ISSI NEWSLETTER

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

#53 / VOLUME 14 NUMBER 1
MARCH 2018

CONTENTS

LETTER TO THE EDITOR

H. Memisevic: In Web of Science We Trust—A Case of a Hijacked Journal Indexed in SCOPUS *page 1*

CONFERENCE CALLS

STI2018 Leiden *page 4*

CARMA 2018: Internet and Big Data in Economics and Social Sciences. *page 6*

SCIENTOMETRICS' CALL FOR PAPERS

Special Issue on Sciento-Network-Mining and Smart Altmetrics for Advanced Scientific Collaboration and Knowledge Utilization *page 8*

ARTICLES

R. Rousseau: Blockchain Technology: A Bibliometric Analysis *page 10*

N.J. van Eck, L. Waltman, V. Larivière, C. R. Sugimoto: Crossref as a New Source of Citation Data: A Comparison with Web of Science and Scopus *page 23*

LETTER TO THE EDITOR

IN WEB OF SCIENCE WE TRUST

A CASE OF A HIJACKED JOURNAL INDEXED IN SCOPUS

ABSTRACT: SCOPUS database is one of the major bibliographic and one of the most respectable scientific databases. I present here a case of a hijacked Journal that managed to be indexed by major databases—SCOPUS and EBSCO. This fake journal does not charge for the publication which made it more difficult to discover. Certain security patches need to be applied in order to prevent these practices for happening in the future. This article should serve as a warning to authors submitting their work to journals, but more importantly to databases administrators who should pay more attention to these issues.

KEYWORDS: Scopus, Web of Science, fake journals, bibliographic data

HARIS MEMISEVIC
University of Sarajevo, Bosnia and Herzegovina

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.

- **Wolfgang Glänzel**, *Editor-in-Chief:* wolfgang.glanzel[at]kuleuven.be
- **Balázs Schlemmer,** *Managing Editor:* balazs.schlemmer[at]gmail.com
- **Sarah Heeffer,** *Assistant Editor:* sarah.heeffer[at]kuleuven.be

- **Judit Bar-Ilan**: barilaj[at]mail.biu.ac.il
- **Sujit Bhattacharya**: sujit_academic[at]yahoo.com
- **María Bordons**: mbordons[at]cchs.csic.es
- **Juan Gorraiz:** juan.gorraiz[at]univie.ac.at
- **Jacqueline Leta**: jleta[at]bioqmed.ufrj.br
- **Olle Persson**: olle.persson[at]soc.umu.se
- **Ronald Rousseau**: ronald.rousseau[at]kuleuven.be
- **Dietmar Wolfram**: dwolfram[at]uwm.edu

Accepted contributions are moderated by the board. Guidelines for contributors can be found at http://www.issi-society.org/editorial.html. Opinions expressed by contributors to the Newsletter do not necessarily reflect the official position of ISSI. Although all published material is expected to conform to ethical standards, no responsibility is assumed by ISSI and the Editorial Board for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material therein.

© International Society for Scientometrics and Informetrics

When you conduct a study, analyse your results, write a paper, you then begin the search for the appropriate Journal that will publish your masterpiece. Ideally, for the scholars from scientific periphery (Marusic & Marusic, 1999), that Journal should be indexed by SCOPUS, if not by Web of Science Current Contents base. Good research published in good journals is a must for scientists as it increases their chances for getting grants and tenure (Memisevic, Taljic, & Hadziomerovic, 2017). Most bibliometric analysis use two data sources for research evaluation: Web of science and SCOPUS (Mongeon & Paul-Hus, 2016). Thus, the need to publish in the Journal indexed by SCOPUS and Web of Science is of paramount importance.

This introduction serves as a basis for a case description and questionability and accuracy of the bibliographic databases, particularly SCOPUS.

My colleagues and I recently submitted a manuscript to the *Turkish Journal of Psychology*, a well-respected journal, published by Turkish Psychological Association and indexed in the Web of Science and SCOPUS databases, with the ISSN number of the Journal 1300-4433. Or that was where I thought we submitted the manuscript. What I did not know at that time was that there was a fraudulent website impersonating the real *Turkish Journal of Psychology* journal. I mistook the fake/hijacked journal (or website) for the true one, and of course, I sent my manuscript to the fake one. The fake journal (or I do not know if we should refer to it as a Journal!?) uses the Open Journal Systems software for the management of the "manuscripts" they receive, and the whole appearance looks pretty professional. So, the last step in the submission process was the question: Do you want to submit your manuscript to the Turkish Journal of Psychology and I clicked OK in hope my paper will be accepted and published in that Journal.

