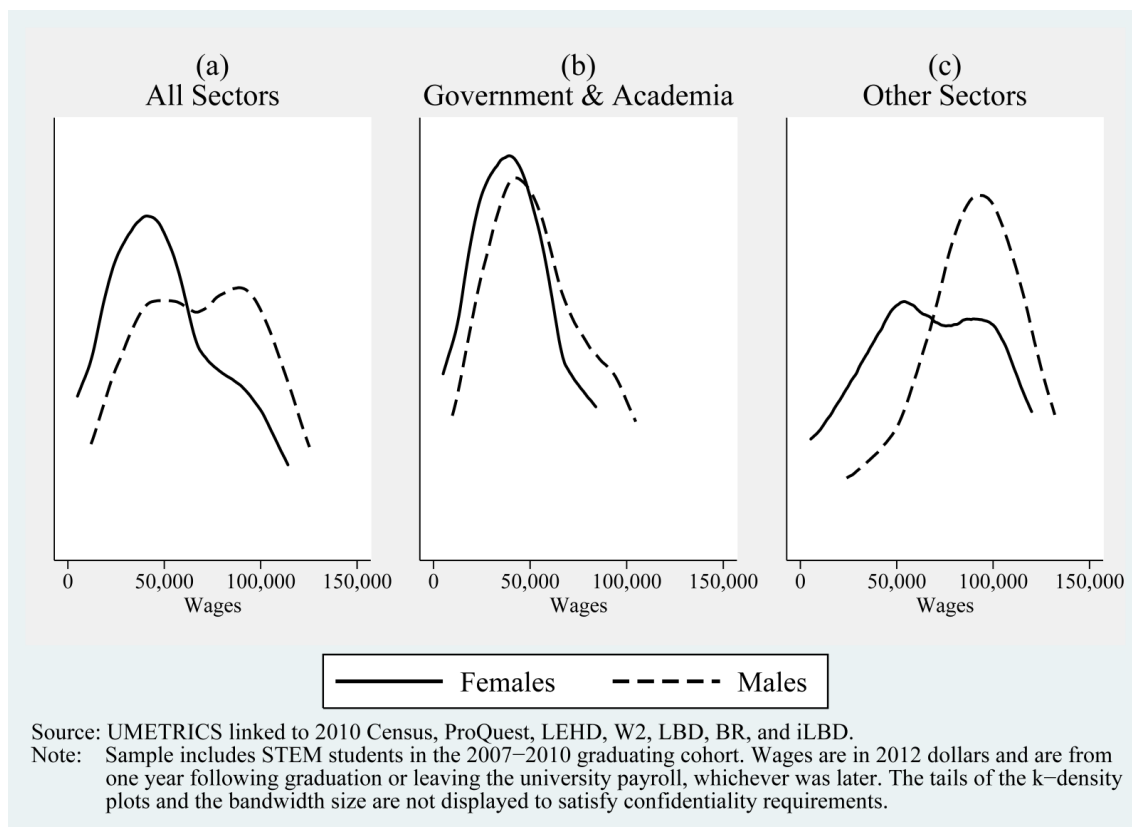
大学に対してアプローチをするときに、そんなことできないよと最初いわれるのですが、6ヵ月後ぐらいになりますとデータをいただけるのです。ですから、ほとんどの大学にとってはそんなに難しいことはありません。そして現行のシステムから出てくるために何かやる必要があると思います。そのようなことは決して不可能ではありません。

データが語ることですけれども、それぞれの企業の所在地です。余り細かいエコノメトリックスに関するようなことは申し上げませんが、ここで申し上げたいのは、企業が大学と商売をする場合、新しい事務所を設立するときに大学に近いところに開く傾向があるということです。それがデータでも厳密に裏づけられました。そして、そのビジネスの規模が大きい場合には、大学により近いところに事務所を開くという傾向が顕著にみられました。事業規模、それからどれぐらいの支出が行われているのかということと新たな事業所設立との相関は強い関係性がみられました。

これは非常に興味深い内容でありました。我々のところに投資、資金提供してくれるのであれば、例えば技術主導型の事業を創出しますとってくれることの裏づけになることです。本当かどうかはわかりませんが、ある意味、そのようなことが現実起こっているということがわかります。

また、米国内において、新たな事務所を設置している企業についてのデータをみるわけですが、やはり新規に開設される場所は関係性がある大学の近くである、また、ある程度の事業の規模をもっているところの近くに立地しているということがみられます。

性別による賃金の格差についても研究しております。



こちらでおみせしておりますのは男女の賃金格差です。これは卒業1年後の年収の違いです。男性は点線です。そして女性は実線です。ごらんのとおり、かなりの違いがみられます。女性と男性

の収入の差です。学术界、そして政府においてはギャップが縮まっているという傾向がみられます。しかしながら、産業界においてはかなりの差がみられるということがわかります。男性のほうが女性よりもかなり稼いでいるというのがわかります。

多くの女性は、男性も同様なのですが、年収レベルとしては大体4万ドル台後半が、ほとんどの米国のポストドクの人たちが得ている卒業直後の年収になっています。女性で学术界に進んだ人は、ポストドクで顕著にみられるこのような収入の傾向であります。

性別の差について、余り細かくは申し上げませんが、緊要している要素について次の表では述べています。女性は男性よりも33%、34%程度年収が低いということがわかっています。

この背景になっておりますのは2つの理由です。女性が専攻している分野であります。例えば生物学の専攻が多い、また男性の場合にはエンジニアリング、工学、コンピュータサイエンス、数学などの専攻の比率が高いです。ですので、このような要因が33%から11%ということで、この11%のギャップは、女性が既婚であるか、また子供をもっているといった場合には年収が低いということが示されています。よって、この2つの要素、専攻分野、そして家族構成について、このような条件が適用されない女性については、年収は平均値よりも高いものとなっております。これは非常に強力なデータです。

このような分析を行うに当たって、どれぐらいの量のデータを生データとして使っているかということですが、何を勉強したのか、どのような年収レベルなのか、家族構成について、このような情報が必要なわけです。ただ、このような情報は全て行政から取得したデータに基づいているわけです。

こちらの表に関しては、深くは申し上げません。この研究で行った別のことは、女性の年収は低い。これは、受けてきた訓練の内容について異なるのではないかとということで、女性が訓練を受ける環境、男性が訓練を受ける環境についても関連するデータをみたわけです。

Table 1. Training Environments of Male and Female Graduate Students Participating in STEM Research

Dependent Variables ↓	(1)			(2)	(3)	(4)	(6)	(7)	(8)
	(a) Females	(b) Males	(c) Diff						
Share of Faculty that are Female	0.2 {0.02}	0.1 {0.01}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1* {0.03}
Share of Graduate Students that are Female	0.1 {0.01}	0.1 {0.00}	0.0*** {0.01}	0.0*** {0.01}	0.0*** {0.01}	0.0 {0.01}	0.0 {0.01}	0.0 {0.01}	-0.0 {0.02}
Ln Team Size	1.7 {0.04}	1.9 {0.03}	-0.2*** {0.05}	-0.2*** {0.05}	-0.2*** {0.05}	-0.1** {0.06}	-0.1* {0.06}	-0.1* {0.06}	-0.1 {0.09}
Faculty to Student Ratio	0.9 {0.06}	0.6 {0.03}	0.3*** {0.07}	0.2*** {0.07}	0.2*** {0.07}	0.2*** {0.08}	0.1** {0.07}	0.1* {0.07}	0.3** {0.13}
Total Number of Awards	2.2 {0.07}	2.7 {0.06}	-0.5*** {0.09}	-0.3*** {0.09}	-0.3*** {0.09}	-0.2** {0.09}	-0.2*** {0.09}	-0.2*** {0.09}	-0.1 {0.15}
Number of Months Participating on the Award	21.0 {0.69}	21.6 {0.45}	-0.6 {0.82}	-1.1 {0.79}	-1.0 {0.79}	-1.0 {0.82}	-1.4* {0.82}	-1.4* {0.82}	-0.9 {1.18}
Years from First Observation to Degree	3.2 {0.08}	3.2 {0.06}	-0.0 {0.10}	-0.1** {0.06}	-0.1* {0.06}	-0.1 {0.06}	-0.1** {0.06}	-0.1** {0.06}	0.0 {0.10}
University, First Year Trend, Left-Censored				✓	✓	✓	✓	✓	✓
Race, Hispanic Origin, Age, Age-squared					✓		✓	✓	✓
Dissertation Topic						✓	✓	✓	✓
Funding Agency							✓	✓	✓
Married or Partnered, Children								✓	✓
Female x (Married or Partnered + Children)									✓
Observations	370	867	1,237	1,237	1,237	1,237	1,237	1,237	1,237

37

そして、ここで皆さんがご確認いただけますのは、右側に向けてですけれども、いろいろなコントロールが行われているのですが、男性と女性が受ける訓練の環境について、余り多くの違いがな

いということがわかりました。

女性は、女性のファカルティーがいるようなところで仕事を得る傾向があるということがわかりました。女性が多い学部に関しては、卒業した後で女性が就職する傾向が高いということがわかりました。また、規模の小さなプロジェクトに関しても、女性が多く雇用されるという傾向がみられました。ただ、最終的な結果としては、より多くの緊張関係があるわけです。全ての補助金、グラントの数ですけれども、何ヶ月間カバーされているかということに関しても、余り多くの違いはありませんでした。

訓練を受ける環境をみるに当たって、そして、それがどのように年収とひもづいているのかという相関をみますと、行政データだけに依存しては、そのようなことはわからないわけです。例えばあなたのプロジェクトは何名雇用されていますか、あなたは大学院生ですか。規模の小さな研究室であれば、そのようなことも可能でありましょうが、もっと大規模なところになってしまいますと、一々それを調べるということはほぼ難しいです。ですから、このようなデータも大学に提供することによって、どのような形で状況を改善することができるのかということに資するデータになると思います。ですから、規模の小さいようなところ、そして、学部により近いようなところで女性が職を得ているということがわかりました。

そして、これらのデータについて、どこが出身国か、ひもづけを行いました。これもまだ公表されていないデータ、研究の内容になりますけれども、こちらについても、生データに関しては国勢調査局からの提供を受けて行った研究です。午前中の話し合いの結果として出ているものではないのですが、余り細かい内容については御紹介いたしません、ここでごらんいただけるのは、異なる国の出身者、世界中から来た学生、博士人材がどのような分野に進んだかということの色分けしています。

そして、ここで驚くべき内容は、グレーの色の多さです。これはエンジニアリング、工学を主に専攻した学生の数です。米国内において博士号を取得した外国人留学生の多くが工学の学位を取得したということがわかります。これは米国人の学生よりも比率が多い。米国人の学生は、例えば生物学などについて学んでいるということで、余り工学には進んでいない。こちらはインドです。こちらは中東です。外国人留学生が米国内で勉強している専攻の内訳を示したものです。

また、実際の博士号取得者がどの程度米国内にとどまっているのかという割合についてもみました。そして、年収の割合をプロットしています。それぞれの国です。米国内にいたときに幾ら稼いでいたかということと、米国から出てしまった割合について縦軸にとっています。もはや米国内で仕事をしていない、または国外に出たということです。

そのほとんどは国外にもう既に出てしまったということを示しているのですが、ここでわかることは、プラスの相関がみられるのです。米国内において最も価値がある分野で学業を修めた人が米国外に出ているということ。例えば中東の出身者、そして欧州の出身者についても同様であります。では、ここでわかることですが、これはすばらしいデータです。もしこのような人材が米国内にとどまっていたらということなのですが、実際に出てしまったかとかというのがわからないわけです。日本人の大学院生、恐らくは日本に戻ってポスドクに進んでいるか、日本に帰国をしているのではないかと思います。それから、東欧出身の学生、途上国の学生が実際に自国に戻ったのかはわかりません。カナダ、英国など。

ただ、その問いには答えることができません。ここがこのような国別の行政データのよい点でもあり、また難しい点でもあるわけです。日本人の学生に関してはこのようなデータを持ってませ

んから、わからないわけです。皆さんのほうが、例えばJGRADの、日本の大学院生がどの程度米国に来ているのかというデータの一部については情報をいただけるかもしれません。

それから、起業に関しても一部研究を行っております。ここはそれぞれの分野ごと、スタートアップということで、立ち上がってから5年以内の企業でのシェアをみております。スタンフォード、MITなどですと、恐らくこの割合は高いのではないかと思います。ここでごらんいただけるのは、多くの方が設立から比較的日の浅い企業に勤めているわけではないということです。これは完全に驚くには値しないことだと思います。

若い企業というのは小さなところが多いです。なので、雇用者も少ないのです。ただ、学部生をみていただきますと、院生とかポストドクあたりと比べていただくと、グリーンはスタッフなのですけれども、やはりファカルティー、教員は非常に比率が少ないということです。スタートアップ企業への就職率が低下傾向にあるというのはほかの文献でもみられております。必ずしも米国のサンプルに特異な現象でもございません。

最後に御紹介したいのが、簡単ではありますが、高齢化におけるイノベーションであります。冒頭に述べましたが、これは国立老化研究所、それから行動社会科学、また科学財団などの協賛もいただいているのですが、ここでの問題は、米国の学術界コミュニティの高齢化が進んでいるということ。その示唆するところは何か。

3つポイントがありまして、どういった力が働いてくるかということです。

David Blau、それからオハイオ州立のJohn Ham、あとシンガポールの方。確定給付型の年金を使いまして、例えば30年とか35年働けば、多額の退職給与がもらえるわけですが、多分、いろいろ大学から大学へ移籍することもあるかと思いますが、給付を受けるということとキャリアの後段階の生産性の関係をみようと思っているわけです。

もう1つのプロジェクトは、Pierre Azoulay、MIT、それからIna Ganguli先生、それからJosh Graff Zivinさん、UCLAのサンディエゴ校ですけれども、こういった方々が手がけているプロジェクト。これは研究者の移動が生産性に与える影響でありまして、やはり学齢期のハイスクールに通っているお子さんがいらっしやると動かないということでもあります。地理的な移動が生産性に与える影響をみています。

それから、研究の革新性、あるいは健康上のインパクトを測定できないかということでもあります。

2人の研究者によるもの。データは使えないものもあるのですが、多くのデータは公表されて使えるものであります。もちろん皆様の大学にもそういった研究者の方はいらっしやると思いますが、みていただけるかと思います。

3. まとめ

○Weinberg教授： 私は、エコノミストであります。午後もかなり遅い時間になってまいりましたけれども、まとめをさせていただきたいと思っております。

多くの国におきまして、もちろん米国もそうですけれども、やはり絶対的に必要なのが研究の価値を測定するということでもあります。研究の価値をはかるということは、研究の資金を得るためにも必要でありますし、また、政策立案者にとって、我々にとっても重要です。

そして、よりよい仕事をするために、という意味でも、研究者にとってもこれを理解することは

必要であります。また、これから研修を受ける人たちにもこれは必要。つまり、研究が終わったらどうなるのか、ノーベル賞をとれるのか、あるいはグーグルで仕事をするのか、どういったことが起きるのかということを知るのも重要だと思います。恐らく、学位を取得した後、何が起きるのかということを知るのは、学生にとっても重要だと思います。

また、冒頭から繰り返し申し上げておりますが、科学、イノベーションというのは国際的なものです。日本のイノベーションから我々は利益を享受することができますし、皆様にも米国のイノベーションの利益を提供することができるかもしれません。また、日本のポスドクから米国に、あるいは米国から成果が共有されるということで、真の研究の価値を理解するためには、国際的なフロー（流れ）も理解しなくてはなりません。

なので、第1段階といたしまして、私が思うには、UMETRICSであったり、JGRADであったり、JD-Proであったり、名前は何でもよろしいのですが、各国による取り組みです。ただ、究極的なゴールとしては、やはりデータ統合だと考えています。といいますのも、研究の価値は実際何にあるのか。国際連携が重要。国際的な流れ、国境を越えてデータを統合しないことには、そういった価値の提供を伝えることはできないと思います。

既にかかなりの作業が米国では行われております。また、日本でもいろいろな試みがされていると理解しております。こういったデータのプラットフォームの相互運用、相互連携がさらに必要になってくると思います。これがさらにますます重要性を増すと考えております。ご清聴ありがとうございました（拍手）。

4. 質疑応答

○司会者（松澤総括上席研究官、NISTEP）： 大変興味深いお話をありがとうございました。いい機会ですので、会場の皆様からの御質問等をお受けしたいと思います。いかがでしょうか。

○三須（広島大学）： プレゼンテーションをありがとうございました。広島大学からまいりました。コメントと質問です。おっしゃっていましたように、まず、日本の状況を理解するというのはいいと思いますが、問題は、米国では国勢調査局の行政データにアクセスできるということであり、現時点におきまして、我々はデータ制限という問題がございます。これができないということです。

そこで質問ですが、米国では国勢調査局とのパートナーシップがあるということ、データアクセスにどういった規制があるのかということ、また、何か規制の変化はあったでしょうか。つまり規制が変更されることによってデータアクセスができるようになったということはあるでしょうか。

○Weinberg教授： 非常に重要なトピックです。御質問ありがとうございました。

まず、コラボレーターで御協力をいただいた方々のお名前をここに掲載させていただいております。国勢調査局はデータを収集し、米国の人口、また経済への理解ということで、できるだけ極力データを使えるようにするというマンドートをもっているのですけれども、セキュリティのかかった形でのリサーチ・データ・ネットワークがありまして、これは資格をもった担当者がセンサスデータを使ってリサーチをできるようにということなのです。我々は幸いなことに、このインフラと接続することが認められまして、我々が構築したのではなくて、既存しているインフラに接続することができるということなのです。

ただ、こういったデータを取得するのが簡単だと過小評価するつもりはないのです。といいます

のは、Jason Smithという人間が実はこれをやっていて、私、直接はかかわっていないのですけれども、今、50の大学、まだデータ収集の段階にある大学もあって、全てを入手できているわけではないのですが、我々が各大学にアプローチするときに、誰かがしたら、それはすばらしいけれども、こんな法律があるから、自分たちではできないですというところもあるわけです。法律の専門家を携え、大学一つ一つごとに、何ができるのか、法の例外規定とかセキュリティーについて十分な説明をする。では、可能かどうか検討しましょうといってくれたりするわけです。

ピープルソフトからデータは入手できません。では、ピープルソフトからどうやってデータを得るのでしょうかということで、もちろんデータ取得の懸念を極小化するつもりはありません。ただ、そういった価値を理解していただけますと、それを実現できる方法を何とか見出せないかということで、そういった取り組みは強みと弱みと両方あって、趣旨はわかります。例えば自分たちができないことを日本でやっているとか、日本を出て米国やそのほかの国に行くリサーチャーに関して調査ができるかできないかという問題もあります。実名は出しませんが、例えば国防省との契約の関係で機密扱いなのでデータの提供はできませんとか、そういった場合もあります。非常に困るわけです。そういったデータは難しいということがあるわけです。ですので、もちろん各大学は、問題を抱えていて、この部分はデータの提供はできかねますということもあるわけです。

パートナーシップでありますので、これをすべきというか、できることをやっていく。そして、できるところから始めて、では、ネクストステップでもう少し貴重な情報の提供。そのときとともに、そういった自分たちがやっていることの価値を認識していただけるということだと思います。やはり価値を見出せないからには、相手は何もしてくれないと思います。

○司会者（松澤、NISTEP）： いかがでしょうか。

○林（科学技術振興機構）： 非常に触発を受けるご発表をありがとうございました。2つ質問があります。先ほどUMETRICSの国際的な対応について御説明をいただきました。フランス、スペインやオーストラリアも数年前にUMETRICSのパイロットスタディを試みたということでしたけれども、国際連携の動きはあるのでしょうか。またUMETRICSの米国版と国際版で何か相違点はあるのでしょうか。

それから、家庭ごとの意思決定における不均質性について質問があります。例えば私が大学院を卒業し新しい職を見つけようとする時、私の妻が気に入っている仕事に就いているとしたら、私の職を選ぶ範囲も制限されるわけです。そのような制約が雇用市場、また性別の格差、そして賃金格差などにも影響を与えると思うのですが。

○司会者（松澤、NISTEP）： 所属とお名前をよろしく申し上げます。

○林（科学技術振興機構）： JSTの林信濃です。

○Weinberg教授： 大変すばらしい御質問をありがとうございました。済みません、ここには入れていませんが、オーストラリアもこの方向性で一定額の投資を行っています。これは人または集団として、このようなことを行うに当たっての価値を見出すかどうかということによるのです。さまざまな障壁がある場合は、そのような障壁を取り除き、コンソ（コンソーシアム）を組む、または機関連携を行う能力によります。

