

Data sources and indicators to study science & innovation II

Introductory Data Science for Innovation (995N1)
Week 6 – 8th November 2021

Frédérique Bone



BUSINESS
SCHOOL

SCIENCE POLICY
RESEARCH UNIT

Lecture objectives

Data to study science and innovation:

After having acquired knowledge about how to manipulate and clean data to obtain a tidy dataset; and also introduced you to a wide range of data.

Today we are going to **focus on specific data to study science and innovation**. These have been traditionally used in innovation studies.

What data sources have been used?

Data sources

Understanding dynamics in science and technology

Publications and patents

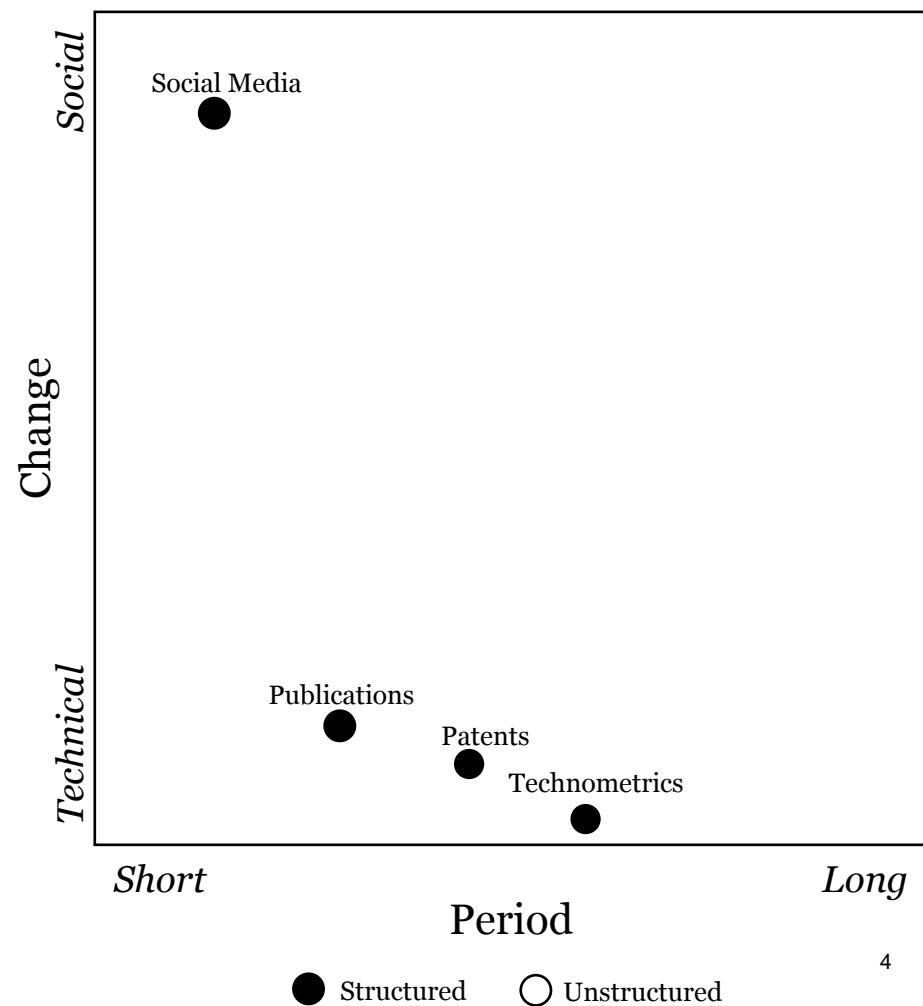
- Structured data
- Covering 30-50 years
- Focus on
 - Scientific knowledge
 - Technological artefacts

Technometrics

- Measuring technological level / quality of products.

Social media

- How science and technologies feature in social medias.



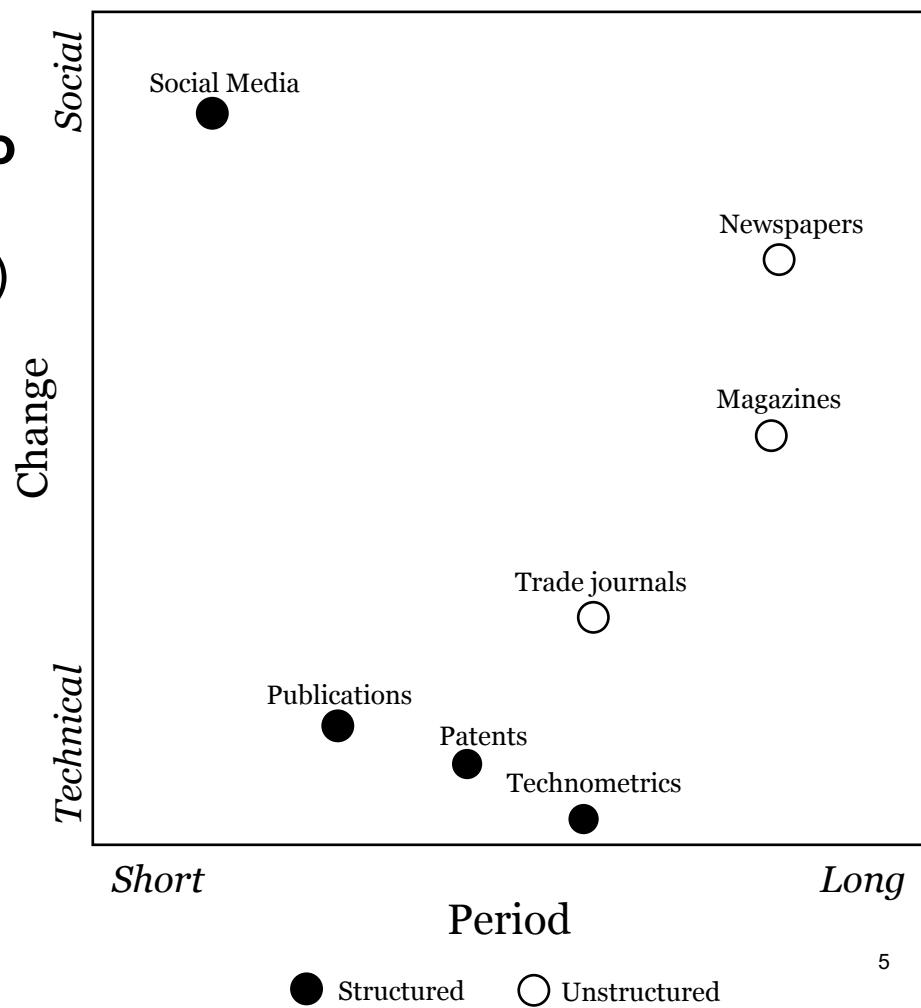
Data sources

Understanding dynamics in science and technology

Other documents have been used to study innovation:

- Historical documents (case studies)
 - Magazine
 - Newspapers
 - Trade journal

→ Full text data (Text mining now presents an opportunity)



Lecture objectives

Let's focus on the main sources:

1) Science

The use of scientific publications to study Science.

2) Technology and Innovation

The use of patents to study technology as artefacts.

How to study publications and patents?

The use of scholarly publication.



BUSINESS
SCHOOL

SCIENCE POLICY
RESEARCH UNIT

Measuring science

There is a whole field about measuring Science:

- **Scientometrics**: “the quantitative methods of the research on the development of science as an informational process”(Nalimov, 1971)
 - **Bibliometrics**: ‘the application of mathematics and statistical methods to books and other media of communications’ (Pritchard, 1969). Bibliometrics is mainly applied to written communications.
- Bibliometrics is not only focused on science, but many of its techniques are used in scientometrics
- **Altmetrics**: “the study and use of scholarly impact measures on activity in online tools and environments” (Priem, 2014). It looks mainly at impact in social networking tools.
- Many –trics in researchers’ bag..., you can also encounter informetrics or webometrics.

Measuring science

These have a few common advantages:

- Available in structured datasets
- Many variables can be used
- Free text data also starts to be explored

→ What data can be used?
(Many of these are relevant for both
patents and publications)



WO 2017/040797 A1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)
 (19) World Intellectual Property Organization International Bureau
 (43) International Publication Date 9 March 2017 (09.03.2017)



(10) International Publication Number

WO 2017/040797 A1

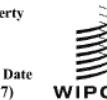
(51) International Patent Classification: C08K 5/17 (2006.01) C08K 5/18 (2006.01)
 (21) International Application Number: PCT/US2016/049928
 (22) International Filing Date:

(25) Filing Language:

(26) Publication Language:

(30) Priority Data:
 62/12,793 1 September
 62/251,379 5 November
 15/254,354 1 September

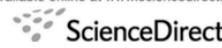
(71) Applicant: UNITED STA
 resented by THE ADMI
 TIONAL AERONAUTIC
 TRATION [US/US]; 300
 20546 (US).



