ISSI Newsletter

QUATERLY E-NEWSLETTER OF THE INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS
ISSN 1998-5460

#60 / VOLUME 15 NUMBER 4
DECEMBER 2019

CONTENTS

WORKSHOP REPORT

THE 2019 WORKSHOP ON OPEN SCIENTOMETRIC DATA INFRASTRUCTURES AT LEIDEN UNIVERSITY

The Open Scientometric Data Infrastructures Workshop took place at CWTS (Centre for Science and Technology Studies), Leiden University on 28 February and 1 March 2019. Over the course of two days, 14 researchers from CWTS and other research institutes and universities came together to discuss current projects and initiatives regarding open scientometric data infrastructures.

GRISCHA FRAUMANN

LUDO WALTMAN

DAY 1

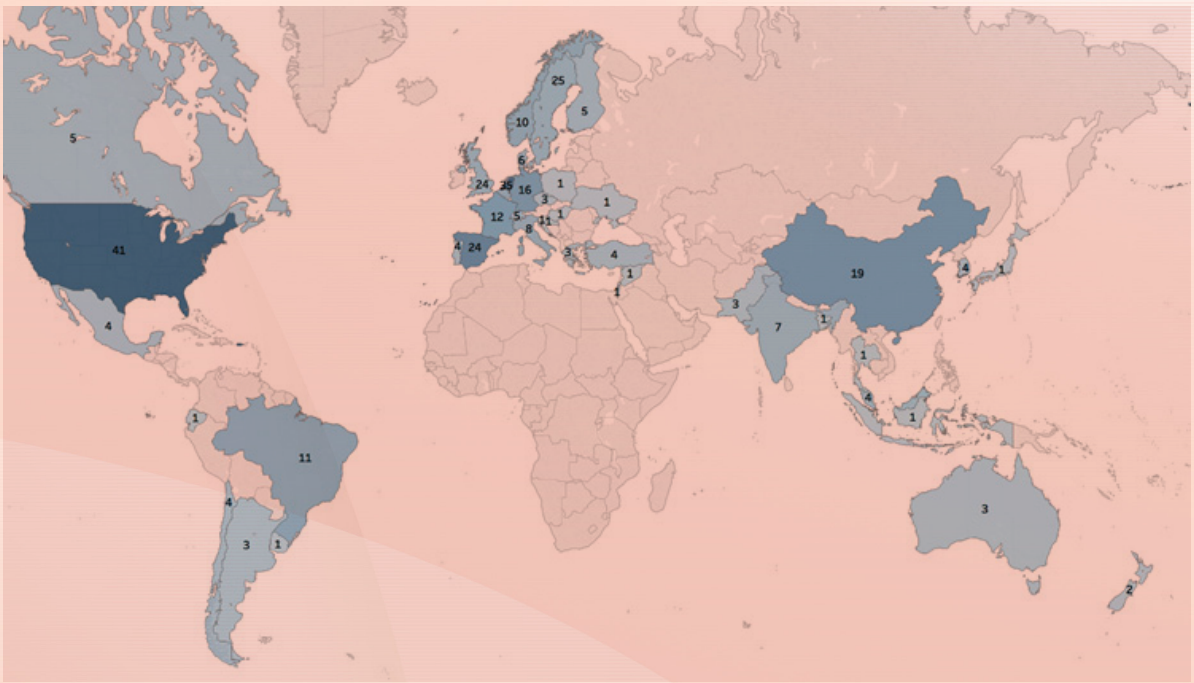
The workshop started with an **introduction** on the historical trajectory of open scientometric data by **Ludo Waltman** (CWTS). The OpenCitations project can be regarded as a starting point, which was followed by the Initiative for Open Citations (I4OC). Whereas OpenCitations provides a technical

ISSI e-Newsletter (ISSN 1998-5460) is published by ISSI (http://www.issi-society.org/).
Contributors to the newsletter should contact the editorial board by e-mail.

infrastructure for open citation data, I4OC is a lobby group for promoting openness of citation data. In December 2017, a number of scientometricians published an open citations letter to support I4OC. In September 2018, a Workshop on Open Citations was held at the University of Bologna. One of the reasons to establish the journal *Quantitative Science Studies (QSS)* in January 2019 was that Elsevier, the publisher of *Journal of Informetrics*, the predecessor of QSS, was unwilling to make citation data openly available. As of March 2019, almost 50% of the citation data in Crossref still needs to be opened. In addition, some publishers do not deposit any reference data in Crossref. Opening up citation data also benefits scientometric software, such as VOSviewer, which provides functionality to query Crossref for bibliometric visualisations.

Nees Jan van Eck (CWTS) and **Ludo Waltman** (CWTS) presented their work on **comparing bibliometric data sources**. The Dimensions database, established in 2018 by Digital Science, is mainly based on Crossref data but also benefits from additional data made available by publishers. Web of Science (WoS) and Scopus have the advantage of providing document types, while Crossref and Dimensions are unable to distinguish between different types of documents published in scientific journals. Comparisons of the different data sources also revealed differences in citation links. Crossref has fewer citation links than WoS and Scopus because of publishers that do not make citation data openly available. Dimensions enriches Crossref data through agreements with publishers and therefore provides more citation links than Crossref. Only a limited number of abstracts are indexed in Crossref. Even open access publishers do not always make abstracts available in Crossref. Other metadata elements, such as affiliations, are also often missing. There was broad support among the workshop participants for the idea that metadata of scientific publications should be made openly available.

Jochen Gläser (Technical University of Berlin) presented the **COPSSH (Communication Patterns in the Social Sciences and Humanities)** project (a summary in German is available here), which investigates communication patterns in the Social Sciences and Humanities (SSH). The project is part of the funding line quantitative research on the



Signatories of open citations letter, as of April 2018 (source: Sugimoto, Cassidy R.; Murray, Dakota S. and Larivière, Vincent (2018). Open citations to open science. Retrieved from http://issi-society.org/blog/posts/2018/april/open-citations-to-open-science/)

science sector of the German Federal Ministry of Education and Research (BMBF) and is carried out in collaboration with CWTS. The way in which SSH researchers publish research differs from Science, Technology, Engineering and Mathematics (STEM). Furthermore, SSH are not well represented in traditional bibliographic databases, and within SSH publications there is a higher proportion of negative citations compared to STEM publications. The main research question to be investigated in the project is: What can we learn from communication practices in SSH by overcoming the coverage problem and combining citation analysis with citation context analysis? The project will create manually a near-complete publication database which includes publication lists, citation databases or national databases, including citing and cited literature from Google Scholar. Art history and international relations in Germany will be compared to the Netherlands. Finally, the project includes interviews with researchers to validate the findings. COPSSH is a challenging project. For example, for some publications, there may be no PDF files available or these files may not have a clear structure (e.g. side notes instead of footnotes and endnotes in certain articles from art history). The project is currently testing citation grabbers and their machine learning capabilities. The research data is to be published as an open dataset including citation context analysis.

David Shotton (University of Oxford) and **Silvio Peroni** (University of Bologna) are the Directors of **OpenCitations**. They presented recent developments regarding OpenCitations. Open Citation Identifier (OCI) equals DOI for citations. The EU-funded FREYA project recognizes OCIs as persistent identifiers for citations, and the identifiers are used for citations in Wikipedia articles with data from Wikidata, and for open DOI-to-DOI citations defined by open references in Crossref. This has made it possible to publish COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. The COCI Index employs live calls to the Crossref Application Programming Interface (API) to pull publication metadata not stored in the Index. WOCI, the OpenCitations Index of Wikidata citations, will be published soon. In addition, CROCI, the Crowdsourced Open Citations Index, has been released. The community is able to submit citation data to CROCI, but CROCI has yet to develop significant content following the call for crowdsourcing data to CROCI made in February 2019. All three of these indexes are accessible through a unified OpenCitations API.

