1. 13:20 – 13:25  Web of Science data
2. 13:25 – 13:30  A research and Education dataset
3. 13:30 – 13:35  A research management dataset
4. 13:35 – 13:40  Q&A
Web of Science data
“Researchers suffer from an overload of poorly structured information that slows discovery and leads to poor investments. Web of Science creates trust through its selectivity, balance, independence, thorough curation, and historical reputation as an authoritative data provider. This enables faster research and better research management.”

Jonathan Adams
Visiting Professor of Higher Education and Research Policy
The Policy Institute, King’s College London
Web of Science data has been used globally by thousands of academic researchers, government, funding agencies, and industry analysts to gain insight and derive results in a myriad of scholarly, economic, and sociological pursuits.

In 1964, the Institute for Scientific Information (ISI), the organization that Dr. Eugene Garfield originally founded and the forerunner of Clarivate Analytics, released the first Science Citation Index. This resource recorded the bibliographical and citation contents of the world’s most influential journals, allowing users to follow citation links to find the specialized research most germane to their own work.

The Science Citation Index was released nine years after he published the concept of citation indexing for the sciences in 1955.
Journal management criteria are **publisher-independent**

**SELECTIVE**

*Selectivity* aids search and analytics by focussing on work identified as particularly significant.

Final selection on citation frequency focusses on the 50% of journals that attract 97% of all citations.

**Research culture** influences behaviour so metrics are adjusted by discipline and geography.

**BALANCED**

We select *‘the best’* globally, then by main discipline, then by sub-discipline and region.

A **unique balance** avoids over-dominance by any sub-set: more is not better if it is more in some areas and less in others.

Web of Science data provide a complete map of the current research territory.

**COMPLETE**

This includes every journal that passes our tests and none that fail to meet our thresholds.

In those journals, every document of all types is indexed.

Every reference in those documents is extracted and traced to verify links to earlier documents.

**ENHANCED**

Metrics are category-specific, derived from citation analysis to enable rapid and early discovery and to support informed management decisions.

Researchers can choose category options from national schemes, journal groups and topical clusters.

Publication records are enriched with search terms, disambiguated geographical tags, funding sources, and links to related data such as patents.
Web of Science Core Collection overview

Web of Science Core Collection covers research as it progresses from conference presentations, to peer-review journal articles, through to books and book chapters. Coverage is multidisciplinary—covering over 250 disciplines in sciences, social sciences, arts, and humanities.

- **SCIENCE CITATION INDEX EXPANDED** 1900-present >52 million
- **SOCIAL SCIENCES CITATION INDEX** 1900-present >9 million
- **ARTS & HUMANITIES CITATION INDEX** 1975-present >4.9 million
- **EMERGING SOURCES CITATION INDEX** 2005-present >2.8 million
- **CONFERENCE PROCEEDINGS CITATION INDEX** 1990-present >9.6 million
- **BOOK CITATION INDEX** 2005-present >1.1 million

**Comprehensive coverage:** >20,900 journals, >101,000 books, >203,000 conferences

**Massive citation network:** Explore over 1 billion Cited References back to 1900 to form an extensive network of scientific connections.

**Historical analysis:** Data start in 1900, enabling long-term trend analysis.

**Complete author data:** Conduct extensive analysis of researchers and their collaborations with indexing of all author names, along with Web of Science ResearcherID and ORCID identifiers, and algorithmic author disambiguation.

**Complete address information:** Complete indexing of all author addresses enables analysis of institutions, cities, states, countries, and regions. Over 13,000 institutions around the world are standardized via Organization-Enhanced.

**Funding acknowledgements:** Connect research to its funding sources from around the world and across sectors.

**XML format:** Data delivered in XML format to enable large-scale analysis, integration with other datasets, and use in statistical and visualization programs.
Broader Web of Science Group
Clarivate Analytics

Clarivate Analytics is a global leader in providing trusted insights and analytics to accelerate the pace of innovation.
A research and Education dataset
Web of Science Data used in Research Output
Factors contributing to this growth:
- Advances in technology enabling large-scale network analytics
- Growing interest in studying the landscape of science and innovation
- Big data projects that mash up bibliometric data with other data types
Web of Science Data in diverse disciplines (top 10 shown)

# Web of Science indexed articles analyzing Web of Science data

- INFORMATION SCIENCE LIBRARY SCIENCE: 2,231
- COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS: 1,102
- COMPUTER SCIENCE INFORMATION SYSTEMS: 540
- MEDICINE GENERAL INTERNAL: 307
- MANAGEMENT: 235
- SOCIAL SCIENCES INTERDISCIPLINARY: 230
- ENVIRONMENTAL SCIENCES: 227
- MULTIDISCIPLINARY SCIENCES: 207
- EDUCATION EDUCATIONAL RESEARCH: 165
- PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH: 164

