

## Abstracts for the 17th Nordic Workshop on Bibliometrics and Research Policy 11.-12.10.2012

### **Keynote speech: The *New York Times* as a resource for scholarship**

*Diana Hicks, Georgia Institute of Technology, USA*

The *New York Times* is one of the highest impact scholarly periodicals as measured by citations in the Web of Science. Somehow this has escaped the notice of the bibliometrics, science & media and science studies communities, even in the USA. With reference to American scholarship, this talk will lay out the facts concerning this mysterious phenomenon: who cites NYT (answer: everybody), what do they cite (answer: new stuff, old stuff), trends (more and more) and why (top 10 reasons to cite NYT). This curious phenomenon does not fit well into existing frameworks. Bibliometrics is not strong on theories. Public understanding of science and media studies look at the reverse direction – media's use of science not science's use of media. And the science studies idea of boundary work is not much use because most NYT citers show no awareness of any boundary. Audience ideas will be welcomed.

### **Swedish and Nordic institute collaboration networks over time – between institute ego networks and fields**

*Katarina Larsen, KTH – The Royal Institute of Technology, Division of History of Science, Technology and Environment and Peter Sjögarde, KTH – The Royal Institute of Technology, Library Services, Language and Communication, Sweden*

The study represents an analysis of a group of institutes included in RISE (Research Institutes of Sweden) representing a wide range of activities from research in forestry, ICT, and also areas such as surface chemistry. The aim was to examine publication patterns over time and also analyze international collaboration in the broader context of technical-industrial institutes in Nordic countries.

A bibliometric analysis was conducted by searching Web of Science for publications containing the addresses of RISE institutes published 2002-2011. A publication count was done for the individual institutes, for an aggregated level dividing the institutes into four groups (Swedish ICT, SP, SWEREA, Innventia/STFI), and as total counts for the whole RISE group. To analyze the collaboration network of the RISE group, frequencies of organizations and countries in the publication set were calculated and co-occurrence visualizations conducted.

The preliminary results show an increasing publication frequency at the aggregated level. Also the international collaboration, measured by the share of internationally co-authored publications, has been augmenting, from about 31 percent in 2002 to about 43 percent in 2011. Nevertheless, in Sweden, universities are still the most common collaborators and the network analysis shows that there are extensive knowledge networks between local research partners.

## **Analysis of the ‘volume’ of researchers active across the different scientific fields in the Netherlands**

*Rodrigo Costas and Thed van Leeuwen, CWTS-Leiden University, the Netherlands*

### **Introduction**

In this exploratory analysis we seek to create information about the ‘volume’ of research fields, in terms of the number of people working in the Netherlands in these fields, as this is quite often an unknown component in science studies. We want to explore the possibilities of bibliometrics with these more “demographic” purposes, something that has been unexplored until now. As a field level we will concentrate on the sub-fields (composed of the so called WoS Journal Subject Categories) used in the former NOWT-reports as used in Dutch policy circles. We will focus on the publications of 2008-2010.

### **Methodology**

From WoS we collect email addresses ending with ‘.nl’. These email addresses generate lists of publications, thus conforming an initial ‘seed’ of publications. From the author names related to these publications, we will algorithmically expand the data collection on publications in order to collect as much as possible the oeuvres of the researchers. Based on their collected ‘oeuvres’ and through the relation publication↔journal↔journal category↔NOWT subfield, we are able to estimate the number of researchers active in the subfields. We also put our results in perspective based on verified data of actual active researchers in the Netherlands.