I waited for awhile and no changes in the status of the manuscript appeared. That made me suspicious and I wrote an email to the „Editor" asking for the manuscript status and of course, received no reply. Then I checked „the Editorial board" list of names and wrote an email to a professor whose name was in that list. I received an email in which the professor asked me whether I sent my manuscript to the real journal or the fake journal and provided me with the genuine website for the real *Turkish Journal of Psychology*. Wait a minute, there are two of them!? Well, actually, there is only one real journal but the other one is impersonating the real one with the same ISSN and all other details that could easily trick someone who is not familiar to this fraud scheme. So I checked the websites:

The real journal is on the webpage: http://www.turkpsikolojidergisi.com/

And the fake one is on this webpage: http://www.turkpsikolojidergisi.org

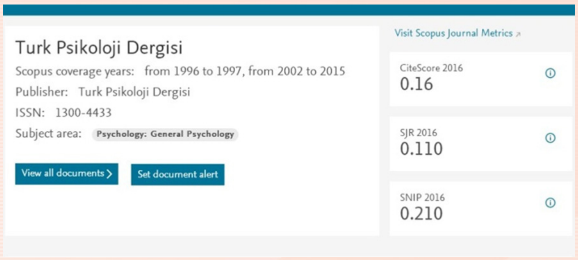
Ok, I accepted the fact that I sent my manuscript to the fake/hijacked Journal (again, should we refer to it as a Journal!?) and I wrote an email to them saying that I am retracting my paper from further consideration. Then I sent my manuscript to another journal, *Cogent Psychology*, where it was eventually accepted and published (Memisevic, Biscevic, & Pasalic, 2017). BUT... A couple of weeks after my paper was published by *Cogent Psychology*, I saw it again on Google Scholar, but this time it was published by *Turkish Journal of Psychology*. The fake *Turkish Journal of Psychology*. Of course I immediately sent an email to the *Cogent Psychology*, telling them about this scheme. I also pointed out, that only *Cogent Psychology* had a right to publish my paper and what appeared on website impersonating the *Turkish Journal of Psychology* does not have my consent. Then I wrote an email to the Editor of the real *Turkish Journal of Psychology* letting him know about this. His reply was that they are aware of the fake Journal and that they have taken some legal measures against the fake Journal. Exploring a little further, I saw an article in EBSCO database, which was published in the fake *Turkish Journal of Psychology* and not in the real one. I asked myself, how is it possible that the fake journal is able to upload manuscripts on the behalf of the real journal? Of course, I sent an email to

EBSCO asking them about this and they said they will look into it. I thought EBSCO does not have strict criteria and that is why such omission could happen. It can probably never happen to SCOPUS or Web of Science. And this morning (01. Feb, 2018), I see my paper, published on the website impersonating the *Turkish Journal of Psychology* appearing in the SCOPUS database. I immediately sent an email to SCOPUS letting them know about this. Exploring further, I discovered a number of articles on SCOPUS published by this fake journal Turkish Journal of Psychology. SCOPUS shows the genuine ISSN number for the Journal but somehow allows the fake Journal to upload the manuscripts (Figure 1).

So far, only Web of Science (WOS) resisted this sneaky "attack" by the fake Journal. I have not found any article "published" in the fake Journal that is covered by WOS. In Figure 2 we can see that WOS covers this Journal.

This article's purpose was to warn authors about this particular fraudulent website http://www.turkpsikolojidergisi.org and possible similar ones, that can literally take over the entire infrastructure of the legitimate journal and present their own content. What makes this Journal different from other hijacking Journals is that this one does not charge the authors submission/publication fees. It seems the sole purpose of the hijackers is to harm the real Journal and the authors who submit articles to their platform.

Given the high stakes as to getting the research funds, the information contained in databases need to be as correct as possible. Not to mention biases it can introduce to the scientific literature, as most studies, say meta-analysis, involve SCOPUS search. And we can see some serious flaws with SCOPUS indexation. As a conclusion, SCOPUS and EBSCO need to find security patches for these issues as soon as possible.