国際的な要素はもう少しややこしいです。やろう、やりたいとは思っているのですけれども、日米とかがもしかしたら最初に実現できるかもしれませんし、できたら素晴らしいと思うのですが、法的な制約がありまして、ここがより難しい点です。ですから、ステップワンは重要なキーマン、主要なプレーヤーに内部からアプローチをするということです。

疑いようのない明らかなことは、ちょっと考えてみていただきたいのですけれども、ミシガン州はオハイオ州のデータをもっていません。オハイオはミシガンのデータをもっていません。カリフォルニアのデータは誰ももっていません。ですから、何が起きているのかということに関して、像は不完全なのです。これを国際レベルでやろうとすると、国内でやるよりもっと難しいです。ただ、最終的には、我々の加盟大学は日本から大学院生、ポスドクの学生が来ています。何名かは戻りましたけれども、みてください、ポスドクのこれだけの人たちは残っているのではないですか。ポスドクの人たちは日本から来ていますよというようなことを語るようにしたいわけです。ミシガンから人材が流出しましたけれどもとか、そのようなこともわかるようにしたいわけです。

どの国も同じような環境で同じようなことができるとは思いませんが、このようなデータをとることによって有用性を認められるようになりましたら、このような種類のプロジェクトの価値が示されることによって、恐らくは少し進捗をみることができないかと思います。

2つ目の質問です。それぞれの世帯に関するものでしたけれども、ぜひ我々もやりたいところですが、これは家族構成にかかわらず、世帯を構成する人たちにに関する状況です。既婚であれ、婚姻関係にないとしても、パートナーが何をやっているのかということについて把握ができるということは大変素晴らしいと思います。

例えば私は既婚で、ドクター、または法曹、航空会社のパイロット、コンサルタントであると。ただ、スキルが国際的、またはそうでない場合もありますが、これがどのような形で自分のやっている研究、また仕事に影響するのかというようなところについては、ぜひこれから研究を深めたい、また研究を開始したいと思います。恐らくは、そのようなことを進めるに当たっては大学院生、ポスドクなどの人たちが必要だと思いますけれども。

○斎藤（総務研究官、NISTEP）： NISTEPの斎藤です。示唆に富む発表をありがとうございます。本日の参加者の方は非常に興味をもってお聞きになったのではないかと思います。

具体的な質問が幾つかありますが、まず、こういった個人データをインプット、アウトプット、出版データとか、その他研究、あるいは納税データと組み合わせられるというのは、日本ではまず想像できないような試みだということで、非常に素晴らしいと思います。

やはりターゲットを絞らなくてはいけない。つまりグリーンカードの取得者のみとか、米国市民権をもっている人だけ、そういった人であれば納税記録も測定できる。本当に最近でありますけれども、日本でも個人のマイナンバー制度が導入されました。ただ、人々は一般に、個人データがナンバリングで特定されますので、こういったものでも使えるということで、これはプライバシー侵害の問題で懸念が高まっているわけなのです。そういった懸念は米国にないのでしょうか。こういった分析をすることによって、目にみえるインターベンションはないということで、対象人口の間で懸念がないのかということを知りたいのが1つです。

2番目は、科学、イノベーションは国際的なものだということに異論はないのですけれども、そういったリサーチを訴求する中で、やはり国際データも必要になってくると思うのです。例えば米国の大学は海外から学生を多く受け入れていって、博士号取得を目指したり、学位を取得

して母国に帰国される、あるいは別の国に行くかもしれませんけれども、学术界に就職するなり、あるいは起業したりと、新しい価値をつくっていくと思うのです。

その一方で、当然米国には高い授業料も払っているわけです。少なくとも米国の国民だけを見る時に、大学に向けた公的な支出、投資効果、あるいは外国人留学生に対する投資、公的支出の合理性、どういった効果があるのか、ざっくりとした見積もりでもよろしいかと思えますけれども、あと、国際データも必要なのかどうかについても御意見を伺えればと思うのです。

○Weinberg教授： 米国の個人のIDというのは社会保障番号になります。氏名を唯一のベースとしてデータリンクしているのですけれども、名前だけで7割、8割、9割は取得できると思います。韓国の名前ですとか中国の名前とかはちょっと違いますよね。日本の名前もまた違うと思いますけれども、特に、アルファベットの名前というのは多くあるわけですが。まず氏名、それから生年月日があれば99%以上捕捉できるということで、社会保障番号は使っていません。このリンクのために社会保障番号は利用していません。それは必要ないです。

率直に申し上げて、社会保障番号が手に入っていないほうがよく寝られます（笑声）。社会保障番号を部分的に、例えば3桁とか4桁とか、そういった場合、加盟大学とも話していたのですけれども、トランスクリプトデータで、学部生なども含めて、1つ提案しましたが、ランダムに選ばば、6のうちの2つとか3つとかわかる。思うに、今は社会保障番号のデータはないです。そういったデータを使うのはできるだけ少なくしたいと思っています。それは今質問にあった全く同じ理由なのです。

それから、国勢調査局は基本的に守られたIDで社会保障番号がありますけれども、データ移送のときに、あちらに社会保障番号があつて、実際その部分は欠落いたしますので、そこにはデータ移送の段階で価値がなくなると思います。それから、社会保障番号は必ずしもいつも正確というわけではないので、問題の解決になるとも限らないわけです。それから、生年月日と申しましたけれども、年月までで、何日はないです。それでマッチデータが99%です。やはりフルで生年月日を出すことには非常にナーバスなところもあるので、年月までを使うということです。

そこは簡単な質問だと思いますが、難しいほうの質問です。投資収益率の問題、ROIの質問だと思いますが、正直に申し上げます。実はその質問に対するお答えはないのではないかと。データの問題というよりも、実際具体的な概念フレームワークがあるとは思っていません。国債を購入するのと比べて投資収益率がどのくらいなのかなど。

もう少し概念的なフレームワークが必要かと思っています。これは私の理解するところ、日本は、ほかの国は違うかもしれませんが、米国に来る留学生の多くは米国の機関が支払いをするわけです。母国が支払いをするわけではなくて、米国の制度のもとで支払いを受けるということです。多くの卒業生はリサーチプロジェクトで支援を受けるということです。また、リサーチもやったり、教えたりもしますので、そのときにコストベネフィットを考えなくてはいけません。つまり支払いもしていますけれども、その報酬として提供もしてもらっているわけです。

ちょっとまだデータの統合までは行っていないのですけれども、1つできると思うのが、論文の執筆者とか発明者、米国外のそういったデータをとって、第1執筆者、そしてプロジェクトに関して何か報酬を受けたデータを見るとか、あるいは、例えば特許の申請をした場合に、そこに米国の発明者がかかわっていたとか、そういったデータの追跡をするということもできるかなと思っています。最重要とはいわないまでも、非常に重要な意味のあるものかと思っています。もちろん重要ではありませんけれども、現実的に考えると、すぐにできるというものでもないと思います。

○齊藤（大阪大学）： 非常に内容の濃い、またエキサイティングなご発表をいただきましてどうもありがとうございます。齊藤と申します。大阪大学です。

非常に興奮しましたのは、私の博士論文は教育訓練の費用便益分析だったのです。大学院生に対して調査を行いました。そして、マレーシアのトレーニングセンター、タイなどでも行いました。現在は税務のデータについてもおもちであると同いまして、税務のデータはキラーコンテンツになるということを私は十分承知しております。

規制について三須教授からもご発言がありましたが、さまざまな法律が省庁にまたがる形で適用されています。とりわけ厚労省が非常に閉鎖的という印象をもっています。

2つ質問があります。博士課程の学生に関して、研究の金銭的な価値についてお話がありましたけれども、それ以外の価値もあると思うのです。金銭的な価値だけを考えますと、研究結果、また教育の結果を過小評価してしまうことになると思いますが、ビッグ・データを活用した形で、金銭的な価値以外の部分に関して行われた調査はありますか。そのような概念をおもちでしたら御紹介ください。

もう1つは、主に絞っていらっしゃるの、さまざまな対象群の違いによつての比較研究と拝察しました。もし可能であれば全体像を伺いたいです。米国の教育はいいのか悪いのか。費用便益分析に基づいてどうなのかというのをご教授ください。

○Weinberg教授： 全くもっておっしゃるとおりで、財政的な価値、金銭的な価値が唯一の価値ではないということです。そして恐らくは最重要、または最も大きな部分でもないということです。

現在進行中の高齢化に関するプロジェクトの一環として行っているものですが、この3つのうちの最後のものなのですけれども、MD-Ph.D.経済学で、スタンフォードでやっている人、それからウオータールーでやっている経済学者がもう一人いるのですけれども、彼らの研究内容は非常にエキサイティングな形で現在進めておりまして、これは我々も用いているのですが、やろうとしていることは、新たな考え方を特定しようとしています。新世界の導入ともいうべきものでありまして、これは研究者に関しての新世界の紹介というようなものです。DNAとかPCRなどを参照しています。これによって研究論文を薬事承認、また最終的には健康効果にまで結びつけるものです。

非常にご丁寧な形でご発言いただきましたけれども、価値を定量化する。NIHの資金提供を受けて研究を行うことでどれだけの金銭的な価値を生むのかということです。ただ、根本的な目的は、病を治すということです。研究活動がどのようにバイオメディカルのブレークスルーにつながっていくのかというところをみたいわけです。ですから、これは非常にエキサイティングなアプローチだと思います。

ただこれは、より難しいことです。定量化するということが決して簡単だといっているわけではありませんけれども、より難しいと思います。また、追跡が難しいです。テキストをみて、そのテキストがどのような形で文章の中で表現されるのかというような追跡の方法もあると思いますけれども、UMETRICSのデータとは一線を画するものです。

雇用主または雇用側からの収入について追跡をするに当たって、こちらとして関心を払っているのは、人々がどのようなアイディアの方向性をもって進めていくのかということです。人の考えが経済の方向を決めていくと思います。この研究室、この会社、この論文は、この研究プロジェクト、この研究機関からの支援を受けている、このアイディアが知財にあらわれているのか、そして特定の機器メーカー、または製薬メーカーが製造しているものにつながっているのかという関連性をみているわけです。既に申し上げましたが、今そのような方向に進もうということによってやっております。

ただ、まだ概念設計の段階です。ただ、そのようなことも方向性としては計測の対象に入りたいと考えています。

経済学者として予測するのは、例えば、がんを治癒することができるのか。もちろん実際にそれに資するだけの研究であれば、そこからもたらされる金銭的な価値も非常に大きいと考えます。ただ、おっしゃるとおり、最終的に本当にやりたいことは、どのようにアイデアが現実になっていくのか、研究が形になるのかということです。

もうちょっとお答えが難しい、また、大変すばらしい質問でもあった部分は、米国として取り組む大学がいいのか悪いのかということですが、それは、JGRADに参画をして、また構築をするということで、興味深いのは、大学のうちの何個かは、他の大学に比べて、うちの大学の仕事ぶり、パフォーマンスはどうなのだという事、自分たちのことについてはよそに比べてほしくないけれども、よそのことは知りたい、ということを知りたい、ということを知りたいです。大学と一緒に仕事をしていますので、正しいモデルで、例えば6校ぐらいのところは相対的にどうなのかということについて、ただ、明確にどこの大学ということにはわからないような形にして、もう少しぼかした形で相関をみるというようなことに関しては、確認ができると考えています。

基本的には、我々にやってほしくない加盟大学が考えていることをやるつもりはありません。やりたくはありません。非常に質問の深い部分ですけれども、仕事ぶりはどうなのか、また、どうやって改善していくことができるのかということですが、エビデンスで示されていることは、ある程度いい仕事をしているということ、そして学術界においても、また産業界においても、どのようなスキルがより多く用いられているのかということに関しては、もう少し改善の余地があると思います。それ以外の部分としては、どのように自分たち自身のパフォーマンスを上げていくか、よりよい質の取り組みにしていくかということです。

女性の数が参画者として少ないようなプログラム、例えば大学院のクラスに関して、もっと女性の数が多ければ、1名、2名、それか全く女性のいない大学院のクラスと比較してどうなのか。また、外国人だったらどうなのか。

御質問いただいたところに厳密に直でご回答することは難しい、答えの方法がわからないのですが、御紹介をしたさまざまな取り組みを進めていくことで答えに近づけるのではないかと思います。

○司会者（松澤、NISTEP）： ありがとうございます。時間も押しているのですが、最後に短い質問を1つだけ、もしあればと思いますけれども、何か特にこの機会に御質問したいということがあれば、よろしいでしょうか。司会の立場から1つだけ……いらっしゃいますか。よろしくお願ひします。

○Larsen（NISTEP客員）： 大変興味深いプレゼンテーションをありがとうございます。私の質問は、トランスフォーマティブなサイエンスのインパクトという最後の部分に非常に興味がありました。さまざまなバリエーションがあるので、タイポロジー、タイプごとにトランスフォーマティブ、変容性で何かクラスターが見つかったのかどうか伺いたいということ。あと、時系列について。サイテーション、引用のウィンドウ。Katarina Larsenと申します。NISTEPの客員研究員をしております。

○Weinberg教授： こちらのスライドです。余り時間は割きませんでしたけれども、幾つかはス

ライドを抜いておりました、これは簡単ではあるのですが、何をやったかといいますと、マトリックスを策定いたしまして、研究の変容性を測定しようとしたわけです。

学際的な非常に幅広いインパクトがあるものということで、トランスフォーマティブリサーチの7つの特色を取り上げまして、それを引用と自然言語プロセスで分析を行いました。どの論文が新しい重要なコンセプト、あるいはユーズということで、例えばDNAを使っているか、あるいはポリメラーゼ連鎖反応を使っているかというようなもの、あるいはRNAインターフェアランスといった用語に引用があるかをみました。

また、トランスフォーマティブネス、縦軸に変容性を、それから標準的な尺度であるインパクトをとりましてプロットいたしました。相関係数がおよそ0.4でありまして、インパクトのあるものが、変容性があるのかどうかということで、ただ、相関係数1からはかなりかけ離れております。非常にインパクトはある、少しずつ進歩はあるけれども、必ずしも革命的ではない。完全に思考を変えるものではない、ゲームチェンジャーではないということです。

古典的な例といたしましては、トランスフォーマティブリサーチで、例えば物理学における量子革命です。20世紀初頭の話ですけれども、インパクトと変容性に相関関係はありますけれども、1からはほど遠いということです。つまり完全な相関があるとはいえないということです。

プロットをいたしまして、こういった形で、赤で3つハイライトしています。これはステムセルズです。文献をみていただくと、トランスフォーマティブな、変容的な研究領域です。ヒトのゲノムプロジェクト、DNAのメチレーション、メチル化にかかると。これらが専門家からみてトランスフォーマティブである、変容性の高いものと評価されたものになります。

御質問はタイムフレームですけれども、非常に興味深かったのは、トランスフォーマティブリサーチという考え方は評価に非常に時間がかかるということなのです。非常にラジカルなアイデア。ただ、なかなか評価を得られない。サイエンスの進化は時間がかかるということで、有名な引用句があるのですが、ちょっと今思いつかないのですけれども、若い人は新しい考えを思いつく、年配は古い考えにこだわる。しかし、私たちはそう思っておりませんで、トランスフォーマティブなアイデアは、長い期間でも使われるし、短期間でも使われるということです。

実際、長期間以上に、トランスフォーマティブは最初の5年が非常に重要であるということもわかっています。我々のデータは1983年から2012年までであります。これが限界なのか、非常にラジカルなアイデアがあつて、83年に出たものが30年ということはいえないのですけれども、より短い期間で影響があるということはいえるのではないかと思います。

○司会者（松澤、NISTEP）： 大変興味深いお話をありがとうございました。また、先生のお話の中で、我々（NISTEP第1調査研究グループ）としても、例えばエイジングですとか幾つかのデータがありまして、そういったものの将来的な統合がどういう夢を描くかということもよくわかりました。最後に、当方の総務研究官から一言ご挨拶させていただきます。

○斎藤(NISTEP)： Weinberg教授、大変すばらしい、また触発を受ける講演をありがとうございました。ほとんどの聴衆は日本人ですので、日本語で申し上げます。

きょうはお忙しい中、この講演会にも御出席をくださりましてありがとうございました。

当初は、アメリカの進んだ取り組みのお話ということで、私自身も、実はアメリカは日本以上にソーシャル・セキュリティ・ナンバーとかそういうものを使って、ある程度の行政データとデータベースのデータを接続したりしやすいからこそ、こういう分析ができるのかなと聞いておりました。

たけれども、今日の質疑応答の中では、そんなナンバーを使わなくても、名前と生年月日を使って結構な名寄せができるというお話も伺いました。

我々も実は、博士人材データベース（JGRAD）と同時に、博士人材の追跡調査(JD-Pro)も3年おきにやっております。その中の質問の1つは、やはり所得水準という質問もあって、これが結構現場では、えっ、こんなことまで答えなければいけないのかという当事者の方の抵抗感も強い項目でございます。

ただ、それはマイナンバーを使って国税当局からデータがもらえるかというのと、これはなかなかハードルが高いわけでありましてけれども、将来的にはそういったデータ接続の可能性も考えながら、民間の雇用者を含めた長期的な追跡がどういった形で可能になるかということもやはり考えていく必要があると感じた次第でございます。

きょう伺ったお話を我々もぜひ今後のJGRADの設計ですとか運用に生かしていきたいと思っておりますし、皆様自身も、こういったいろいろな使い方ができる、あるいはいろいろな見せ方をすれば、もちろん短期的には、JGRADの登録者、あるいは大学側のメリットもありますが、社会に対するアカウントビリティという意味でも、こういったデータの使い道はいろいろあるのではないかとヒントをきょうは持ち帰っていただけるのではないかと思っております。

もう1つ、彼のお話の重要なメッセージは、やはり国際的な比較可能性とか、データの必要性ということだったと思います。もちろん皆様の大学にもたくさんの留学生が来ておられると思います。今のJGRADの仕組みですと、留学生の追跡はなかなか難しい状況ではございますけれども、日本で学位をとった方がいかにして活躍しているかを追いかけていくことも、広い意味での日本の高等教育に対する投資とか、あるいは大学の重要な役割としてみていく必要もあると思います。

一方で、グローバルにみますと、今アメリカもトランプ政権で科学技術イノベーション政策が厳しい転換期を迎えていると思っておりますし、日本の大学に関しても、非常に財政状況が厳しい中で、基盤的経費ですとか、あるいは博士を目指す若者自身が減っているといういろいろな厳しい課題に直面しております。そういう中で、国を越えて博士人材の役割なり価値が大変高いのだというデータを、きちんと国際的にも比較可能な形で出すことによって、しかも国際的な場でそれを共有することによって、我々の仕事の重要性なり意義をしっかりと国内外に訴えていくということも、特にトランプ政権のもとでは重要なのではないかと考えた次第でございます。

我々も実は今少し構想していることがございまして、例えば、毎年AAASの年次大会が2月に開かれますけれども、来年はテキサス州オースティンで開催予定になっております。そこで、こういった課題を国際的に比較、分析するとしたら、どういう課題があるかということも、できればセッションの提案などでもしたいと思っておりますし、その際に、ぜひ日本のJGRADを初めとした取り組みを、世界を相手に紹介しつつ、今後のあり方について議論ができればと。まずは企画競争を勝ち抜かなければいけません。これをご縁に、ぜひWeinbergさんとも御相談しながら、そういった国際的な場での発信もしていきたいと思っております。

今後、引き続きJGRADの関係者の皆様におかれては、御協力を期待しておりますし、逆に皆様からの提案もぜひお待ちしております。本日は最後まで本当にありがとうございました（拍手）。

○司会者（松澤、NISTEP）： 本日は長い時間どうもありがとうございました。もう一度Weinberg博士に拍手をよろしくお願ひします。（満場拍手）本日の講演会は、これにて終了させていただきます。ありがとうございました。——了——

講演録（英文）

(MC: Matsuzawa)

Good afternoon, Ladies and Gentlemen, I would like to start this seminar on UMETRICS. Thank you very much for your time indeed, for being with us. My name is Matsuzawa. I am the research director, heading the Policy-oriented Research Group One, and today we have the honor to have the speaker Dr. Bruce Weinberg from the United States, and before we get started, I would just like to give you a little bit of background. The project UMETRICS...