In the future, OpenCitations hopes to harvest references extracted from the arXiv corpus as part of the EXCITE project that is carried out at GESIS – Leibniz Institute for the Social Sciences and WeST – Institute for Web Science and Technologies. Furthermore, a collaborative initiative with OpenAIRE is planned. During the workshop discussion, it was mentioned that it is almost impossible for one organisation to host all the information, and that data needs to be enriched from other databases, for example via live API calls, or by database federation.

Following the discussion of the developments relating to OpenCitations, **Silvio Peroni** presented the **Open Biomedical Citations in Context Corpus** funded by the Wellcome Trust as part of the Open Research Fund programme, which started in July 2019. The project is about harvesting the textual context of individual in-text reference pointers in the full text of publications in the biomedical literature. The data will be derived from an open access subset of Europe PMC (Pubmed Central) by using the EPMC API to harvest XML (Extensible Markup Language) documents. Europe PMC is a comprehensive database of life sciences and biomedical research. Finally, the project will also provide a description of the ingestion workflow.

Gianmarco Spinaci (University of Bologna) presented an **ongoing research project** that aims to **analyse Arts and Humanities (A&H) publications in major bibliographic databases**, such as WoS, Scopus, Crossref, Dimensions and Microsoft Academic Graph. One of the goals is to iden-

tify, count and cluster A&H publications in the different databases. Furthermore, all A&H fields of studies were retrieved from Microsoft Academic Graph. The clustering will be visualised with VOSviewer. One preliminary result is the relatively small amount of books and book series within WoS. Potential further use cases are still being explored.

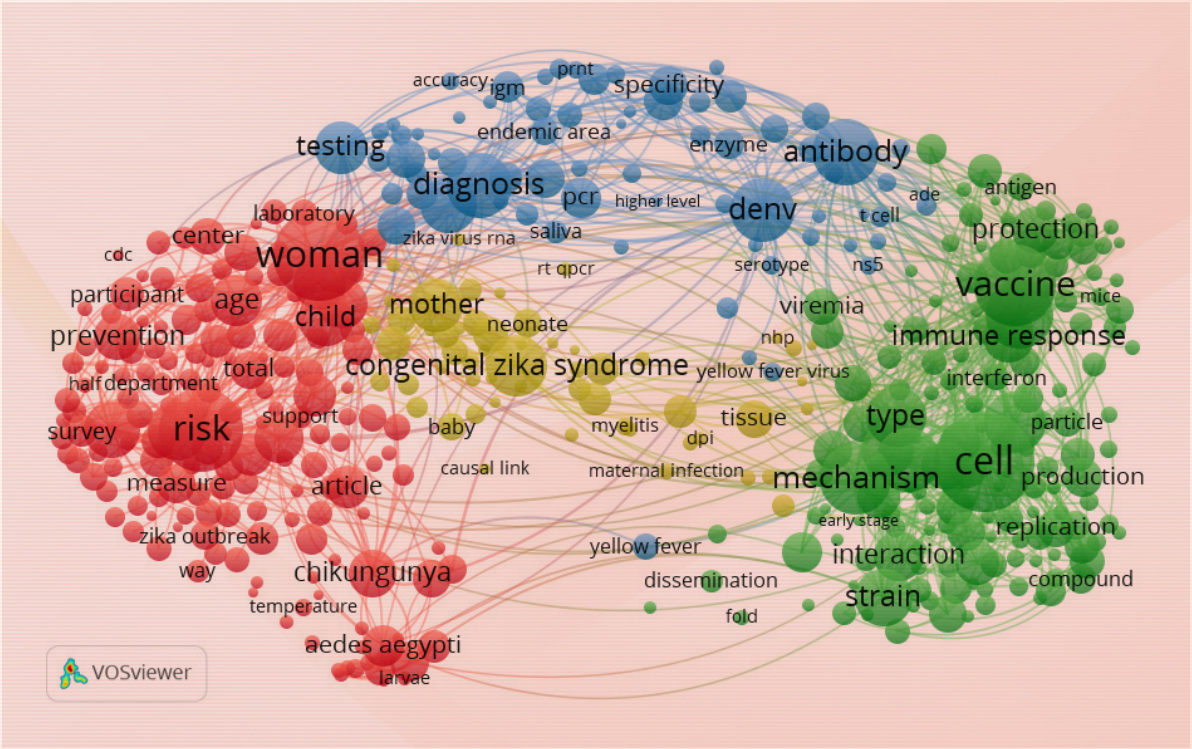
Giovanni Colavizza (University of Amsterdam and CWTS) presented the **Scholar Index**, which includes a citation index for the Arts and Humanities from the Arts and Humanities. Information retrieval in A&H is challenging. Scholar Index is to be integrated in the OpenCitations corpus as well as to Europeana, which "provides access to over 50 million digitised items – books, music, artworks and more". Currently, a prototype is being developed to improve information retrieval by connecting several systems. The prototype is focused on the History of Venice, and it is also possible to add tools. The coordinators of Scholar Index are currently looking for pilot partners around Europe (e.g. libraries and archives).

Thomas Franssen (CWTS) presented an overview on the **RISIS2 project** funded by Horizon 2020, which is the follow-up of the recently concluded RISIS (Research Infrastructure for Science and Innovation Policy Studies) project. Compared to the first RISIS project, the frontend (e.g. core facility) is to be developed further. Several research infrastructures are available to researchers as part of the RISIS2 project, and the consortium partners include various research institutes from all over Europe.

DAY 2

The second day started with a presentation by **Rodrigo Costas** (CWTS) on the use of **Mendeley** readership statistics to develop an **open classification scheme for Crossref**. Crossref has certain limitations that this project aims to tackle, such as lack of metadata on affiliations, funding acknowledgements and particularly a homogeneous classification for journal publications. This lacking of metadata hinders research efforts based on Crossref data, including for example the monitoring of the disciplinary uptake of open citations. Mendeley is a reference manager and an academic social network by Elsevier that provides free access to its data for research purposes (which can be freely queried using the Mendeley API). For the study, a free global classification of Crossref based on all available DOIs in Crossref was carried out. This classification is based on the 28 academic fields as defined by Mendeley. Mendeley users classify themselves in these subject areas when they create their profile on Mendeley. The main idea of the project is to classify Crossref publications in the field(s) of the Mendeley users that are saving them in their individual libraries. Thus, it is possible to develop a sort of 'crowdsourced' classification of Crossref publications, independently from their indexing in other databases (e.g. Scopus or WoS). The study first investigated journal classifications, leaving classifications of publications as a next step. The potential of the Mendeley dataset is to provide a global free classification for all Crossref publications. During the workshop discussion it was suggested to explore the open reference manager Zotero as an alternative, while Microsoft Academic Graph could also be tested to develop open classifications of publications.

Grischa Fraumann (TIB Leibniz Information Centre for Science and Technology) provided a summary on the **ROSI (Reference Implementation for Open Scientometric Indicators)** project which is carried out at the TIB Leibniz Information Centre for Science and Technology. The project aims to develop a prototype that visualises open scientometric indicators, for example in an online dashboard. This prototype will be tested with researchers in interviews and workshops. The ROSI project is also part of the funding line quantitative research on the science sector of the BMBF. The workshop discussion focused on the



Example of a visualisation created by VOSviewer for the query 'zika' via the Europe PMC API

public registry of data sources as part of the project, and on the sustainability of research infrastructures.

Nees Jan van Eck, Ludo Waltman, David Shotton and **Silvio Peroni** discussed recent developments regarding the **VOSviewer** software and open data sources that can be queried via APIs. Historically, VOSviewer supported WoS and Scopus, which both require a subscription. More recent data sources include Dimensions, Crossref, Wikidata and the OpenCitations Corpus (OCC). Dimensions has a limited edition that is freely accessible. Crossref is open but about half of its citation links are closed. OCC is open but currently provides only a very limited coverage of the scientific literature. Crossref, OCC, Europe PMC, Wikidata, and Semantic Scholar, a data source created by the Allen Institute for Artificial Intelligence, can all be queried via APIs. Support for these APIs was recently added to VOSviewer. Compared to downloads from WoS or Scopus, working with APIs is more convenient. However, the APIs of the different data sources all have limitations, for instance in their speed and flexibility. In the end, the discussion focused on the best way for VOSviewer and other scientometric tools to support data sources via APIs. Ideas were developed for improved interoperability between scientometric data sources and scientometric tools.

