COST TD1210 workshop

Alternative metrics or tailored metrics? Science dynamics for science policy

Warsaw, November 9-10, 2016, Palace of ancient Zaluski Library, Hipoteczna 2

Science and technology indicators have a long tradition. They cover traditional input (human capital and expenditures) with output statistics (publications, patents, citations). Knowledge production of Mode 2, globalized team science and the use of new ICT technologies have added to the traditional set of macro-level indicators. Bibliometrics offers tailored toolsets, and altmetrics offers indicators on the fly. We have more metrics than ever, more metrics than probably data can be collected for, or so it seems.

This workshop brings together specialists in STI, science and technology scholars, bibliometricians and altmetrics specialists and researchers with extensive research management experience to reflect about current metrical practices.

This workshop will address questions such as: How much metrics do we need and how much metrics can we afford? Do we have the right metrics? How can data bases with STI relevant information made interoperable? What is the role for Linked (Open) Data information gathering and management? What new metrics do we need to develop to capture new institutional forms in science? How can visualisations and infographics help to map and measure science?

We will pay special attention to

- metrics suitable for social sciences and humanities

- the role of metrics for libraries,

- the relationship between metrics concerning the input of science (e.g., OECD statistics) and output of science

- the role of different classification systems (KOS, folksonomies) for collecting metrical information

- the role of visual analytics

- the relationship between metrics and science policy

- Big Data

Day 1 – November 9

First session 9.00-9.30
(Chair A. Scharnhorst)

Welcome and invitation (15 minutes)

Jan Kozłowski,
Ministry of Science and Higher Education, jan.kozlowski [at] nauka.gov.pl

Where we are – Zaluski Library

Key note speech 9.30-10.30

Kevin Boyack,
SciTech Strategies, currently visiting scholar at the Centre for Science and Technology Studies, CWTS Leiden
k.w.boyack[at]cwts.leidenuniv.nl

Grand Challenges and National Bias: Altruism and Economics

Coffee break 10.30-10.40

Second session 10.40-12.20
(Chair: A. Akdag)

Peter Van den Besselaar, Vrije Universiteit Amsterdam, Department of Organization Sciences & Network Institute, p.a.a.vanden.besselaar[at]vu.nl, *Metrics, altmetrics and theory based metrics* (20 minutes)

Dominik Rozkrut, Chairman of the Main Statistical Office, d.rozkrut[at]stat.gov.pl, *Big Data and Official Statistics* (20 minutes)

Ray Harris, Department of Geography University College London, ray.harris[at]ucl.ac.uk, *Data policy and data metrics* (20 minutes)

Discussion (10 minutes)

Student paper rush **(11.50-12.20)**

Miss Erifilli Kokkalidou (eri.kokkalidou[at]gmail.com) and Miss Christina Fotiou (fotiouxristina[at]gmail.com), *Current citations and a future with linked data*

Mr Panagiotis Gioulekas (panagiotisgkioul[at]gmail.com) and Mr Theodoris Chalkidis (theod.chalkidis[at]gmail.com), *The new age of altmetrics: the case of Scimago Journal & Country Rank*

Mr Dimitrios Iliadis (ddiliad[at]gmail.com) and Miss Suzana Maragga (smaranga22[at]hotmail.com), *A systematic study of metrics and evidence suitable for institutions which operate in areas of high inter-disciplinarity*

Lunch 12.20-13.40

Third session 13.40-15.10
(Chair: Y. Tonta)

Rob Koopman, OCLC Research, Leiden, The Netherlands, Rob.Koopman[at]oclc.org; *Semantic indexing and bibliometric analysis* (20 minutes)

Janusz A. Hołyst, Julian Sienkiewicz, Krzysztof Soja, and Peter M. A. Sloot, jholyst[at]if.pw.edu.pl, *Categorical and Geographical Separation in Science* (20 minutes)

Julian Sienkiewicz, Eduardo G. Altmann, julasms[at]gmail.com *Impact of lexical and sentiment factors on the popularity of scientific papers* (20 minutes)

Andrea Scharnhorst, Royal Netherlands Academy of Arts and Sciences, *Bibliometrics, Webometrics, Altmetrics, Alternative metrics. A possible Zeno effect for science metrics, and why we nevertheless look for metrics* (20 minutes)

Discussion (10 minutes)

Fourth session 15.30-17.00

(Chair: P. Polydoratou)