Turk Psikoloji Dergisi

Scopus coverage years: from 1996 to 1997, from 2002 to 2015

Publisher: Turk Psikoloji Dergisi

ISSN: 1300-4433

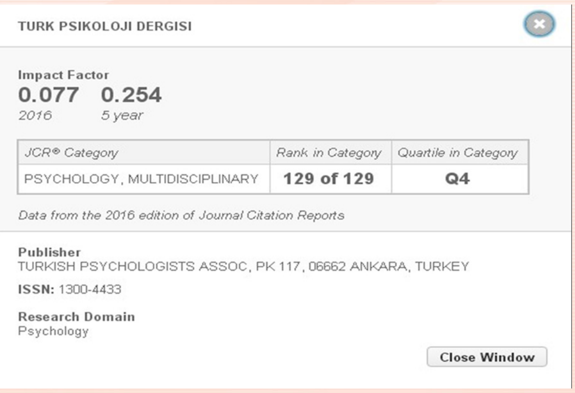
Subject area: Psychology: General Psychology

CiteScore 2016
0.16

SJR 2016
0.110

SNIP 2016
0.210

Figure 1. Screenshot of the Turkish Journal of Psychology metrics in SCOPUS



TURK PSIKOLOJI DERGISI

Impact Factor

0.077 2016

0.254 5 year

JCR® Category	Rank in Category	Quartile in Category
PSYCHOLOGY, MULTIDISCIPLINARY	129 of 129	Q4

Data from the 2016 edition of Journal Citation Reports

Publisher
TURKISH PSYCHOLOGISTS ASSOC, PK 117, 06662 ANKARA, TURKEY

ISSN: 1300-4433

Research Domain
Psychology

Figure 2. Screenshot of the Turkish Journal of Psychology in Web of Science

P.S. Although I first learned about my article in the fake *Turkish Journal of Psychology* through the Google Scholar, Google Scholar did not incorporate that article in my profile, although it automatically adds articles (or citations) to the profile. In addition to this, the Google does not "recognize" references in that "paper" in the Google Citation Index. Usually, Google is much faster to include papers and citations into someone's profile than SCOPUS. However, in this particular case, it seems that Google has better "filters" to recognize fake papers. We will see if the Google will keep ignoring the fake paper.

REFERENCES:

1. Marusic, A., & Marusic, M. (1999). Small scientific journals from small countries: breaking from a vicious circle of inadequacy. *Croatian Medical Journal*, 40(4), 508–514.
2. Memisevic, H., Biscevic, I., & Pasalic, A. (2017). Developmental trends in semantic fluency in preschool children. *Cogent Psychology*, 4(1), 1403064.
3. Memisevic, H., Taljic, I., & Hadziomerovic, A. M. (2017). Making Use of H-index: the Shape of Science at the University of Sarajevo. *Acta Informatica Medica*, 25(3), 187-191.
4. Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228.

STI2018 LEIDEN

23rd INTERNATIONAL CONFERENCE ON SCIENCE & TECHNOLOGY INDICATORS

12—14 SEPTEMBER 2018
LEIDEN, THE NETHERLANDS

CALL FOR PAPERS

Leiden – Photo courtesy of © Balázs Schlemmer :: schlemmerphoto.com ::

"SCIENCE, TECHNOLOGY AND INNOVATION INDICATORS IN TRANSITION"

#STI18LDN

The STI/ENID 2018 conference will be held 12-14 September 2018 in Leiden, The Netherlands. This edition will have a special focus on the discussion of "indicators in transition" as a driving force for more comprehensive, broader and socially oriented forms of Science, Technology and Innovation indicators and evaluations.

We welcome contributions on the general topics covered by the conference as well as contributions to the special tracks.

Contributions on, but not limited to, the following general topics:

- Altmetrics & social media (theoretical foundations, validation studies, Data sources, etc.)
- Careers in science (Gender and diversity, careers outside academia, early career researcher experience, etc.)
- Indicators of Science and Technology (responsible use of indicators, societal impact of research, systemic and behavioral effects of indicators, etc.)
- Innovation (gendered innovations, public-private interactions, industrial R&D dynamics, etc.)
- Open Science (Open access, Open data, Open science and the academic reward system, etc.)
- Research evaluation (responsible research evaluation, methods in research evaluation, case studies, etc.)
- Research integrity (policies promoting research integrity and their effects, misconducts in scientific , publishing, studies of other types of misconduct in research, etc.)