PUBLIC PRESENTATIONS AND OUTLOOK

The workshop concluded with short presentations of the above-mentioned initiatives and projects. These presentations were open to the public and a significant number of researchers from CWTS and other research institutes attended. The discussions led to useful feedback to develop the projects and initiatives further.

The workshop provided a summary on recent developments in open scientometric data infrastructures from some selected initiatives and projects. It will be interesting to see the next steps in these important developments. For example, at the ISSI Conference 2019 at Sapienza University Rome in September 2019, a workshop on Open Citations: Opportunities and Ongoing Developments was organised.

NEW BOOK ON RANKING BY PÉTER ÉRDI

Péter Érdi: Ranking – The Unwritten Rules of the Social Game We All Play.
Oxford University Press, New York, 2019
256 pages, ISBN-13: 9780190935467. £22.99

https://global.oup.com/academic/product/ranking-9780190935467

The main title of the book "Ranking" is a buzzword hardly requiring detailed explanation. It is, nevertheless, supplemented with a subtitle that is certainly flashy and alluring, but an earlier manuscript version had a different one maybe more relevant to the content of the book: "The reality, illusion and manipulation of objectivity."

A book review has been published recently [Schubert, 2020]; some excerpts may serve as teaser for the book.

"It is an informative and amusing book that can be recommended for everybody except for those who expect some readily usable recipes for ranking exercises. The author collected a treasury of stories and reflections connected with comparison, rating and ranking from the widest possible area of sports, arts, sciences, politics, media and shopping, just to mention a few. The book's main concern is not how to rank, but rather how and in what extent ranking can be avoided."

"The chapter Ranking games [...] highlights a topic of great interest for scientometrists: university ranking. Tracing back such efforts as early as 1863, the author takes account of the best-known ranking systems (ARWU, THE, QS), and depicts a wide panorama of views and opinions of those affected on either side of the evaluation process, as well as of neutral analysts."

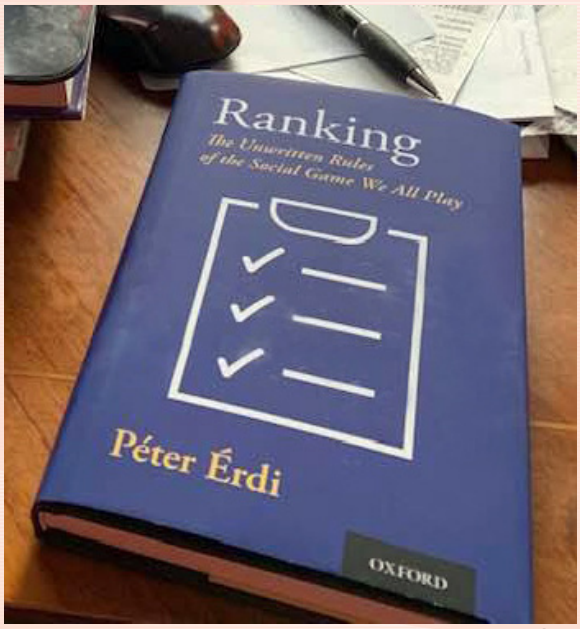


Photo copyright: © Péter Érdi

"The author emphasizes that the objectivity of ranking is often illusory and subject to manipulation. Yet, most rankings, whether based on human opinion or computer algorithms, are far for being random, therefore they capture some element of reality."

András Schubert
Hungarian Academy of Sciences, Library & Information Center, Budapest, Hungary

REFERENCES

Schubert, András (2020)
DOI:10.1007/s11192-019-03335-1
(to be published in Scientometrics, Vol. 122, No. 3)

25th NORDIC WORKSHOP ON BIBLIOMETRICS AND RESEARCH POLICY 2020

NIFU & OSLOMET
15—16 OCTOBER 2020

Oslo. Photo courtesy of © Balázs Schlemmer

The 25th Nordic Workshop on Bibliometrics and Research Policy is organised by the Nordic Institute for Studies in Innovation, Research and Education (NIFU) and Oslo Metropolitan University (OsloMet).

LOCATION & DATE

The event is going to take place in Oslo, Norway on **15—16 October 2020**.

WORKSHOP DETAILS

The objective of the workshop is to present recent bibliometric research in the Nordic countries. Furthermore to create better links between the bibiometric research groups and the research policies.

The workshop language is English and the workshop is open to participants from any nation. More information about the workshop will be announced in early May at www.nifu.no. Some of the workshop themes: Research strategies • Open acces • Databases • Evaluation of reserach • Research policy • International collaboraion.

IMPORTANT DATE

Deadline for submission of abstract will be **14 August 2020**.

MORE INFO

http://www.nifu.no/news/nwb2020/

COUNTING THE UNCOUNTABLE
ANECDOTES FROM THE TECHNOLOGICAL HISTORY OF SCIENTOMETRICS

GUNNAR SIVERTSEN
Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway (NIFU)

INTRODUCTION

This contribution to the ISSI Newsletter has two stories. One of them will entertain you with anecdotes from the technological history of scientometrics. The other story will explain my main title. The stories will sometimes interrupt each other. The chronology and my styles of writing will be shifting as well, but I will lead you as we go.

WORD PERFECT

The year is 1993. Five years have passed since I was asked to start up scientometrics in Norway at my research institute in Oslo. We are still using *Word Perfect* for computer-based writing. Microsoft *Word* has not taken over yet, but a new revolutionary printer has been installed, a *laser* printer! For the first time in history, we can select the typography ourselves:

Times New Roman!
Verdana!
Century Gothic!

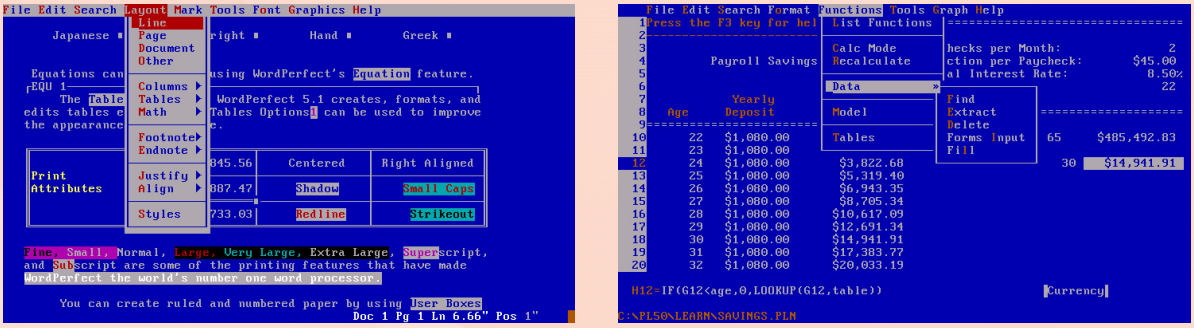
Still, most of us continue to use the `Courier` font, which is standard in *Word Perfect* and looks like all the typewritten documents we produced before we got computers in the eighties. In 1993, we still feel more at home with the `Courier` font.

Typewriters were a very efficient technology before the slow computers took over. We did not need

to go to the printer in the next room at that time, and there was no queue. Pressing one key on the keyboard caused the character to be printed on the paper instantly because a ribbon with dried ink was struck against the paper.