## **Visualization of research activities based upon scientific journals that are co-cited in articles published by different organizational units of Wageningen University and Research**

*Marco G.P. van Veller, Wageningen UR Library, the Netherlands*

The main topics for the research activities of Wageningen University and Research (or Wageningen UR) are concentrated around the following three themes: food and food production, living environment and health and lifestyle and livelihood. Although Wageningen UR concentrates on the life sciences, its various organizational units deal with different scientific disciplines in order to perform research in one or more themes. In general, chair groups (university) and business units (research institutes) work in agrotechnology and food sciences, plant sciences, animal sciences, environmental sciences or social sciences. In order to visualize the scientific landscape in which Wageningen UR and its organizational units are active, an analysis is made based upon the scientific journals that are cited in articles that are published by staff of Wageningen UR between 2006 and 2011. For the analysis a link is established between scientific articles for Wageningen UR -that are collected from Web of Science- and organizational information on the scientific output that is collected from the Wageningen UR digital repository on scientific output (WaY). The link enables to select sets of scientific articles per organizational unit of Wageningen UR. For these sets of scientific articles analyses on the cited scientific journals are made. The lists with most cited research journals per organizational unit of Wageningen UR are helpful to identify journals that are important for research staff or students working in a particular research field. Moreover, comparative analyses on these journal lists enables visualization of similar interest by organizational units of Wageningen UR in (groups of) scientific journals and maybe used as a proxy for (putative) collaboration in research activities. Besides a presentation on how the analysis is performed on Wageningen UR research articles I will discuss the place of several research units

in Wageningen UR in the visualized scientific landscape and possible collaborations with other organizational units of Wageningen UR that result from this analysis.

### **Publishing orientations in public research institutes**

*Elina Late, University of Tampere, School of Information Sciences, Finland*

18 public research institutes operate in Finland in fields of natural sciences, technology, bio and environmental sciences, social sciences and humanities. Research made in public research institutes is mainly applied and the proportion of external funding is often high. However, practices vary between research institutes. Tight collaboration with society and private companies has effect on publishing practices in public research institutes. Research is naturally published for academic audience but professional audience, state government and general public are also important and sometimes the main audiences for research. Survey was used to gather information about publishing practices in public research institute in spring 2010. Principal component analysis was used to study more detail publishing orientations.

Three orientations were discovered: professional orientation, academic orientation and industrial orientation. Professional orientation covers 55 percent, academic orientation 40 percent, and industrial orientation five percent of publications made by researchers working in public research institutes in Finland. Study indicates significant differences between disciplines in publishing orientations. Also other factors such as research type, research funding, audience, nature of research projects, and working practices has significant impact on publishing activity in different publishing orientations.

### **Using Scaling Exponents for Citation Analysis**

*Haiko Lietz, Mathias Riechert and Daniel Sirtes, Institute for Research Information and Quality Assurance, Germany*

Citation analysis has classically been a part of bibliometric analyses and evaluations of the science system and its objects. In order to compare across disciplines or subjects with their different citation practices, field normalizations have been introduced. Comparability across fields is supposed to be achieved by dividing the observed citation rate in a field by the expected citation rate of the same field. The most common of these normalized indicators are calculated on the basis of the arithmetic mean of all the raw citation scores. However, as citation distributions are highly skewed, the validity of such indicators has been repeatedly challenged. In contrast, it has been demonstrated that a power law plausibly fits this distribution above a certain threshold. As an alternative to the use of averages we introduce an indicator that is based upon the scaling exponent. Field normalization is achieved by dividing an observed citation exponent in a field by the expected citation exponent of the same field. Using *Scopus* and countries as objects of study, we discuss the statistical properties of this indicator, its comparison to the average-based indicator, and its usefulness.

## **Do different bibliometric measures for a research center follow the same trend? A case study**

*Marianne Gauffriau, Danish National Research Foundation and Jesper W. Schneider, Danish Centre for Studies in Research and Research Policy, Aarhus University, Denmark*

Bibliometric measures range from results of simple to more complex counting methods. The simple methods are often considered biased. In this case study, we calculate bibliometric measures for a research center at a department of chemistry. The data covers the period 2002-2009. We focus on a selection of widely used bibliometric measures and as a first step, show the trend over time for each measure. Thereafter we compare the trends and examine the consistency between them.