Eight years ago, the Obama Administration tried to explore the effectiveness of the research project so it has to do with the Star Metrics and the Obama Administration, and then the universities in the US, mainly universities located in the Midwest, so the consortium of the US universities were set up and using the metrics, there was wide-ranging research initiative started, and last year in December 2015, the research result was publicized in science

and which brought about lots of the... drew lots of attention from all over the world and we also got a lot of the inquiry about this project, including the similarity of what are doing to the UMETRICS in the States, and so that is why we wanted to have this symposium by inviting the registered members of JGRAD

and we understand today's speaker is very busy and he has to leave for the US tomorrow, before getting into the symposium per se, just a little background. In 1996, he got a PhD in the University of Chicago for economics, and so majoring in the innovation competitiveness among different countries and so forth, and as for UMETRICS, he's one of the main researchers, and he traced the outcomes of doctoral recipient, and also he was invited to the Federal Reserve Bank of Cleveland as well as Hoover Institution and the National Bureau of Economic Research.

We are very interested in this project because in the States, mainly NSF are taking the lead and there is a program called the SED to track the doctoral recipients, so the project is already ongoing in the States, and this is something we need to turn to. We also have this way to track the doctorate recipients, so as we think about the doctoral human resources data base, and I think there will be a very informative presentation with lots of implications, so Dr. Weinberg, the floor is yours

(Dr. Weinberg)

Thank you very much. I'm really honored to be here today

You know it is, I think, probably obvious to everyone and will become obvious as we continue, that innovation and science and research are international and so while we'd like to think that we've made tremendous progress on understanding research and its value in the United States, even that is very incomplete without a data and an understanding of research and its value internationally.

Ideas produced in the US flow across national boundaries to Japan and to other countries. People or researchers move from one country to another country. The US, if we look at the value that we get from research and we don't look at the value that we get from research produced in other countries, we're missing a huge portion of the picture, so I think it's very valuable and very important if we're doing these types of studies to be thinking about the international connections and the international linkages, and I am, as they say, really honored to be able to be here and to be able to talk about what we're doing and maybe show it as something that could be relevant for the JGRAD Initiative

So the focus of what we've been doing is to think about how to measure the value of research. A lot of work trying to understand the value of research has focused on bibliometrics-- counting publications, counting citations, looking at the impact factors of the journals in which research is published, and you know there are a million different things that people have done to try to get better measures, but what we've been trying to do is to think more and more deeply about the value of research in two ways

One thing that we've tried to do is to think about people as the primary or one of the primary products of the research enterprise – that the graduate students and the undergraduates and the post-docs who were employed on research projects are one of, if not the single most important product of the research that they are doing

And so if we want to see the value that they're obtaining from that experience, we can look at the outcomes that they have after they graduate, we can look the industries and the businesses that hire them, and we can try to understand the value of their skills from that... from looking at what happens to them

The other thing that we've been doing this morning is that we were talking about connections between what we're doing at UMETRICS and things that are happening here at NISTEP, and so it seemed obvious that there were some slides that I had that might be less relevant and some material that I had that might be more relevant than the slide deck that you had, so I'll mention this work but I made some tweaks to the slide deck to try to make it most relevant for you

So the second thing that we've done, and this is the piece that I'll de-emphasize as we talk today, is to try to develop metrics for the newness and the novelty of research, and those are obviously related to the impact of research that we've studied in many cases, but are distinct from it in terms of trying to capture really novel, innovative, revolutionary aspects of research

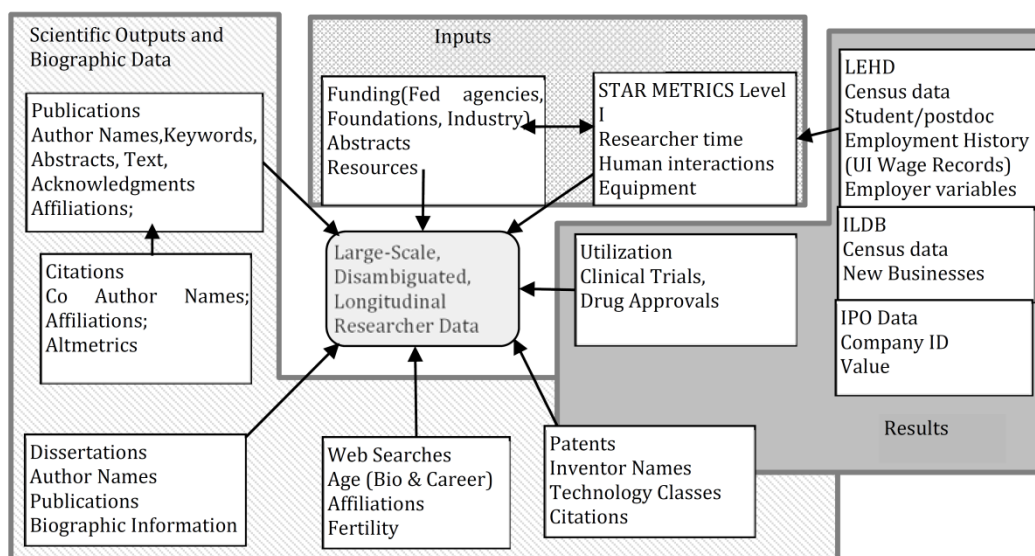
So there are two initiatives that I'm involved in. The first of those is the Institute for Research on Innovation and Science, or IRIS. The primary product, the primary initiative of IRIS is UMETRICS, and UMETRICS is a partnership between universities, the Census Bureau, and the University of Michigan. There are nodes at Ohio State University and New York University, and potentially more nodes to come, and we've been very fortunate to have support from a wide range of really wonderful funders,

without whom this work would not be possible

The second initiative is Innovation in an Aging Society, and I'll talk about this one a little bit, not as much as in the slide deck, and the idea here is that researchers are widely seen to be most innovative early in their careers. Our data show that that's not quite as true as the conventional wisdom is, but what is happening is that the research workforce is aging and if people are most innovative relatively early in their careers, that potentially... that aging workforce is potentially going to reduce the productivity of the scientific enterprise in countries like the US and Japan where the research enterprise is aging

and both of these initiatives are producing large-scale data that is being made available to the research community, so we're using the data but we're also trying to make sure that other people can access the data

So this is a schematic of our data infrastructure



I'm not going to go through it in detail. You're probably pleased to hear that, but what I do want to point out is some of the key data sets and key links that we're making, so at the heart of our data are the UMETRICS data. The UMETRICS data, for people who aren't familiar with it, came out of the Star Metrics Initiative and basically they're the backend of payments on sponsored research projects

So if you think about it being like a checkbook or a credit card account, basically you need to know who the money is going to, you need to know how much they're getting, you need to know what you're getting for it, and you need to know what account the charge should be billed to, right?

And so that's really powerful, because if you think about a research project, what does

it do? It pays the people who were employed on that project, so you can identify every single person who's employed on that research project, whether they are a faculty member who's the last author on the publication or some undergraduate or staff person who never appears on a publication, you can identify them

You can also identify all the purchases that are made on those projects. Every piece of equipment, every supply, every computer, so you can actually identify all of the inputs into the production of the research, both the personnel and the equipment and materials and supplies

So then we're linking these data out to a number of different things. We're linking them out to publication data, so we can see not just everything that went into that lab. We can measure all the publications that came out of that lab

And we have, as I said, citation data and metrics for novelty and innovativeness and impact of that research

You can measure the patents that come out of that research

We're trying to find ways of linking in a lot of the work that we're doing supported by the National Institute of Health, so there's a focus on biomedical research, and so you can try to link in to drug approvals and medical device approvals, and so forth

So that's one set of linkings, to the scientific outcomes, but I think increasingly people are interested in the economic and societal value of research, and so what we've also done is we've matched our data to Census data that include both the decennial census that's conducted in the US, it's a federally-mandated survey of everyone in the US every 10 years, but the Census Bureau also houses a wealth of data from across the federal government, and most notably, it has data from tax returns, so from these tax data, you can match every person who's employed legally in the US to their employer

You can measure their earnings, and then you have a wealth of data on those employers, from their tax filings and data that the Census Bureau collects about those employers, so we have the ability to see an individual, their employed on a research project, they graduate, they go to work at some firm, we can see how much they earn, what industry that firm is in, whether that firm conducts R&D, the size of that firm, the age of that firm, what the typical level of wages that firm pays, and so forth, and I'll show you some of the results that we've generated using those data

So let me talk to you a little bit about IRIS and UMETRICS, this... and these are slides that I... some of these slides you have and then some of these slides were added today, so not all of them will you have

So IRIS and UMETRICS are a partnership between the Big Ten Universities, the AAU (the Association of American Universities), and the Census Bureau, and other universities have joined. It's a hub at the University of Michigan and nodes at Ohio State and New York University that house the data that we collect and distribute it

securely. The infrastructure was very generously supported by the Sloan and Kauffman foundations, and then research using the data has been supported by the National Institutes of Health, the National Science Foundation, the Small Business Administration, the Department of Agriculture, and the Patent and Trademark Office

IRIS is a consortium, ok, of member of universities. The universities provide data to IRIS that IRIS houses securely, and they also pay annual membership dues that support IRIS operations

In exchange, they receive reports and they participate in the governance of IRIS through electing members of a board of directors that governs IRIS. Participation in IRIS is completely voluntary, so it's critical that all the stakeholders-- the universities, the Census Bureau, every single entity that's participating in IRIS has to benefit from it, which, you know, sometimes, you know, that's part of the art of making this type of initiative flourish but it does mean that we're very responsive to our members

What do the members actually get? They get information, research that can help them to improve their programs, their activities. I'll show you some of that. They receive access to data for their researchers, most of the data provided free of charge, some of the data... the data at the Census Bureau is... the Census has an official policy on charging researchers, that's not free of charge, but our data unlinked to Census are. And they receive reports that are valuable, we hope, to them in documenting the value of what they do

So this is... this is a map of the state of Ohio, every county in the state. Ohio is where I'm based

Ohio State is here in Columbus, Ohio, and this map is showing the amount of research money that Ohio State University spends in every county in the state of Ohio, and that's Carol Whitacre up there. She's the vice president for research at OSU, and she has the job of having to go to state legislators and US congress people and senators, and try to convince them to support research at OSU, and so we generated this map for her to show where the money that goes to OSU gets spent, and we have a version of this map that covers the entire US. What she's saying in the quote that is probably too small for you to read is basically this is exactly what people in congress and in the state legislature want to know. They want to know, "If we give you money, what is going to do for the people in my district?" That's what we're providing her with in a level of detail and accuracy that she has never had before

(Naoki Saito)

Sorry. Among the maps colored, what means the... districts colored in white?

(Dr. Weinberg)

The different counties are colored according the amount of spending that occurred in those districts and so... I mean this is probably more than people wanted to know

about the geography of the state of Ohio, but this is Cleveland, Ohio, this is Cincinnati, Ohio, this is Athens, Ohio, where Ohio University is located, this is Toledo, so the amounts are shaded according to the amount spent, and white means that less than \$15 was spent in that county

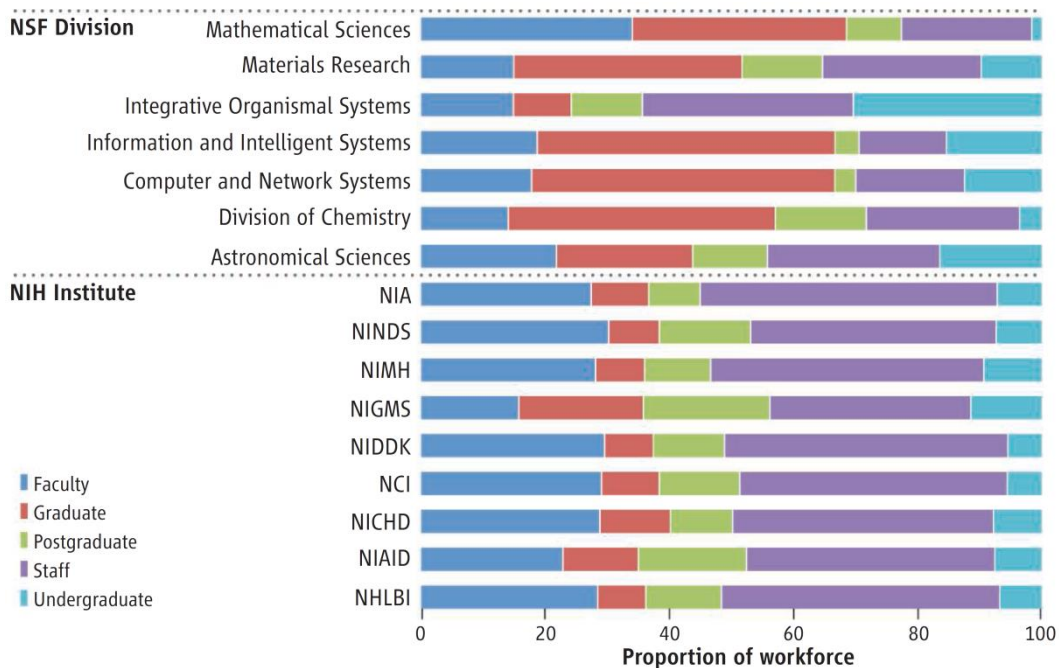
Other questions?

As I said, the Big Ten Universities provided a large portion of the initial support for us. This is a report that talks about spending at those universities. I'll go through it in more detail. This is a report that we generated for Purdue University, one of our earliest and strongest supporters in Indiana, and you can see, here's the state of Indiana with a very similar map

Purdue's colors are black and gold. They made their map black and gold, so you can see this is the type of things that universities are getting, and this is going to their state legislators, to the governor, who used to be Mike Pence, this is the type of thing that we're producing for our member universities

In addition, we're producing data, so the member universities are providing data, it gets processed, it gets sent to the Census Bureau, where there's secure access that's provided to the research community, and then researchers are also able to access the data themselves, so this is also, as I said, producing research on the campuses that are involved with us, and, we had this article in *Science*, and naturally, OSU and our member institutions like that sort of activity

So let me go through and try to show you some different things that have come out substantively from this work. This is a piece that a bunch of us had in *Science* in 2014 in the Policy Forum. This is showing the composition of the workforce employed on research projects.



Differences in workforce composition in projects funded by NSF divisions and NIH institutes. NIA, National Institute on Aging; NINDS, National Institute of Neurological Disorders and Stroke; NIMH, National Institute of Mental Health; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NIAID, National Institute of Allergy and Infectious Diseases; NHLBI, National Heart, Lung, and Blood Institute. (See SM.)

The top are for divisions at the National Science Foundation and the bottom are for institutes and centers at the National Institute of Health. There's the sense that a lot of research in the US gets spent on faculty, you know, sort of welfare queens in white lab coats, if you will. The faculty is actually these blue bars at the left of the graph, and the point that we're trying to make here is that a very small portion of the people touched by research projects are faculty. The vast majority of the people supported by research projects are people who are in some form of the training pipeline. They're undergraduates, they're graduate students in red, especially at the National Science Foundation, they're post-docs in green here, and at the National Institutes of Health, there is a fair number of staff in purple, but relatively few, maybe 20% of the people are faculty

Because understanding the outcomes, the experiences of graduate students are so fundamental, we've focused a lot of our research on looking at graduate students/doctoral degree recipients, and this is a study that's at the moment not yet published, that matches the UMETRICS data to data from the survey of earned doctorates that's collected by the National Science Foundation

And I think they're two reasons to do this, to talk about this to you. One is, I think, we get some interesting things from it, but the other thing is it shows the value of combining different types of data, and that that's even doable, so if you're trying to think

about how to build a data infrastructure, there are, I think... someone will probably think of a fourth or a fifth, but I can think of three pretty easy big categories of data

One category is administrative data, a second category is what I'll call sort of naturally occurring data, like publications and citations and data like that, patents, and the third type of data that sort of are survey data

And I'm trying to give you some examples of the costs and the benefits of these different approaches. The benefits of administrative data are that it's low burden for the people involved. We don't go to anyone and ask them to spend time telling us stuff. Everything we're doing, everything I'm showing you aside from the Survey of Earned Doctorates linked data is coming from administrative sources. The people that we're studying don't do anything. They don't have to do anything. They don't even necessarily know that this is occurring, unless they read the articles that we're writing or see the reports that we're generating. It gives us complete coverage.

You know the person who gets the survey and doesn't want to fill out the survey, this is administrative data. If you filed your tax returns, and most researchers file their tax returns, we have data on you. It's accurate, ok? And it can give you incredible richness, just I'll show you some of the data that we can do, but, you know, we can reconstruct complete networks of researchers, we can look at all the purchases, all the equipment, all the supplies they used, and we don't ask them anything, so that's a really powerful thing.

What are the problems with this approach? Well one problem is it's costly up front to write the algorithms to clean and link these data, and then the second thing is there are some things that it's very hard to get from an administrative data set, so one of the things I think people would like to know is we see people going into industry after they get their doctorates

Is that something they wanted to do? Is that something that they were planning on doing all along? Maybe they went to graduate school because they didn't know what they wanted to do and this was something to do that would, you know, get their parents from bothering them. That would be incredibly hard to learn from an administrative data set

It's not so hard to ask someone on a survey. People like Henry Sauerman and Mike Roach have done it and people are very happy to answer those questions on surveys, so there are some things that you really can't get from administrative data. Naturally occurring data are low burden also, they have great coverage, they're accurate, they also involve a fair amount of linking and cleaning

And then the third approach, which is the one that the Survey of Earned Doctorates is using, is a survey where you could ask people, you know, "Why did you go to graduate school? What did you plan on doing after you got a job as a post-doc? Are you happy with that?" You know, that's stuff that I can't get, as I say, from an administrative source

The problem with it is that it's costly and it's costly every single time you run that survey, and if you want to get high response... response rates are never going to be perfect because, you know, people are busy and you're not gonna know the most busy people may be the least likely to respond to the surveys

And the higher response rate you want to get, the more you've got to spend tracking people down and hassling them to answer your survey. The other problem with survey data is accuracy. You know, first of all, you know, if you ask people questions like, "How much money did you make last year?" You know, in a lot of countries, a lot of cultures, people don't want to answer that question, but they tell the Internal Revenue Service how much they made on their tax returns, so that's a question that it's very hard to get people to answer accurately, that you can get extremely quickly from an administrative data set

And there are other things like the complete networks of researchers that they collaborated with, I mean you could never remember everyone you've worked with, even if you wanted to remember everyone, you'll never get that out of people accurately. I can't remember everyone who I've worked with, but if you look at my publications, I can write an algorithm to tell you exactly who I've worked with, so each of these approaches has strengths and weaknesses.

So you want to use administrative data and naturally occurring data wherever you can, and then you want to get the stuff that you can't get using those approaches from a survey, and that's part of what we did with this research on the Survey of Earned Doctorates that I'll show you some slides from

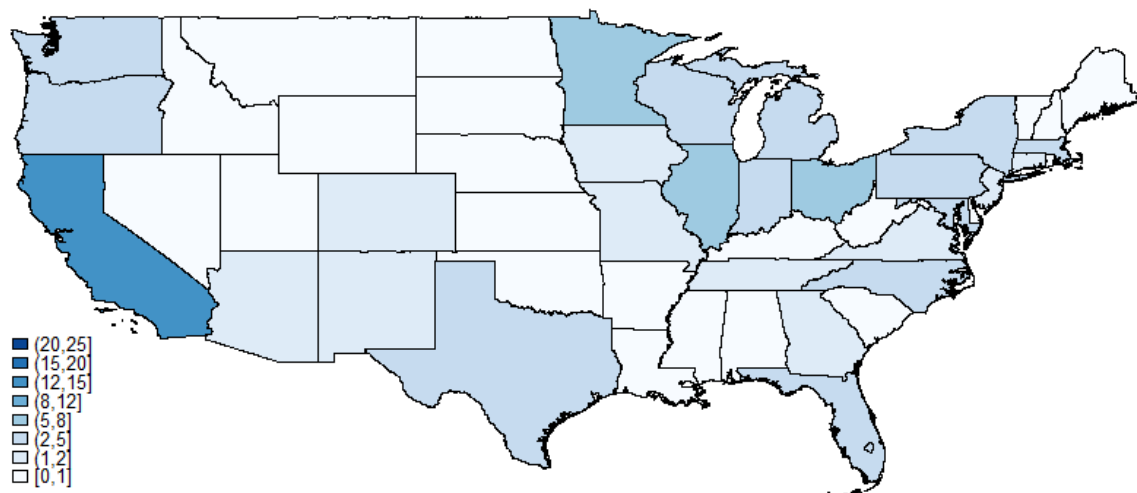
So we tried to see what share of people were funded by research field here, and what you can see is that for most of the fields, the share that are receiving federal funding is in the 60, 70, 80 plus percent range, it's lower in Psychology, second from the right, and then substantially lower in the Social Sciences and the lowest in the Humanities. That's probably not surprising, right?

This is something that I think is very powerful. It's a little complicated, so I'm not going to go through every single number in this table, but what we did was we said, "Let's look at all of the agencies that supported 3,329 graduate students at two of our member universities." And I've highlighted the Defense Department here, 7% of these graduate students were supported by the Defense Department alone.