SPECIAL TRACKS

A brief description of the tracks can be found at http://sti2018.cwts.nl:

1. A closer look into corporate science and publishing | Roberto Camerani; Nicola Grassano; Laurens Patricia; Daniele Rotolo; Antoine Schoen; Robert Tijssen and Alfredo Yegros
2. Challenges in establishing macro-level effects of macro-level interventions | Jochen Gläser; Carolina Cañibano; Thomas Franssen; Grit Laudel and Jesper W. Schneider
3. Reproducibility in scientometrics | Sybille Hinze; Jason Rollins; Andrea Scharnhorst; Jesper W. Schneider; Theresa Velden and Ludo Waltman
4. Research assessments as participatory explorations on content, missions, methods and indicators | Noortje Marres; Ismael Rafols and Sarah de Rijcke
5. Assessment of Responsible Research and Innovation (RRI) – beyond indicator development | Ingeborg Meijer; Susanne Bührer; Erich Griessler; Ralf Lindner; Frederic Maier; Niels Mejlgaard; Viola Peter; Jack Stilgoe; Richard Woolley and Angela Wroblewski
6. Studies in the sociology and history of the sciences, social sciences, arts and humanities | Matteo Romanello; Giovanni Colavizza and Thomas Franssen
7. Scientific and technological novelty: impact and determinants | Jacques Mairesse; Fabiana Visentin and Michele Pezzoni
8. Determining and steering research quality in practice: the institutional research perspective | Cathelijn Waaijer; Ad Scheepers and Nynke Jo Smit
9. Open scholarship | Thed Van Leeuwen; Clifford Tatum and Paul Wouters
10. Public-private interactions in business innovation | Hugo Hollanders and Lili Wang
11. Challenges of social media data for bibliometrics | Katrin Weller; Astrid Orth and Isabella Peters
12. Rethinking the research agenda on the internationalization of the scientific workforce | Eric Welch; Julia S. Melkers; Nicolas Robinson-Garcia and Eric van Holm

IMPORTANT DATES

1 April 2018:
Deadline for submissions (papers and posters)
June 15, 2018
Notification of acceptance

SUBMISSIONS

Short paper (max 3,000 words) with a comprehensive description of a completed study; Poster (max 1,000 words) with an abstract of the study

ORGANIZATION

Conference chair: Paul Wouters | Local organizers: Rodrigo Costas, Thomas Franssen, Suze van der Luijt, Petra van der Weel, Alfredo Yegros

CONTACT / MORE INFO

e-mail: sti2018@cwts.leidenuniv.nl
web: http://sti2018.cwts.nl/

CARMA 2018

INTERNET AND BIG DATA IN ECONOMICS AND SOCIAL SCIENCES.

2nd INTERNATIONAL CONFERENCE ON ADVANCED RESEARCH METHODS AND ANALYTICS

JULY 12—13, 2018
VALENCIA, SPAIN

CALL FOR PAPERS

Research methods in economics and social sciences are evolving with the increasing availability of Internet and Big Data sources of information. As these sources, methods, and applications become more interdisciplinary, the 2nd International Conference on Advanced Research Methods and Analytics (CARMA) aims to become a forum for researchers and practitioners to exchange ideas and advances on how emerging research methods and sources are applied to different fields of social sciences as well as to discuss current and future challenges.

TOPICS OF INTEREST

Topics of interest include, but are not limited to, the following topic areas:

Internet and Big Data sources in economics and social sciences

- Google Trends and Search Engine data
- Web scraping
- Social media and public opinion mining
- Geospatial and mobile phone data

Big Data methods in economics and social sciences

- Sentiment analysis
- Internet econometrics
- Information quality and assessment
- Crowdsourcing

Internet and Big Data applications

- Official statistics
- Tourism forecasting
- Business analytics with social media
- Social behavior and mobility patterns
- Consumer behavior, eWOM and social media marketing
- Politics and social media
- Bibliometrics and sciencetometrics

Digital transformation and global society

- Privacy and legal aspects
- Electronic Government
- Smart Cities
- Industry adoption
- Gender bias

Participants from all over the world are expected to present their latest and unpublished research findings. The program committee encourages the submission of articles that communicate applied and empirical findings of interest to social sciences researchers.

CONFERENCE VENUE

The CARMA 2018 conference will be held on July 12-13, 2018 at the Faculty of Business Administration and Management of the Universitat Politècnica de València (UPV), which has been recently ranked as the best technical university in Spain by the Academic Ranking of World Universities (ARWU) 2017.

Valencia is the third largest city in Spain and is located on the shore of the Mediterranean Sea. It is the capital city of the Comunitat Valenciana region, which is major tourist destination in summer.