# Publications

- INFORMATION SCIENCE LIBRARY SCIENCE: 2,231
- COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS: 1,102
- COMPUTER SCIENCE INFORMATION SYSTEMS: 540
- MEDICINE GENERAL INTERNAL: 307
- MANAGEMENT: 235
- SOCIAL SCIENCES INTERDISCIPLINARY: 230
- ENVIRONMENTAL SCIENCES: 227
- MULTIDISCIPLINARY SCIENCES: 207
- EDUCATION EDUCATIONAL RESEARCH: 165
- PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH: 164
“The Web of Science contains a remarkable treasure of data on scientific content, impact, and collaborations from 1900 to the present day on a global scale. Its comprehensive coverage has been an indispensable resource for the study of science, technology, and knowledge – enabling breakthroughs that would have been impossible without it, including not just new insights about the physical and social world but first facts about how the very practice of great science can be advanced.”

Brian Uzzi
Northwestern University
Richard L. Thomas Professor of Leadership and Organizational Change
Faculty Director, Kellogg Architectures of Collaboration Initiative
The nearly universal link between the age of past knowledge and tomorrow’s breakthroughs in science and technology: The hotspot

Satyam Mulherjee\textsuperscript{1,2}, Daniel M. Romero\textsuperscript{1,2,3,}, Ben Jones\textsuperscript{1,4} and Brian Uzzi\textsuperscript{1,2,7}

\textsuperscript{1}Northwestern University, Evanston, IL 60208, USA,
\textsuperscript{2}Northwestern Institute on Complex Systems and Data Science, Evanston, IL 60208, USA,
\textsuperscript{3}University of Michigan, Ann Arbor, MI 48109, USA,
\textsuperscript{4}National Bureau of Economic Research, Cambridge, MA.
*Corresponding author. Email: uzzi@northwestern.edu
+ See all authors and affiliations

Science Advances 19 Apr 2017,
Vol. 3, no. 4, a1601315.
DOI: 10.1126/sciadv.a1601315

MATERIALS AND METHODS

Data sources used in the analyses

Scientific papers database.

We examined all 28,436,345 research articles indexed in the Thomson Reuters WOS database that were published from 1945 to 2013. The sub-field designation of science and engineering (171 subfields), social sciences (54 subfields), and arts and humanities (27 subfields) was defined by the WOS and covers research publications in science and engineering since 1945, social sciences since 1956, and arts and humanities since 1975. These data are described in the Supplementary Materials and are available from Thomson Reuters.

Abstract

Scientists and inventors can draw on an ever-expanding literature for the building blocks of tomorrow’s ideas, yet little is known about how combinations of past work are related to future discoveries. Our analysis parameterizes the age distribution of a work’s references.
Do Nobel Laureates Create Prize-Winning Networks? An Analysis of Collaborative Research in Physiology or Medicine

Caroline S. Wagner, Edwin Horlings, Travis A. Whetzel, Pauline Mattsson, Katerina Nerdeqvist

1 Battelle Center for Science and Technology Policy, John Glenn College of Public Affairs, The Ohio State University, Columbus, Ohio 43210, United States of America. 2 Rathenau Institute, The Hague, Netherlands. 3 Department of Learning, Informatics, Management and Ethics (LIME), Tekniska högskolan 18 A, Karolinska Institutet, Stockholm, Sweden. 4 Nobel Museum, Stockholm, Sweden.

Abstract

Nobel Laureates in Physiology or Medicine who received the Prize between 1969 and 2011 are compared to a matched group of scientists to examine productivity, impact, coauthorship and international collaboration patterns embedded within research networks. After matching for research domain, h-index, and year of first of publication, we compare bibliometric statistics and network measures. We find that the Laureates produce fewer papers but with higher average citations. The Laureates also produce more sole-authored papers both before and after winning the Prize. The Laureates have a lower number of coauthors across their entire careers than the matched group, but are equally collaborative on average. Further, we find no differences in international collaboration patterns. The Laureates coauthor network reveals significant differences from the non-Laureate network. Laureates are more likely to build bridges across a network when measuring by average degree, density, modularity, and communities. Both the Laureate and non-Laureate networks have for all models a negative but the Laureates present much stronger negative links. More analysis is needed in understanding the differences in the research network structure.
Example 3

Goal: to capture the interplay of scientific collaboration and transport connectivity on a global scale