Almost the same share were supported by the Defense Department and the National Science Foundation. Two percent were supported by Defense and the National Institutes of Health, and then you can see other combinations here – almost 5% were supported by 3 or more agencies including the Defense Department, so what's powerful about this is that if the Defense Department had an accurate list of everyone that it paid, which I don't think it does, they would have no idea about what's happening to their own researchers at NIH, NSF, Energy, and all the other federal agencies that touch their graduate students

So this is something where the administrative data are incredibly powerful, at a level that you simply can't get, and by the way, if you tried to ask these questions in a survey, these graduate students couldn't answer this question. I employ a bunch of graduate students and post docs and undergrads, they have no idea what agency supports them, right? I move them from project to project depending on what work they're do... "Oh, they're working on this project. Pay them on that account this semester," but they don't know that, unless the agency has some specific requirements. For instance, NIH has specific requirements for training and human subjects, but basically they don't know this. This is the only reliable source of this data

This is the piece that we in *Science a little bit over a year ago, Nick Zolas took the lead on, this is all UMETRICS data



The pins in this map show the member universities that we work with, and what you can see here is the share of the doctoral degree recipients going to each state in the US.

More of the doctoral degree recipients go to California than any other single state. But if you look at the states and the universities, for six of the eight universities in our study, more people stayed in the state that they graduated in than went to any other single state, and again, I don't want to like give you a course in US geography, but this is Iowa, this is Indiana

These may not be the states that you think about really attracting lots of tech workers, right? But more people are staying in these states than going to any other single state for six of the eight universities, even with this set of states, and if you're a research vice president, that's incredibly important. You know, our elected officials understand why they support Ohio-born children going to Ohio State and hoping that they stay in the state of Ohio when the graduate, but I don't know that they really appreciate why the state of Ohio supports doctoral candidates from outside of Ohio or other countries to go to OSU. This is really powerful for them to make the case for why the state is

supporting this kind of research

This, and I'm not going to go through all the tables just in the interest of time, but I will go through some of the results

Table 1. Postgraduation employment of UMETRICS doctoral recipients who were paid by research grants and left the university between 2010 and 2012. The national workforce distribution is calculated from all employment in all establishments covered by the Census's LBD between 2010 and 2012.

Locale and small	Doctoral recipients placed in sector (%)				
	Industry		Academia	Government	All
	R&D firms	Non-R&D firms			
Placed within sector	17.0	21.7	57.1	4.1	100.0
National sample (<i>M</i>)	10.8	75.0	10.7	3.5	100.0
Of those in sector, percent placed:					
Within 50 miles	10.1	23.5	8.9	18.2	12.7
Within state	16.6	36.0	18.0	25.8	22.0

The top of this table is showing where people are getting jobs. From our discussions today, I know that it's important to see how people are flowing from academia into industry and back. We've only studied one portion of that picture so far. We could do the other one but haven't done it yet. This is the flow of doctoral degree recipients to sectors: Industry, academia and government. As you can see, very few doctoral degree recipients get jobs in government. That's probably not surprising. This is one year after they graduate and leave the university that they earned their degree at

About 57% go into academia, many of those it seems are going into post docs, based on other data that we have, but about 40% of these students are going into industry, and of the people who are going into industry, a very large share of them are going into R&D performing firms, so they're going to work at the types of businesses that it looks like would benefit from having them work there, (actually, this would be a great survey question). We don't know what they do on their jobs, but it looks like they're working at the types of businesses that would use their skills.

If you look down here on the second line, this is the workforce in the US. Very few workers as a whole in the US work at R&D performing firms, so these workers are unique in terms of the share that are going to R&D performing firms, and you can break this by high-tech firms and in other ways. And other cuts through this data that will show a similar picture

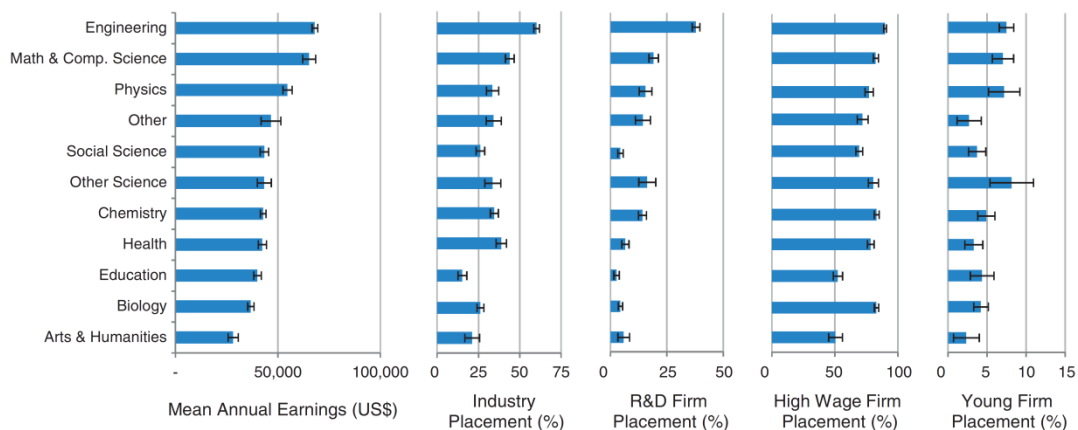


Fig. 3. The annual earnings and placement of doctoral recipients supported by grants vary by field. Young firm are defined to be those <5 years old. High-payroll per worker establishments are defined as those with a payroll per worker above the median for the establishments within their six-digit industry. Mean annual earnings are stated as U.S.\$1 ×1000. Means and standard errors for each variable.

ProQuest, has a database of dissertations, international, and we matched every doctoral recipient in our data to ProQuest dissertation data base, and from that, we know exactly what that student wrote their dissertation on and we actually have the abstracts and could, in principle, get the dissertations if we wanted to do to do some kind of natural language processing, which I think we'll be doing over time, but right now all we're doing is we're looking at what subject that was in

and if you do that, and we're looking here in the left panel at annual earnings the year after people graduated and left their university, you see that people who studied Engineering and Math & Computer Science have the highest earnings/ Probably not surprising but now we have that, and it goes down, Arts & Humanities are the lowest, again probably not surprising. Biology is the second lowest, Chemistry is somewhat low also, these appear from our data to be people who are going into post docs. We know that a lot of Biology students, a lot of Chemistry students are taking post docs, and so that's presumably depressing their earnings. The Biology students frequently stay in their post docs for a long time. The Chemistry typically moves on more quickly, but it's affecting their wages at this point - right after graduation.

We can look at the share of people who are getting employed in industry, that is to say outside of academia. Health, Chemistry, they tend to be very high in industry employment relative to where they are in terms of earnings. Next is whether they get a job at a firm that conducts R&D, then the share that are going to high-wage firms, and to get at some sense of entrepreneurship, lastly, we have the share of people that are going to young firms, firms that are under five years old

This is a deeper look into the funding, where people are... where the spending on research projects is going, so we have another map for these universities showing where the money that they spend goes across the US.

Not surprisingly, a lot of it is going in the Midwest, given that's where the universities are located, but a fair amount of the spending is going to East Coast and to the West Coast, and even Florida, Texas, so forth, actually have a fair amount of spending going to those states

We did a similar study for data from a French university, and from two Spanish universities, there's nothing special about the US. This type of data is pretty common. Basically everyone needs to know where they're spending money, who the checks are going to, what they're getting, what have you

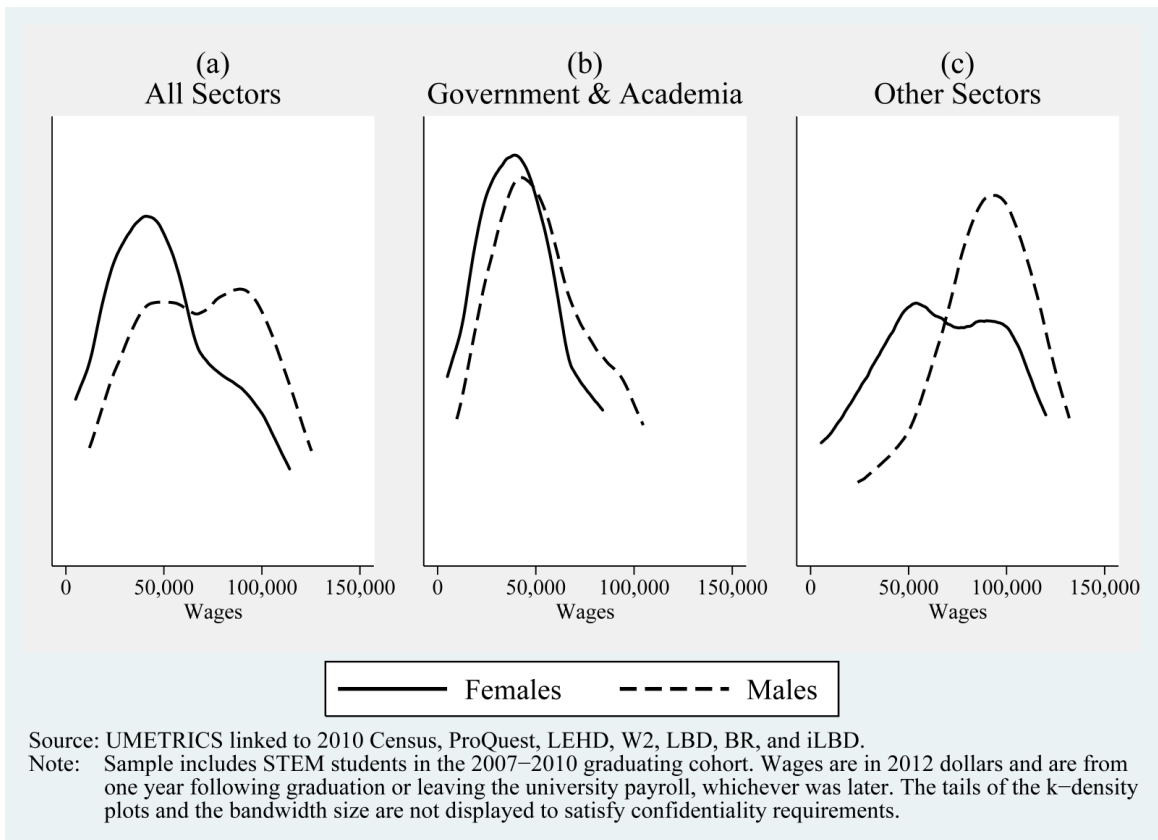
Wouldn't it be cool to have something like this for Japan? It's the type of thing that's very doable using most universities' data. When we approach a university, they say, "Oh, we could never do this."

Six months later, we get the data, it's not that hard for most universities, and they have to do something to get it out of their systems, but it's usually not impossible

The data show us every location, every establishment that a company has, and so we can identify when companies open new establishments, new locations, and that is what we did here, I'm not going to go through all of the econometrics? But what we did here asks if a company does business with a university, whether it is more likely to open an establishment close to that university. And that's exactly what we found, and the more business they do with that university, the more likely they are to open their new establishment close to the university. The strength of that relationship between spending, the amount of business and where they open their new establishment is stronger for R&D performing firms, so these are the types of firms that, people are really interested in attracting. And this is one of the arguments that universities make, "If you invest in us, we will generate businesses, R&D-intensive businesses, tech-oriented businesses close to us. We'll be a hot spot. We'll be the next Silicon Valley." I don't know about that but you can see that happening to some extent here

The second part of this table looks at only the firms that opened a new establishment somewhere in the US and we show that among the ones that are opening new establishments, they're also more likely to site that new location closer to the universities that they do business with and that that's a stronger relationship the more business they do with that university

We have also done some work on gender wage gaps



This figure shows earnings for men and women the year after they graduated and left the university. The men are the dashed series and the women are the solid line, and you can see there's a pretty substantial difference between the earnings of women and men

The gap is still present for people, who get jobs in academia, but it's substantially smaller than it is for people who get jobs in industry

In industry, men earn considerably more than the women do

A lot of the women are, and frankly a lot of the men too, are earning in the high \$40,000 range, that's what post docs, most post docs in the US get paid, so... and we have other evidence that suggests that a lot of the people who are going into academia, and especially the women going into academia are in fields that where post docs are prevalent

Two factors explain this gender gap, and that's the table

So we find a one-third gender earnings gap, basically women earn about 33-34% less than comparable men do. Two things explain that. One, field of study - women study Biology and Health, and men disproportionately study Engineering and Math and Computer Science. That explains about 2/3 of the 1/3. So we've taken the gap from 33% to about 11%. The remaining 11% of the gap is explained by marital status and presence of children. Married women and women with children earn substantially less than unmarried women or women without children. Those differences don't show up

for men

And so if you control for those two factors-- Field of Study and Family Structure, actually women earn slightly more than men do, so that I think is a pretty powerful finding, and it's powerful in part because, of the amount of data that goes into doing this

We needed to know what these people studied, we needed to know their earnings, we needed to know their family structure, and we never talked to anyone to do that study. All of that came from administrative data linked to the Census

I'm not going to go through this table in detail either, but another thing that we did in this study was to try to say, "Well, ok, women earn less. Maybe they earn less because of something different in terms of their training experiences. We have these data. They're incredibly rich. Let's look at the environments in which the women trained and the men trained."

Table 1. Training Environments of Male and Female Graduate Students Participating in STEM Research

Dependent Variables ↓	(1)		(2)	(3)	(4)	(6)	(7)	(8)	
	(a) Females	(b) Males	(c) Diff						
Share of Faculty that are Female	0.2 {0.02}	0.1 {0.01}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1*** {0.02}	0.1* {0.03}
Share of Graduate Students that are Female	0.1 {0.01}	0.1 {0.00}	0.0*** {0.01}	0.0*** {0.01}	0.0*** {0.01}	0.0 {0.01}	0.0 {0.01}	0.0 {0.01}	-0.0 {0.02}
Ln Team Size	1.7 {0.04}	1.9 {0.03}	-0.2*** {0.05}	-0.2*** {0.05}	-0.2*** {0.05}	-0.1** {0.06}	-0.1* {0.06}	-0.1* {0.06}	-0.1 {0.09}
Faculty to Student Ratio	0.9 {0.06}	0.6 {0.03}	0.3*** {0.07}	0.2*** {0.07}	0.2*** {0.07}	0.2*** {0.08}	0.1** {0.07}	0.1* {0.07}	0.3*** {0.13}
Total Number of Awards	2.2 {0.07}	2.7 {0.06}	-0.5*** {0.09}	-0.3*** {0.09}	-0.3*** {0.09}	-0.2** {0.09}	-0.2*** {0.09}	-0.2*** {0.09}	-0.1 {0.15}
Number of Months Participating on the Award	21.0 {0.69}	21.6 {0.45}	-0.6 {0.82}	-1.1 {0.79}	-1.0 {0.79}	-1.0 {0.82}	-1.4* {0.82}	-1.4* {0.82}	-0.9 {1.18}
Years from First Observation to Degree	3.2 {0.08}	3.2 {0.06}	-0.0 {0.10}	-0.1** {0.06}	-0.1* {0.06}	-0.1 {0.06}	-0.1** {0.06}	-0.1** {0.06}	0.0 {0.10}
University, First Year Trend, Left-Censored				✓	✓	✓	✓	✓	✓
Race, Hispanic Origin, Age, Age-squared					✓		✓	✓	✓
Dissertation Topic						✓	✓	✓	✓
Funding Agency							✓	✓	✓
Married or Partnered, Children								✓	✓
Female x (Married or Partnered + Children)									✓
Observations	370	867	1,237	1,237	1,237	1,237	1,237	1,237	1,237

37

And what you'll see is that if get all the way to the right, where we have all the controls in this, there are not that many differences in terms of the environments in which women and men trained

Women tend to work for women faculty. That's probably not surprising, but if you're a woman graduate student, you're more likely to be employed by a woman faculty member than a man is, about 10% more likely. Female graduate students tend to be employed on smaller projects than male graduate students are, but as a consequence, there are more faculty per graduate student, so they get more attention, and then the

other things that we looked at: Number of months they were on a grant, total number of grants, number of months covered by project, there are not really differences. So we have again an incredible ability to look at these training environments and relate them to earnings again using administrative data. It's the type of thing you could never get, you could never do on a survey

"How many people were employed on your project when you were a graduate student?" Ok, maybe if I was at a small lab, I could count them up, I mean maybe, but say I was employed by some center grant. I have no idea how many people I was employed with on that center grant. We know it's the kind of the thing that could be helpful to a university in thinking about what things they can do and can do better

This is another unpublished study, we've linked these data to data on country of origin

These data are also housed at the Census Bureau

What you have here are different countries or regions of the world, and then the share of graduate students, doctoral degree recipients in different fields, and I think what strikes you if you look at this is the amount of grey on the figure, and those are people who were studying Engineering

Foreign doctoral degree recipients in the US study Engineering in far greater numbers than native-born Americans do. Native-born Americans study non-Stem fields. They study the biological sciences. They don't go into engineering

But we don't know what happens to you if you leave to the US. If I have a Japanese graduate student and that student leaves the U.S., I'll bet you that student comes back to Japan, or maybe they go do a post doc but come back to Japan. The students from Eastern Europe, the developing world, I don't know that they go back to their home countries. Canada, England, are probably a much more likely destination, but we can't answer that question. This is the beauty and the frustration of this kind of country-based administrative data

You could probably tell us if they come to Japan. We can't do it. You actually have some ability to tell us from JGRAD if they come to the US. We could tell you a lot about them if they are in the US

We did some work on an entrepreneurship. We looked at the share of people in different categories who are working for start-ups. A start-up is five or fewer years old, and maybe, we don't have the right universities – these numbers might be higher if we had Stanford and MIT

But what you can see here is that there are not a lot of people who are going to work at young establishments... young firms. That's probably not completely surprising. They are young firms, which are small so they don't employ as many people as you might think would, but if you look at them, the under grads are the most likely to go, graduate

students and post docs, intermediate, staff are in green and then faculty are the least likely to go, and you can see that there's a decline over time in the share going to work at start-ups. That's not unique to these data - the share of the US workforce employed at start-ups has been declining over time, that's a whole other literature that we can talk about, but it's not unique to this sample

The last thing that I'll talk to you about just briefly is our work on Innovation in an Aging Society.

As I said at the outset, this is supported by the National Institute on Aging, the Office of Behavioral and Social Science Research, and the National Science Foundation. The US scientific workforce is aging and the question is, what are the implications of that? So we have three broad thrusts here. One thrust, and this is work that I'm doing with David Blau at OSU, and John Ham, who is at National University of Singapore, is looking at productivity late in the career, and I don't know if this is true in Japan. In the US, many universities have defined benefit pension plans where if you work there for say 30 years or 35 years, you get a big bonus relative to leaving a year or two earlier. We're looking to see if people who reached that bump later in their careers, because they started graduate school late, because they moved universities a few times, or for whatever reason - because the university happens to have a system where that bump is further out in the career - we're looking to see if the people who kind of get that bump later in their lives are more productive while in the years that they're trying to kind of get to that bump

Another project that's being run by Pierre Azoulay at MIT, Ina Ganguli at UMass Amherst, and Josh Graff Zivin at UC San Diego, is looking at moves on the productivity of researchers, and they're using the fact that, at least in the US, nobody moves when they have a kid in high school. If they move, the kid gets to a new school, doesn't integrate very well, they're miserable, people won't move geographically while they have a kid in high school. They're using that to look at the productivity of people.

Lastly, we're also generating measures of the innovativeness, the biomedical health value of research, and as I've said, we're producing a large amount of data that we're making publically available to researchers

Some of the data are protected, they're proprietary, but many of the data elements are openly available and we're allowing researchers to have the data. You have researchers on your campuses, we're happy to let you have the data that aren't protected.

So let me thank you for letting me talk to you for a while so late in the afternoon, I'm an economist, nobody likes hearing hours of economics late in the afternoon!