IMPORTANT DATES

Submission deadline:	23 March, 2018
Acceptance notification:	11 May, 2018
Camera ready due:	28 May, 2018
Conference:	12-13 July, 2018

SUBMISSION GUIDELINES

Authors from all over the world are invited to submit original and unpublished papers or extended abstracts, which are not under review in any other conference or journal. All submissions will be peer reviewed by the program committee based on their originality, significance, methodological soundness, and clarity of exposition.

Submissions (extended abstracts or full papers) must be written in English and should be in PDF format. They must follow the instructions in the template file, available in Microsoft Word format at: http://www.carmaconf.org/template.docx

Full-paper length must be between 4 and 8 pages (A4 size), incorporating all text, references, figures and tables. Extended abstracts (which will not receive a DOI) should not exceed 3 pages.

These guidelines are strict: papers failing to adhere to the guidelines (by being more than 8 pages, altering margins or not following the template) will be rejected without consideration of their merits. Submissions imply the willingness of at least one author to register, attend the conference, and present the paper.

ORGANIZING COMMITTEE

General chair

- Josep Domenech,
  Universitat Politècnica de València

Scientific committee chair

- María Rosalía Vicente,
  Universidad de Oviedo

Local arrangements chair

- Desamparados Blazquez,
  Universitat Politècnica de València

Scientific committee:

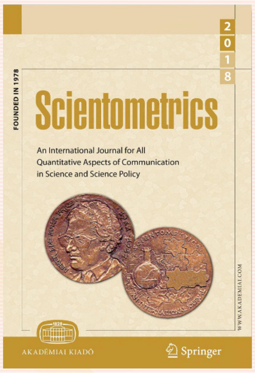
- Concha Artola, Nikolaos Askitas, Petter Bae Brandtzaeg, Jonathan Bright, José Luis Cervera, Piet Daas, Pablo de Pedraza, Giuditta de Prato, Rameshwar Dubey, Enrico Fabrizi, Juan Fernández de Guevara, Jose A. Gil, Felix Krupar, Caterina Liberati, Juri Marcucci, Rocio Martinez Torres, Esteban Moro, Enrique Orduña, Bulent Ozel, Ana Pont, Ravichandra Rao, Pilar Rey del Castillo, Anna Rosso, Vincenzo Spiezia, Pål Sundsøy, Sergio L. Toral Marin, Antonino Virgillito, Sang Eun Woo, Zheng Xiang

CONTACT / MORE INFO

e-mail:	secretariat@carmaconf.org
web:	http://www.carmaconf.org/

The organizing committee looks forward to welcoming you all to a fruitful conference with open discussions and important networking to promote high quality research.

SCIENTOMETRICS SPECIAL ISSUE—CALL FOR PAPERS



SCIENTOMETRICS

An International Journal for
All Quantitative Aspects of the Science of Science,
Communication in Science and Science Policy
Editor-in-Chief: Wolfgang Glänzel
ISSN: 0138-9130 (print version)
ISSN: 1588-2861 (electronic version)

SPECIAL ISSUE ON

SCIENTO-NETWORK-MINING AND SMART ALTMETRICS FOR ADVANCED SCIENTIFIC COLLABORATION AND KNOWLEDGE UTILIZATION

CALL FOR PAPERS

AIMS AND SCOPE

Many people these days are doing research in multidisciplinary areas. Scientometrics is the study of quantitative aspects of scientific research, library and information science using methods from other fields, such as, mathematics, statistics, computer science and network science. The main focus is usually on bibliometrics, webometrics and altmetrics. These days many social networks (e.g. Academic Social Networks (ASNs)) have emerged from professional interactions between academic people. Specifically, Sciento-Network-Mining is focused on using data mining techniques for dealing with scientometric tasks in ASNs.

We encourage submission of papers especially that are utilizing datasets of Academic Social Networks, such as, researchgate.net, mendeley.com, academia.edu and linkedin.com, but not limited to it. Bibliometric datasets, such as, scopus.com, dblp, aminer.org or similar can also be used to perform various types of analysis in academic domain.