The Courier font was developed by IBM for their typewriters in 1956. It was designed to give each letter the same horizontal space for each touch on the keyboard, e.g. with the use of serif at the top and bottom of the 'l', as in the first word in the example below. The types were slender and mono-

The most widespread version (ver 5.1) of the 'WordPerfect' was released in 1988. Its splash screen proudly announces its cutting-edge formatting features that "have made WordPerfect the world's number one word processor". Surprisingly enough equation editing (see the uppermost row) was already supported by this early version. Screenshot courtesy of © WinWorldPC.com.



Left: WordPerfect 6.x (released in 1993) was the last version programmed for DOS. Right: PlanPerfect 5.x. It was the WordPerfect's spreadsheet managing counterpart. Version 3.0 had not even had a menu system – all commands had to be executed via function keys. According to the WinWorld software preservation portal's somewhat malicious remark on the user interface's usability, "If you enjoy hitting yourself over the head with a hammer, you might also enjoy PlanPerfect". On the positive side, the installation files of the whole WordPerfect—PlanPerfect bundle occupied less than 7 megabytes altogether. Screenshots courtesy of © WinWorldPC.com.

tone in weight to make a clear impression as the character was hammered into several sheets of paper with blueprints in between to provide instant copies. Here is my example of the Courier font, which starts the second story. We are in Oslo in January 1993:

> It is dark outside, mid-winter cold and freezing without snow. The window at the end of the corridor mirrors an insecure person in a strange place: I am walking towards myself in the cold neon light. Behind half-open doors in the side corridors, low voices are heard, but I see nobody – or maybe I try to avoid it. The visiting hours are gone. Patients are left alone with their cancer. I'm trying to find an office, searching for a name on a door, but am I in the wrong corridor?

I knew this hospital from before. *Ultrastructural Pathology* is the title of a scientific journal in the field of cancer research. It was launched by the Scandinavian University Press (SUP) in 1980. Five years later, it was still running with deficits. As the Editorial Director of the Scientific Journals Department at SUP, I therefore handed the journal over for free to Hemisphere Publishing, Inc. in New York. Taylor & Francis acquired it later, now making profits from much higher subscription prices than we had dared to set. In 1993, the editorial address of the journal was still in Oslo: *Editor-in-chief: Jan Vincents Johannessen, The Norwegian Radium Hospital*. But I was not there to visit him.

Radiumhospitalet (The Norwegian Radium Hospital) Imagery ©2020 Google, Map Data ©2020)

> I am uncertain where to go. The instructions they gave me in the hospital reception are already forgotten. This place is too overwhelming. I select a corridor at the side, only to find laboratories, lifeless between working hours. I return. Did I choose the right floor? I am in a corridor of offices now, but the names on the doors are different from the one I am looking for.
>
> I am nervously excited. I never met this person. We only had one call to set up this appointment. A late night at work, an unknown author name from Science Citation Index had suddenly beamed back at me from the top of a ranking list on the screen of my computer. It took a couple of days before I found out who the person was. I then made the call.

Science Citation Index. My first encounter with the potential of this database had been in April 1988 in Elsinore, Denmark, at a conference about scholarly publishing organized by the Nordic Council of Ministers. The scientific field of scientometrics suddenly personalized as the floor was taken by a bearded Swede in tweed and jeans. Olle Persson demonstrated empirically the need for further internationalization of the social sciences in Scandinavia by using articles in scientific journals as a data source (Persson, 1988). I was deeply amazed. I attended the conference because I was invited as an Editorial Director at SUP to speak about international marketing of scholarly journals (Sivertsen, 1988). I knew nothing about scientometrics, but I should have known. My own Journals Department at SUP – for a reason I could not understand at the time – had for several years subscribed to an obscure journal from Eastern Europe

with an orange cover and a futuristic title. It had been there all the time on the shelf in my office, completely ignored by me. I did not know then that *Scientometrics* would become my most frequent scientific publication outlet in the years to come.

PLAN PERFECT

Only a few weeks after I met Olle in Denmark, I shifted over to my present position as a researcher at NIFU in Oslo and started with scientometrics from day one. My institute wanted to introduce this field of research to Norway and expected me to know something about it since I came from the scholarly publishing industry. Waiting for me in my new office was fifteen meters of computer paper print-outs with large columns of numbers, bundled together in a big dark green plastic folder with the acronym 'SPRU' written on top of it. Fortunately, an assistant at the institute had already transferred the paper print-outs to the type of spreadsheets we used at that time – *Plan Perfect*.

I soon learned that 'SPRU' meant the Science Policy Research Unit at the University of Sussex. One of the pioneers of quantitative science-policy studies in the UK, professor Ben R. Martin, had provided the data for my director, Hans Skoie, who had been on a research visit to SPRU in the late seventies. Soon after, Ben came to Norway with his colleague John Irvine and helped introduce indicator-informed research evaluation.

The 'SPRU data' was based on Science Citation Index and contained two datasets of country-level statistics in eight major areas of research. One dataset covered the years 1973-80 with a fixed journal set of 2,300 journals. The other covered 1981-86 with a fixed journal set of 3,100 journals. One set of columns showed the number of articles per country in each area of research. Another set of columns showed the number of citations per country. Citations??? Citation analysis was something new in Norway. So far, we had only contributed with references.

Together with the big data file came evidence – a thick report – that the data originally came from the United States: M.P. Carpenter: *Updating and maintaining thirteen bibliometric data series through 1982*, published in 1985 by Computer Horizons, Inc., Cherry Hill, N.J. I soon understood that I was harvesting the results of Francis Narin's wise science-based development of scientometric indicators for the National Science Foundation's earliest *Science and Engineering Indicator Reports*. Among the references in the report was also Eugene Garfield's book *Citation Indexing: its theory and application in Science Technology and Humanities* (1979). I read it and became at the same time humble and enthusiastic. Scientometrics appealed to me as a well-established field. I particularly liked the connections to the history, philosophy and sociology of science. Soon, I was on my way to Philadelphia to visit Narin and his group. I was nicely received, and I learnt a lot!

Period photos taken in the Nordic Institute for Studies in Innovation, Research and Education (NIFU) in 1993. Photo courtesy of © NIFU.

TRANSPARENT PLASTIC SHEETS AND SOMETHING CALLED E-MAIL

After a while, I was ready to publish Norway's first report with international comparisons of national research performance based on Science Citation Index (Sivertsen, 1991). With an ambition also to join the scientometric community at the international level, I then prepared my first paper for the STI Leiden conference series (Sivertsen, 1992).

I was accepted for presentation and came to Leiden in 1991. At this time, chalk and blackboard had been replaced by transparent plastic sheets which we would put one after the other on an overhead projector. It was a fragile technology. To be prepared, the plastic sheets had to run through a xerox machine. They could easily melt and crumble if they did not find their way. The projector needed very bright light for the projection of our black and white transparencies. A very audible fan was needed to cool down the light bulb, but it sometimes exploded anyways.

I was extremely nervous as I put my first plastic sheet on the overhead projector in Leiden. At that time, all of us presented in the same plenary session. A big new auditorium was full of international seniors who had never seen me before. Henk Moed chaired the session with warmth and understanding, he noticed my stiffness. After I started talking, many strange things happened at the same time. The screen started going up and down. The lights went off and on again. All the time. It was like a discotheque. I looked up – the audience was laughing. I looked down and suddenly understood that my trembling knees had unwittingly steered the control panel for the auditorium. It was placed behind the speaker's desk.

Transparency on the OHP (overhead projector)
© Bundesarchiv, Bild 183-T0617-007 / CC-BY-SA 3.0

At this time, I was collaborating with Terttu Luukkonen from Finland and Olle Persson from Sweden (Robert Tijssen from the Netherlands joined us in the next phase) on developing methods for understanding and measuring the dynamics of international collaboration in science. Our data came from Francis Narin in spreadsheets and consisted of numbers of co-occurring countries in authors' addresses in Science Citation Index 1981-86. There was no such thing as an online edition of the database. The CD-ROM version was yet to come. We used fax or letters to exchange analysis and text for the paper that would appear in *Science, Technology & Human Values* (Luukkonen, Persson & Sivertsen, 1992).[1] While working on it, we visited Olle at his university in Northern Sweden. It was a sunny day. We had a rest on the lawn in front of the Department of Sociology. Terttu suddenly said: – Have you heard about something called e-mail? I think it could ease our communication.