The bibliometric measures in this study can be divided into three groups going from simple to complex counting methods. The first group includes publication and citation counts in raw numbers. The second group includes impact factors; a calculation of the average number of citations per publication. The third group includes impact factors compared to the world average for a similar research field. The trends which are the results of the simple counting methods may show a first indication of which results can be expected from more complex counting methods. The results presented here are preliminary and should not be used for generalization.

## **Using Monte Carlo simulations to assess the impact of author name disambiguation accuracy on different bibliometric analyses**

*Jan Schulz, TU Bergakademie Freiberg, Germany*

Bibliometric analyses depend on the quality of datasets and the author name disambiguation process (ANDP) which matches authors of papers to real persons. Errors of the ANDP result in wrong numbers of papers matched to the actual individual or groups of authors, which in turn might change the results of the subsequent analysis, e.g. publication performance rankings. A manual ANDP is very time consuming and therefore not realistic for large datasets. Automatic ANDP are applicable to such datasets but are found to be error prone (Heath et al. 2009; Levin et al. 2012; Torvik and Smalheiser 2009). It is therefore important to assess the impact of errors in the ANDP, but this requires itself a manually verified error free dataset.

Monte Carlo simulations are a way to avoid this time consuming step. By simulating lists of authors with attributed papers and adding known errors to the distribution of papers, the effects of such errors on the results of the analysis can be assessed. This paper presents a basic algorithm for such a Monte Carlo simulation and explains how different errors can change the results of different bibliometric analyses (e.g., rankings, regressions with number of papers as dependent variable).

## **The Flemish BOF-key: further refinement and impact on co-authorship practices**

*Tim C.E. Engels, Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium*

The Flemish BOF-key has included publications and citations indexed in the Web of Science since 2003 (Debackere & Glanzel, 2004, 2008; Engels et al, 2012). In this presentation we review the gradual refinements that have been made, including (1) the extension to non-WoS publications through the VABB-SHW and (2) the latest 2011-2012 round of revision. We then zoom in on the

whole counting per university that is used. In particular we analyse the potential impact on copublication practices. Few researchers in Flanders are enthusiastic about the BOF-key, initially because it only provided productivity incentives and more recently because the weighting system of WoS publications as well as the criteria applied by Thomson Reuters were deemed to lack transparency. Policy makers, however, often claim that the BOF-key has a substantial positive effect on university research. It is credited with stimulating competition, fostering excellence and boosting productivity. Indeed, over the last decade Flanders has been one of the few European regions that succeeded in increasing its share in WoS indexed papers. However, productivity increases vary widely between disciplines and the link with changes in co-authorship practices remains unclear

### **Journal Citation Cartels: Can they be detected?**

*Clara Calero-Medina and Rodrigo Costas, CWTS-Leiden University, the Netherlands*

The possible manipulation of self-citations to increase journal impact factors is a well known problem in scientific communication and evaluation. However, the effect of self-citations can be analyzed through bibliometric indicators allowing the detection of possible fraudulent behaviours. Currently some scientific forums are starting to talk about the importance of detecting other fraudulent practices aiming also to blow up the impact of journals. This is what is known in the literature as 'Citation Cartels' – groups of editors and journals working together for mutual benefit (Franck, 1999). In this study we propose a methodology aiming to detect potential Citation Cartels. In our approach a 'Journal Citation Cartel' is a group of editors and journals working together for citation benefits. Our approach uses as starting point the members of the Editorial Boards of around 300 scientific journals in five scientific fields. Then we identify journals that share members and create a 'co-editorship journal network'. After that and through network analysis we will detect clusters of journals sharing editors members. This is what we call 'Potential Journal Cartels'. We will then define the 'self citation' of a journal not only in the traditionally way (i.e. a journal citing its own papers) but also the citations received from the journals of the same potential cartel. Finally we make a correlation analysis between the new defined self citation of a journal and its normalized impact (MNJS). The result of this analysis will determine whether a journal uses the journal cartel to increase its citations.