TOPICS OF INTEREST:

Researchers are invited to submit papers focusing on Data Mining techniques such as:

- Frequent Pattern Mining or Association Discovery
- Classification and Prediction
- Clustering
- Time Series Analysis
- Ranking
- Ontologies
- Social Network Analysis and Mining
- Deep Learning

To deal with following scientometrics tasks:

- Anomaly Detection / Group Anomaly Detection
- Author Order Patterns Mining /Author Contribution Detection
- Finding Rising Stars / Predicting Rising Stars / Rising Research Area Detection
- Finding Author Collaboration Patterns
- Influential Authors Finding / Influential Authors Prediction
- Expert Finding / Author Ranking / Author Indexing / Expert Prediction
- Sub Community Detection
- Author Name Disambiguation
- Link Prediction / Reciprocal Link Prediction
- Author Profiling
- Citations Prediction / Finding the correct references cited in papers / Reference Prediction

Note: We encourage submission of papers that are utilizing open data or that make their datasets available online.

SUBMISSION PROCESS

Submissions to this special issue should follow the journal's guidelines for submission and can be submitted to the guest editor via email (nraljohani@kau.edu.sa) for initial review process. Initial submissions should be in PDF format but the final paper (after acceptance) should be submitted via Journal Submission System. All papers submitted to this special issue will be reviewed by three reviewers who will be experts of the domain of submitted paper. Every submitted paper should clearly state the purpose and technical issues addressed in the submitted paper. Also, if a paper is an extension to a prior workshop or conference publication, the journal submission must demonstrate substantial differences/contributions.

KEY DATES

Submission deadline:	March, 31, 2018
Notification:	May, 25, 2018
Revision Due:	June 30,2018
Estimated Publication:	Q4 2018

GUEST EDITORS

- Naif Radi Aljohani,
  King Abdulaziz University,
  Saudi Arabia
  (nraljohani@kau.edu.sa)

- Ali Daud,
  King Abdulaziz University,
  Saudi Arabia
  (adfmohamad@kau.edu.sa)

- Miltiadis D. Lytras,
  The American College of Greece,
  Greece
  (mlytras@acg.edu)

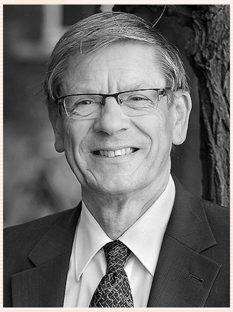
- Ahtisham Aslam,
  King Abdulaziz University,
  Saudi Arabia
  (maaslam@kau.edu.sa)

- Saeed Ul Hassan,
  Information Technology University,
  Pakistan
  (saeed-ul-hassan@itu.edu.pk)

CONTACT

Contact for further information:
(nraljohani@kau.edu.sa)

BLOCKCHAIN TECHNOLOGY: A BIBLIOMETRIC ANALYSIS



RONALD ROUSSEAU
University of Antwerp, Belgium
ECOOM, KU Leuven, Belgium
ronald.rousseau@kuleuven.be

Abstract: In this contribution we perform an elementary citation analysis related to the blockchain technology, the technology underlying the bitcoin currency. In order to sketch the framework we first provide a basic introduction to this technology. More importantly we point out that this technology has the potential to transform ownership, traceability, incentives and policymaking. As such its potential influence on research and publishing cannot be underestimated.

Keywords: bitcoin, blockchain technology, cryptocurrencies, informetrics

1. WHAT IS BLOCKCHAIN TECHNOLOGY: AN INTRODUCTION

Before going into details, let me first say that I am not a specialist in computer science or blockchain technology. Consequently, most of the information in the first five sections is taken from the sources acknowledged at the appropriate places. That said, I am very interested to find out how this technology will further develop in the real world and how scientists will use it in their investigations.

Let me begin with a few words about the history of the blockchain technology and the bitcoin. Satoshi Nakamoto (2008) – probably a pseudonym – invented the blockchain technology as the underlying technology for the bitcoin. The bitcoin itself was launched in 2009. This technology deals with the distribution of value, such as money or property rights, without a trusted third party, such as a bank, a governmental office, a lawyer or a notary.

The blockchain technology is as an application of cryptography, the practice and study of techniques for secure communication in the presence of adversaries. Cryptographic techniques allow for the protection of sensitive information (organizational, institutional or personal), either in storage or in communication. A blockchain itself is a continuously growing list of records, called blocks, which are linked and secured using cryptography. Roughly speaking a blockchain is an accountancy system of transactions. This accountancy system has two properties which makes it different from a traditional accountancy system. First, what is added to the system can never be removed, and second, there does not exist a unique copy (or a limited number of copies) of a transaction but a large and decentralized number of identical copies. For this reason there can never be any discussion about the contents of the blockchain and, if for whatever reason, one or a few copies are destroyed, there are still plenty of them so that information never gets lost. This decentralized nature is one of the key aspects of the blockchain technology.