(72) Inventors: WOHL,^a Christopher J., Jr.; 5831 Woodside Lane, Portsmouth, VA 23703 (US); SMITH,^a Joseph G., Jr.; 208 Smith Drive, Smithfield, VA 23430 (US).
 (74) Agents: EDWARDS, Robin, W. et al.; NASA Langley Research Center, Office of Chief Counsel, Mail Stop 30,



Available online at www.sciencedirect.com



Research Policy 36 (2007) 1107–1127



www.elsevier.com/locate/researchpolicy

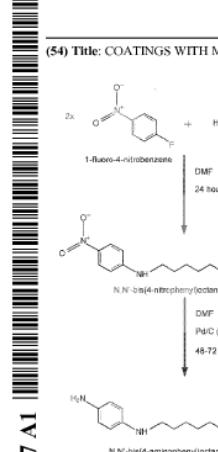


FIG. 1

Inventors and invention processes in Europe:
 Results from the PatVal-EU survey

Paola Giuri^{a,*}, Myriam Mariani^b, Stefano Brusoni^b, Gustavo Crespi^{c,1},
 Dominique Francoz^d, Alfonso Gambardella^e, Walter Garcia-Fontes^f, Aldo Geuna^c,
 Raul Gonzales^f, Dietmar Harhoff^g, Karin Hoisi^g, Christian Le Bas^h, Alessandra Luzzettiⁱ,
 Laura Magazzini^a, Lionel Nesta^{c,2}, Önder Nomaler^j, Neus Palomerasⁱ,
 Pari Patel^c, Marzia Romanello^k, Bart Verspagen^j

^a Scuola Superiore Sant'Anna, P.zza Martiri della Libertà 33, 56127 Pisa, Italy

^b IEP and CESPI, Bocconi University, Via Sarfatti 25, 20136 Milan, Italy

^c SPRU, The Freeman Centre, University of Sussex, Brighton, East Sussex BN1 9QE, United Kingdom

^d Ministère de l'Education Nationale et de la Recherche DPD-C3, 3-5, Boulevard Pasteur, 75015 Paris, France

^e IESEI and CESPI, Bocconi University, Via Filippetti 9, 20136 Milan, Italy

^f Departament d'Economia, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

^g Institute for Innovation Research, Technology Management and Entrepreneurship (INNO-tec), LMU Munich School of Management, University of Munich, Kaubachstraße 45, D-80539 Munich, Germany

^h LEFI, Université Lyon 2, 14 Avenue Berthelot, 69363 Lyon Cedex 07, France

ⁱ Department of Business Administration, Universidad Carlos III Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain

^j Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

^k University of Pisa, via Ridolfi 8, 56100 Pisa, Italy

Received 13 February 2006; received in revised form 28 February 2007; accepted 4 July 2007

Available online 11 September 2007

Abstract

Based on a survey of the inventors of 9017 European patented inventions, this paper provides new information about the characteristics of European inventors, the sources of their knowledge, the importance of formal and informal collaborations, the motivation to invent, and the actual use and economic value of the patents.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Patent; Inventor; Collaboration; Licensing; Invention process

1. Introduction

This paper provides new information, not available from other sources, on the characteristics of the invention processes in Europe, and on the economic use value of European patents. Our data are drawn from the survey (PatVal-EU, or PatVal for short) of 9017 patents granted by the European Patent Office (EPO) between 1993 and 1997, located in France, Germany, Italy,

* Corresponding author. Tel.: +39 050 883 359; fax: +39 050 883 344.

E-mail address: giuri@ssupi.it (P. Giuri).

¹ Present address: International Development Research Centre, 250 Albert Street, PO Box 8500, Ottawa, Ontario Canada K1G 3H9.

² Present address: OFCE, Department of Research on Innovation and Competition, 250, rue Albert Einstein, Valbonne, Sophia Antipolis 06560, France.

Publication data

What do the data look like?
What are the main databases?

Measuring science

What data can be used (e.g. bibliometric metadata)?

Metadata give information about other data (i.e. scientific papers or patents)

- Title of the paper
 - Abstract
 - Authors
- Author addresses
 - Journal
- Author keywords
- References



Inventors and invention processes in Europe: Results from the PatVal-EU survey

Paola Giuri^{a,*}, Myriam Mariani^b, Stefano Brusoni^b, Gustavo Crespi^{c,1},
 Dominique Francoz^d, Alfonso Gambardella^e, Walter Garcia-Fontes^f, Aldo Geuna^c,
 Raul Gonzales^f, Dietmar Harhoff^g, Karin Hoisl^g, Christian Le Bas^h, Alessandra Luzziⁱ,
 Laura Magazzini^a, Lionel Nesta^{c,2}, Önder Nomaler^j, Neus Palomerasⁱ,
 Pari Patel^c, Marzia Romanelli^k, Bart Verspagen^j

^a Scuola Superiore Sant'Anna, P.zza Martiri della Libertà 33, 56127 Pisa, Italy

^b IEP and CESPRI, Bocconi University, Via Sarfatti 25, 20136 Milan, Italy

^c SPRU, The Freeman Centre, University of Sussex, Brighton, East Sussex BN1 9QE, United Kingdom

^d Ministère de l'Education Nationale et de la Recherche DPD-C3, 3-5, Boulevard Pasteur, 75015 Paris, France

^e IEGI and CESPRI, Bocconi University, Via Filippetti 9, 20136 Milan, Italy

^f Departament d'Economia, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

^g Institute for Innovation Research, Technology Management and Entrepreneurship (INNO-tec), LMU Munich School of Management, University of Munich, Kaulbachstraße 45, D-80539 Munich, Germany

^h LEFI, Université Lyon 2, 14 Avenue Berthelot, 69363 Lyon Cedex 07, France

ⁱ Department of Business Administration, Universidad Carlos III Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain

^j Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

^k University of Pisa, via Ridolfi 8, 56100 Pisa, Italy

Received 13 February 2006; received in revised form 28 February 2007; accepted 4 July 2007

Available online 11 September 2007

Abstract

Based on a survey of the inventors of 9017 European patented inventions, this paper provides new information about the characteristics of European inventors, the sources of to invent, and the actual use and economic © 2007 Elsevier B.V. All rights reserved.

Keywords: Patent; Inventor; Collaboration; License

References

Alcacer, J., Gittleman, M., 2006. How do I know what you know? Patent examiners and the generation of patent citations. *Review of Economics and Statistics* 88, 774–779.

Allison, P.D., Stewart, J.A., 1974. Productivity differences among scientists: evidence for accumulative advantage. *American Sociological Review* 39, 596–606.

Almeida, P., Kogut, B., 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Science* 45, 905–917.

* Corresponding author. Tel.: +39 050 883 355.
fax: +39 050 883 344.

E-mail address: giuri@ssup.it (P. Giuri).

¹ Present address: International Development | Albert Street, PO Box 8500, Ottawa, Ontario Ca

² Present address: OFCE, Department of Research and Competition, 250, rue Albert Einstein, Valbonne 06560, France.

Measuring science

What data can be found in datasets?

Example Web of Science

- AU: Authors
- AF: Authors' full name
- TI: Title
- SO: Source (journal name)
- DT: Document type
- DE: Authors' keywords
- ID: keywords assign by WoS algorithms
- AB: Abstract
- C1: Authors' affiliations
- TC: Total citations
- PY: Publication year
- DI: DOI
- SC: Subject categories (put in place by WoS)

→ How are these datasets made?



| | |
|---|--|
| AU Giuri, P | |
| Mariani, M | |
| Brusoni, S | |
| Crespi, G | |
| Francoz, D | |
| Gambardella, A | |
| Garcia-Fontes, W | |
| Geuna, A | |
| Gonzales, R | |
| Harhoff, D | |
| Hoisl, K | |
| Le Bas, C | |
| Luzzi, A | |
| Magazzini, L | |
| Nesta, L | |
| Nomaleri, O | |
| Palomeras, N | |
| Patel, P | |
| Romanelli, M | |
| Verspagen, B | |
| AF Giuri, Paola | |
| Verspagen, Bart | |
| TI Inventors and invention processes in Europe: Results from the PatVal-EU survey | |
| SO RESEARCH POLICY | |
| LA English | |
| DT Article | |
| DE patent; inventor; collaboration; licensing; invention process | |
| ID RESEARCH-AND-DEVELOPMENT; PATENT CITATIONS; RESEARCH PRODUCTIVITY; PROPERTY-RIGHTS; TECHNOLOGY; SPILLOVERS; KNOWLEDGE; FIRMS; INNOVATION; MOBILITY | |
| AB Based on a survey of the inventors of 9017 European patented inventions, this paper provi | |
| C1 Scuola Super Sant Anna, I-56127 Pisa, Italy. | |
| Bocconi Univ, IEP, I-20136 Milan, Italy. | |
| Bocconi Univ, CESPRI, I-20136 Milan, Italy. | |
| Univ Sussex, Freeman Ctr, SPRU, Brighton BN1 9QE, E Sussex, England. | |
| Minist Educ Natl & Rech, F-75015 Paris, France. | |
| Univ Pompeu Fabra, Dept Econ, Barcelona, Spain. | |
| Univ Munich, LMU Munich Sch Management, Inst Innovat Res Technol Management & Enterpreneur | |
| Univ Lyon 2, LEFI, F-69636 Lyon 07, France. | |
| Univ Carlos III Madrid, Dept Business Adm, Madrid 28903, Spain. | |
| Eindhoven Univ Technol, NL-5600 MB Eindhoven, Netherlands. | |
| Univ Pisa, I-56100 Pisa, Italy. | |
| RP Giuri, P (reprint author), Scuola Super Sant Anna, P Za Martiri Libertia 33, I-56127 Pisa | |
| EM giuri@ssup.it | |
| RI Giuri, Paola/K-9986-2015; Geuna, Aldo/H-5049-2013; Garcia-Fontes, Walter/D-6958-2013 | |
| OI Giuri, Paola/0000-0002-5633-0848; Garcia-Fontes, Walter/0000-0002-6919-3985; PALOMERAS, NEUS/0000-0002-7871-3671; Verspagen, Bart/0000-0001-6975-8141; MAGAZZINI, LAURA/0000-0003-2482-1789; Geuna, Aldo/0000-0001-8587-840X | |
| CR Alcacer J, 2006, REV ECON STAT, V88, P774, DOI 10.1162/rest.88.4.774 | |
| ALLISON PD, 1974, AM SOCIOl REV, V39, P596, DOI 10.2307/2094424 | |
| Almeida P, 1999, MANAGE SCI, V45, P905, DOI 10.1287/mnsc.45.7.905 | |
| Anand BN, 2000, J IND ECON, V48, P103 | |
| Arora A, 2006, MANAGE SCI, V52, P293, DOI 10.1287/mnsc.1050.0437 | |
| Arora A, 2004, TND CORP CHANGE, V12, P451, DOI 10.1080/1062454041281192 | |