Since then, e-mail has eased communication to the extent that we can co-publish with authors that we never met. I tend to be a bit conservative in this situation. I need to meet and talk. In 2012, I published a paper in *Scientometrics* with our Danish colleague Birger Larsen (Sivertsen & Larsen, 2012). We did not meet during the research process. We used e-mail. However, all ideas had been developed already by walking and talking together, back and forth several times, on a Brazilian beach. We were there for the ISSI

1 The article was covered as a news item in *ISSI Newsletter* no. 20, December 2009, p. 60.

conference in Rio de Janeiro in 2009. We walked and talked, continuously and totally engrossed by our plans for the paper, completely ignoring the presence of beautiful young Brazilians on the beach. Nerds...

CD-ROM AND FLOPPY DISKS

At the beginning of the nineties, Institute for Scientific Information made the SCI, SSCI and A&HCI available in CD-ROM editions that could be purchased by libraries. This was the basis for the ranking list of Norwegian scientists that I produced a late night at the office in 1993.

In the morning, I had collected data from the CD-ROM version of the Science Citation Index at the Library of the Faculty of Mathematics and Natural Sciences at the University of Oslo. I loaded the CD discs for the three years 1990-1992 and filtered by selecting "Norway" in the author address field. Thus, I received bibliographic information on 9,500 articles from the three-year period with at least one author from Norway. I transferred the files with search results to a large stack of floppy disks. Completing one floppy disk could take more than half an hour. Finally, I transferred the data to my computer. There was so far no internet connection between the university and my institute. I transferred the data with the help of my bicycle.

After loading the data in the floppy disks to my computer, I merged them in one regular document in the Word Perfect format. It had a little over 3,000 pages. I created macro commands for the document

Left: Floppy discs came in 3 main sizes (8", 5¼", 3½"), and a huge variation of colours and capacities. Many believes that the word 'floppy' comes from the two earlier (and larger) versions' flexible casing, whereas others explain the etymology by mishearing the word 'flippy'. The early models were manufactured one-sidedly and the disc drives could also handle one-side reading only. However, rumour has it that due to the high price of the early discs many cut openings on the other side of the casing as well and started to use the other side of the discs (which also had the magnetic coating), effectively doubling the discs' capacities by this makeshift hardware hack. To read both sides with a one-sided reader these discs often had to be 'flipped' in the drive, hence their potentially original name: flippies. Photo © George Chernilevsky @ WikiMedia Commons, public domain. Right: 5¼" floppy disc with a modern day gag. Back in those days a complete operating system, like the infamous Windows 3.1 required 20MB free space and was packed to no more than 6 installation discs with storage capacity of 1.2 MB each. This fictional floppy's label reads as "Windows 8 Installation Disk 1 of 5,972", which shows the insane growth in our data usage in just three decades. Photo provided by pxhere.com, public domain.

that allowed the machine to read, sort, and count. Thus, the text became names and numbers. The result was a ranking of Norwegian researchers according to productivity. It showed how many articles each author had published in the last three years, and for each author, which journals had published their articles.

I soon observed the same skewed pattern of productivity that I knew from the research literature: A small proportion of the researchers are authors of a large part of the publications, and at the upper end, the difference from the average is very large.

There were few surprises at the top of the ranking. I recognized the most prominent Norwegian researchers in the sciences. However, one author's name was completely unknown to me. The name was at the top of the list with a significantly higher number of articles than the second author on the list.

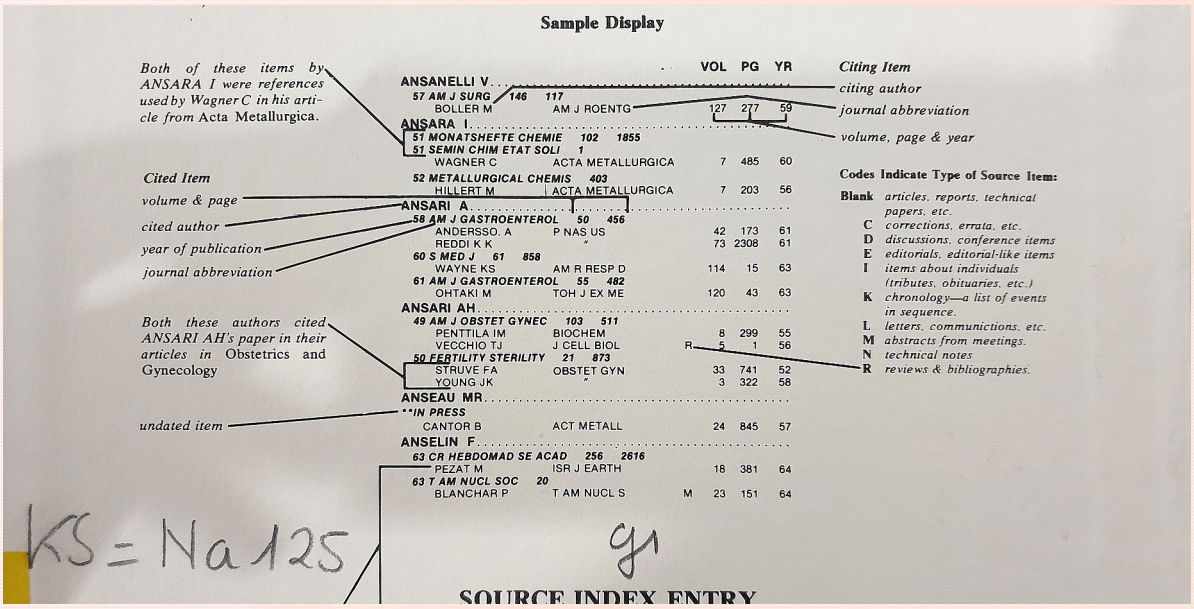
The person in question had published 81 articles during the three years, i.e. one article every other week in leading international journals. I was curious, also because the number increased during the three-year period. I had recently read in the magazine Science Watch from Institute for Scientific Information that the world's most-publishing researcher, a British immunologist, had published 35 articles registered in Science Citation Index in one year. In the Norwegian scientist's most productive year, there were 33 articles.

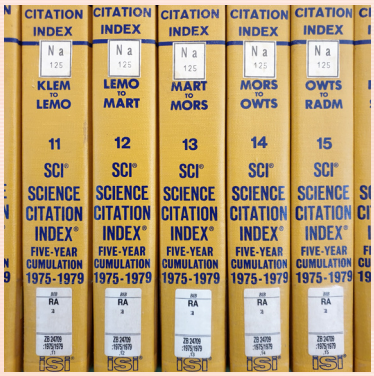
The author name at the top of my ranking was 'Fossa-SD'. Who was this scientist? Our institute had recently started using e-mail, but internet was still on its way. I had to shut down the computer and proceed with my investigations the next day. None of my colleagues recognized the name. I called some people at the Research Council of Norway, but no one with a name like 'Fossa-SD' had ever applied for research funding.

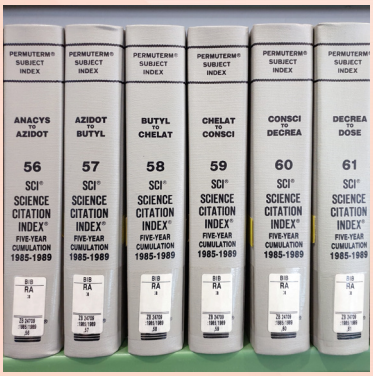
I still had more information to start from. Most of the 81 articles had appeared in British Journal of Cancer (10), British Journal of Urology (9), Journal of Urology (9), European Journal of Cancer (7) and Annals of Oncology (6). There were also articles in the most prestigious medical journals, The Lancet and the New England Journal of Medicine. Moreover, I had the author's address: "The Norwegian Radium Hospital". I called professor Per Ottar Seglen at the Norwegian Institute for Cancer Research, which is located next to the same hospital. I knew him because at that time, he was also contributing to the beginnings of scientometric studies in Norway. I asked him: – I have an author whose name is recorded as 'Fossa-SD' in Science Citation Index. Do you happen to know him?