I think in many countries, certainly in the US, it's absolutely critical that we understand the value of research. It's important if we want to be able to continue getting funding for research, to communicate this to the public, to policy makers, and it's important for us as researchers to understand what we can do to do what we do what better and make

it more impactful

It's not listed here, but I think it's also valuable for potential trainees to see what's going happen to them when they get done. Everyone goes to graduate school, "I'm gonna win a Nobel Prize." "I'm gonna start the next Google." Most people don't. They may still do really good stuff. They should probably have some sense of what's going to happen to them backed by research when they start

As I said at the beginning and I've said a couple of times through the talk, science and innovation are international. We benefit from innovations in Japan just as I'd like to think you benefit from innovations from the US. We get students and faculty and post docs from Japan, you get them from the US, we need to know if we want to fully account for the real value of the research that we're conducting, how people and ideas flow internationally

The first step is, I think, to have individual countries do something, whether it's UMETRICS or JGRAD or JD-Pro or whatever it is, to do something for their own countries, but ultimately, I think the goal has to be to try to integrate these data because we're not going to be able to quantify the real value of this research and to communicate the importance of international collaborations and the importance of international flows unless we can integrate ultimately these data across countries

That's going to take a lot of work. We've already done a lot of work in the US. I know you all are doing a lot of work here. It's going to take even more work to make these data platforms interact with each other, but it's incredibly important. Thank you

(MC: Matsuzawa)

Thank you very much indeed, Dr. Weinberg, for your wonderful presentation. I think it's a very great opportunity probably for the audience to ask questions so I'd like to take questions from the floor. Any questions

(Misu)

Thank you so much for your presentations. I'm Max Misu from Hiroshima University. I have a comment and a question for you. As you said, it's definitely cool to understand the situation in Japan, but the issue is that, I believe in the US you have access to all those administrative data and national census... national survey and census data, and so you can match the data, and at this moment, you know we have quite the restriction to access to all those data so I don't think we can do the same at this moment, so my question is that in the US, you said that you have a partnership, like a Census-University partnership, what kind of regulations do you have to access all those data and also where there any change in regulation in the past so that you can access all these data and match the data? Thank you

(Dr. Weinberg)

Sure. Thank you. It's a great question and it's an important topic.

And while I'm answering the questions, maybe I'll put up the names of my collaborators - these are the people who are actually doing the work while I'm here!

The Census Bureau has a mandate to collect data and to make it maximally available to understand the US population and economy

And the Census Bureau has a research data center network of secure facilities across the US where qualified researchers can get approved to do research projects using census data. We were very fortunate to be able to connect into that infrastructure. We didn't build it. We took advantage, tapped into what already existed

Having said that, I wouldn't underestimate the difficulty of getting these data and getting them in one place.

I'm not the person who does this - Jason Owen-Smith is the person who does this. We have agreements at this point from 50 universities to provide data, not all of them, most of them still pulling the data, they haven't given it to us yet, but they will

Every campus that we approach says, "Oh, that's great that someone else has done this! We could never do this. In our state..." I don't care, whatever state it is, "We have this law that says that this can't be done." I think what it takes is to sit down with the lawyers and explain what you're doing and explain the security and university by university, people say, "Well, ok. We can see that this is doable. There are exemptions to laws for certain activities, we see that this is possible." Every university that we approach, "We use People Soft. We could never get this data out of People Soft." Well, other universities have figured out how to get the data out of People Soft, so while I don't want to minimize the concerns and the difficulties of it, but the point is, when people start seeing the value of this, they figure out ways of making it happen, and each country is going to have strengths and weaknesses. There are things that you're not going to be able to do in Japan that we can do. They're things that I understand that you currently do in Japan that we can't do.

I understand that you have the ability to survey researchers who leave Japan and go to the US and other countries. Our data don't really allow us to do that.

We had a university, I won't mention names that said, "We won't give you on data that's on Defense Department contracts because it's potentially classified. We don't want to be giving you data on classified research." I had a meeting at the Defense Department and they said, "Which university was that? They were incredibly upset that we don't have that data in our system. They said, "We'll talk to the university." So I mean everyone is going to have their issues. Universities come in and they say, "Well, we can't give you this portion of the data." That's fine. Give us the other portion. We're a partnership. No one is telling anyone what to do. You give us what you have. We'll do with it what we can. If you can figure out how to take next steps, we'll give you more valuable things, and we've found that over time, people have really figured out

ways of doing things that they thought they couldn't... wouldn't be able to do, because they started seeing the value of what we were giving them

I mean no one is going participate if they don't see some value from that process, right?

(MC: Matsuzawa)
Any other question?

(Hayashi)
Thank you so much for inspiring presentation. I have two quick questions. One is, you earlier introduced us international adaptation of UMETRICS like France and Spain, and I believe that Australia also conducted part of study a few years ago, I guess, and then did you find any international linkage of the United... US-version and international-versions? And the second one is related to the heterogeneity of that household decision-making. Suppose that I'm about to graduate school and going to find a job, but my wife has got a job and she loves, and it's very difficult for me to have a, you know, flexible to find a job, you know, so that kind of restriction may related to job market and also gender gap in the wage as well. Thank you

(Dr. Weinberg)
Those are both...

(MC: Matsuzawa)
So before asking question, please specify your name

(Hayashi)
My name is Shinano Hayashi, from that JST Japan Science and Technology Agency

(Dr. Weinberg)
Great. Thank you. That's a great... those are great questions too. Australia is making some investments in this direction. I think it really depends on someone or some group seeing the value of doing this, and having the ability to navigate all the barriers and generate a consortium, a coalition of institutions that are willing to do this. The international component is going be more tricky. We've not been able to do that. We haven't sought to do it. We'd love to do it. Maybe Japan and the US, we could be the first countries to do it. That would be fantastic

But I do think, you know, you can see for legal reasons why that's going to be a more difficult thing, and so I think that the first step has to be just getting important players to have these data internally, and then the next step is to share internationally... and

there's no question, I mean there's no question that's critical.

Let me go back here

Think about it this way. If Michigan didn't have data on Ohio, and Ohio didn't have data on Michigan, and nobody had data on California, we'd have an incredibly incomplete picture of what was going on here, so again it's going to be more difficult to do this internationally than it is going to be to do it domestically, but ultimately, our member universities want to be able to say, "Look. Yes. We had graduate students come here or post docs here from Japan, and, yes, some of them went back, but look at all the graduate students and post docs that stayed, and look at all the graduate students and post docs from Japan who came to the US," just as we want to be able to say, "Ok, we lost people to Michigan, but Michigan lost people to us." You know, and that's, I mean, ultimately, you know...

I believe... not every country is going to be able to play in that environment, but once people start seeing the value and the potential of it, I think you're going to see more countries doing that. It's going to take some demonstration. I mean this Science piece in some sense was a demonstration of the value of this type of project. It's going to take something like that, but I think ultimately it's something that will be done.

In terms of your second question about households, that's a great question. That's actually something we'd love to do a paper on. We have data on both partners, regardless of the structure of the family, we have data on all the people in the household, and whether they're married, whether they're unmarried, whether they're same sex, we have all of the data, and we know what their partners are doing, and so I think it would be really fascinating to be able to say, for instance, "I'm married to a doctor or a lawyer, whose practice is very local," versus, "I'm married to an airline pilot or a consultant, who's equally skilled but whose practice is national or international. How does that affect my mobility? How does that affect my outcomes? How does that affect the research I do?" I think would be a fascinating, fascinating set of questions, and that's something we hope to work on. We've not started that project. Maybe we need a graduate student or a post doc to do that paper, but I think it would be a fabulous thing to do

(MC: Matsuzawa)

Yes, please

(Naoki Saito)

Saito from NISTEP, and thank you very much for your very inspiring presentations, and I believe that today's audience is very much impressed by the very advanced type of practice now going on in the US, but I have some specific questions. One is I think it's very great that you can combine those personal data with their inputs and outputs like publications or research fundings, and also with the taxation data. In Japan, it's almost unimaginable, and so that may mean that you should limit the target people to

only those persons who have the Green Card or US citizenship who can be measured through the tax record; But also in Japan, we have a very good personal number that is called 'My Number' that is very recently introduced but there is a very growing concern that the general public thinks that if those kind of personal data or numbers being used by any kind of researchers for administrative purpose, without their interventions, then that's kind of an invasion of privacy and such kind of things they're always thinking, so is there any kind of concern in the US that this kind of analysis can be made without any kind of visible permission or intervention by the target people themselves? That's one question.

And another thing is that I totally agree that science and innovation is actually international, so... if you try to pursue the full accounting of this kind of research, then you need to have the international data. For example, your US universities are accepting many overseas students and they get a doctorate degree and after graduation, they go back to their home country or go to another country to get the position of academic society, or starting a new business, then they create some kind of new values or jobs, but on the other hand, they are paying very expensive tuition fee to the US universities, so at least if you can count or measure only the US domestic people or US citizens, do you think that you can rationalize the public investments to the US universities for future impact of original investments or outcomes, or you have to count those investment to the foreign students in order to rationalize or make accountable of the public investment to the US universities, so that rough guess will be ok, but do you really need the international data or not? That's my primary question.

(Dr. Weinberg)

Right. So the US personal identifier is the Social Security Number. We have done all of our linkages based on names only. You can get 70-80-90% of the people using only names, but that's not a representative sample

Korean names and Chinese names, at least in English, are very ambiguous, that's not so much the case with Japanese names or names from other countries, especially countries that that are natively in the English alphabet. We get over 99% of researchers when we have names and partial dates of birth, and we do not use Social Security Numbers to make these linkages.

We don't need them, and frankly, I sleep better at night knowing that we don't have any Social Security Numbers. We have done some experiments where we vary the amount of information that we use to make matches. One of the things we've talked about is we're going to member universities and we're asking them to give us transcript data for all students including their undergraduates and so forth, and one of the things that I've suggested was could we use three randomly-chosen numbers from your SSN. If it happens to have a 6 in it, that tells me something. It's much, much less sensitive than knowing that the third digit is a 6. I just know that there's a 6 in it or there are two 6s in it, or whatever. We haven't gotten that far, so that's more speculation, but currently, we don't have any SSN data and I think we're going to do our best to use as little of that data as possible for the exact same reasons that you have identified.

One thing that I would say is that the Census Bureau basically has a parallel SSN system. It's called the Protected Identification Keys or PIKs, and when you transfer data to the Census Bureau, whether it has a SSN or doesn't have a SSN... If it has a SSN, the SSN gets stripped off, and either way, you get assigned a PIK, which basically is sort of like a parallel universe SSN that has no value outside of the Census Bureau, and nobody knows what anyone's PIK is... I mean I don't even know what my PIK is so it has no value once you leave the Census environment

I think the other thing that's worth noting is that the SSNs aren't always accurate, so, you know, people say, "Oh, well if you have the SSN..." and we say, "Well, we'd rather not have SSNs," but even if we did, it doesn't solve all the problems, and other thing I should restate is that when I say that we have the date of birth. We have the year and the month of birth, we don't even have the day of birth, and with a name that is enough to get us over 99% match rates. Some universities were nervous about giving us full birth dates, and so only took the year and the month of birth from the universities that were willing to provide it. We get over 99%, and there are measurement errors in those too, so those aren't perfect

Ok, that was the easy question. Now is the difficult the question... which I think of is basically being the R.O.I. question, the Return on Investment question, and I'll be honest with you, I don't think we have the answer to that question and I don't think that the problem is entirely one of data. I don't know that we have a fully specified conceptual framework to help us answer that question. I think we have bits and pieces that we know are important, but to turn it into a Return on Investment-type of number - is it positive or negative... is it better than I can do by buying Treasury bonds or worse than I can do buying Treasury bonds or tech stocks or whatever. I think we need a richer conceptual framework than what we have.

I will note, and I think this this varies substantially by country and my understanding is that other countries are different from Japan at least in degree, that many of the foreign graduate students who come to the US are being paid by US institutions and are not paying. People who are coming... being sent by employers in their home countries, frequently are paying in to the US system, and that's defraying costs or even subsidizing others, many graduate students are being supported on research projects or other projects, and that's a cost. But they're also doing research and they're also teaching classes, and so you need to make that calculation of what the costs and the benefits are. We're paying people but they're also working in exchange for that, and conducting research and so forth

And just go one step further, one of the things that we could do is to take authors on articles or inventors on patents that are outside of the US and see a graduate student, who was maybe a first author on a paper as a graduate student that they were paid on a project in the US, now they're in Tokyo, what's... what's happening to those papers? Are they still authoring papers with the people that they were involved in the US. If they file for a patent, does that patent have US inventors on it as well? I think that those are things that we could do. We haven't done that yet, but this is an incredibly important

part of the picture and it is going, I'd like... I'm glad that what I said was inspiring. I think that's important to do but I think we also need to be realistic. That this one isn't happening tomorrow

(Takahiro Saito)

Thank you very much for your informative and exciting presentation, so I am Saito, so it's the same name with him from the Osaka University, so I was excited because my doctoral thesis is Cost Effectiveness and Analysis of the Education and Training, so I did a questionnaire survey of university graduates and the graduates of the training centers in Malaysia and Thailand, I spent many months to do the survey, so but now, so you have the data of the tax payment, so wow! I was so very... I understand that is an exact contents of this kind of research, so I... what do I say for the regulation and the this kind of phenomenon, the professor Miso said that there is a very high barrier/wall between the ministries, yeah, so but I can say that you or we can... what we can do is the knocking the Minister of Labor and Social Welfare, so actually so they're very... you know it's very close, so anyways, I have two questions

So you talk about the value of the research and doctoral student, and you talk... mainly talking about monetary value, but I think that there is other value, so if you think about only the monetary value, so it will be the underestimate of the outcome of the research and educational activities, so if you have any a proving (?) or a kind of a survey of this kind of the... using this kind of big data, so if you have some kind of concept, please let me know

So another one is that you... you talk about you're mainly focusing on the comparison of the result among the different groups, but I... if possible, I would like to know, in the totally speaking, so education in the US... I mean the education in the universities in the US is good or bad based on the result of the cost effective analysis? Thank you

(Dr. Weinberg)

So I couldn't agree more that... that the financial values are really only one part of the picture, and probably not even the most important or largest portion. We have a project now. It is on the aging... part of the Aging project that we're doing. It's the last of the three listed here, Jay Bhattacharya, who's an MD PhD at Stanford, MD PhD in Economics at Stanford, and Mikko Packalen, who is an economist at Waterloo, are trying to measure the health value of research, and they have, I think, a pretty exciting way of doing it that we're using as well in my research group

So they're trying to identify new ideas by looking at the introduction of new words in research papers, so which was the first paper to use "DNA" or "PCR" or what have you, and so that gives them a way of linking a research paper to patents, to drug approvals, and ultimately linking it to health benefits, and while you didn't quite say this (you were more polite), but it is a little bit bizarre to try to quantify the value of research that's supported by the National Institutes of Health by saying, "How much money do the people who do this research make?" We're supporting that research so they will

cure diseases and improve health treatments. We want to be able to figure out how their activities are leading to biomedical breakthroughs that will improve health, so that's the approach that they've been doing. I think that's a very exciting approach. I think that's also a harder thing to do, not that this is easy, but that's a harder thing to do

Because it's hard to trace those ideas, I think the idea of saying, "Well, can we follow the text and see how that text moves around different documents?" Is one very powerful way of doing it. I think there's another way of doing it that would rely more on the UMETRICS data and when we measure the movements to employers and the earnings and employers, in some sense, what I care about is of people as being the idea vectors. People are the ways that ideas move throughout the economy and throughout society, and so the question is, if someone goes from this lab to that company, and they wrote this dissertation and they were an author on these papers and they were supported on these research projects, when they get to that company, can you see those ideas showing up in the IP, the intellectual property that they're working on, that they're patenting, the drugs that that Pharma company or the devices that that medical device company is producing...

As I say, we're a step or two from being able to do that. I think we're missing some of the conceptual framework in order to know how to do that exactly, but that's ultimately what we have to do. That's how we're really going to measure the value of research. The economics is a partially a proxy. As an economist, I'm going to speculate that if they ever cure cancer, whoever cures cancer is going to make a lot of money doing it, and we can get some sense of the value of curing cancer from how much money the person who does it makes

But I agree, what we really want to be able to do is say, you know, "How do we cure cancer?" I mean god willing, we will. How did that happen, and how did those ideas, move around?

Again, the harder of the two questions, and another great question is, are we doing well or are we not doing well? I think that raises something that is interesting, you know, as you think about JGRAD and participating in and building it. We've had some universities say, "We want to know how we're doing relative to peers."

I think we've worked with universities to say, "Maybe the right model is that if you want to know how you're doing, you can specify half a dozen universities and we'll tell you how you look relative to those universities. We won't tell you how any other university does, but we'll average a bunch of them and you'll get to see how you compare to a group of institutions," but again, we're... we're a partnership. We're a consortium. We don't want to do anything that our members don't want us to be doing, so I think the deep question you asked is are we doing a good job? How can we do that job better?

I think the evidence suggests that in many ways we are doing a good job. It does look like these people are getting jobs at the types of establishments... either in academia or in industry at the types of establishments where we think that their skills will be used. I think we can learn more about how we're doing that

and I think the other piece of it is understanding how we can do what we do better? You know we have a paper that we're working on, where we look at women who are in programs that have relatively few women, and saying, "There's another woman in you're their starting graduate class. If I'm a woman and I'm in a graduate class with no other women compared to one or two other women, do I do better? Am I more likely to complete my degree?" What have you. "If I'm foreign born, how does that work if there are more or fewer people who are foreign-born in the class?" So I think the other piece of it, since we don't know how to answer the exact question that you asked, is to say how can we learn what we can do better? Those are questions that I think that we're closer to being able to answer here

(MC: Matsuzawa)

Well, I'm afraid we are running out of time, but I think we can take one more question, if there are any questions. Any other questions? Yes. Yes, please

(Larsen)

Thank you very much for your very interesting presentation, and my question is related to the one of the final parts about high-impact and transformative science, it's a very interesting topic, and one of the things I read from the outline is that it is... it includes a great variation, so my question is if you have found some clusters or if you can see some typology of different types of transformative-ness? That would be really interesting to hear, and also, the time dimension here. What... what is the citation window that you're working with here? Thank you. And my name is Katarina Larsen and I'm a visiting researcher at NISTEP

(Dr. Weinberg)

Great. Thank you. So this is the part of the slide deck that I talked less about. As I said at the outset, I took some slides out that you have.

Very briefly, what we did was we tried to develop a set of metrics to understand, to measure the transformative-ness of research. We identified seven features of transformative research. It's radical, it's multi-disciplinary, it has a very broad impact, these are... should be in the slide deck that you have, and then we measured them using both citation analysis and natural language processing

So we looked to see which papers use important new concepts, important new terms, which papers were the first papers to use, for instance, DNA or polymerase chain reaction, or RNA interference, and then we also constructed those into a metric for transformative that combined all of those, and then we have a standard measure of Impact

And this is a plot of how the two are related. The correlation between them is about 0.4, so work that's more transformative is more impactful on average, and work that's more impactful is more likely to be transformative on average, but the correlation is far from

one. There's work that's very impactful that's sort of an incremental advance, maybe a very substantial incremental advance, maybe very widely used, but it's not revolutionary. It doesn't completely change the way people think. It's not game changing

The classic example of transformative research that we've thought about is the quantum revolution in physics at the beginning of the early 20th Century, so... so there's impact and there's transformative-ness, and the two are related, but far from one, far from perfectly correlated

Here, what we've done, as I say, is plotted them. You can see things, three things we've highlighted. We've highlighted in red, work on stem cells

That is if you just read the literature. That seems to have been a transformative concept, a transformative area of research. The Human Genome Project, and nuclear reprogramming and DNA methylation – all things that have been taken by experts to be seen to be transformative

So your question was about the time frames over which these happen. One of the things that we thought was very interesting was that there's this idea that transformative research takes a long time to be appreciated, and you have a really radical idea and nobody even appreciates it. Max Planck may have been the... one of the foremost proponents of this idea.

He was very quotable. He said, "Science advances one funeral at a time." He also said that, I'm forgetting his other great quote on this, but his view that, young people pick up new ideas, but old people stick to what they know. We actually find that's not the case. We also find that the most transformative ideas are used more in the long run, but they're used more in the short run too. They're just more important and they get used more, and in fact, they get used even more in the short run, which we define as the first five years, than they do in the long run relative to work that work that's not transformative, so that's the time frame.

Now our data cover 1983 to 2012. That's a limitation of our study, I can't tell you if there's some idea that was so radical that they had it in 1983 and it's not going to be until 2050 that it really gets used. That piece of it I can't answer, but at least in the 30 years that we have, the transformative work gets used more in the short run as well as more in the long run.

Thank you

(MC: Matsuzawa)

It's quite insightful and quite interesting, and in the remark, there are a number of data available, for example in relation to aging, so... and what sort of divisions that can be named. This is even that those data are incorporated, so that's the learning, so in

closing, there will be the remark offered by the general, director general

(Naoki Saito)

Thank you very much for your very nice and inspiring presentations again, but I'm sorry, but a lot of the audience is Japanese so I will make comments in Japanese

Thank you very much indeed for coming to the lecture despite your busy schedules. At the beginning, we thought... we originally thought that there will be the advanced initiatives shared by making use of the Social Security Number and referring to the administrative data. I thought the environment is easier to get that data connected by making use of SSN, but without using SSN, only using the Identi... to identify the name and the date of birth and linkages are not possible

That's what we have learned, so we've got the JGRAD and we've got the similar tracing survey to be conducted in every three years and one of the questions we ask is income level

Then the typical reaction from the respondent is, "Do we have to answer that much?"