2. SOME DEFINITIONS AND EXPLANATIONS

Before dealing with the blockchain and the bitcoin we first explain some terms used in this context: ledgers, public-key cryptography, hash functions, nonces and proof-of-work.

TRADITIONAL VERSUS DIGITAL, DISTRIBUTED LEDGERS

We recall that in common word use a ledger is a book containing accounts to which, e.g., debits and credits are posted. As such, ledgers have been at the heart of commerce since ancient times and are used to record many things, most commonly assets such as money and property. Walport (2016) notes that through history they were recorded on clay tablets, papyrus, vellum or paper. However, one may say that in all this time the only notable innovation has been computerization, which initially was simply a transfer from paper to bytes. Now, however, algorithms enable the collaborative creation of digital, distributed ledgers with properties and capabilities that go far beyond traditional paper-based ledgers.

A distributed ledger is essentially an asset database that can be shared across a network of multiple sites, geographies or institutions. All participants within a network can have their own identical copy of the ledger and any changes to the ledger are reflected in all copies. The assets can be financial, legal, physical or electronic. The security and accuracy of the assets stored in the ledger are maintained cryptographically through the use of keys and signatures to control who can do what within the shared ledger. Entries can also be updated by one, some, or all of the participants, according to rules agreed by the network (Walport, 2016).

This leads us to a short discussion of different types of digital ledgers. On the one hand we have public, decentralized ledgers which are accessible to every Internet user. We will see that the bitcoin belongs to this category. On the other extreme we have the fully private ledger, where write-permissions are monitored by a central locus of decision-making. Besides write-permissions there are also read-permissions involved, which may either be public or restricted. A private blockchain amounts to a permissioned ledger, whereby an organizational process enables the whitelisting (or blacklisting) of user identities. The difference between public and private blockchains is the extent to which they are decentralized, or ensure anonymity. Between the two extremes, there exists a continuum (Brown, 2015; Allison, 2015) of "partially decentralized" blockchains rather than a strict public/private dichotomy. Although the bitcoin belongs to the public part, many future applications will probably belong to the private or partially decentralized part.

KEYS, HASHING AND PUBLIC-KEY CRYPTOGRAPHY

In cryptography, a key is a piece of information that determines the output of a cryptographic algorithm (https://en.wikipedia.org/wiki/Key_(cryptography)). For encryption algorithms, a key specifies the transformation of plain text into cipher text, and vice versa for decryption algorithms. Keys also specify transformations in other cryptographic algorithms, such as digital signature schemes. An attacker who obtains the key (by, for example, theft, extortion, assault, torture, or social engineering) can recover the original message from the encrypted data, and issue signatures. Encryption algorithms which use the same key for both encryption and decryption are known as symmetric key algorithms. A newer class of "public key" cryptographic algorithms was invented in the 1970s. These asymmetric key algorithms use a pair of keys, a public key and a private one. Public keys are used for encryption or signature verification; private ones decrypt and sign. The design is such that finding out the private key is extremely difficult, even if the corresponding public key is known. The best known public-key cryptographic algorithm is the RSA algorithm (Rivets et al., 1978).

The result of an encryption is often called a hash and the action of performing encryption is often referred to as hashing. Recall that the verb 'to hash' means to chop something up. A hash function is a mathematical algorithm that takes an input and transforms it into an output. A cryptographic hash function such as the one used in the RSA encryption scheme, is characterized by its extreme difficulty to revert, in other words, to recreate the input data from its hash value alone.

THE RSA PUBLIC-KEY ENCRYPTION SCHEME

In this section we recall the RSA public-key cryptosystem, largely taken from the original source (Rivest et al., 1978). In this way we provide a simple example of hashing and of signing. This is useful to understand the blockchain.

The RSA algorithm involves four steps: key generation, key distribution, encryption and decryption. But first we explain the notion of a public-key cryptosystem. In a public-key cryptosystem each user places their encryption procedure E in a public file, hence the name of this cryptosystem. However, the user keeps the details of the corresponding decryption procedure D secret. The whole procedure has four properties:

(a) Deciphering an enciphered message M yields M, i.e., D(E(M)) = M.
(b) E as well as D are easy to compute.
(c) Publicly revealing E does not reveal D.
(d) First deciphering and then enciphering a message M returns the original message M, i.e. E(D(M)) = M.

Properties (a) and (d) mean that the operations E and D are each other's inverse.