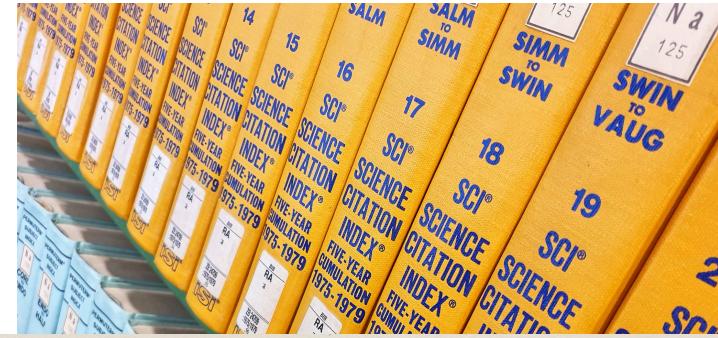
Bibliometric datasets

What do the data look like?
What are the main databases?

Bibliometric databases

Many databases used for bibliometric research today come from Indexing systems

- These databases of scientific publications include meta-data about scientific articles
- Historically linked individual publications to references.
- In the 1960s these were disciplinary at first
- The first interdisciplinary one was the Science Citation Index.



CITATION INDEX

The Citation Index of the SCI® is an alphabetical listing by author of all the references (cited items) found in footnotes and bibliographies of the cited items covered in the SCI. To search the Citation Index, look up the name of an author known to have published material relevant to the subject area of your interest. If any author's previously published works have been cited during the period covered by the SCI edition you are using, the item will appear in the Citation Index and the authors citing the material will be listed. Using the names of these citing authors you can then enter the Source Index for complete descriptions of their articles. (See sample Source Index entry below.)

Sample Display

| Citing Item | citing author | journal abbreviation | volume, page & year |
|--------------|-----------------------|----------------------|---------------------|
| ANSANELLI V. | BOZ M. | AM J ROENTG | 127 277 56 |
| ANSAR A. | MONATSHFETE CHEMIE | 166 111 | 1855 |
| ANSARI AH. | SEARCH CHIM ETAT SOLI | 162 | 1 |
| ANSARI AH. | METALLURGICAL CH | 52 | ACTA METALLURGICA |
| ANSARI AH. | HILLERT M. | 50 | ACTA METALLURGICA |
| ANSARI AH. | MAGNETONTEROL | 456 | |
| ANSARI AH. | ANDERSO A. | P NAS US | |
| ANSARI AH. | REDDI K.K. | 60 | 73 2368 |
| ANSARI AH. | WAYNE KS. | 58 | 42 173 61 |
| ANSARI AH. | OBSTET GYN | 55 | 114 15 63 |
| ANSARI AH. | COOK M. | TOH J EX ME | 120 43 63 |
| ANSARI AH. | PEZAT M. | 49 AM J OBSTET GYN | 103 51 |
| ANSARI AH. | VECCO T.J. | BIOCHEM | R 8 299 55 |
| ANSARI AH. | STRUVE FA. | J CELL BIO | 5 1 56 |
| ANSARI AH. | YODER M. | OBSTET GYN | 33 741 52 |
| ANSARI AH. | ANSELM F. | undated item | 3 322 58 |
| ANSARI AH. | **IN PRESS | CANTOR B. | 24 845 57 |
| ANSARI AH. | ANSSELIN F. | ACT METALL | |
| ANSARI AH. | 63 CR HEDOMAD SE ACAD | 256 | 2016 |
| ANSARI AH. | PEZAT M. | ISR J EARTH | 18 381 64 |
| ANSARI AH. | 63 T AM NUCL SOC | 20 | |
| ANSARI AH. | BLANCHARD P. | T AM NUCL S | M 23 151 64 |

SOURCE INDEX ENTRY

Source Index entry for article by Pezat M which makes reference to the 1963 paper by Anselin F.

| | | | |
|----------|---|------------------------------|-----|
| PEZAT M. | TANGUY B. VLASSE M. PORTIER J. HAGENMUL P.-R. | RARE EARTH NITRIDE FLUORIDES | 64 |
| | ISR J EARTH | 18(4):387-390 | 28R |

Codes Indicate Type of Source Item:

- Blank: articles, reports, technical papers, etc.
- C: corrections, errata, etc.
- D: discussions, editorials, etc.
- E: editorial, editorial-like items
- I: items about individuals (biographies, obituaries, etc.)
- K: chronological list of events in sequence.
- L: letters, communications, etc.
- M: abstracts from meetings, etc.
- R: reviews & bibliographies.

Consult the Source Index section of the SCI for bibliographic information on all citing items in the Citation Index.

Bibliometric databases

The ISI was subsequently bought by Thomson Reuters and then by Clarivate analytics – and called now the Web of Science.

There are also new companies providing databases about scientific publications and citations (among others: Scopus, Dimensions, or even google scholar)



Google Scholar



WEB OF SCIENCE™

Scopus®



Bibliometric databases

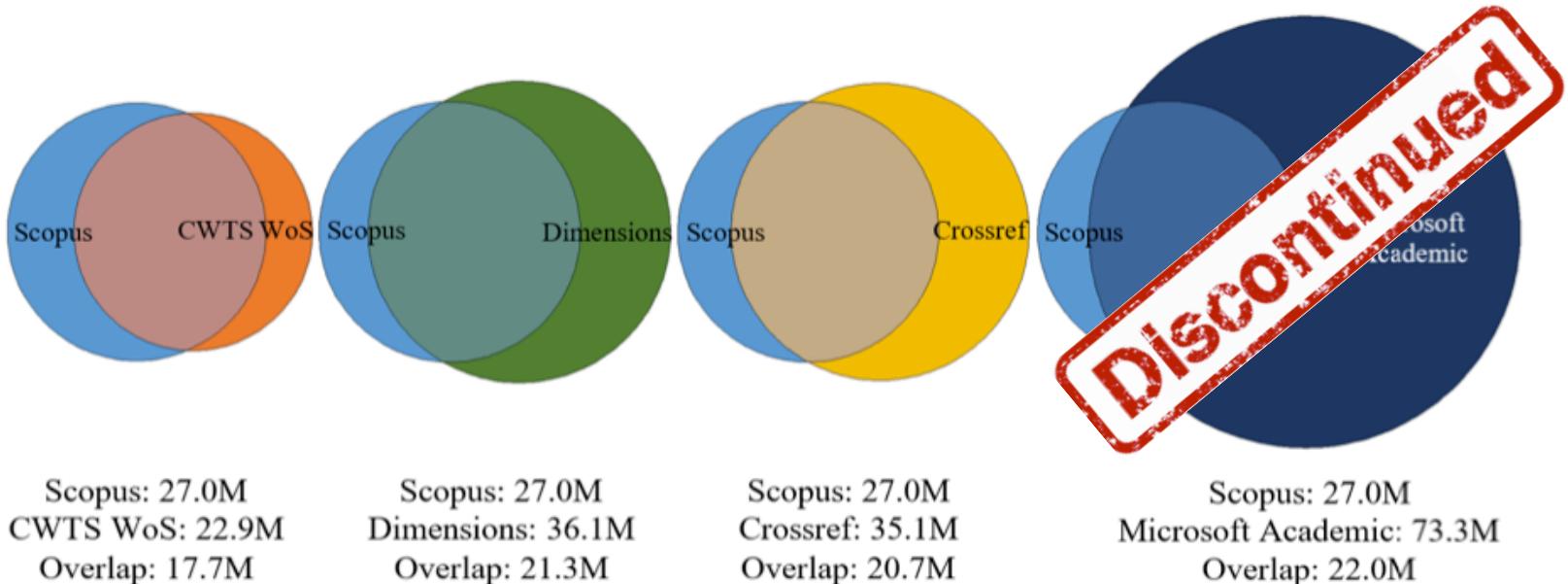


Figure 1. Overlap of documents between Scopus and the other data sources.

Comparison of articles indexed in each database between 1996-2017 (Visser et al. 2019)

Bibliometric databases have may not index the same articles

Bibliometric databases

These datasets have grown to include more and more fields:

1) To add more data:

- Add more journals over the years
- Funding information have been added in the late 2000s (e.g. 2008 for the Web of Science)

2) To include a set of indicators

(Journal Impact factor, Citation count, h-index ...)

- There is an increasing link between datasets, so publications in one dataset can be matched in others. This is done through unique identifiers such as the doi (digital object identifier), WOS id, Pubmed id are available in a range of dataset so you can combine fields from specific datasets.
- Other Ids are now introduced: author id (ORCID classification), discussion of grant id etc.

Bibliometric databases

How do I choose the right dataset?

Check the coverage:

- Language (WoS / Scopus ... for English literature)
- Disciplines (WoS for Science, Scopus for Social Sciences, Pubmed for medical Sciences, CAB for agriculture).
- Classification:
 - Pubmed (expert classification)
 - WoS categories
- Meta-data:
 - Authors (WoS, Scopus)
 - Citations (WoS, then scopus)
- Free to use? (Dimensions, Microsoft academic).