Laura Elien Ridenour, University of Wisconsin, Milwaukee, redienour[at]uwm.edu, *Boundary objects as interfield phenomena: From sociological phenomena to information system artifacts* (20 minutes)

Alesia Zuccala, Copenhagen University, a.zuccala[at]hum.ku.dk, *Developing Metrics for the Humanities* (20 minutes)

Veslava Osińska, Nicolaus Copernicus University, Torun, wieo[at]umk.pl, *The role of Visual Analysis* (20 minutes)

Alberto Martín-Martín, Emilio Delgado López-Cózar, Evaluación de la Ciencia y de la Comunicación Científica, Facultad de Comunicación y Documentación, Universidad de Granada, albertomartin[at]ugr.es; edelgado[at]ugr.es, *The Social Sciences & Humanities in the light of Google Scholar: tapping the bibliometric potential of an academic search engine* (20 minutes)

Discussion (10 minutes)

Dinner 20.00

Second Day, 10 November

Key note speech 9.15-10.15
(Chair: K. Boyack)

Loet Leydesdorff, University of Amsterdam, loet[at]leydesdorff.net,

Three Anecdotes about the Validity of Bibliometric Constructs

Discussion

Coffee break 10.15-10.30

Fifth session 10.30-12.00
(Chair: J. Holyst)

Giulia Ajmone Marsan, Economist/Policy Analyst, Directorate for Science, Technology and Innovation OECD, giulia.ajmonemarsan[at]oecd.org, *Policies to promote open science – evidence from OECD countries* (20 minutes)

Christopher Watts, Freelance consultant, *The complexity of knowledge production*, cjwatts1007[at]gmail.com (20 minutes)

Jan Fazlagić, *Measuring Creativity: Bridging the Psychology and Economic Perspectives*, Poznan University of Economics and Business, jan.fazlagic[at]ue.poznan.pl (20 minutes)

Jan Kozłowski, Ministry of Science and Higher Education, *Distant cataloguing – step towards digital and qualitative humanities*, jan.kozlowski[at]nauka.gov.pl (20 minutes)

Discussion (10 minutes)

Lunch (self-paid) 12.00-13.00

Sixth session 13.00-14.35

(Chair: Jan Kozlowski)

Ray Harris, Department of Geography University College London, ray.harris[at]ucl.ac.uk, *The UK Research Assessment Exercise* (20 minutes)

Krzysztof Klincewicz, *Inequality from bibliometric perspective*, Warsaw University, kklincewicz[at]wz.uw.edu.pl (20 minutes)

Pawel Sobkowicz, National Centre for Nuclear Research, pawelsobko[at]gmail.com, *Agent Based Model of dynamic research landscape: effects of funding, fashion, and lobbying* (20 minutes)

Katarzyna Szkuta, DG JointResearchCentre European Comission, katarzyna.szkuta[at]ec.europa.eu; *Tension between researchers reputation driven metrics and the R&I policy goals and how through funding you can nudge behavioural changes in science* (20 minutes)

Discussion (10 minutes)

Coffee break 14.35-14.45

Seventh Special session – meeting the coordinator of research team preparing a proposal for the act of law on the higher education 14.45-16.00

Arkadiusz Radwan – 45 minutes

Discussion (20 minutes)

Closing of the conference (10 minutes)

Participants (without a speech)

Sandor Soos soossand[at]gmail.com

Janusz Hołyst jholyst[at]if.pw.edu.pl

Marcin Kardas, Ministry of Science and Higher Education, marcin.kardas[at]nauka.gov.pl

Miłosz Rojek milosz.rojek[at]nauka.gov.pl

Anna Dawidziuk akd[at]twarda.pan.pl

Marcin Zarod marcin.zarod[at]gmail.com>

Milena Damaynova milenadam[at]gmail.com

Przemyslaw Kazienko Przemyslaw.Kazienko[at]pwr.edu.pl

Rajmund Klemiński rajmund.kleminski[at]pwr.edu.pl;

Abstracts

Kevin Boyack, *Grand Challenges and National Bias: Altruism and Economics*

What motivates the research strategies of nations and institutions? We suggest that research primarily serves two masters – altruism and economic growth. Some nations focus more research in altruistic fields while others focus more research in fields associated with economic growth. What causes this difference? Are there national characteristics that would suggest why they are more aligned with altruism or economic growth? To answer this question, we have identified nine grand challenges in research by analyzing the publication activity of 4429 institutions using Scopus data. Three grand challenges are more altruistic (there is little involvement by industry) and three grand challenges are clearly aligned with economic growth. The altruistic vs. economic nature of nations based on their publication profiles across these grand challenges is correlated with national indicators on wealth, education, capitalism, individualism, power, religion, and language. Previous research has suggested that national research strategy is aligned with national wealth. However, our analysis shows that national wealth is not correlated with the trade-off between altruism and economic growth. Instead, the trade-off is largely captured by a culture of individualism.