But by making My Number, so the individually assigned number for the national, where we can receive that data, now that's a quite hurdle for us, but in the future, we also would like to scope the possibility of collecting that sort of data, and also we would like to make use of data available in the private sector in the long term to trace the subjects going forward

So what we have heard today is good input for... to be made use for the future research activities of JGRAD, and the way your representing data in the short run, then learn the benefit and that can be shared among the universities and the post graduate students, but broadly speaking, this is something that we aren't able to share to the open public as well

and the necessity of the data for the international comparison purpose, that's another thing we talked about. Currently referring to the available of JGRAD, the mechanism is difficult to trace international students, but the international students are taking the greater role after graduation, now that is something that is worth tracing and there is a necessity of tracing such data

And globally speaking, US is under the Trump Administration experiencing the drastic change

And universities do face the quite stringent budget situation

And physically speaking, the number of youth has been on decline in Japan, so beyond the border. The doctor and the recipients, the study needs to be deepened going forward by tracing the data and also the aggregating data on an international basis so that we can find the significance and share that significance to the broader

world

And what we've been thinking so far is for example next year, the Triple As annual meeting is held in February, but the next year, there will be the meeting scheduled in Austin, Texas, so there we want to make an international comparison and analysis of these kind of topics, so... and I would like to make a proposal for the possible session to be held and on that occasion, if that is materialized, then I would like to share the Japanese initiatives including JGRAD

So I would like to continue having a discussion with Professor Weinberg to look ahead in collaborating each other going forward

So I would like to ask for your continued support from the people concerned. And I would like to hear your active input from yourselves. Thank you very much indeed

(MC: Matsuzawa)

Thank you very much for your times. Once again, please give a big hand to Dr. Weinberg

So, that's the end of today's symposium. Thank you very much again

<<END

謝辞

本報告書は、2017年3月3日に開催された講演会及び講演者との意見交換を元に日米のデータベースプロジェクトの比較を行ったものである。講演者のWeinberg教授におかれては、大変ご多忙の中、貴重なご後援を賜り、心からお礼を申し上げたい。

また、文部科学省科学技術・学術政策研究所企画課国際協力調整官 堀野功氏におかれては、米
国側との調整等において、また、同研究所第一調査研究グループ上席研究官 相馬りか氏には当日
の講演会の設定・準備にご尽力いただいた。また、報告書作成に当たって株式会社 日立コンサル
ティング マネージャー岡山将也氏、株式会社コスモピアの平尾慶氏に資料作成・チェック等、ご
尽力をいただいた。この場をお借りして皆様に感謝の意を表したい。

調査体制

本報告書は、文部科学省科学技術・学術政策研究所第1調査研究グループが主催した講演会（第
2章）の成果を踏まえ、講演者との意見交換や文献調査により情報収集し、「UMETRICSと博士
人材データベース（JGRAD）の比較研究」（第1章）を追加してまとめたものである。

第1章は筆者の責任でまとめたものである。なお、JGRADの現状について説明するため用いた
図表作成データは、以下の委託事業の調査研究成果を用いている。

平成28年度科学技術調査資料作成委託事業「博士人材データベースの利用者に対する支援及びシ
ステムの運用・データ管理に基づくシステム評価に係る業務」株式会社 日立コンサルティング

また、第2章の講演会の開催及び資料作成に当たっては、文部科学省科学技術・学術研究所企画
課及び第一調査研究グループの協力のもと実施した。

參考資料

参考資料1 講演会開催案内

【開催案内 2017/3/3】科学技術・学術政策研究所講演会のお知らせ

科学技術・学術政策研究所による講演会を、下記のとおり開催いたしますので、御案内申し上げます。お手数ですが、参加を希望される方は御氏名・御所属を平成29年2月20日（月）17時までに、下記の参加申込先に、メールにてあらかじめ御連絡いただきますようお願い申し上げます。

なお、会場の都合により参加者を調整させていただく場合がありますので、御了承ください。

記

○演題：「博士取得者の初期キャリアの設定：米国 UMETRICS プロジェクトからのエビデンス
－Early Career Outcomes of Doctoral Recipients: Evidence from the UMETRIC－」

○講師： Dr. Bruce Weinberg
オハイオ州立大学 経済学研究科 教授
全米経済研究所 (NBER) 上席研究官

○日時： 2017年3月3日（金） 15時00分～16時50分 （受付開始14時45分）

○場所： 文部科学省 科学技術・学術政策研究所会議室（16B）
（東京都千代田区霞が関3-2-2 中央合同庁舎第7号館東館16階）

○講演趣旨：

UMETRICS (Measuring the Impacts of Research on Innovation, Competitiveness, and Science) は、2015年1月に、米国オバマ政権が進める STAR METRICS から分離独立したデータベースプロジェクトであり、米国中西部の主要大学が設立した大学間コンソーシアム CIC (Committee of International Cooperation) の連携協力により進められています。UMETRICS は米国国勢調査局の持つ雇用と家計データや職歴データや、企業データ等の社会的データ、出版・引用データなどを連結させることで、研究開発投資の社会における成果を多角的に捉えようとしているのが特徴です。また、UMETRICS は、米国の大学で博士号取得者の動向を毎年追跡調査している「Survey of Earned Doctorates (SED)」との連携に向けて作業中であり、政府が博士号取得者を支援することの長期的な社会影響を測定することを目標としています。UMETRICS に関する人材政策の論文は Science 誌等に発表され話題となりました。

今回の講演会では、当該論文の主著者 Weinberg 教授をお招きし、UMETRICS プロジェクトから得られた 米国の博士人材研究の最新情報についてご講演いただきます。我が国の進める博士人材データベース (JGRAD) プロジェクトにも大変参考となると思われまますので是非ご参加ください。

○言語： 英語（同時通訳あり）

○講師経歴：

1996年博士号取得（経済学、シカゴ大学）。イノベーション創出人材の流動性や国家間のイノベーション競争力等についての実証研究を行い、合理的な科学イノベーション政策立案のために必要な科学的根拠となる活動を推進。また、研究人材の賃金差異などの労働経済学にも携わっている。UMETRICSでは主要研究者の一人として、対象大学における博士号取得者の動向の追跡調査などを通じて、研究の社会的影響を測るための定量的アプローチに携わっている。また、クリーブランド連邦準備銀行やフーヴァー研究所等にも招聘され共同研究を行ってきた。

○備考

本講演会は、博士人材データベース事業参加大学への情報提供の一環として、「博士人材データベースのパイロット運用に関する連絡協議会」に引き続き開催されるものです。原則、参加大学の関係者を対象としておりますが、これから参加を検討される大学（オブザーバー大学）関係者の聴講も可能です。なお、席に限りがあるため、事前の参加登録が必要です。

○講演内容についてのお問合せ先：

科学技術・学術政策研究所 第1調査研究グループ（担当：松澤）
Tel：03-3581-2395（直通）
e-mail：lpg@nistep.go.jp

○講演会の参加申込先：

科学技術・学術政策研究所 第1調査研究グループ（担当：笠掛）
Tel：03-3581-2395（直通）
e-mail：seminar-lpg@nistep.go.jp

○申込み締切り： 平成29年2月20日（月） 17時00分

○ 会場へのアクセス：

駅名	出口	所要時間
銀座線 虎ノ門駅	11	徒歩 2 分
丸の内線 霞ヶ関駅	A13	徒歩 5 分
千代田線 霞ヶ関駅	A13	徒歩 5 分

文部科学省の地図



※省外から参加される方は、身分証明書をお持ちいただき、開催案内を文部科学省 2 階受付にて提示していただきますようお願いいたします。

参考資料 2 講演会スライド

UMETRICS: The Database of Doctoral Holders and their Impacts

Bruce A. Weinberg
OSU Economics, IZA, NBER
March 3, 2017

Measuring the Value of Research

1. People-centric approach to valuing research
 - People, not disembodied publications, are the vectors by which ideas are transmitted
 2. Metrics for newness, novelty, transformativeness of research
 - Correlated with, but distinct, from impact
 3. Measure economic activity
- Move beyond counting Pubs, Patents, Citations...

Two Initiatives

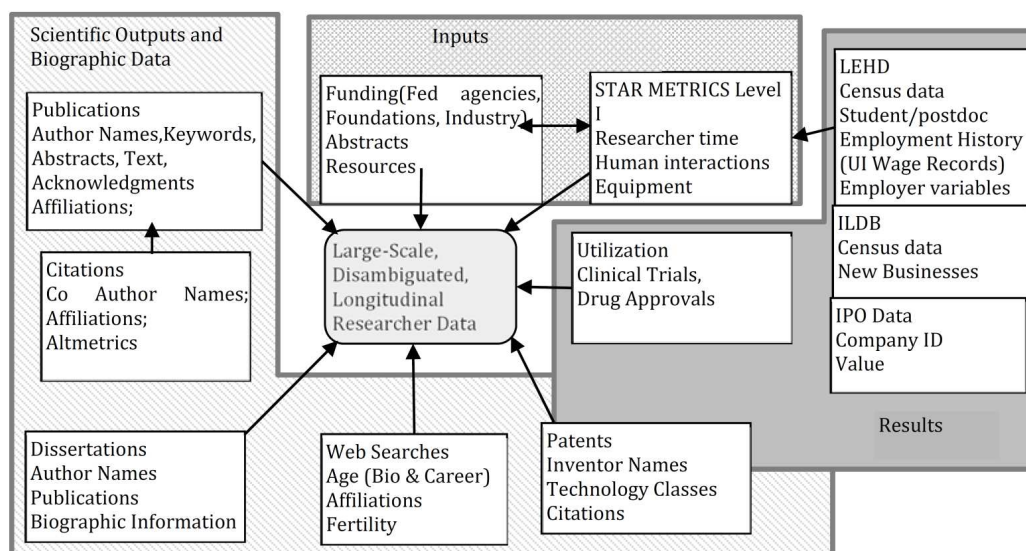
Institute for Research on Innovation and Science (IRIS)

- UMETRICS, Census-University (CIC / AAU) partnership
- Hub at U Michigan; Nodes at OSU, NYU, ...
- Sloan, Kauffman, NIH, NSF, SBA, USDA, USPTO

Innovation in an Aging Society

- NIA, OBSSR, SciSIP-funded Program Project to NBER
 - Researchers typically viewed as most innovative early in their careers; the U.S. scientific workforce is aging rapidly
 - Studies how the aging biomed. research workforce affects the production of innovation, health, and the economy
- Both producing data for the research community

Lee Giles (PSU), Nate Goldschlag (Census), Madian Khabsa (Microsoft), Kunho Kim (PSU), Ron Jarmin (Census); Julia Lane (NYU), Jason Owen-Smith (Michigan), Neil Smalhesier (UIC), Vetle Torvik (UIUC); Nik Zolas (Census)



IRIS / UMETRICS

- Partnership with B1G, AAU, and other Universities and Census
- Hub at the University of Michigan
- Nodes at OSU, NYU,...
- Data housed securely

- Infrastructure supported by Sloan and Kauffman
- Research supported by NIA, NSF, SBA, USDA, USPTO

The IRIS / UMETRICS Partnership

- IRIS is a consortium of member universities
 - Provide data
 - Pay membership dues
 - Receive reports
 - Participate in governance: elect board of directors

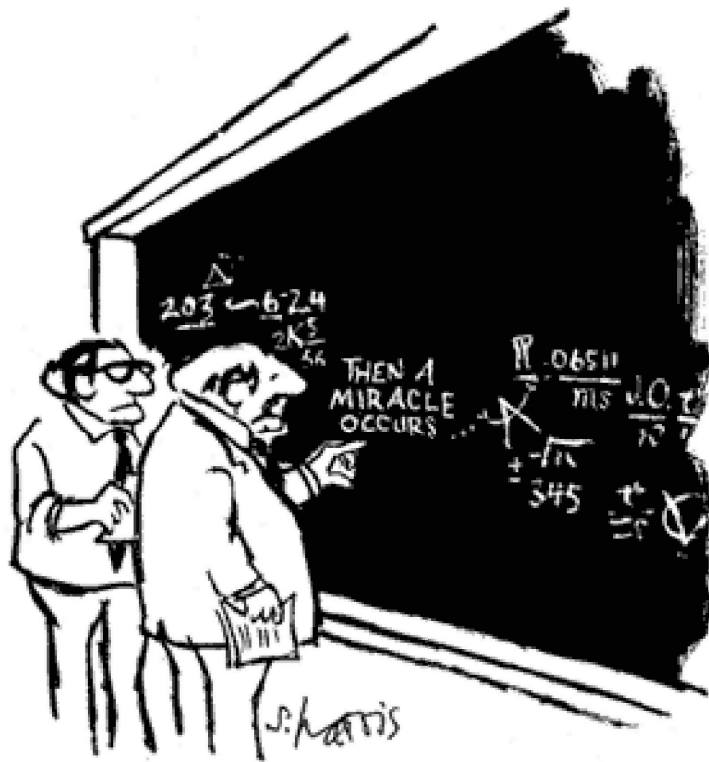
- Voluntary participation; *all* stakeholders benefit

What do Member Universities Get?

- Information to improve programs
- Access to data for their own researchers
- Reports showing their activities

In 2014, U.S. Universities invested \$210 in academic research for every man, woman, and child in the country

- We make those investments to develop human knowledge and to improve quality of life and well being.
- How do we understand, explain and improve those effects?



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

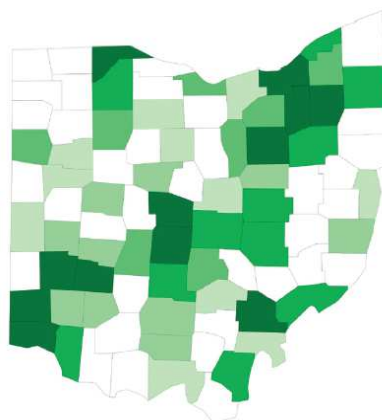


My congressional representative asked me about output measures of Ohio State research.

Was that research making an impact and how did we measure it? The traditional measures of publications and patents were not the answer he was looking for. I showed him UMETRICS data on the economic impact of federally sponsored research. This data was shown as a map of Ohio with vendors both within and beyond the state. This was exactly the type of data he wanted, showing both short-term and long-term gains from research. I find the UMETRICS data increasingly useful for discussions about the value of research.

—CAROLINE C. WHITACRE, VP FOR RESEARCH, OHIO STATE UNIVERSITY

REGIONAL DISTRIBUTION OF EXPENDITURES



State-level map of vendor and subaward expenditures from 3 years of federal awards to one research institution

\$15 - \$1,000	
\$1,001 - \$10,000	
\$10,001 - \$20,000	
\$20,001 - \$100,000	
\$100,001 +	

Federal Research Funding: A Detailed Analysis of Expenditures at U-State University



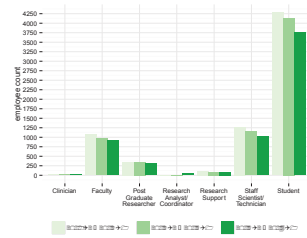
This report documents current Federal research funding and expenditures at U-State University. The report is based on UMETRICS data from U-State University. These are actual financial and payroll records for the University from 2011 Q2 – 2014 Q1.

Individuals Employed by Federal Research Funding

Between the second quarter of 2011 and the first quarter of 2014, Federal research awards supported a yearly average of 6,658 individuals at U-State.

Students constituted an average of 61% of individuals supported by federal research funding every year, while faculty employees comprise an average of 15%.

Yearly counts of total individuals on U-State University Federal research awards, broken down by occupational category (2011 Q2 – 2014 Q1)

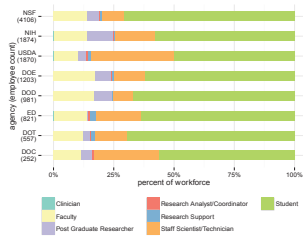


Organization of the Scientific Workforce

Between 2011 Q2 – 2014 Q1, students at U-State University constituted over 70% of the research workforce supported by NSF awards and over 58% of employees supported by awards from NIH.

Post graduate researchers comprise 11%, 5%, 7%, and 8% of the employees on awards from NIH, NSF, DOE, and DOD, respectively.

Breakdown of employment patterns on Federal research awards to U-State University, aggregated across 2011 Q2 – 2014 Q1

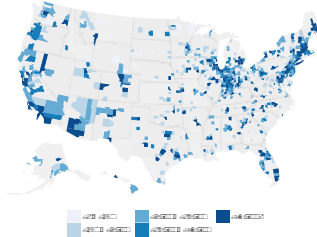


National Distribution of Research-Related Expenditures

The production of science requires the purchase of scientific equipment and technology as well as collaboration with private/public research organizations.

University research expenditures exceeded \$2.11 billion from 2011 Q2 – 2014 Q1 and included transactions with vendors in almost 850 US counties.

Total vendor & subaward expenditures on Federal research awards to U-State University, aggregated across 2011 Q2 – 2014 Q1



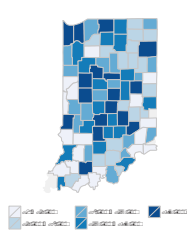
For more information about UMETRICS as well as methodology and data sources, visit www.dcn.net/projects/umetrics
Federal Research Funding Analysis

Regional Distribution of Research-Related Expenditures

Between 2011 Q2 – 2014 Q1, U-State University research generated over \$24 million in expenditures in Indiana counties alone.

Purchases from Marion county vendors exceeded \$8 million.

Federal research award expenditures on vendors & subawards in Indiana counties, aggregated across 2011 Q2 – 2014 Q1



FEDERAL RESEARCH FUNDING: A DETAILED ANALYSIS OF EXPENDITURES AT PURDUE UNIVERSITY

This report documents current federal research funding and expenditures at Purdue University. The report is based on actual financial and payroll records for the University for 2010, 2011 and 2012 as well as published government data for 2010, 2011 and 2012.

SCOPE

Research funding represents an injection of external funds to the university and the academic community.

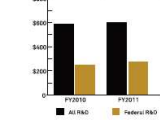
- Researchers at Purdue University generated over \$601 million in research activity in 2011 (the latest year for which figures are available).
- \$270 million of that research & development was funded by the federal government.

EMPLOYMENT

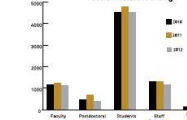
Scientific research both creates new scientific knowledge and trains the next generation in the scientific method.

- The research enterprise also employs many technicians, clinicians and other support staff.
- In 2012, more than 2,340 individuals (equivalent to more than 2,250 FTE positions) were directly employed at Purdue University by federal research funding.

University Research Funding



Number of Individuals Employed by Federal Research Funding

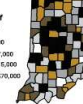


EXPENDITURES

The production of science requires the purchase of scientific equipment and technology as well as collaboration with private/public research organizations.

- Purdue University research generated over \$14 million in expenditures in Indiana counties alone.
- In 2012, federal research funding to Purdue University supported the purchase of almost \$96 million of supplies and subcontracted services from the nation as a whole.
- Vendors in over 700 US counties do business with researchers at Purdue University.
- In 2012, vendors in each of more than 145 of those counties derived combined revenues of over \$60,000.

Regional Distribution of Expenditures



National Distribution of Expenditures



Federal Research Funding Analysis

November 2013

Michigan

- State Legislature, Federal Delegations, Regents

OSU

- State and Federal Delegations

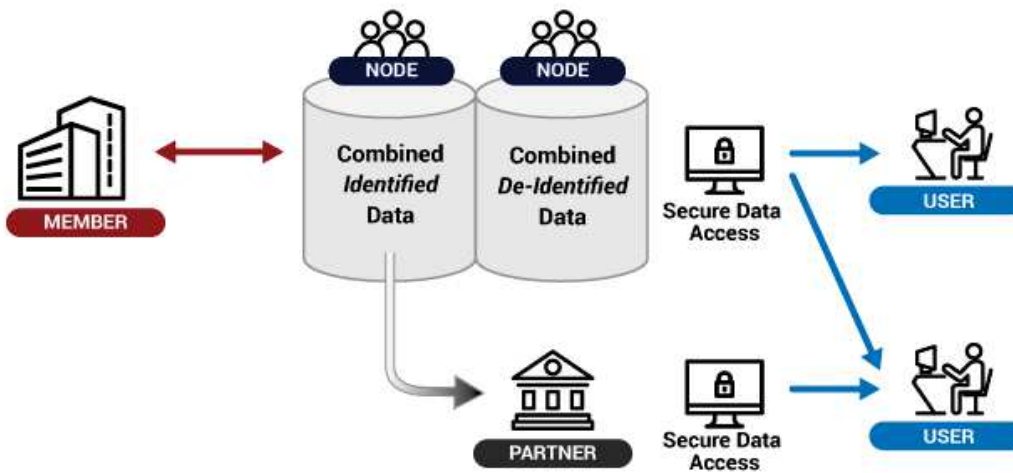
Purdue

- State and Federal Delegations, General Outreach

MEMBERS: Universities contribute data, support infrastructure and receive campus-specific and aggregate reports

NODES: Approved nodes materially improve data, develop products, and expand user communities

USERS: Approved users securely access de-identified aggregate datasets



PARTNERS: Approved partners receive data from IRIS which they improve and make accessible through their own secure systems

Weinberg et. al., *Science*, 2014

POLICYFORUM

RESEARCH FUNDING

Science Funding and Short-Term Economic Activity

Expenditures from grant funds support many different types of workers and vendors across the nation.