Web of Science Core Collection

BIOSIS Citation Index

BIOSIS Previews

Current Contents Connect

Data Citation Index

Derwent Innovations Index

KCI-Korean Journal Database

MEDLINE®

Russian Science Citation Index

SciELO Citation Index

Bibliometric databases

How to search for publication data:

Search in: **Web of Science Core Collection** ▾ Editions: All ▾

DOCUMENTS

AUTHORS

CITED REFERENCES

STRUCTURE

All Fields

Example: liver disease india singh

+ Add row

+ Add date range

Advanced Search

Clear

Search



Scopus

Search Sources Lists SciVal ↗



Start exploring

Discover the most reliable, relevant, up-to-date research. All in one place.

Documents

Authors

Affiliations

Search tips ↗

Search within

Article title, Abstract, Keywords

Search documents *

+ Add search field

Add date range

Advanced document search >

Search Q

Search History

Saved Searches

New

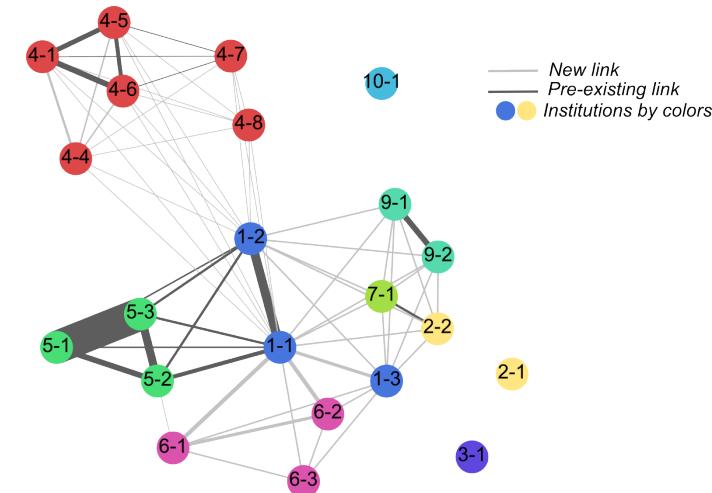
What indicators can we use?

First to understand how science works...

Scientometric indicators

Using publications to understand research collaborations:

- Use **authorship, addresses** to understand how individuals, organisations, cities, countries are linked.
- Looking at networks of collaborators for a technology, within a field, etc.



Co-publication network working on a tropical neglected disease.

Scientometric indicators

Using publications to understand research collaborations:

- Use **authorship, addresses** to understand how individuals, organisations, cities, countries are linked.
- Looking at networks of collaborators for a technology, within a field, etc.

Detecting relationships, implies that all entities are reported in a consistent manner:

- Authors name (e.g. A. Smith, M. Hopkins)
- Affiliation: some department are part of two organisations (e.g. university, CNRS)
- Addresses: part of addresses are reported

→ Think about disambiguating names (see side for tools)



Scientometric indicators

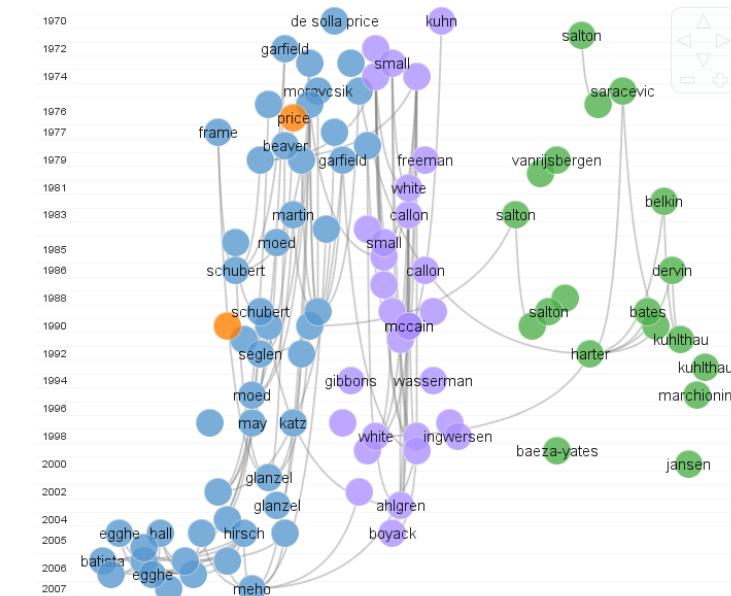
What can I use citations for?



Scientometric indicators

Direct citation analysis:

- Looking how a paper/journal/discipline build on previous knowledge
- Looking how a paper/journal/discipline is important to future knowledge



Scientometric indicators

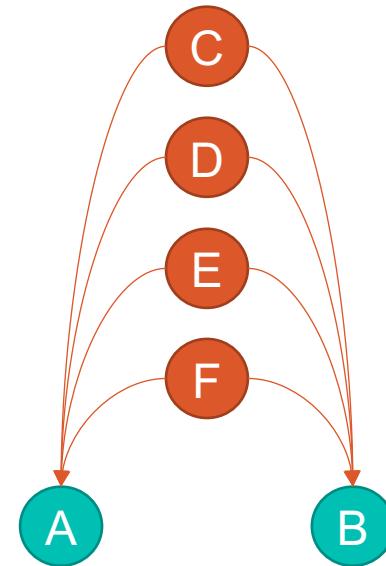
Direct citation analysis:

- Looking how a paper/journal/discipline build on previous knowledge
- Looking how a paper/journal/discipline is important to future knowledge

Relatedness of papers:

- Co-citation analysis (Small, 1973; Marshakova 1973)

Co-citation



The more articles are citing the same two works, the more these documents are related.

Scientometric indicators

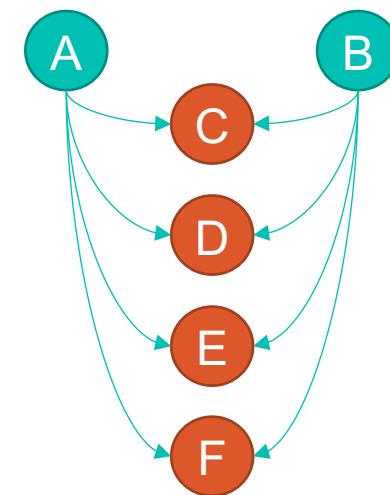
Direct citation analysis:

- Looking how a paper/journal/discipline build on previous knowledge
- Looking how a paper/journal/discipline is important to future knowledge

Relatedness of papers:

- Co-citation analysis (Small, 1973; Marshakova 1973)
- Bibliographic coupling (Kessler 1963)

Bibliographic coupling



Infer the strength of the relationship between two documents based on how many citations they share

Scientometric indicators

Direct citation analysis:

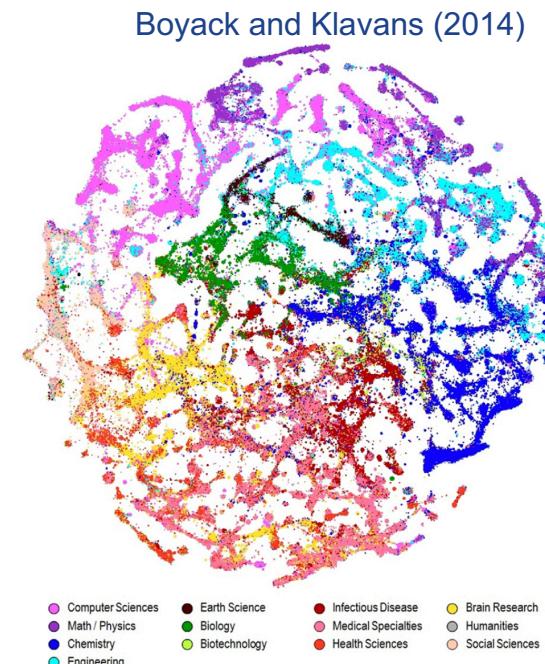
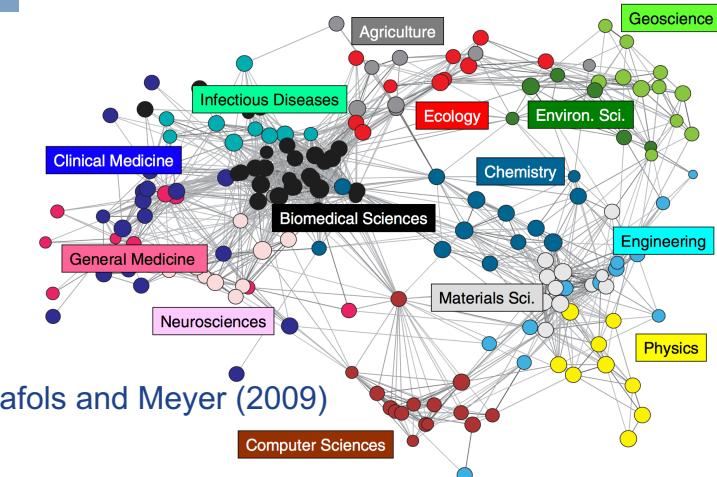
- Looking how a paper/journal/discipline build on previous knowledge
- Looking how a paper/journal/discipline is important to future knowledge

Relatedness of papers:

- Co-citation analysis (Small, 1973; Marshakova 1973)
- Bibliographic coupling (Kessler 1963)

Understanding the overall structure of science:

→ using the above techniques to map the structure of science.