References:

Franck G. Science communication—A vanity fair? *Science*. 1999;286(5437):53–55.

### **Size matters – also in scientometric studies and research evaluation**

*Jesper W. Schneider, Danish Centre for Studies in Research and Research Policy, Aarhus University, Denmark*

What constitutes an important difference between two universities in a ranking? What difference in impact factors between journals would be considered an important one? And more general what is an important finding in a scientometric analysis? In empirical studies, importance is most often determined mechanically by a statistical significance test or not considered at all. The major predicament is that such tests are close to useless when it comes to importance. Decisions of importance is a matter of informed subjective judgment based on research design, theory, former findings, application, cost-benefit etc. Importance of empirical findings lies in "differences that make a difference". The problem for scientometrics and research evaluation is that we generally

have weak or no theories to guide us in these matters. The problem is specifically dire in research evaluation, where the practical importance of differences in rankings or indicators are not considered. Our field is basically a-theoretic. We discuss these matters and their consequences through examples and present some best practices that can improve the current state-of-affairs, including the use of effect sizes.

### **Something old, something new, something borrowed...and a new scientific topic in her shoe?**

*Carolin Michels, Fraunhofer Institute for Systems and Innovation Research ISI, Germany*

The assessment of research topics according to their development stage can be used for different purposes, most importantly decisions concerned with (financial) support of research groups and regions. For such objectives, it might also be necessary to decide whether the respective topics have the necessary prerequisites to persist.

In this work, I try to determine the influencing factors of new scientific topics during their early development stage. Eight pre-defined topics are analyzed according to their growth rate and citation and publication numbers. Also, the characteristics of the involved researchers, the publications they refer to and the diffusion of the topics are studied. The distributions of these features for the topics are compared to the more general distributions in the respective research fields.

The most important observed qualification for the breakthrough of a new topic is the involvement of experienced scientists. It can be shown that they have to rely on recent references, thus newer instead of more fundamental findings, and accordingly their own know-how. Further-more, in some topics, the set of references used by the various researchers seems to converge to a common denominator, while in other topics new references appear to act as an activator for the breakthrough.

### **The differences between journal articles and monographs in political science**

*Pei-Shan Chi, iFQ – Institute for Research Information and Quality Assurance, Germany*

This study investigates the publication, citation, and reference characteristics of publications in political science with special attention to journal articles and monographs. By analyzing the complete 2003-2007 publication list of 33 professors of two top-ranking institutions of political science in Germany, Department of Political Science at Mannheim University and Institute of Political Science at University of Muenster, this study explores the publication behavior of German political scientists and also analyzes their citation and reference characteristics. Language and nation analyses in both publication output and citation impact are also used to indicate the localization of German political science. The results of analyses on journal articles and monographs are compared to point out the huge differences between these two main types of publications in political science, and how their own characteristics form the framework of a field. According to the results of this study, we will be able to provide important knowledge to the process of evaluating research performance in such a local language oriented subfield in the social sciences.

## **Bibliometric analysis for information scientists at the University of Tampere in 2012: some results and discussion on information sources**

*Merja Hyödynmaa, Eija Poteri and Saija Tapio with the assistance of Esa Hakala, Heli Vanamo and Tarja Wisakanto, Tampere University Library, Finland*

The project was a part of library's services for researchers and ties with tools for and assessment of research. The partner was the School of Information Sciences (SIS) at the University of Tampere. The data was drawn from the publication register of the University and it was supplemented with publication information from Web of Science (WoS), Scopus and Google Scholar. Firstly, the aim of the project was to find out how many publications the researchers of Information Sciences of Tampere University had 2008-2010, which forms of publication they preferred, how many times the publications were cited, and how their favourite publications were located in the categories of Julkaisufoorumi (The Finnish Publication Forum). Secondly, the aim was to increase the know-how regarding citation databases and bibliometric analysis in the Library, as well as to strengthen the contacts to scholars at the University. In this presentation, we depict some results including the amount of citations and the list of most popular scientific journals with the categories of Julkaisufoorumi. Moreover, we try to discuss, how WoS, Scopus and Google Scholar performed as citation sources for information scientists?