Abstract: This work in progress compares the strength of existing co-affiliations and collaborations with the number and capacity of air flights at the city level. We use a novel algorithmic approach that transforms distances to capture the interplay of collaboration and air traffic. Accompanying visualizations communicate research strengths and major bridges to international partners in support of data-driven decision making. More than 50,000 publications by IU faculty published between 2008 and 2016 are used to exemplify the method and to showcase the global reach of IU's research (collaborations).
Support a wide range of research projects

What can we learn about gender gap from WoS?
Sifan Zhou
Harvard University

Quantifying scientific success: when, why and how
Junming Huang
Northeastern University 14
Nov 2016

Correlating Air Transportation with Co-affiliation and Collaboration Data
Xiaoran Yan, Adam Ploszaj, Katy Börner
Open Science Forum, 19 October 2016

FORTHCOMING IN ROUTLEDGE IN 2017
The geography of scientific collaboration: theory, evidence and policy
Authors: Agnieszka Olechnicka, Adam Ploszaj, Dorota Celinska-Janowicz

Hot Hands in Science: Quantifying individual career dynamics
Lu Liu
with Y. Wang, C. Song, R. Sinatra, L. Giles, D. Wang
Sample of implemented research institutions

**University of Chicago**
- WoS XML now implemented in AWS enclave.
- Activity focused on extending reach to Metaknowledge network.
- Major NSF Grant secured

**University of Washington**
- Specific Licenses for R&D
- License agreement completed for ongoing use of WoS.

**Northwestern**
- Activity focused on establishing research agenda around entity disambiguation

**University of Mass Amherst**
- Data center access established & research started; Data Science Center started highlighting research from WOS dataset in early 2017

**Indiana University**
- Hosted Web of Science as Research Dataset.
- Variety of ongoing collaborations

**Penn State University**
- Exploring usefulness of tools and techniques developed for use with WoS.

**Swinburne University of Technology**
- Research with WoS XML at Centre for Transformative Innovation
A research management dataset
Web of Science Data used for strategy and outreach
Research Analytics
Advanced indicators

- InCites publications and citations indicators
  - Collaborations
  - Normalised indicators (Category normalised citation Impact – CNCI, percentiles)
  - Open Access
  - Journal Impact Factor
- Data used in rankings like Leiden Ranking (Including fractional counting of authorship)
Using XML to support collections and acquisitions decisions

Accessing XML:
• Through Big Ten Academic Alliance
• Processed XML into PostgreSQL database using open source parser code from Indiana University.
• Provide access to both raw XML and mediated access to PostgreSQL database for researchers.

Bibliometric analysis:
• Mathematics and Science are often studied together, but how different are they from a bibliometric standpoint?
• Analyzed citing and cited behavior in Mathematics, Physics, and Computer Science WoS Categories.
• Able to determine publication growth over time, median reference age for publication, median citation age per publication, and more

Conclusions for collection development:
• Mathematics research stays relevant.
• Can anything be weeded?
• Stay on top of old material for acquisitions.

Figure: Median Reference Age for Publication

Samuel Hansen
Mathematics & Statistics Librarian
Example 2

VIVO

VIVO is an open source web application and information model for representing scholarship and supporting research discovery.

“VIVO creates a connected, integrated record of the scholarly work of your institution, ready for reporting, visualization, and analysis.”

- Rich data model focusing on relationships and establishing links
- Shared across organizations
- Captures all phases of research
- Open network of information
VIVO Example: Fred Hutchinson Cancer Center
Converis with VIVO and implementation services
Example 3
2018 Research Fronts report
Chinese Academy of Sciences, Clarivate Analytics

100 hot Research Fronts and 38 emerging Research Fronts were identified based on co-citation analysis that generated 10,143 Research Fronts in the Clarivate Analytics database Essential Science Indicators (ESI).

• Web of Science core collection is unique by its ongoing selectivity
• Can be used anytime Research Output need to be analysed
• Can be used anytime Research Activity needs to be monitored
• Linking to other Clarivate resources or external information
Important links
Web of Science data and services

• Web of Science Data Integration: https://clarivate.com/webofsciencegroup/solutions/xml-and-apis/

• InCites: https://incites.clarivate.com/

• The Institute for Scientific Information: https://clarivate.com/webofsciencegroup/solutions/isi-institute-for-scientific-information/

• Web of Science Editorial selection process: https://clarivate.com/webofsciencegroup/solutions/editorial/

• Clarivate is an official registered service provider of VIVO: https://clarivate.com/webofsciencegroup/solutions/vivo/

• VIVO demo site: https://clarivatevivo.com/
Thank you

Contact your Local Account Manager

Benjamin Clark
Benjamin.clark@clarivate.com
+61401504223

Brian Syron
Brian.syron@clarivate.com
+61411653291