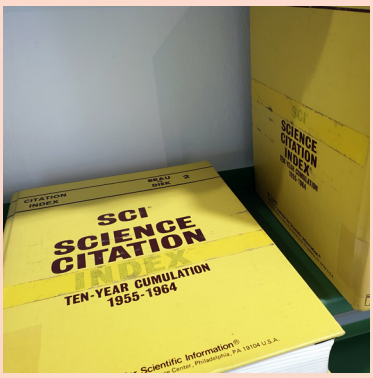
THE PRINTED EDITION OF SCIENCE CITATION INDEX

Per O. Seglen was and still is one of Norway's leading scientists in cancer research. He has also contributed significantly to international scientometric research and

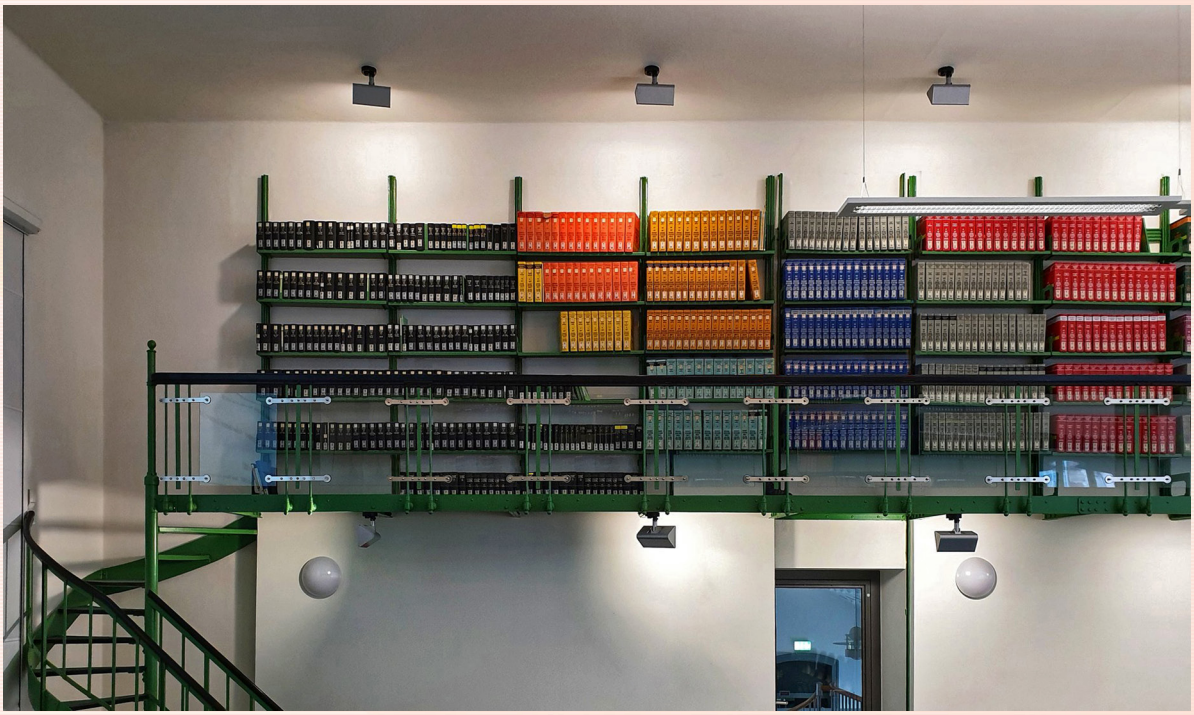








Scientometrics before the computerisation. Literally hard science. Photos courtesy of © Dr. Katrin Weller



Heavy volumes of the printed Science Citation Index in the Göttingen State and University Library (SUB Göttingen). Photo courtesy of © Dr. Katrin Weller

is famous for his pioneering critique of the use of the Journal Impact Factor (JIF) for evaluating individual contributions to research. His article in *BMJ* (Seglen, 1997) has received 1,200 citations and is the oldest among the eleven references cited in the DORA declaration on research assessment.[2] The creator of the JIF, Eugene Garfield, acknowledged Seglen's work in an article in *Nature* (Garfield, 2001, p. 522):

> *It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the journal impact factor as a surrogate. The latter practice is fraught with difficulties, as Seglen and others have pointed out.*

Seglen visited my institute in 1988 and told us about the strange results of an internal evaluation at the Norwegian Institute for Cancer Research: Scientists were ranked according to the JIF that could be attributed to each of their articles. Seglen consulted the printed edition of Science Citation Index and found a mismatch between the JIF of the journal and the actual number of citations received by each article that had been evaluated. He detected the skewness of citation distributions and wanted to study the dynamics of citation indicators as well as their use in research evaluation. We provided him with an adjunct professorship which led to a series of major contributions to scientometric research.

In 1988, Seglen needed scientometric data at the individual level for his first critical studies of the JIF. CD-ROM was not available yet. Only the printed edition of the Science Citation Index could serve this purpose in Norway at that time. Seglen collected his citation data by hand. Due to the laborious method, his first studies of the relation between journal impact and article impact were based on a sample of only eighteen active Norwegian principal investigators in biomedicine. Since his studies became so influential in retrospect, Lin Zhang, Ronald Rousseau and I did a replication study (Zhang, Rousseau & Sivertsen, 2017). We came to similar results by examining more

2 https://sfdora.org/read/

than eighteen thousand articles authored by nine hundred Norwegian researchers.

Per Seglen knew who 'Fossa-SD' was: – Did you say 'he' or 'she', Gunnar? There is no one at our research institute with a similar name. But I know of a doctor who is head of a medical department at the adjacent cancer hospital, and her name is Sophie Dorothea Fosså.

HANDWRITTEN MANUSCRIPTS FOR TYPEWRITING DURING NIGHTTIME

I still remember the uneasy feeling I got after the phone call with Per Seglen. Was I detecting someone's secret? Yes, indeed I was. She told me already when I called her. I knew her secret as I went to visit her at the hospital on the cold and dark mid-winter day. In the neon-light:

A door is open to a small office. I finally see her name on the door. But there is no one in there. I hear something rolling behind me in the corridor. It's a big trolley full of patient reports. Behind it, smaller, is a pleasant smiling woman in the white appearance of a doctor.

She offers a cup of coffee and invites me to sit between tall stacks of manuscripts in a small sofa. I ask, and she confirms: The stacks contain her own manuscripts as well as those she has agreed to review for journals. The book shelf has folders, almost no books. The desk calendar is a gift from the British Journal of Urology. Most parts of the desk are covered by patient treatment reports, administrative documents, messages. Where is the research? Covering almost the whole inside of her door, she has set up a detailed weekly schedule for her clinical hospital department. She is head of the department with the daily responsibility from 8am to 4pm for 24 in-bed patients plus 30-35 outpatients every week. In addition, she lectures as an adjunct professor at the University of Oslo and supervises PhDs, preferably between 6:30 and 8 in the morning.

Both of us are a bit nervous. I am here to publish her secret. It will happen by her participation in an interview (Sivertsen, 1994). As she said when I called her:

- I knew someone would detect me someday. I am not afraid myself that being active in clinical research is in conflict with clinical work, but I have been afraid that the leadership would think so. Hence, I have failed to report my publications to the hospital's 'boasting list'. I'm employed as the head of clinical services in my specialization, not as a researcher.

But it was no surprise to her when I told her on the telephone that she was Norway's most frequently publishing scientist.

I see a picture of her family on her wall. She prefers to talk about her family initially. She has four sons and a husband, also a medical scientist, who support her. Three sons have

recently moved away from home and are studying abroad. The youngest goes to high school and does fine.