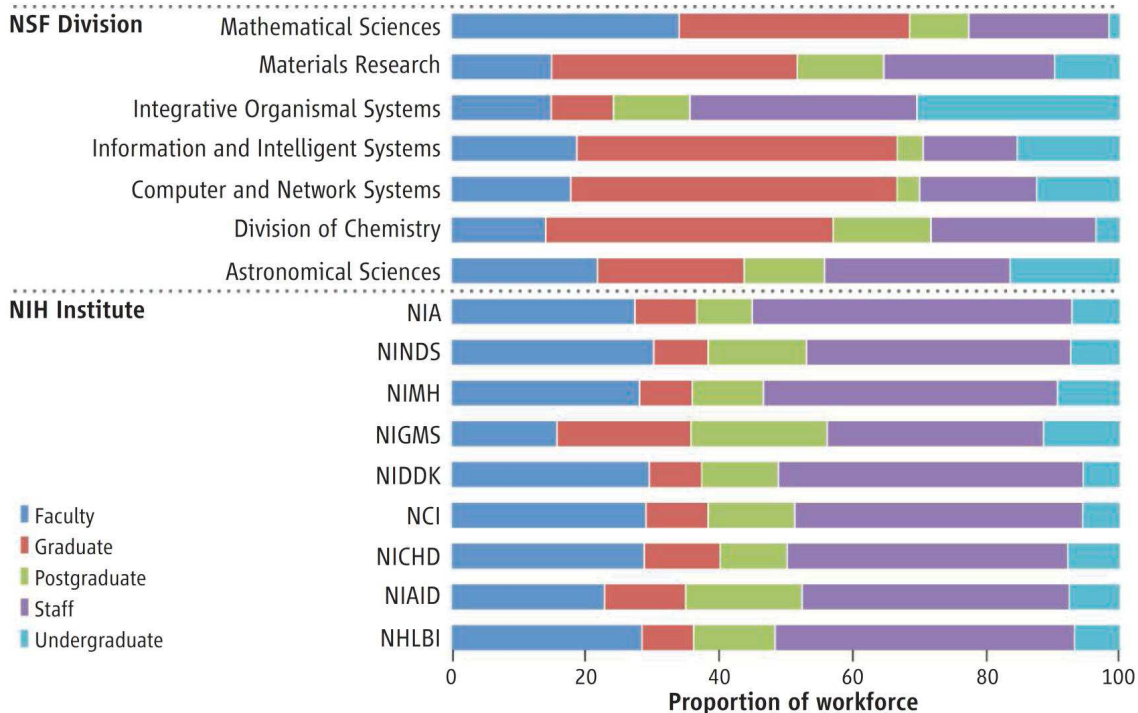
Bruce A. Weinberg,^{1,2,3} Jason Owen-Smith,⁴ Rebecca F. Rosen,⁵ Lou Schwarz,⁶ Barbara McFadden Allen,⁷ Roy E. Weiss,⁸ Julia Lane^{5,9,10,11*}

There is considerable interest among policy-makers in documenting short-term effects of science funding. A multiyear scientific journey that leads to long-term fruits of research, such as a moon landing, is more tangible if there is visible near-term activity, such as the presence of astronauts. Yet systematic data on such activities have not heretofore existed. The only source of information for describing the production of most science is surveys that have been called “a rough estimate, frequently based on unexamined assumptions that originated

automated approaches to do so (2). The first tranche of rich data are drawn directly from university personnel and financial administrative records that track actual expenditures of all active federal projects. These data provide project-level information about the occupations of the part-time and full-time workforce paid on each funded grant and about the purchases made from vendors who supply scientific researchers. Neither of these types of information have reliably been available before (3, 4).

The results reported in this paper represent

aggregate, the 15 institutions that make up the CIC receive 8 to 10% of all federal research dollars. The majority of these institutions are large, Midwestern public universities. Hence, they are not representative of all recipients of federal funds. However, it is unlikely that the type of science that is conducted with those funds is markedly different from that conducted at many other major research universities, and they provide a window into a large portion of federal research activity in the era of tightening federal budgets. Moreover, the 2012 data we analyzed reflect expenditures



Differences in workforce composition in projects funded by NSF divisions and NIH institutes. NIA, National Institute on Aging; NINDS, National Institute of Neurological Disorders and Stroke; NIMH, National Institute of Mental Health; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NIAID, National Institute of Allergy and Infectious Diseases; NHLBI, National Heart, Lung, and Blood Institute. (See SM.)

Match to Survey of Earned Doctorates

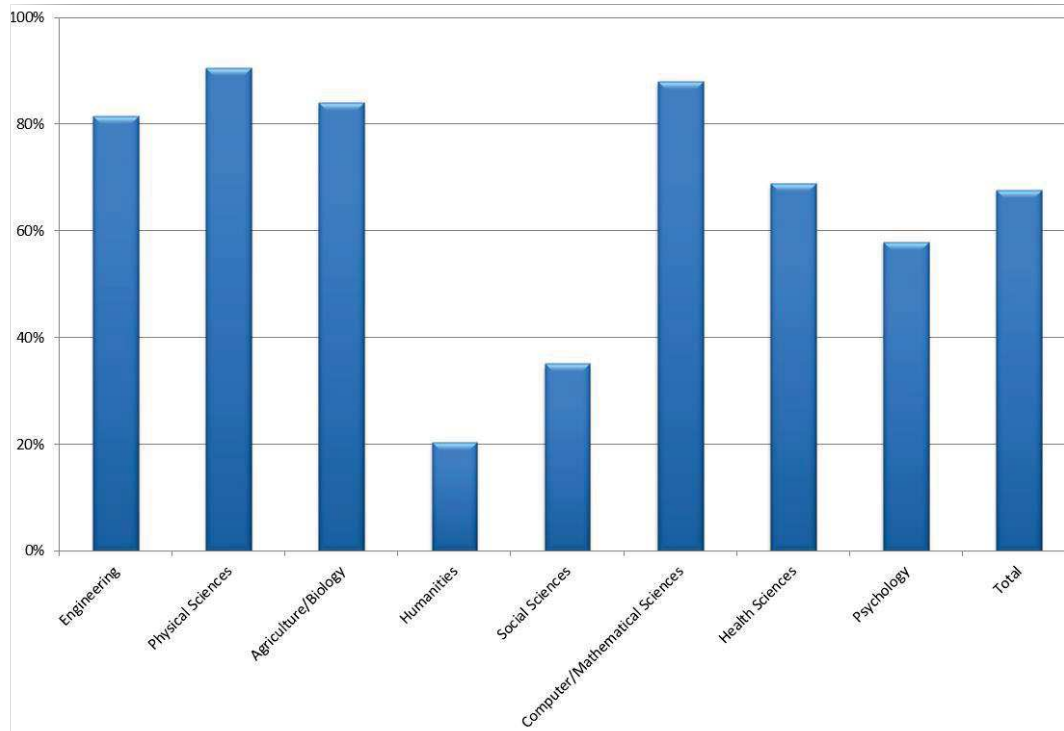
Federal Funding of Doctoral Recipients: Results from new linked survey and transaction data

- Wan-Ying Chang, National Science Foundation
- Wei Cheng, Ohio State University
- Julia Lane, New York University
- Bruce Weinberg, Ohio State University

Data Options

	Administrative	Naturally Occurring	Survey
Example	UMETRICS: HR Records, Tax Data	Publications, Patents, Citations	Survey of Earned Doctorates
Benefit	Low Burden Full Coverage Accurate Incredible richness	Low Burden Full Coverage Accurate	Topics more flexible
Cost	Initial Cleaning and Linking Only some topics possible	Initial Cleaning and Linking	Ongoing costs Accuracy Response rates

Federal Funding by Field



Mix of Federal Support

(3,329 Students @ 2 Universities)

	NIH	NSF	DOD	DOE	Other Fed	Left agency and 2 more
NIH	29.62	7.60	1.89	0.48	1.20	3.24
NSF		23.67	6.79	2.37	3.60	5.50
DOD			6.94	0.90	0.72	4.81
DOE				3.60	0.36	2.40
Other Fed					3.78	3.00

RESEARCH INVESTMENT

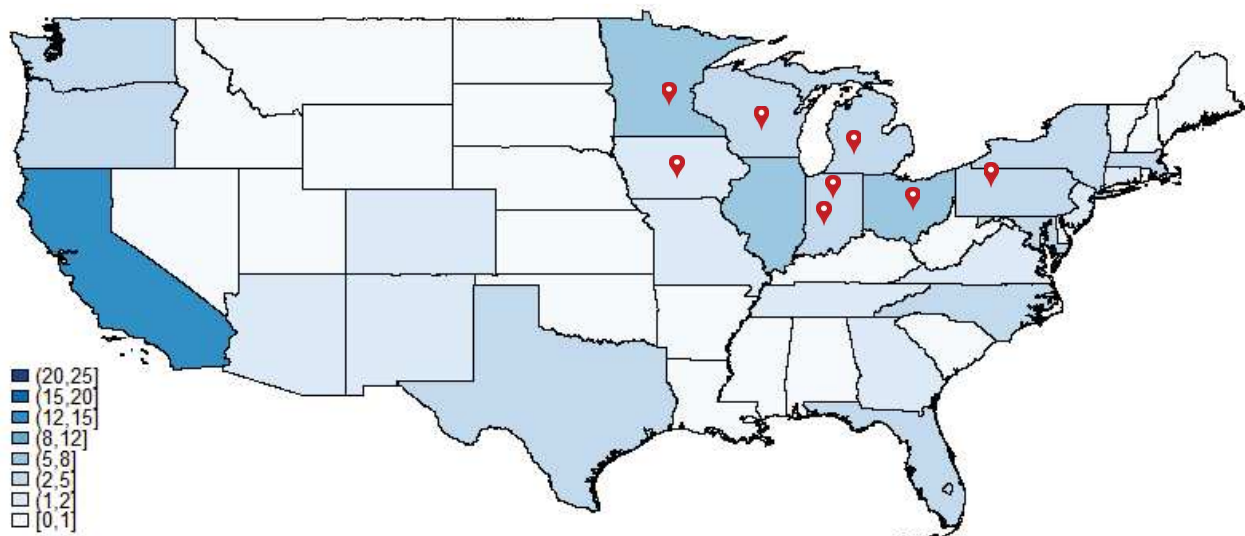
Wrapping it up in a person: Examining employment and earnings outcomes for Ph.D. recipients

Nikolas Zolas,¹ Nathan Goldschlag,¹ Ron Jarmin,¹ Paula Stephan,^{2,3}
Jason Owen-Smith,⁴ Rebecca F. Rosen,⁵ Barbara McFadden Allen,⁶
Bruce A. Weinberg,^{7,3,8*} Julia I. Lane^{1,5,8,9,10}

In evaluating research investments, it is important to establish whether the expertise gained by researchers in conducting their projects propagates into the broader economy. For eight universities, it was possible to combine data from the LIMETRICS project, which provided

all
Uni
data
v
plac
in 2
iden
be
hou
con
by l
unic
resu
info
dist
tain
the

Placements



Zolas et. al. 2015

Table 1. Postgraduation employment of UMETRICS doctoral recipients who were paid by research grants and left the university between 2010 and 2012. The national workforce distribution is calculated from all employment in all establishments covered by the Census's LBD between 2010 and 2012.

Locale and small	Doctoral recipients placed in sector (%)				
	Industry		Academia	Government	All
	R&D firms	Non-R&D firms			
Placed within sector	17.0	21.7	57.1	4.1	100.0
National sample (<i>M</i>)	10.8	75.0	10.7	3.5	100.0
Of those in sector, percent placed:					
Within 50 miles	10.1	23.5	8.9	18.2	12.7
Within state	16.6	36.0	18.0	25.8	22.0

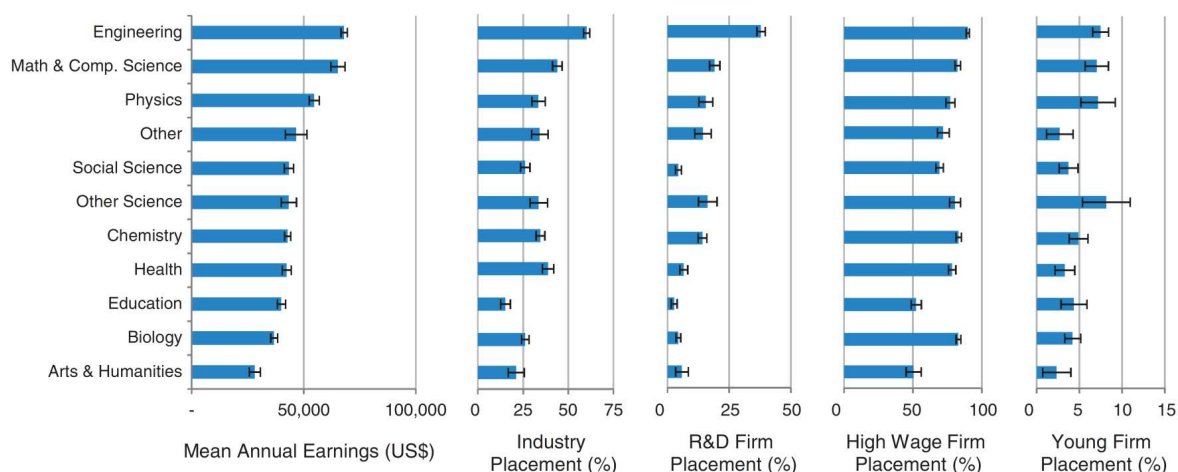


Fig. 3. The annual earnings and placement of doctoral recipients supported by grants vary by field. Young firm are defined to be those <5 years old. High-payroll per worker establishments are defined as those with a payroll per worker above the median for the establishments within their six-digit industry. Mean annual earnings are stated as U.S.\$1 ×1000. Means and standard errors for each variable.

Research Funding and Regional Economies

Nathan Goldschlag
Stefano Bianchini
Joseba SanMartin Sola
Bruce Weinberg
Julia Lane

Goldschlag, et. al., 2016

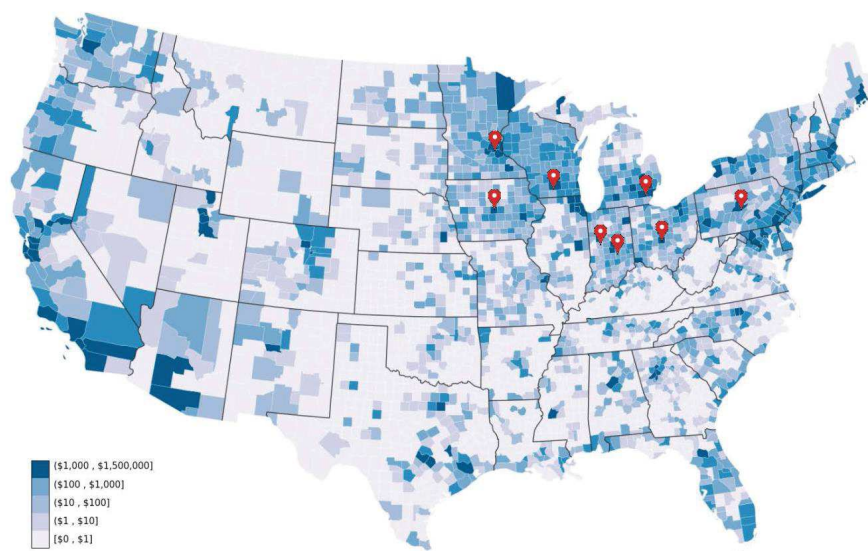
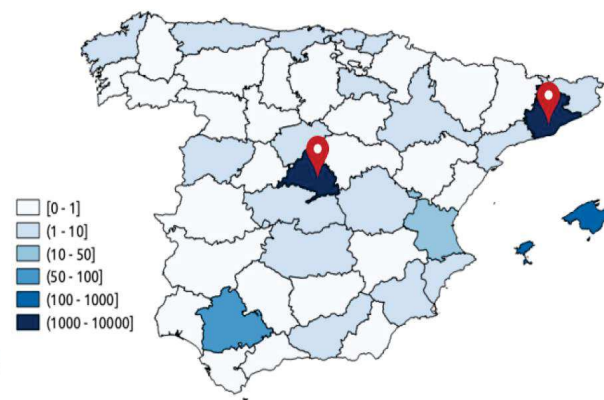
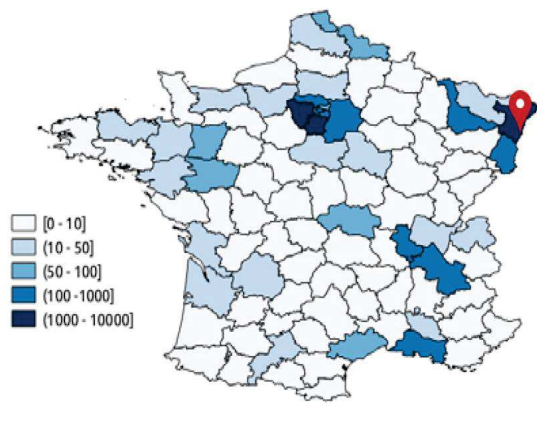


Figure 1a: The geographic distribution of vendor purchases in the US

Nothing Special about the US!