Scientometric indicators

Direct citation analysis:

- Looking how a paper/journal/discipline build on previous knowledge
- Looking how a paper/journal/discipline is important to future knowledge

Relatedness of papers:

- Co-citation analysis (Small, 1973; Marshakova 1973)
- Bibliographic coupling (Kessler 1963)

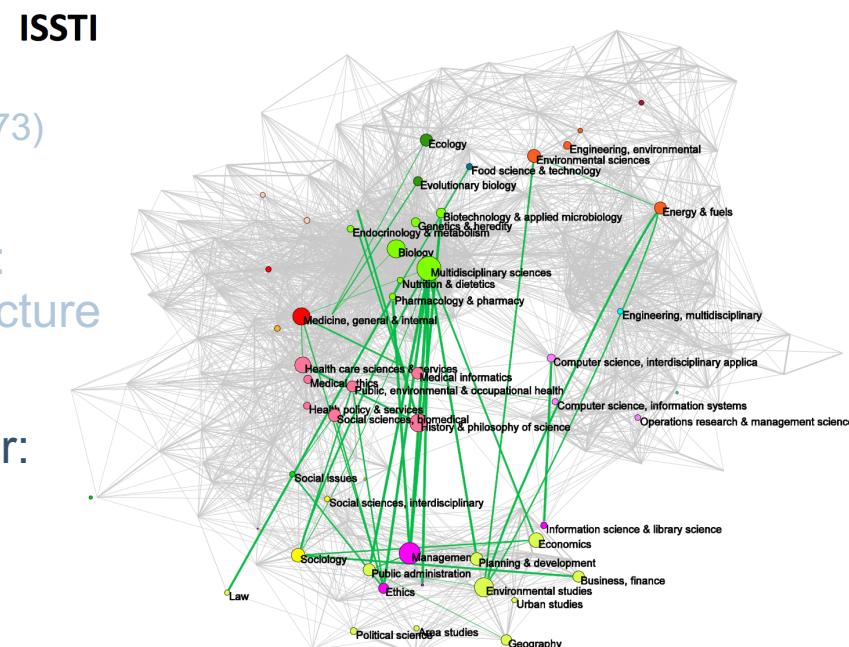
Understanding the overall structure of science:

→ using the above techniques to map the structure of science.

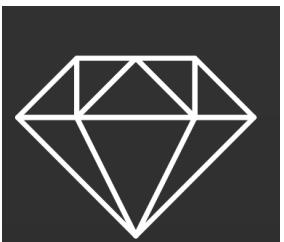
Understanding how communities work together:

(Rafols et al., 2012)

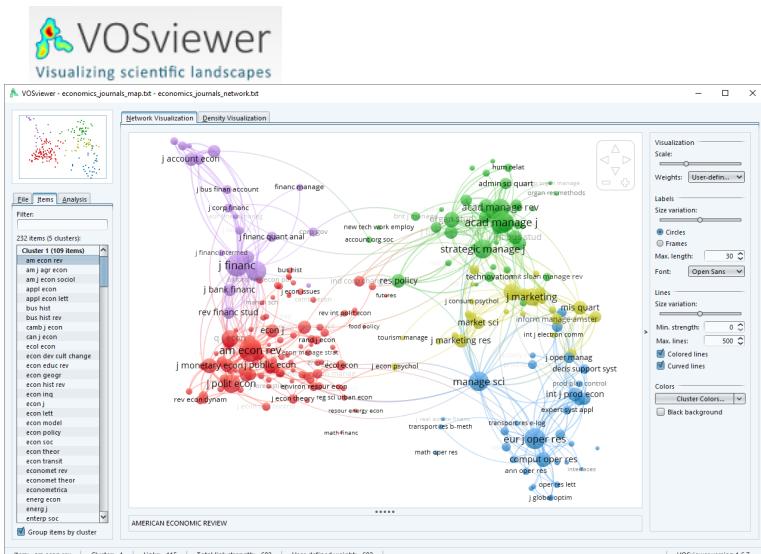
- Understanding interdisciplinarity



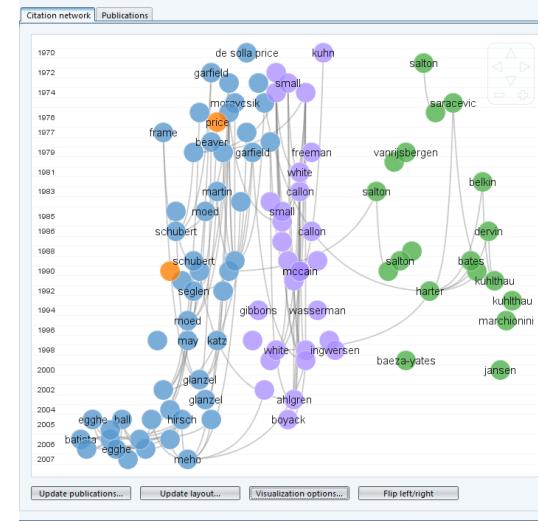
Number of tools that can help



OpenRefine



CitNetExplorer
Analyzing citation patterns in scientific literature



<https://www.citnetexplorer.nl/>

US
UNIVERSITY
OF SUSSEX

BUSINESS
SCHOOL

Let's Practice!



Let's code!

1. Load the dataset

```
#-----
```

- Set a directory
- load the file (this is specific to bibliometrix)
- Use ? to know how to input the data into the function

Let's code!

2. Descriptive bibliometrics analysis

#-----

- Don't forget to inspect the commands and all the graphs to answer the questions

Second to understand science's
impact...

Science's impact

There are two main ways to look at impact, first looking at the productivity (output count), and second, looking at how these outputs have been received (citations count).

Research unit considered: researcher, research group, research institution, country, journal ...

Productivity:

- nb of publications over a period of time
In a selected set of outlets (e.g. REF)
- nb of collaborators (type of collaborators may be important)
- nb of patents
- nb of spin-offs
- nb of mentions in social media (see altmetrics)
- Public engagements (dissemination of knowledge or awareness to a variety of actors / publics)



Altmetrics

There are new indicators in development that make use of online resources to give an indication of impact (on a shorter time scale).



“Thousands of conversations about scholarly content happen online every day. Altmetric tracks a range of sources to capture and collate this activity, helping you to monitor and report on the attention surrounding the work you care about.”
altmetric.com)

“Over the last two decades the way people use research and communicate about research has completely changed. People often learn about and find research through their social media network. They access research through a myriad of sites including open access repositories. [...] In early 2012, Plum Analytics was founded with the vision of bringing modern ways of measuring research impact to individuals and organizations that use and analyze research.”
plumanalytics.com)



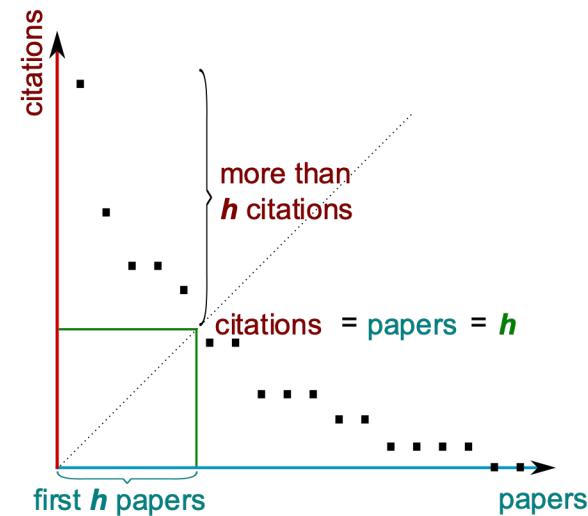
Science's impact

There are two main ways to look at impact, first looking at the productivity (output count), and second, looking at how these outputs have been received (citations count).

Research unit considered: researcher, research group, research institution, country, journal ...

Impact through citations:

- Total number of citations across publications
- Average number of citations per publication
- Number of highly cited publication (pre-set threshold)
- Proportion of highly cited publication (pre-set threshold)
- h-index (see side)
- Journal Impact factor: A journal's impact factor is based on 2 elements: the numerator, which is the number of citations in the current year to items published in the previous 2 years, and the denominator, which is the number of 'citable items' published in the same 2 years. (Garfield, 2006)



Comparability

The production of numbers using a unique recipe can give a sense that these indicators are a basis for comparison between research units. However different fields of study have different characteristics and should not directly be compared (Waltman, 2016).

Counting collaborators on publications:

- Depending on type of projects (e.g. megaprojects in particle physics)
- Increasing nb of authors per publications
- Should we allocate full impact to publications to all authors (i.e. full counting)?

Alternative counting methods:

- fractional counting: each author get a equal share of the credit institution or country sum the share of contributions of their researchers
- Weighted credits: position in author list, corresponding author...

→ The ordering of authorship vary by field, and their meaning is debated.

Comparability

The production of numbers using a unique recipe can give a sense that these indicators are a basis for comparison between research units. However different fields of study have different characteristics and should not directly be compared (Waltman, 2016).

The citation have different patterns depending on :

- Fields / discipline of study
- Years of publications
- Document types (reviews, editorial, letters, articles)

The citation score can be normalised on:

- average citation count
- determine whether a publication can be considered as highly cited within field (e.g. top 1%, 10%)
- Others exist...

→ Need to consider the classification system at hand to delimit the boundary of a field.