Miss Erifilli Kokkalidou and Miss Christina Fotiou, *Current citations and a future with linked data*

The practice of Citation Analysis has been an important and useful tool for research evaluation. Through citations, a new world of creativity and exploration can be opened. This way, a paper that has been cited many times, can become known to the scientific community. Is this, however, a reliable way to measure a paper's quality, within the chaos of the information available on the Internet? Admittedly, they help us in the bibliography selection, but the information they provide is, sometimes, not sufficient for today's research needs. We have to take into consideration that, with the current citation practices, it is impossible to describe the reason that this particular document has been used.

Although citation analysis is widely used and plays an important role in scientific and academic publishing, it still has room for improvement. The current protocols do not provide adequate support for linking data and metadata. The Semantic Web technologies can contribute to this in a number of manners, since they provide robust methods to describe, manage and use information. Linked data, in particular, offer a flexible way to represent citation data and make interoperability between different applications possible. In the poster, we will present the idea of converting citation data to RDF triples, in order to allow for semantic linkage between publications. By doing

this, it will be possible to perform complex searches and provide more advanced metrics.

Mr <u>Panagiotis Gioulekas</u> (panagiotisgkioul[at]gmail.com) and Mr <u>Theodoris Chalkidis</u> (theod.chalkidis[at]gmail.com), *The new age of altmetrics: the case of Scimago Journal & Country Rank*

Recently, the growing interest about the scholarly publications and their impact has led to the conception of the term alternative metrics, in other words altmetrics, due to the interaction with technology and the Social Web.This relatively new kind of assessment of researchers' output is here to overcome any potential problems that have arisen and support existing citation impact metrics. The dissemination of such publications is evaluated through different levels, measuring, for instance, the impact of how many times an article is viewed, saved, discussed, recommended and cited in electronic form. Nevertheless, another feature of the altmetrics is the ability to compute the importance of the source, more specific the authoritativeness of the authors. Also, for the better depiction of the impact in social media, altmetrics use visual representation for the mentions by source type. As always, the academic libraries are the primary repositories of existing knowledge for the scholars. At the same time observed that the scientific research has flooded the social media. And the challenge that arises is if it possible a useful tool, like altmetrics, could be used in the libraries in order to understand the impact of scholarly activities. Like any new tool, there are advantages and disadvantages to using altmetrics. On the one hand altmetrics provide scholarly output from multiple sources of data. On the other hand the tool fail to serve social sciences and humanities as well as medicine and natural sciences. All things considered, the new inquiries that the era of altmetrics brings upon have to be answered in order to be able to inform ourselves correctly and reliably about the impact of such important published information.

Keywords: altmetrics, bibliometrics, scholarly publications, impact, Social Web, advantages, disadvantages, citations, scientific journals

Mr <u>Dimitrios Iliadis</u> and Miss <u>Suzana Maragga</u>, *A systematic study of metrics and evidence suitable for institutions which operate in areas of high inter-disciplinarity*

This paper presents the first results of our Short Term Scientific Mission to The Netherlands, from 9th to 17th of July funded by the COST Action TD1210 KNOWeSCAPE. The research aimed to study metrics and evidence suitable for institutions which operate in areas of high interdisciplinarity at Data Archiving and Networking Services (DANS). Specifically, the main objective of this study was to categorise and analyse the data of six reports published by different scientific organisations, all collaborators of DANS in The Netherlands, Switzerland and Germany. The methodology comprised desktop research, brainstorming sessions with the team members at DANS and the supervisors of the study both in The Netherlands and Greece, observation and empirical analysis using the classification scheme UDC. The scientific and business reports of six organisations which take action in the same interdisciplinary environment categorised by time, place, institutional type and the type of each report (business or scientific). The collection and the categorisation of the reports' elements based on UDC scheme leads us to make interesting observations. One of them is that the production of scientific

knowledge and the study of internal information, such as reports, can interconnect the academic institutions with industries to create more efficient products which will be certificated. Thus, the research acquires a practical scope and the industries obtain a strong ally and useful scientific tools for their healthy development.