## **Gender, age and scientific publishing. A bibliometric snapshot of the slow process towards gender equality in research**

*Ragnar Edgren Pettersen, CRISTin – Current Research Information System In Norway, Agnethe Sidselrud, CRISTin – Current Research Information System In Norway & Gunnar Sivertsen, NIFU – Nordic Institute for Studies in Innovation, Research and Education, Norway*

Norway's Current Research Information System (CRISTin) was extended in 2011. It now covers scientific publishing at virtually all institutions in the public sector: higher education institutions, independent research institutes and hospitals.

A total of 13,630 (fractionalized) scientific publications from the year 2011 have been registered in CRISTin by 17,314 researchers in the age of 27 to 67 years. Of these, there were 40 per cent female researchers contributing to 33 per cent of the publications. Both shares are clearly higher among young researchers, but on the other hand, the difference in productivity between men and women is also larger among young researchers.

The gender difference in productivity is consistent across the research performing sectors, but female researchers and their publications are better represented in hospitals and at research institutes than they are in academia. We will discuss our findings in relation to other studies on gender, age and scientific productivity.

## **Lost in Archives: the role of archival references in the scientific communication**

*Ülle Must, Estonian Research Council, Estonia*

Current study concentrates at one facet of the reference practices of the historians – references to the archival documents. We study the Estonian historians' work, with a selection of 1,027 publications (articles, monographs, collections, commented source publication etc.) published

during 1988 to 2007 (in total 1027). They contained 46,632 bibliographic references citations, including 7558 references to archival documents.

Thanks to the rapidly changing environment the significance of given information has considerably increased. The archival resources, together with intangible heritage, represent the history of individual European countries, our histories have intertwined through the times, and countries have belonged to different empires and even cultural areas at different times. Not only the role of the analysis of references to literature has actualised in the humanities—references to databases, also to archival resources, make it possible to connect sources.

In the case of Estonia's history a large part of "national documentary memory" has been prepared at different times by the foreign rulers, and they are kept in the archives of the respective countries. The same time the materials, preserved for example in Estonian and Russian archives, provide in some cases the different information (a secret policy). To get the complete picture, the fullest possible use of the material is necessary. Because of this, the references of Estonian historians reflect archival resources of Swedish National Archives, Latvian History Archives, British National Archives (former Public Record Office), Vatican Archives, several Russian, Finnish and German Archives, United Nations (former League of Nations) archives, etc.

### **Visibility of scholarly activity in Twitter**

*Kim Holmberg, Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton. United Kingdom*

This research in progress studies scholarly activities in Twitter. The goal of this research is to identify scholarly activities that are visible on Twitter and to study possible disciplinary differences in both use and ways of use of Twitter. To meet the goals of this research three different types of data from Twitter are being collected: a) tweets that mention top journals from a certain research field, b) tweets that contain certain keywords that are unique to the research field, and c) tweets by top researchers in the research field studied. The research fields that are being studied are digital humanities, cognitive science, social network analysis, history of science, biochemistry and molecular biology, economics, sociology, astrophysics, drug discovery and cheminformatics. For the presentation at the 17th Nordic Workshop on Bibliometrics and Research Policy we plan to have some preliminary results to present and we hope to discuss methods for data analysis with the participants.

This research is part of the research project *Cascades, Islands, or Streams? Time, Topic, and Scholarly Activities in Humanities and Social Science Research* (<http://www.diggingintodata.org/Home/AwardRecipients2011/tabid/185/Default.aspx>)