- I found an increasing number of articles by you from year to year?

She confirms. There has been more time to do research recently.

The research starts after working hours at 4pm. In a small room adjacent to the clinical department, named 'the cage', sit two nurses working extra time to punch the results from her clinical examinations. They are paid by the assets from her adjunct professorship at the university. Herself, she now embarks on analyzes of already collected data and manuscript preparation. As she leaves at 8 pm, she delivers manuscripts for language editing and typewriting to a woman in her neighborhood. A new business day starts already at 2am. She collects the edited and typewritten manuscripts at the doorstep where she delivered them a few hours earlier. She goes to her office at the hospital, continues the processing of the articles, and attends the world-wide communication in her field of research. Before 8am, there is even time to swim in the hospital's pool or to supervise PhD students. From 8am, she again attends the responsibilities as the head of the clinical department specialized in treating urological cancer.

She tells me that she is currently working on a quality of life and sexuality study in relation to the treatment of prostate and testicular cancer. Her project design will be taken up by German specialists – it will be an international parallel study. Fosså is also the chair of the committee for quality of life in the urological group under the European Organisation for the Research and Treatment of Cancer. She says:

- Research demands international collaboration. This is also important from the patients' point of view. They need to be confident that Norwegian specialists are at the international frontier. We need to offer – and help develop – the latest in treatment methods and medicines. This is my main motivation for researching and publishing. It is important to systematize clinical experiences. Learning can benefit new patients. Moreover, publishing can provide a correcting and still very stimulating response from my international professional environment.

- But I also need to admit: Research is my lifestyle, publishing is my quality of life. This is me: My family often goes skiing together. As we go, I tumble with ideas for 3-4 new articles. My husband and sons laugh at me, but they also appreciate it. A male colleague of mine once said that publishing a scientific article is almost like getting a new child every time. Yes, so it is!

And yes, I had been counting the uncountable.

REFERENCES

Garfield E. (2001.) Impact factors, and why they won't go away. Nature, 411(6837), 522.

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding Patterns of International Scientific Collaboration. Science, Technology & Human Values, 17(1), 101-126.

Persson, Olle. (1988.) Nordisk forskning på den internationella tidskriftsmarknaden innom samhällsvetenskap. In Lundberg, E. (Ed.), Internationell vetenskaplig publicering i Norden (pp. 40-57). Copenhagen: Nordisk Ministerråd.

Seglen, P.O. (1997.) Why the impact factor of journals should not be used for evaluating research. BMJ, 314(7079), 498–502.

Sivertsen, G. (1988.) Internasjonal markedsføring av nordiske forskningspublikasjoner. In Lundberg, E. (Ed.), Internationell vetenskaplig publicering i Norden (pp. 58-67). Copenhagen: Nordisk Ministerråd.

Sivertsen, G. (1991). Norsk forskning på den internasjonale arena. Rapport 1/91. Oslo: NAVFs Utredningsinstitutt.

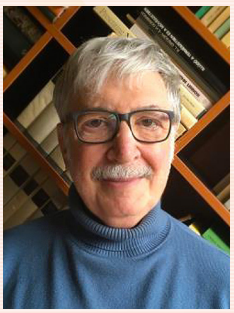
Sivertsen, G. (1992). Should a new bibliometric database for international comparisons be more restricted in journal coverage? In A.J.F. van Raan et al. (Eds.), Science and Technology in a Policy Context (pp. 35-50). Leiden: DSWO Press.

Sivertsen, G. (1994). Forskeren med flest artikler. Forskningspolitikk, 17(1), 14-15.

Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. Scientometrics, 91(2), 567-575.

Zhang, L., Rousseau, R., Sivertsen, G. (2017). Science deserves to be judged by its contents, not by its wrapping: Revisiting Seglen's work on journal impact and research evaluation. PLoS ONE, 12(3), e0174205.

FRIEDRICH ENGELS, THE GREAT-GRANDFATHER OF SCIENTOMETRICS



ANDRÁS SCHUBERT
Hungarian Academy of Sciences, Library and Information Center, Budapest, Hungary

SUMMARY:
It is argued that Price's thesis about the exponential growth of science can be traced back to Engels' 1844 work, thereby, due place is to be credited to him among the ancestors of scientometrics.

The query "father of scientometrics" is responded by Google just over half a second: Derek John de Solla Price (see Figure 1).

On the other hand, one receives the message: «No results found for "grandfather of scientometrics"». It's not as if grandfathers would be totally absent from scientific genealogy. It is interesting to note that one of the authors proposing the above honoring title to Price, Eugene Garfield [Merton & Garfield, 1986], was credited as the "grandfather of Google" [Rumsey, 2010]. Quite intricate family relationships.

The recognition and consistent pursuit of exponential growth in scientific endeavor is considered one of the most prominent contributions of Price to the discipline of scientometrics. He first published on the topic in 1951 [Price, 1951], later he expounded his notions in his milestone books [Price, 1961; 1963; 1975; 1986]. Remarkably, in none of these works Price refers to any direct precedent of his exponential growth concept, although a vast literature is cited by him containing supporting data for his analyses way back to the 19th century.

Nicholas Rescher, the doyen of contemporary American philosophy, recurrently deliberated on the origin and consequences of the exponential growth of science and knowledge [Rescher, 1978; 2006]. He heavily relied upon the works of Price, but he

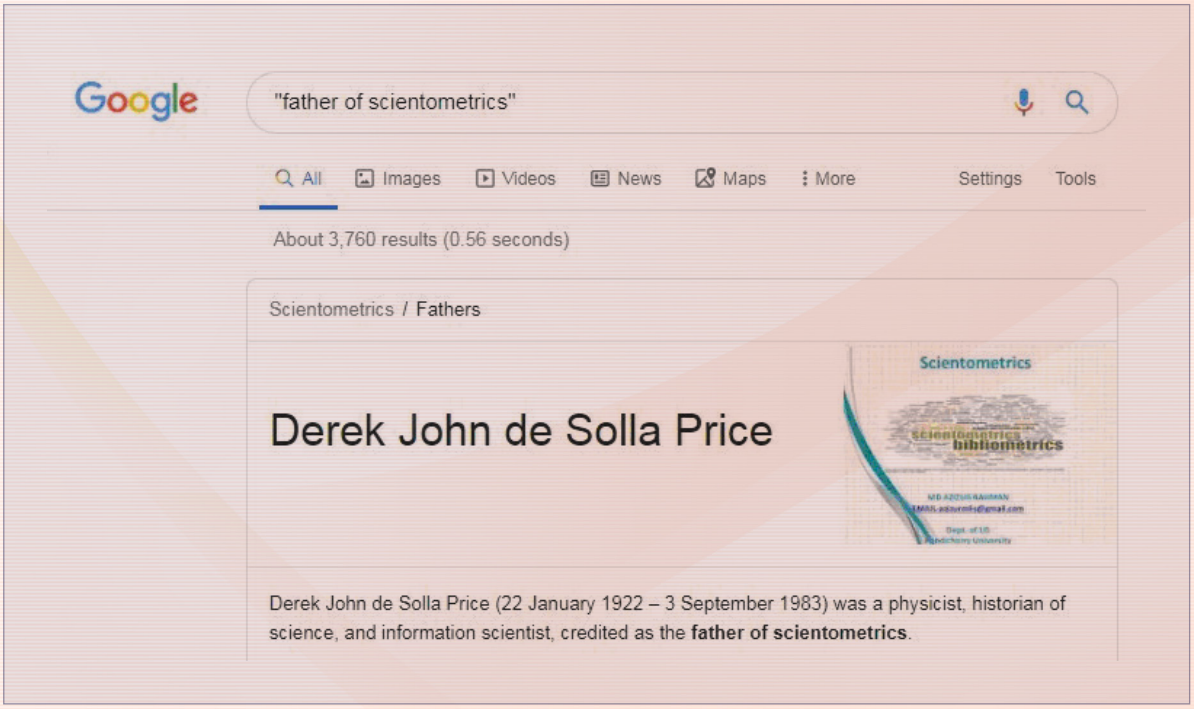


Figure 1 Google's response to the query "father of scientometrics"

definitely refrained from calling him the "father" of the concept. He even stated that by 1946 the idea of exponential growth became commonplace [Rescher, 2006, p. 54, "Bibliographical Apppendix"]. He attributed the idea to Henry Adams, and even eponymized the phenomenon of exponential growth of science as Adams's Law [Rescher, 1978] or later as Adams's Thesis [Rescher, 2006].