Christopher Watts, Freelance consultant, *The complexity of knowledge production*,

If the generation of knowledge and know-how is complex, takes time and is embodied and embedded in institutions' staff and organisation, how easily can higher education institutions boost their capabilities? With the help of data from the UK's 2014 Research Excellence Framework (REF), we explore the scope for applying Hidalgo's method of reflections when assessing institutions' complexity of knowledge production. This method attempts to evaluate both the complexity of products output and the complexity of the producing institution. We invite comments on the transferability of Hidalgo's concepts and methods from economics to research assessments. We then trace some of the authors of publications submitted for the REF. This has the potential to cast doubt on the extent to which REF data reflect the knowledge generating capabilities of the institutions. Some institutions may be importing innovation from elsewhere rather than generating it, and may also be struggling to retain their imported staff. This has implications for the use of REF results as a basis for funding research, as well as for Hidalgo-inspired estimation of knowledge complexity.

Peter Van den Besselaar, (p.a.a.vanden.besselaar[at]vu.nl), *Metrics, altmetrics and theory based metrics*

I will argue that

- most metrics developments are data driven - which also holds for alternative metrics.
- I will also argue, using several examples, that some theory about how the research system works, may be helpful to develop meaningful indicators.

Ray Harris, ray.harris[at]ucl.ac.uk, *The UK Research Assessment Exercise*

The UK Research Assessment Exercise: reports from on the ground. This would look at the history of the UK Research Assessment Exercise and how it changed from its origin in 1984 as a way of praising research to a way of allocating funding to UK universities. It changed dramatically along the way and it has had a significant effect on the behaviour of researchers. This paper relates to metrics in that it describes how assessment and metrics can change research behaviour. The paper would be more of a presentation of ideas than a paper proposing action.

Ray Harris, ray.harris[at]ucl.ac.uk, *Data policy and data metrics*

This paper would look at the ICSU World Data System (WDS), its history and its data policy. In the WDS there is a growing need for data publication to be recognised more clearly. This implies in part some metrics for recognising data publication and more

widely for scientists to participate actively in the development of high quality data. This will become a greater concern as more and more open data become available. So this paper would be a discussion of the challenges facing scientists as Big Data become more prominent in science rather than an examination of alternative metrics.

Julian Sienkiewicz, julasms[at]gmail.com, Krzysztof Soja, Janusz A. Holyst and Peter M. A. Sloot, *Categorical and Geographical Separation in Science*

Using two extensive university ranking lists (ARWU and QS) as well as data from Science Citation Index we infer that the number of publications in specific scientific categories strongly correlate with the university rank. At the same time, through principal component analysis, we find evidence of separation of data with respect to scientific disciplines. Detailed complex network analyses indicate that scientific collaboration is linked to the geographical region and that the number of common papers decays with the physical distance between universities, while the strength of the ties between universities is proportional to product of their total number of publications. We believe that these finding shine a new light on the relevance and interpretation of University ranking. [1] J. Sienkiewicz, K. Soja, J. A. Holyst, P. M. A. Sloot e-print: arXiv:1307.0788v2

Julian Sienkiewicz, Eduardo G. Altmann, julasms[at]gmail.com, *Impact of lexical and sentiment factors on the popularity of scientific papers*

"How should I present my results in order to attract more attention and citations to my paper?" is a question that bothers many scientists. No wonder, scientists are increasingly evaluated by citation counts and recent studies suggest that simple stylistic choices can affect it (e.g., "the shorter the title the more citations"). Is it really that simple? Well, apparently not. In my talk I will show how textual properties of scientific papers relate to the number of citations they receive [1]. The main finding is that correlations are not linear and thus affect differently the most cited and typical papers. According to our research the short title recipe works only for the most cited papers, surprisingly for less popular ones it is even better to have a longer title. Quantile regression analysis of six different factors, calculated both at the title and abstract level of 4.3 million papers in over 1500 journals, reveals the number of authors, and the length and complexity of the abstract, as having the strongest influence on the number of citations. [1] J. Sienkiewicz, E. G. Altmann, Royal Society Open Science 3, 160140 (2016)