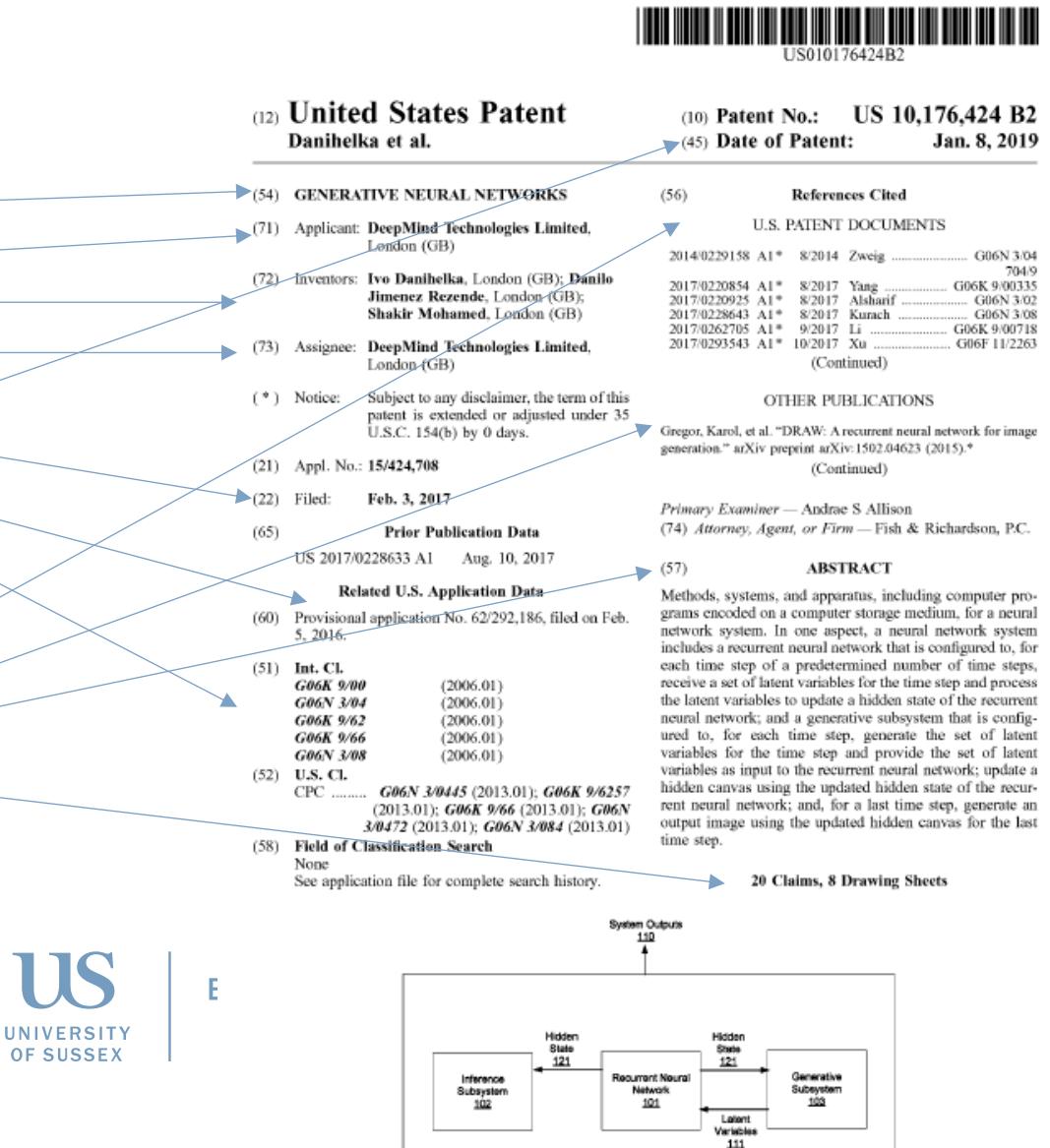
Patent data

What do the data look like?
What are the main databases?

Measuring science

What patent data can be used (e.g. bibliometric metadata)?

- Title
- Applicant(s)
- Inventor(s)
- Assignee
- Publication date
- Filing date
- Priority date
- Classification
- Patent references
- Other references
 - Abstract
 - Claims



Patent data

What data can be found in datasets? Example PatentLens

- Jurisdiction: Country where the patent is filed
- Kind: Classification
- Publication number
- Lens ID
- Publication Date
- Applicants
- Inventors
- ...

→ What are the main patent datasets?

Generative Neural Networks

Published: Jan 8, 2019 Earliest Priority: Feb 05 2016 Family: 6 Cited Works: 36 Cited by: 0 Cites: 10

Additional Info: 4 Cited Works Full text Published

Patent Summary Full-text Cites 36 Works Cites 10 patents Family Info Legal Info Collections Notes

Share Patent Create New Collection Make Note Download Citation

Abstract

Methods, systems, and apparatus, including computer programs encoded on a computer storage medium, for a neural network system. In one aspect, a neural network system includes a recurrent neural network that is configured to, for each time step of a predetermined number of time steps, receive a set of latent variables for the time step and process the latent variables to update a hidden state of the recurrent neural network; and a generative subsystem that is configured to, for each time step, generate the set of latent variables for the time step and provide the set of latent variables as input to the recurrent neural network; update a hidden canvas using the updated hidden state of the recurrent neural network; and, for a last time step, generate an output image using the updated hidden canvas for the last time step.

Claims

1. A neural network system implemented by one or more computers, the neural network system comprising:
 a recurrent neural network that is configured to, for each time step of a predetermined number of time steps, receive a set of latent variables for the time step and process the set of latent variables to update a hidden state of the recurrent neural network; and

Granted Patent

US 10176424 B2

176-584-942-766-320

Download PDF

Document Preview



| Jurisdiction | Kind | Publication Number | Lens ID | Publication Date | Pubcat | Application # | Application Date | Priority Num | Earliest Priority Title | Applicants | Inventors | Owners (US) URL | Type | Has Full | Cited by Pat | Simple Fam | Extended Fa | Sequence | Cc | CPC Classific | IPC Classific | US Classifica | NPL Citation | NPL Resolve | NPL Resolve | NPL Resolve | NPL Citations |
|--------------|------|--------------------|-------------|------------------|--------|---------------|------------------|--------------|-------------------------------------|----------------------------------|-------------------------------|-------------------------------|------|----------|--------------|---|---|----------|-------------|-------------------------|-----------------------------|---------------|--------------|-------------|-------------|-------------|---------------|
| GB | A | GB 2513105 A | 155-687-125 | 22/10/2014 | 2014 | GB 2013047 | 15/03/2013 | GB 2013047 | 15/03/2013 Signal Proce | DEEPMIND | CORNEBIESE JULIEN ROBEI | https://lens.Patent Appli no | 0 | 9 | 9 | 0 | G06N3/045; G06N3/04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| GB | D0 | GB 201304795 D0 | 141-660-366 | 01/05/2013 | 2013 | GB 2013047 | 15/03/2013 | GB 2013047 | 15/03/2013 Signal Proce | DEEPMIND TECHNOLOGIES LTD | https://lens.Patent Appli no | 0 | 9 | 9 | 0 | G06N3/0454; G06N3/084 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| EP | A1 | EP 3360087 A1 | 176-772-125 | 15/08/2018 | 2018 | EP 1680957 | 11/11/2016 | US 2015622 | 12/11/2013 Training Neu DEEPMIND | VAN HASSELT HADO PHIL | https://lens.Patent Appli no | 0 | 4 | 4 | 0 | G06N3/084; G06N3/08 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| EP | A1 | EP 3371747 A1 | 006-086-151 | 12/09/2018 | 2018 | EP 1682046 | 09/12/2016 | US 2015622 | 10/12/2015 Augmenting DEEPMIND | GRAVES ALEXANDER BEN | https://lens.Patent Appli no | 0 | 6 | 6 | 0 | G06N3/082; G06N3/04; G06N3/063 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| EP | A1 | EP 3398118 A1 | 143-509-973 | 07/11/2018 | 2018 | EP 1682347 | 30/12/2016 | US 2016622 | 04/02/2016 Associate I DEEPMIND | DANIELKA IVO; KALCHE | https://lens.Patent Appli no | 0 | 4 | 4 | 0 | G06N3/085; G06N3/04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| EP | A1 | EP 3398119 A1 | 151-084-541 | 07/11/2018 | 2018 | EP 1771181 | 03/02/2017 | US 2016622 | 05/02/2016 Generative I DEEPMIND | DANIELKA IVO; REZENDE | https://lens.Patent Appli no | 0 | 6 | 6 | 0 | G06N3/045; G06N3/04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| KR | A | KR 20180091850 A | 054-926-385 | 16/08/2018 | 2018 | KR 2018701 | 09/12/2016 | US 2015622 | 10/12/2015 T[...] & I[...] DEEPMIND | GRAVES ALEXANDER BEN | https://lens.Patent Appli no | 0 | 6 | 6 | 0 | G06N3/082; G06N3/04; G06N3/063; G06N3/063 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| CN | A | CN 108431832 A | 124-934-521 | 21/08/2018 | 2018 | CN 2016800 | 09/12/2016 | US 2015622 | 10/12/2015 Augmenting DEEPMIND | GRAVES ALEXANDER BEN | https://lens.Patent Appli no | 0 | 6 | 6 | 0 | G06N3/082; G06N3/04; G06N3/063 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| WO | A1 | WO 2018/064591 A | 053-404-726 | 05/04/2018 | 2018 | US 2017/009 | 29/09/2017 | US 2016624 | 30/09/2016 Generating I DEEPMIND | KALCHBRENNER NAL EMIL | https://lens.Patent Appli yes | 2 | 3 | 12 | 0 | G06N3/045; G06N3/04 | 5 | 1 | 073-060-212 | 10.1007/971 JUNHYUK OH | 0 | 0 | 0 | 0 | | | |
| US | A1 | US 2018/0330185 A | 035-678-581 | 15/11/2018 | 2018 | US 2018160 | 20/07/2018 | US 2015621 | 05/06/2018 Spatial Trans DEEPMIND | JADERBERG; GOOGLE LLC | https://lens.Patent Appli yes | 0 | 8 | 8 | 0 | G06K9/527; G06K9/52; G06K9/03; G06K9/04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| GB | D0 | GB 201212518 D0 | 135-221-311 | 29/08/2012 | 2012 | GB 2012125 | 13/07/2012 | GB 2012125 | 13/07/2012 Method And DEEPMIND | TECHNOLOGIES LTD | https://lens.Patent Appli no | 0 | 2 | 4 | 0 | G06F16/5838 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| US | A1 | US 2014/0019484 A | 033-427-691 | 16/01/2014 | 2014 | US 2012135 | 17/08/2012 | GB 2012125 | 13/07/2012 Method And DEEPMIND | COFFIN BEN COPPIN BEN GOOGLE LLC | https://lens.Patent Appli yes | 28 | 2 | 4 | 0 | G06F16/583 G06F17/30; 707/772; X0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| WO | A1 | WO 2014/009751 A | 189-778-991 | 16/01/2014 | 2014 | GB 2013051 | 12/07/2013 | GB 2012125 | 13/07/2012 Method And DEEPMIND | COPPIN BENJAMIN KENN | https://lens.Patent Appli yes | 1 | 2 | 4 | 0 | G06F16/532 G06F17/30 | 1 | 1 | 091-363-765 | 10.1109/vl. CARSON C ET | 0 | 0 | 0 | 0 | | | |
| US | A1 | US 2014/0019431 A | 125-928-191 | 16/01/2014 | 2014 | US 2013138 | 14/03/2013 | US 2013138 | 13/07/2012 Method And DEEPMIND | SULEYMANI | GOOGLE LLC | https://lens.Patent Appli yes | 16 | 2 | 4 | 0 | G06F16/532 G06F17/30; 707/706 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| US | B2 | US 10176424 B2 | 176-584-942 | 08/01/2019 | 2019 | US 2017154 | 03/02/2017 | US 2016622 | 05/02/2016 Generative I DEEPMIND | DANIELKA | GOOGLE LLC | https://lens.Granted Pat yes | 0 | 6 | 6 | 0 | G06N3/044; G06K9/00; G06K9/62; G06K9/04 | 36 | 7 | 053-094-537 | 10.1016/j.vl Gregor, Karol, | 0 | 0 | 0 | 0 | | |
| CN | A | CN 108701754 A | 115-321-154 | 29/10/2018 | 2018 | CN 2016800 | 11/11/2018 | US 2016622 | 05/02/2016 Generative I DEEPMIND | VAN HASSELT HADON PHIL | https://lens.Patent Appli no | 0 | 4 | 4 | 0 | G06N3/044; G06K9/00; G06K9/62; G06K9/04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | |

Patent databases

There is a mix of datasets which enable to search for patents, they vary in access, coverage of regions and time.

There is a mix of free and paying services, the ones shown here are mainly free, but all are accessible to you.

They vary in terms of the search that they offer, data fields provided.. check documentation on individual websites.

Derwent Innovations Index



Technology's impact

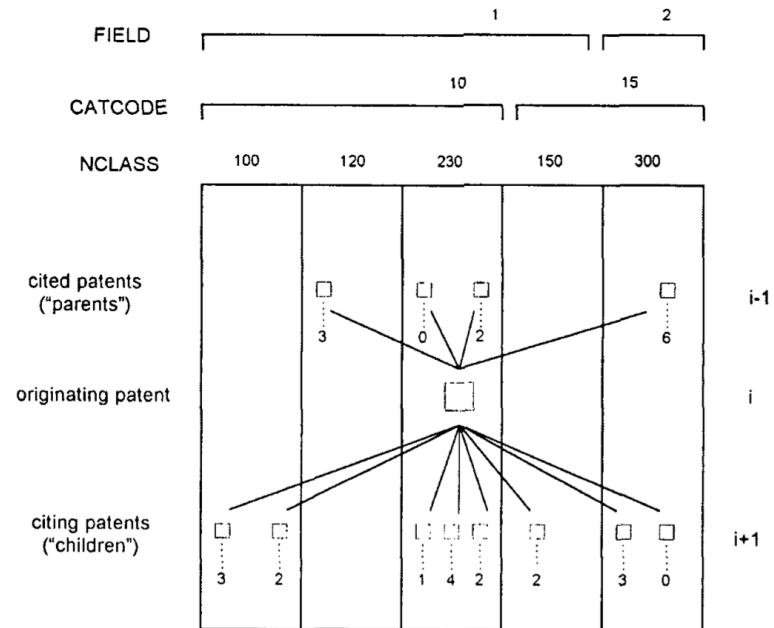
Like publications, patents have also two main way to look at their impact, looking at **patent count**, and **patent citations**.

Trajtenberg (1990)

- Simple Patent counts
- Weighted by citations count (to adjust for the value of an innovation)

Trajtenberg et al. (1997)

- Patent generality (see side)
- Patent originality (see side)
- Patent radicalness: where a patent is dissimilar in citation structures of past patents and concurrent patents, and citation structure will be replicated in the future. (Dahlin, 2005)



$$IMPORTF_i = 8 + 0.5 \cdot 17 = 16.5$$

$$IMPORTB_i = 4 + 0.5 \cdot 11 = 9.5$$

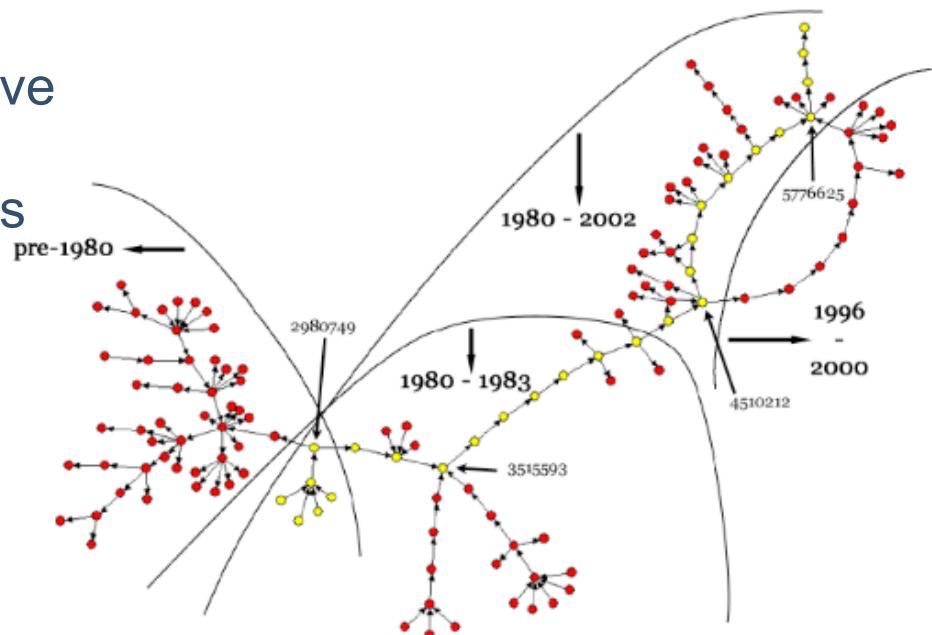
$$GENERAL = 1 - \left[\left(\frac{2}{8} \right)^2 + \left(\frac{3}{8} \right)^2 + \left(\frac{1}{8} \right)^2 + \left(\frac{2}{8} \right)^2 \right] = \frac{23}{32}$$

$$ORIGINAL = 1 - \left[\left(\frac{1}{4} \right)^2 + \left(\frac{2}{4} \right)^2 + \left(\frac{1}{4} \right)^2 \right] = \frac{5}{8}$$

Technology's trajectory

Understanding how technologies evolve
by Verspagen (2006):

- Looking at one specific patent class
'Fuel cells'
- Using citation network
- Using main path analysis



What do the data, and indicators represent?

Publication data

A publication is a piece of research, which aims at communicating to peers results of the research. It has to show that it is relevant to a field, and a specific journal it is published in.

This has an impact on the main contribution and on the references.



Available online at www.sciencedirect.com



Research Policy 36 (2007) 1107–1127



www.elsevier.com/locate/respol

Inventors and invention processes in Europe: Results from the PatVal-EU survey

Paola Giuri ^{a,*}, Myriam Mariani ^b, Stefano Brusoni ^b, Gustavo Crespi ^{c,1},
 Dominique Francoz ^d, Alfonso Gambardella ^e, Walter Garcia-Fontes ^f, Aldo Geuna ^c,
 Raul Gonzales ^f, Dietmar Harhoff ^g, Karin Hoisl ^g, Christian Le Bas ^h, Alessandra Luzzi ⁱ,
 Laura Magazzini ^a, Lionel Nesta ^{c,2}, Önder Nomaler ^j, Neus Palomeras ⁱ,
 Pari Patel ^c, Marzia Romanelli ^k, Bart Verspagen ^j

^a Scuola Superiore Sant'Anna, P. za Martiri della Libertà 33, 56127 Pisa, Italy

^b IEP and CESPRI, Bocconi University, Via Saffiotti 25, 20136 Milan, Italy

^c SPRU, The Freeman Centre, University of Sussex, Brighton, East Sussex BN1 9QE, United Kingdom

^d Ministère de l'Education Nationale et de la Recherche DPD-C3, 3-5, Boulevard Pasteur, 75015 Paris, France

^e IEGI and CESPRI, Bocconi University, Via Filippetti 9, 20136 Milan, Italy

^f Departament d'Economia, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

^g Institute for Innovation Research, Technology Management and Entrepreneurship (INNO-tec), LMU Munich School of Management, University of Munich, Kaufbeckerstraße 45, D-80539 Munich, Germany