Adams, notwithstanding he was a most remarkable personality of the turn-of-the-century America (not necessarily in a likable and laudable way: among others, he was a ranting anti-semite), contributed only marginally to the topic in question. The reference Rescher used with predilection [Adams, 1918, Chapter XXXIV, "A law of acceleration" written in 1904] contains only rather vague and hardly quantifiable hints about the acceleration of progress in the history of mankind: "theory may assume what it likes — say a fifty, or even a five-and-twenty-year period of reduplication for the eighteenth century, for the period matters little until the acceleration itself is admitted". Rescher considers the word "reduplication" a convincing argument to believe that Adams speaks about an exponential growth.

Rescher himself offers at least two more legitimate candidates for the "grandfather" title [Rescher, 1978; 2006]. The first of them may be somewhat surprising: Sir Arthur Conan Doyle. In a short story entitled "The Great Keinplatz Experiment" [Doyle, 1885] the statement "Knowledge begets knowledge as money bears interest" can be found. Although it is just an inconsequential passing remark in the story, it doesn't appear to be inferior to Adams' assertion, nevertheless, to consider it the first printed formulation of the idea [Tague et al., 1981] seems to be a bit exaggerated.

William Thomson (Lord Kelvin), in his Presidential Address [Thomson, 1871] established: "Scientific wealth tends to accumulate according to the law of compound interest". This is a clear and unambiguous quantitative statement from a most authentic person, all the more since Thomson himself contributed to the accumulation of scientific wealth with about 660 papers. And, as Price remarked with proper admiration, "Almost every one of these could be viewed as a major scientific contribution" [Price, 1975, p. 176].

Tracking back one more generation, we get to our title character. In [Engels, 1844],

one finds: *"Die Wissenschaft [...] vermehrt sich mindestens wie die Bevölkerung; diese vermehrt sich im Verhältnis zur Anzahl der letzten Generation; die Wissenschaft schreitet fort im Verhältnis zu der Masse der Erkenntnis, die ihr von der vorhergehenden Generation hinterlassen wurde, also unter den allergewöhnlichsten Verhältnissen auch in geometrischer Progression"* ("Science [...] increases at least as fast as the population; the latter increases in proportion to the size of the previous generation, and science advances in proportion to the body of knowledge passed down to it by the preceding generation, that is, under the most ordinary circumstances in geometrical progression").

Rescher was aware of Engels' work, and even cited it [Rescher, 1978, p. 124], but did it incompletely, leaving out the passage concerning to the geometric progression. From the truncated statement he concluded: "In this respect, the early Engels would clearly qualify as a precursor of Henry Adams and Co. in the anticipation of Adams' Law. But Engels was not a mathematician, and so, perhaps, we are entitled to construe his 'in proportion to' not literally and technically, but rather more flexibly as 'stands in a fixed positive correlation to'." Obviously, the naming of the geometric progression precludes any alternative interpretation.

Instead of giving due credit to Engels for what credit is due, Rescher coined the eponym Engels' Theory based on what he calls "Engels' quadratic law of progress" [Rescher, 1978, p. 129]. This alleged theory was contrived on the basis of a cursory note in the Dialectics of Nature [Engels, 1883]: *"die Entwicklung der Wissenschaften mit Riesenschritten vor sich und gewann an Kraft, man kann wohl sagen im quadratischen Verhältnis der (zeitlichen) Entfernung von ihrem Ausgangspunkt"* ("the development of the sciences proceeded with giant strides, and, it might be said, gained in force in proportion to the square of the distance (in time) from its point of departure", an idea which, in the words of Rescher was left "in a lamentably undeveloped state." On this shaky foundation, Rescher constructed, and then deconstructed and refuted, the "Communist Theory of Scientific Progress". (In his defense, no other communist theories of scientific progress, of whatever origin, were less unfounded than his.)

Rescher considered Engels' 1844 notion completely overwritten by the four decades later fragmentary note. It is interesting to notice that not forty years, but only two paragraphs after the mentioned "reduplication" argument, Adams wrote: "Or better, one might, for convenience, use the formula of squares to serve for a law of mind; [...] the attraction of one century squared itself to give the measure of attraction in the next." Whatever this metaphor was intended to mean, it is certainly quite far from conforming to an exponential law.

"Pater semper incertus est" (the father is always uncertain) advises the ancient wisdom. What can one say then about grandfathers or great-grandfathers? In family genealogy, DNA tests may help to eliminate doubts. There is no similar aid in scientific genealogy. Citation analysis may be a useful tool in some cases, but as we try to dig deeper, bloodlines become more and more blurry. We are not always happy with what is found, but is not prescribed, either, that we have to be proud to our ancestors. As the present author is concerned, a paternal lineage Price–Thomson–Engels sounds more than flattering.

BIBLIOGRAPHIC ACKNOWLEDGMENT

In collecting historical references, I made great efforts to rely upon trustworthy sources, whenever possible, copies of original documents. Experiencing the difficulties to be encountered even in today's almost unlimited possibilities provided by internet services, I could not help but admire those historians who, even just a few decades ago, could survive and succeed

without such aids. Gratitude and appreciation to them, even if some of their inaccuracies can now be corrected.

REFERENCES

Adams, Henry (1918) The Education of Henry Adams: An Autobiography. Houghton Mifflin, Boston

Doyle, Sir Arthur Conan (1885) The Great Keinplatz Experiment. Belgravia Magazine, 57(July), 52.

Engels, Friedrich (1844) Umrisse zu einer Kritik der Nationalökonomie. In: Deutsch-Französischen Jahrbüchern, Ruge & Marx, Paris, pp. 86-114.

Engels, Friedrich (1883) Dialektik der Natur. Unpublished fragments.

Merton, Robert K.; Garfield, Eugene (1986) Foreword to "Little science, Big science... and beyond", New York Columbia University Press.

Price, Derek J. de Solla (1961) Science since Babylon. Yale University Press, New Haven.

Price, Derek J. de Solla (1963) Little Science, Big Science. Columbia University Press, New York. ISBN 978-0-231-08562-5.

Price, Derek J. de Solla (1975) Science since Babylon: Enlarged Edition. Yale University Press, New Haven. ISBN 978-0-300-01797-7.

Price, Derek J. de Solla (1986) Little Science, Big Science... and Beyond, Columbia University Press, New York. ISBN 978-0-231-04957-3

Rescher, Nicholas (1978) Scientific Progress. Blackwell and University of Pittsburgh Press, Pittsburgh. ISBN 978-0-822-91128-9

Rescher, Nicholas (2006) Epistemetrics. Cambridge University Press, Cambridge. ISBN 978-0-521-86120-5

Rumsey, Eric (2010) Eugene Garfield: Librarian & Grandfather of Google. https://blog.lib.uiowa.edu/hardinmd/2010/07/12/eugene-garfield-librarian-grandfather-of-google/ (accessed on Nov 11, 2019)

Tague, Jean; Beheshti, Jamshid; Rees-Potter, Lorna (1981) The Law of Exponential Growth: Evidence, Implications and Forecasts. Library Trends, 30(1), 125-149.

Thomson, William (Lord Kelvin) (1871) Inaugural Address of the President to British Association for the Advancement of Science, Edinburgh, August 2, 1871. The Chemical News, 24(610), 49-56.