New Establishments <50 Miles

	All Firms	R&D Performing Firms	All Firms Opening New Establishments	R&D Performing Firms Opening New Establishments
Vendor in 2011	2.065*** (0.480)	-0.321 (0.704)	6.643*** (1.559)	-1.742 (2.337)
2011 Amount (\$1M)	3.358** (1.548)	5.710** (2.585)	5.575** (2.789)	6.445* (3.393)
Constant	0.104*** (0.00880)	0.243*** (0.0566)	0.832*** (0.0703)	1.632*** (0.371)
Observations	1,183,000	59,000	147,000	9,000
R-squared	0.252	0.370	0.246	0.364

Gender Underrepresentation

American Economic Review: Papers & Proceedings 2016, 106(5): 1–7
<http://dx.doi.org/10.1257/aer.p20161124>

STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data linked to the 2010 Census[†]

By CATHERINE BUFFINGTON, BENJAMIN CERF HARRIS, CHRISTINA JONES,
AND BRUCE A. WEINBERG*

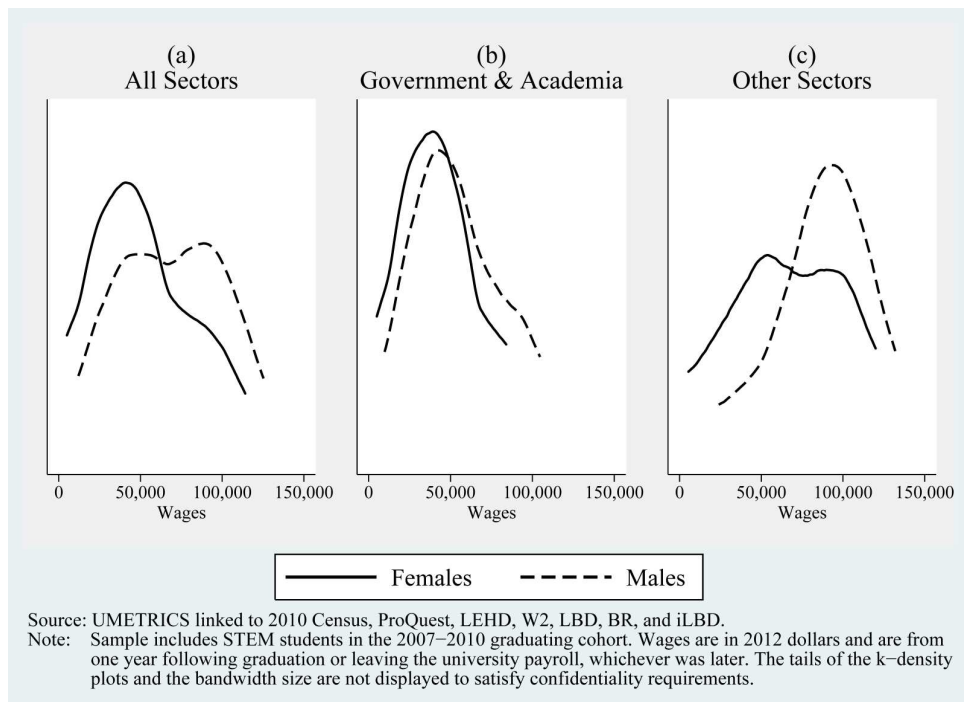
Women are underrepresented in a number of science and engineering fields, and the extent of underrepresentation generally increases in career stage (National Science Board 2014). This arti-

We use unique new administrative data that allow us to identify personnel employed on federally funded research grants at four participating universities over ten years. These data

Table 1. Training Environments of Male and Female Graduate Students Participating in STEM Research

Dependent Variables ↓			(1)	(2)	(3)	(4)	(6)	(7)	(8)
	(a) Females	(b) Males	(c) Diff						
Share of Faculty that are Female	0.2 (0.02)	0.1 (0.01)	0.1*** (0.02)	0.1*** (0.02)	0.1*** (0.02)	0.1*** (0.02)	0.1*** (0.02)	0.1*** (0.02)	0.1* (0.03)
Share of Graduate Students that are Female	0.1 (0.01)	0.1 (0.00)	0.0*** (0.01)	0.0*** (0.01)	0.0*** (0.01)	0.0 (0.01)	0.0 (0.01)	0.0 (0.01)	-0.0 (0.02)
Ln Team Size	1.7 (0.04)	1.9 (0.03)	-0.2*** (0.05)	-0.2*** (0.05)	-0.2*** (0.05)	-0.1** (0.06)	-0.1* (0.06)	-0.1* (0.06)	-0.1 (0.09)
Faculty to Student Ratio	0.9 (0.06)	0.6 (0.03)	0.3*** (0.07)	0.2*** (0.07)	0.2*** (0.07)	0.2*** (0.08)	0.1** (0.07)	0.1* (0.07)	0.3** (0.13)
Total Number of Awards	2.2 (0.07)	2.7 (0.06)	-0.5*** (0.09)	-0.3*** (0.09)	-0.3*** (0.09)	-0.2** (0.09)	-0.2*** (0.09)	-0.2*** (0.09)	-0.1 (0.15)
Number of Months Participating on the Award	21.0 (0.69)	21.6 (0.45)	-0.6 (0.82)	-1.1 (0.79)	-1.0 (0.79)	-1.0 (0.82)	-1.4* (0.82)	-1.4* (0.82)	-0.9 (1.18)
Years from First Observation to Degree	3.2 (0.08)	3.2 (0.06)	-0.0 (0.10)	-0.1** (0.06)	-0.1* (0.06)	-0.1 (0.06)	-0.1** (0.06)	-0.1** (0.06)	0.0 (0.10)
University, First Year Trend, Left-Censored				✓	✓	✓	✓	✓	✓
Race, Hispanic Origin, Age, Age-squared					✓		✓	✓	✓
Dissertation Topic						✓	✓	✓	✓
Funding Agency							✓	✓	✓
Married or Partnered, Children								✓	✓
Female x (Married or Partnered + Children)									✓
Observations	370	867	1,237	1,237	1,237	1,237	1,237	1,237	1,237

Earnings Distributions



38

Table 2. Labor Market Outcomes of Male and Female Graduate Students Participating in STEM Research

Dependent Variables ↓	(1)		(2)	(3)	(4)	(6)	(7)	(8)	
	(a) Females	(b) Males	(c) Diff						
Employed in Industry	0.40 (.022)	0.5 (0.02)	-0.1*** (0.03)	-0.1*** (0.03)	-0.1*** (0.03)	-0.1 (0.03)	-0.1 (0.03)	-0.0 (0.03)	-0.0 (0.05)
Ln Wage	10.50 (.063)	10.9 (0.03)	-0.4*** (0.07)	-0.3*** (0.07)	-0.3*** (0.07)	-0.1** (0.07)	-0.1* (0.07)	-0.1* (0.07)	0.0 (0.10)
Ln Wage (with Industry Controls)	10.40 (.057)	10.7 (0.04)	-0.3*** (0.07)	-0.3*** (0.07)	-0.3*** (0.06)	-0.1* (0.07)	-0.1 (0.06)	-0.1 (0.07)	0.0 (0.10)
University, First Year Trend, Left-Censored				✓	✓	✓	✓	✓	✓
Degree Year				✓	✓	✓	✓	✓	✓
Race, Hispanic Origin, Age, Age-squared					✓		✓	✓	✓
Dissertation Topic						✓	✓	✓	✓
Funding Agency							✓	✓	✓
Married or Partnered, Presence of Children								✓	✓
Female x (Married or Partnered + Children)									✓
Observations	318	731	1,049	1,049	1,049	1,049	1,049	1,049	1,049

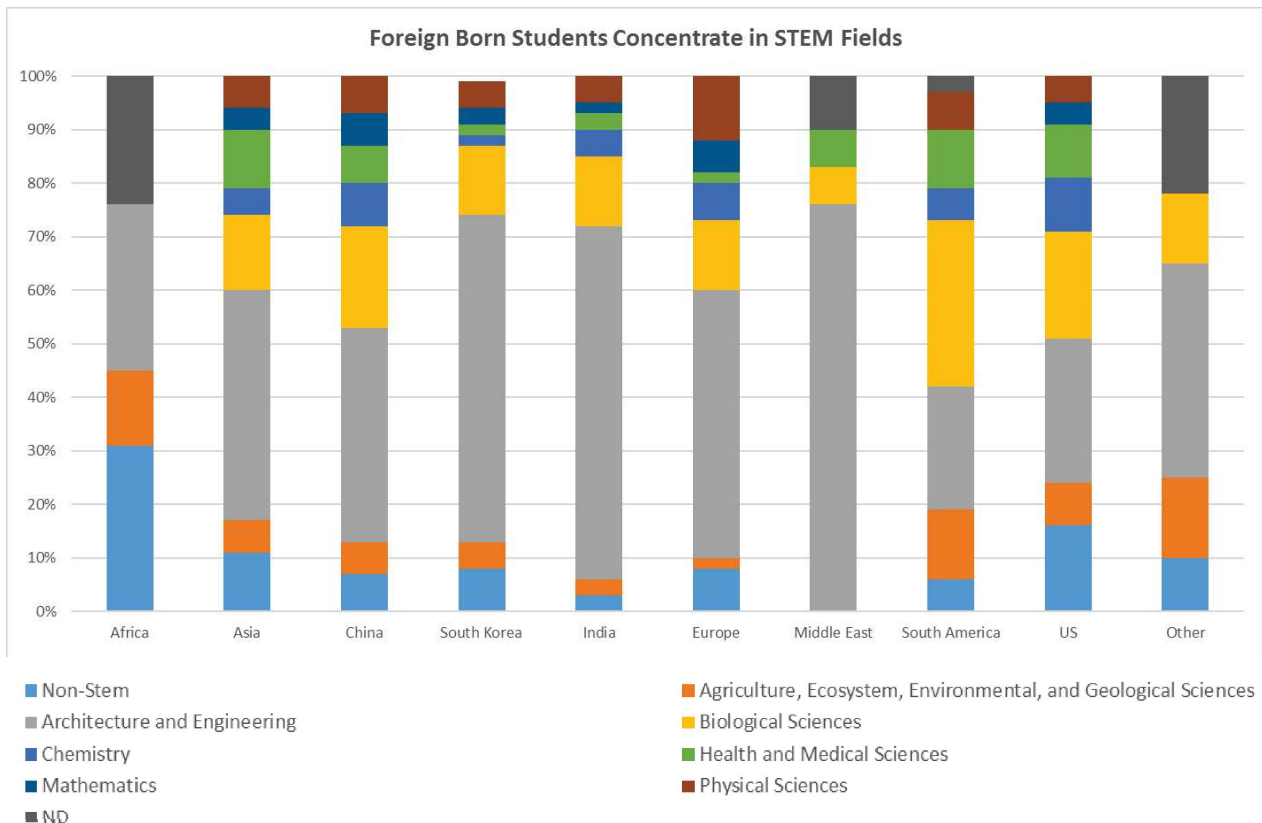
39

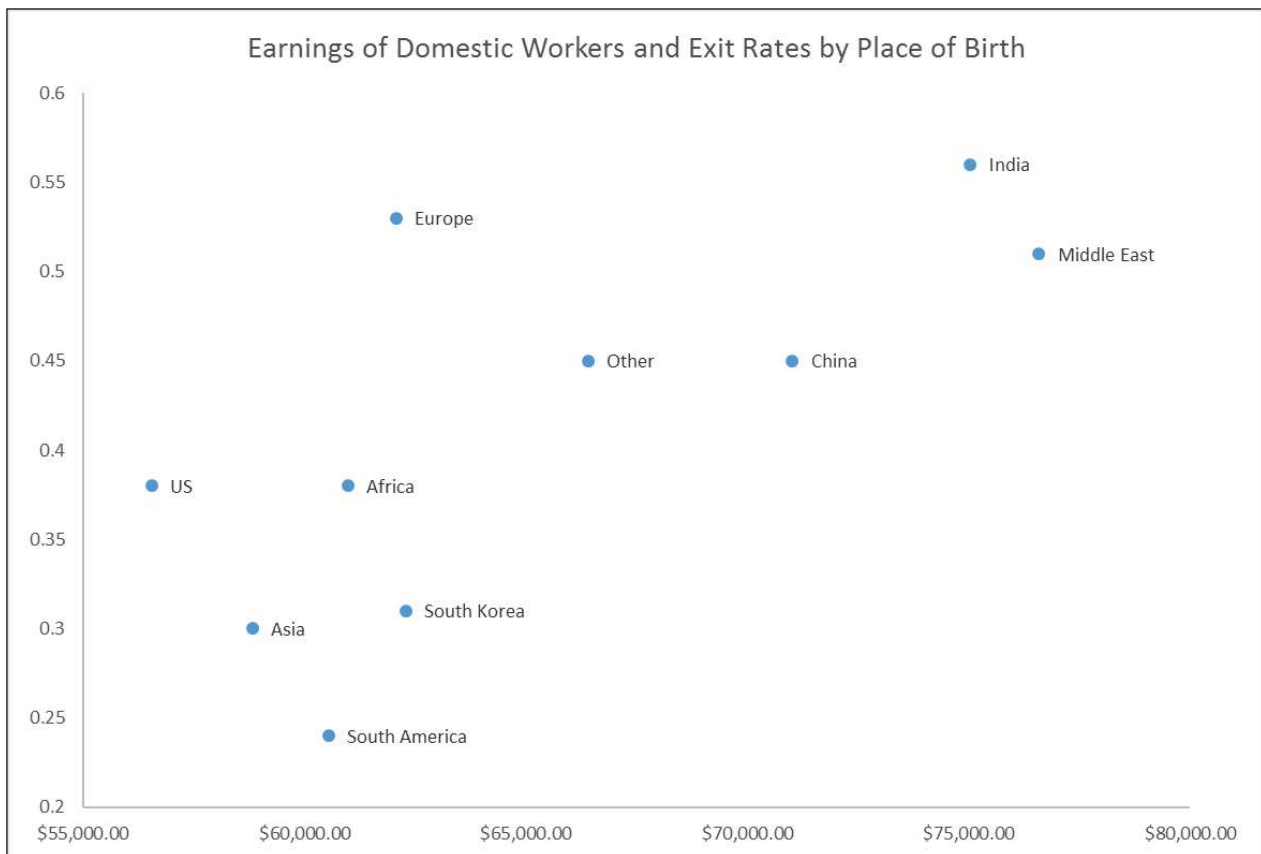
Research funding and the foreign born

Catherine Buffington, US Census Bureau
 Benjamin Cerf, US Census Bureau
 Akina Ikudo, UCLA
 Julia Lane, New York University
 Bruce Weinberg, Ohio State University

ASSA, 2017

Disclaimer: This presentation is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the US Census Bureau, UCLA, New York University, or Ohio State University.





42

Innovation in an Aging Society

1. Look at research productivity in the late career using DB pension as IVs for retirement
 - David Blau, John Ham, Bruce Weinberg
 2. Effects of researcher moves on productivity using age of children as IVs
 - Pierre Azoulay, Ina Ganguli, and Josh Graff Zivin
 3. New measures of research innovativeness and health impacts
 - Jay Bhattacharya and Mikko Packalen and Katy Borner, Jerry Marschke, and Bruce Weinberg
- Data Core
 - Neil Smalheiser and Vetle Torvik

Identifying High-Impact and Transformative Science (HITS)

*Jerry Marschke and Huifeng Yu (SUNY Albany);
Katy Börner and Robert Preston Light (Indiana
U); Bruce Weinberg and Joseph Staudt and (Ohio
State U)*

Metrics

1. Radical - Generates New Paradigms and Scientific Fields

Introduction and use of heavily-used new concepts

2. Radical - Destructive

The number and age of backward citations

3. Risky

Variance of Forward Citations

4. Multidisciplinary

Breadth of articles (citations) and concepts used

5. Wide Impact

Breadth of forward citations and concepts introduced

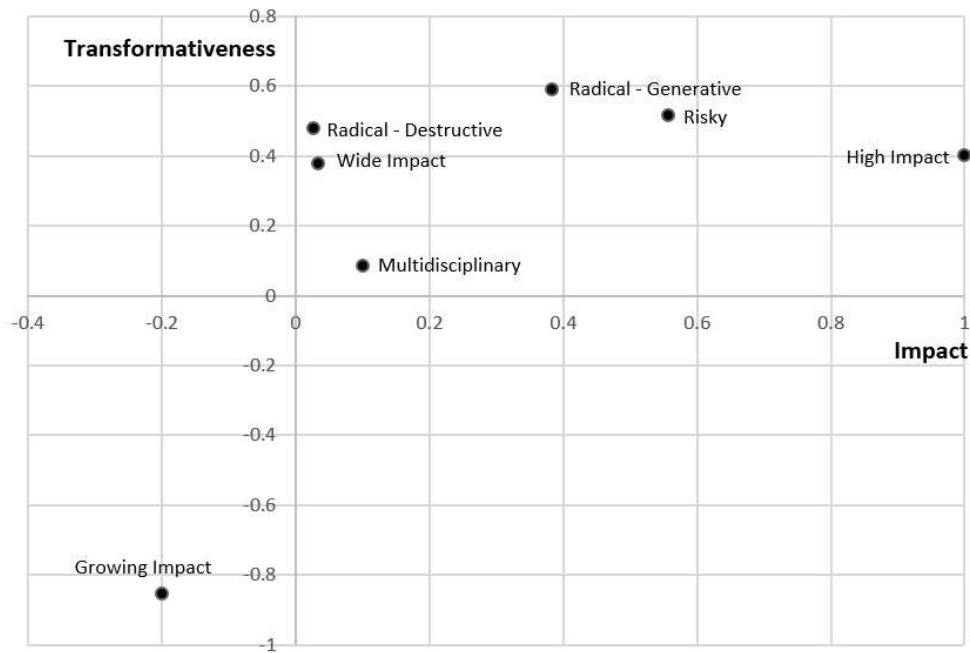
6. Growing Impact

Time to forward citations

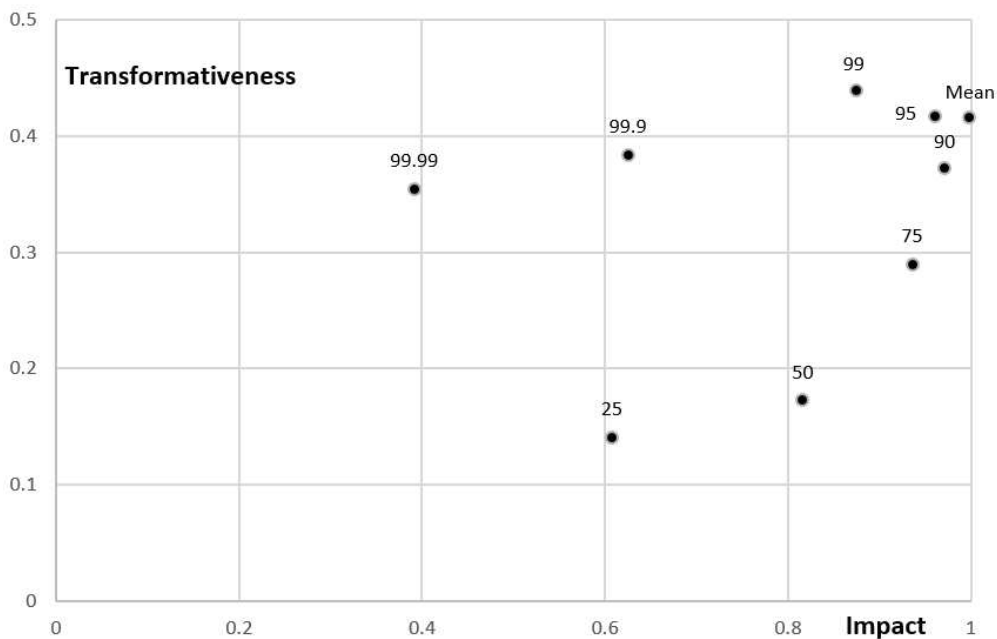
7. High Impact

Forward citation counts

Aspects of Transformativeness Related to Impact and Transformativeness



Citations Related to Impact and Transformativeness



Patterns

- Multi-disciplinarity and wide impact related
- Radical Generative and Destructive weakly related
 - New ideas can be born without obsolescence
- Multi-disciplinarity weakly related to impact
- Transformative work has shorter forward citation ages
 - Impactful over time, but even more so soon

Findings

- Cohesive set of indicators for transformative work
- Correlated (.4) with, but distinct, from impact
- Tremendous variation
- People, especially mid career flow into fields with transformative work (voting with feet)
- Fields with more young people and fewer old are more transformative and lower impact

Concluding Thoughts

- Increasingly critical to measure value of research
 - For policy makers, public, and to improve
- Science and innovation are international
 - Full accounting requires international data
- First step: Initiatives in individual countries
- Ultimately: Integrate data across countries
 - Will take considerable work!

The Teams

Innovation in an Aging Society

- Pierre Azoulay, MIT
- Jay Bhattacharya, Stanford
- David Blau, OSU
- Katy Borner, IU
- Wei Cheng, OSU
- Ina Ganguli, U Mass - Amherst
- Josh Graff Zivin, UCSD
- John Ham, NU Singapore
- Robert Light, IU
- Gerald Marschke, SUNY - Albany
- Mikko Packalen, Waterloo
- Neil Smalheiser, UIC
- Joe Staudt, OSU / Census
- Vette Torvik, UIUC
- Bruce Weinberg, OSU
- Huifeng Yu, Albany

IRIS / UMETRICS

- Stefano Bianchini, Strasbourg

- Cathy Buffington, Census
- Wan-Ying Chang, NSF
- Lee Giles, PSU
- Nathan Goldschlag, Census
- Josh Hawley, OSU
- Ron Jarmin, Census
- Christina Jones, AIR
- Kunho Kim, PSU
- Julia Lane, NYU
- Madian Khabza, Microsoft
- Barb McFadden Allen, CIC
- Christopher Morphew, Iowa
- Jason Owen-Smith, Umich
- Rebecca Rosen, NYU
- Lou Schwartz
- Joseba SanMartin Sola, Fundación Española para la Ciencia y la Tecnología
- Paula Stephan, Georgia State
- Roy Weiss, Miami
- Nik Zolas, Census

調査資料-263

博士人材政策から見た米国 UMETRICS;
UMETRICS と博士人材データベース (JGRAD) の国際比較研究

2017 年 7 月

文部科学省 科学技術・学術政策研究所
第 1 調査研究グループ
松澤 孝明

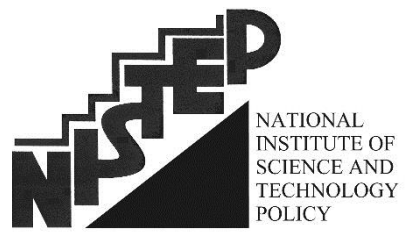
〒100-0013 東京都千代田区霞が関 3-2-2 中央合同庁舎第 7 号館 東館 16 階
TEL: 03-3581-2391 FAX: 03-3503-3996

UMETRICS in term of human resource policy for PhD holders;
International comparative study between UMETRICS in US and JGRAD in Japan

July 2017

1st Policy-Oriented Research Group
National Institute of Science and Technology Policy (NISTEP)
Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan

<http://doi.org/10.15108/rm263>



<http://www.nistep.go.jp>