^h LEFI, Université Lyon 2, 14 Avenue Berthelot, 69363 Lyon Cedex 07, France

ⁱ Department of Business Administration, Universidad Carlos III Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain

^j Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

^k University of Pisa, via Ridolfi 8, 56100 Pisa, Italy

Received 13 February 2006; received in revised form 28 February 2007; accepted 4 July 2007

Available online 11 September 2007

Abstract

Based on a survey of the inventors of 9017 European patented inventions, this paper provides new information about the characteristics of European inventors, the sources of their knowledge, the importance of formal and informal collaborations, the motivations to invent, and the actual use and economic value of the patents.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Patent; Inventor; Collaboration; Licensing; Invention process

1. Introduction

This paper provides new information, not available from other sources, on the characteristics of the invention processes in Europe, and on the economic use and value of European patents. Our data are drawn from a survey (PatVal-EU, or PatVal for short) of 9017 patents granted by the European Patent Office (EPO) between 1993 and 1997, located in France, Germany, Italy, the

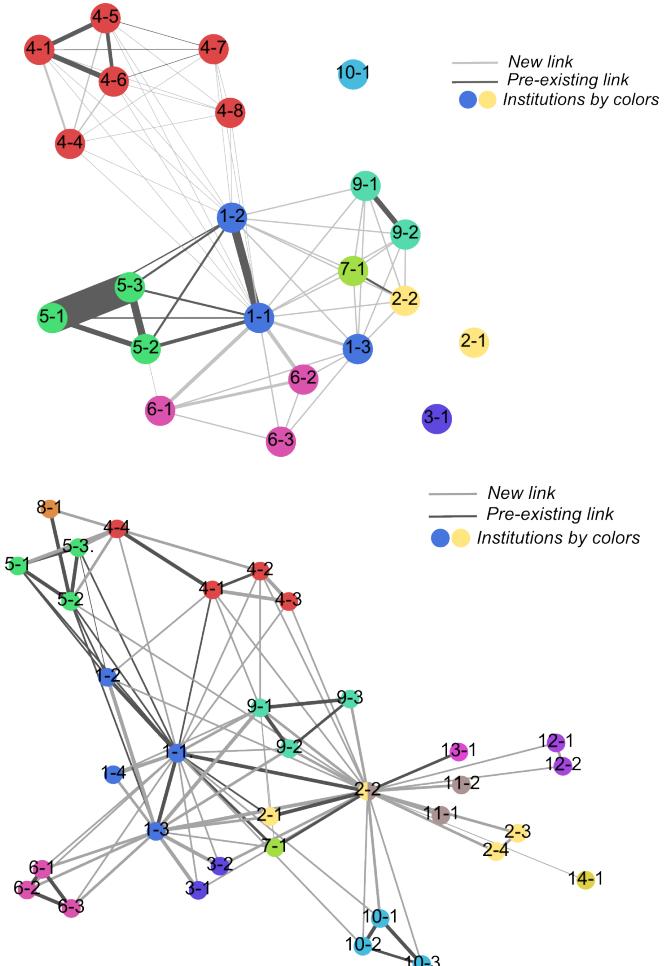
Publication data

A publication is a piece of research, which aims at communicating to peers results of the research. It has to show that it is relevant to a field, and a specific journal it is published in.

This has an impact on the main contribution and on the references.

Other considerations:

- Is co-publication a good proxy for collaborations? (e.g. study of neglected disease)
- Time lag between doing the research, submitting the paper and publication



Patent data

Patent has two main functions:

- to protect: exclusion of exploiting commercially an invention in a given country or region) up to 20 years.
- to disclose: patent application give public access to information about new technologies.

It contains both legal information and technical information...

Considerations on looking at patents:

- Time given on patents from patent filing, priority to granted date. There can be considerable time lag between these.
- Citation as knowledge flow: many citations can come from examiners and not applicants (depends also on the country of application).

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/463319/The_Patents_Guide_2nd_edition.pdf



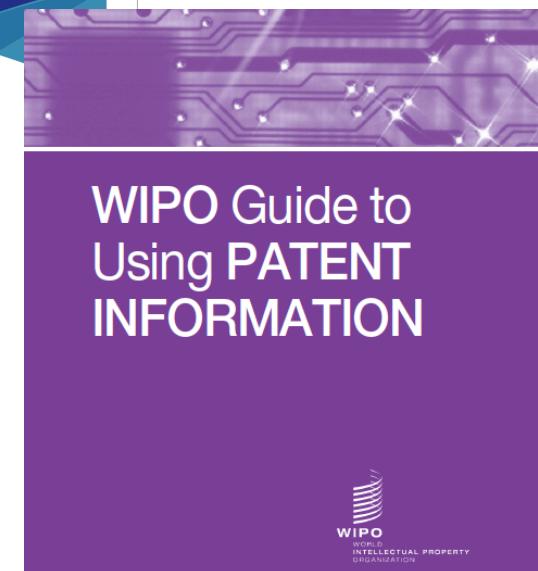
The Patent Guide

A handbook for analysing and interpreting patent data

Second Edition



Intellectual Property Office is an operating name of the Patent Office



https://www.wipo.int/edocs/pubdocs/en/wipo_pub_l434_3.pdf

Using indicators for evaluation:

The two sides of the coin.

Accountability and behaviour

Accountability metrics and behaviour:

- Once a metric is used to assess the performance it may affect the behaviour of the one assessed (individuals and organisations).

Changes can occur in terms of (Wouters, 2014; the metrics tide):

1. Goal displacement: high-score in the given measure becomes an end rather than a means.
 - Higher quantity of publications (salami slicing)
 - Selection of journals to publish in (high impact factor)

Accountability and behaviour

Accountability metrics and behaviour:

- Once a metric is used to assess the performance it may affect the behaviour of the one assessed (individuals and organisations).

Changes can occur in terms of (Wouters, 2014, the metrics tide):

2. Profound changes in the research process itself

- Financial rewards for publications in high impact journals
- Push to a focus on publications from universities
- Abandonment of particular type of work (book reviews, encyclopaedia articles for historians)
- Reduction of interdisciplinary work
- Transfer market for research with publications in highly ranked journals
- Citation clubs

Accountability and behaviour

Accountability metrics and behaviour:

- Once a metric is used to assess the performance it may affect the behaviour of the one assessed (individuals and organisations).

Changes can occur in terms of (Wouters, 2014, the metrics tide):

3. Journals can also pervert their performance:

- Coercive self-citation for journals to increase their impact factor (Faber Frandsen, 2007)

Limitations of the use of metrics

The call of the bibliometric community for a more responsible use of indicators!

COMMENT

SUSTAINABILITY Data needed to drive UN development goals p.432



CONSIDERATION Economics and environmental catastrophes p.434

BIBLIOGRAPHY Questions raised over proposed Anthropocene dates p.438

HISTORY Music Inspired Newton to add more colours to the rainbow p.439



The Leiden Manifesto for research metrics

Use these ten principles to guide research evaluation, urge Diana Hicks, Paul Wouters and colleagues.

Data are increasingly used to govern science. Research evaluations that were once bespoke and performed by peers are now routine and reliant on metrics. The problem is that evaluation is now led by the data rather than by judge-

advice on good practice and interpretation. Before 2000, there was the Science Citation Index on CD-ROM from the Institute for Scientific Information (ISI), used by experts for specialist analyses. In 2002, Thomson Reuters launched an integrated web platform,

were introduced, such as InCites (using the Web of Science) and SciVal (using Scopus), as well as software to analyse individual citation profiles using Google Scholar (Publish or Perish), released in 2007. In 2005, Jorge Hirsch, a physicist at the



[SIGN DORA](#) [READ THE DECLARATION](#) [SIGNERS](#) [BLOG](#) [GOOD](#)

San Francisco Declaration on Research Assessment

There is a pressing need to improve the ways in which the output of scientific research is evaluated by funding agencies, academic institutions, and other parties. To address this issue, a group of editors and publishers of scholarly journals met during the Annual Meeting of The American Society for Cell Biology (ASCB) in San Francisco, CA, on December 16, 2012. The group developed a set of recommendations, referred to as the San Francisco Declaration on Research Assessment. We invite interested parties across all scientific disciplines to indicate their support by adding their names to this Declaration.

The outputs from scientific research are many and varied, including: research articles reporting new knowledge, data, reagents, and software; intellectual property; and highly trained young scientists. Funding agencies, institutions that employ scientists, and scientists themselves all have a desire and need to assess the quality and

Limitations of the use of metrics

A call for an appropriate use of indicators:

Assessment of researcher's performance:

- PhD students and early career researcher are pushed to publish in high impact journals (JIF)
- They are pushed to acquire external funding before they are ready
- Problem of use of these indicators for funding attribution, promotion, hiring....

Quantitative metrics should not be used independently of qualitative or expert assessments.
Research need to be judged on its own merit and not from the journal it is published in.
Measure performance against missions.
Recognise locally relevant research
Be transparent on methodology used for any assessment ...

BUSINESS
SCHOOL



Row erupts over university's use of research metrics in job-cut decisions

Critics say redundancies are being decided using unreliable measures related to funding and citations, highlighting broader unease about the use of metrics in science.

Nature. News. 25 March 2021



Thank you.

Contact:

Dr Frédérique Bone
[\(f.bone@sussex.ac.uk\)](mailto:f.bone@sussex.ac.uk)

Research Fellow at SPRU



BUSINESS
SCHOOL

SCIENCE POLICY
RESEARCH UNIT