This cryptosystem is used for sending messages and for signatures, which is performed as follows. Alice wants to send a secret message, M_A, to Bob. Encryption and decryption functions are denoted as E_A, D_A, E_B and D_B, depending on the owner. Now Alice encrypts her message using Bob's public key, leading to $E_B(M_A)$. Now Bob, who is the only person knowing D_B performs $D_B(E_B(M_A))$ and reads M_A. Suppose now that Alice wants to sign a document C. Then she performs $D_A(C)$. There is only one E, namely E_A, which leads to $E_A(D_A(C)) = C$. As E_A is public this means that anyone can check that indeed Alice has signed the document.

Practically, in the RSA system these properties are realized as follows. The public encryption key is a pair (e,n) of positive integers.

First, the message M is represented as an integer between 0 and n-1. If necessary, the message is broken up into blocks of the required length. Hence, M is now an integer. Encryption is performed as follows:

$$E(M) = M^e \text{ (mod n)} \qquad (1)$$

This means that E(M) is the remainder of the division of M^e by n. If we denote E(M) by C (the enciphered message), then deciphering is done as follows:

$$D(C) = C^d \text{ (mod n)} \qquad (2)$$

The encryption key is the pair (e,n) and the decryption key is the pair (d,n). The public-key cryptosystem works if knowledge of n and e does not help an attacker in finding d. This leads to the problem of choosing the keys. The integer n must be the product of two, randomly chosen, large prime numbers: n = p*q. Recall that n is public, but when n is large enough it becomes practically impossible to find p and q. To make factoring harder (for an attacker) the primes p and q should be similar in magnitude but differ in length by a few digits. The integer d is a large integer which is relatively prime to (p-1)*(q-1). This means that the greatest common divisor of d and the product (p-1)*(q-1) is 1. Finally e is determined as the inverse of d modulo (p-1)*(q-1), i.e.

$$e*d = 1 \ (\mathrm{mod}(p-1)*(q-1)) \qquad (3)$$

In their paper Rivest, Shamir and Adleman show that this method satisfies the four requirements for a safe public-key cryptosystems and provide a simple example. Of course, since the publication of the original paper the basic RSA-algorithm has been refined to protect against many types of attacks.

THE DOUBLE SPENDING PROBLEM

The double spending problem is the following problem. If you have a digital asset, such as digital money, and you want to give it to somebody else, how can one prevent you from giving it to two different people at (almost) the same time? As this asset is digital it is, indeed, easy to make copies. We will show how the bitcoin solves this problem.

A NONCE

A nonce is an arbitrary (random) number that can only be used once.

PROOF-OF-WORK

A proof-of-work (POW) system is a measure to prevent or at least make it difficult to abuse a service. The goal is reached by requiring some work from the service requester. The concept, if not the term, was invented by Cynthia Dwork and Moni Naor (1999) in the context of preventing spam. Indeed, these authors claim that computational costs deter junk mail but do not interfere with other uses of the system. The main idea is for the mail system to require the sender to compute some moderately expensive, but not intractable, function of the message and some additional information.

This idea was further worked out by Black who proposed the so-called Hashcash algorithm (Black, 2002). Computation is performed using a cost-function. Its outcome, in this context referred to as a token, should be easily verifiable, but moderately expensive (in time or in another commodity) to compute. Preferably this function has a parameter so that, if necessary, the difficulty related to its computation can be made to increase. Black calls this cost-function MINT because of the analogy between creating cost tokens and minting physical money. Later, Nakamoto used a similar cost-function to mint bitcoins.

3. WHAT IS A BLOCKCHAIN TECHNOLOGY?

The blockchain technology is a cryptographic process involving a network of computers, referred to as miners. Its main purpose is to record the existence of digital objects and to organize their transactions. We already point out that the basic blockchain approach as used in the introduction of the bitcoin (Nakamoto, 2008), can be modified to incorporate rules, smart contracts, digital signatures and an array of other new tools.

After the introduction of the bitcoin, scientists realized that the essence of the blockchain is actually informational and processual, and does not necessarily relate to the monetary sphere. In this sense, blockchains may exist without an underlying token or coin.

In the blockchain each digital record is turned into a unique string of letters and numbers called a hash (which can be seen as a unique fingerprint) and inserted into a transaction. A transaction is initiated when the future owner of the digital object sends his/her public key to the original owner. The object is transferred with a digital signature. Transactions are broadcasted to a network of miners (the nodes in the network) who check them. Miners turn pending transactions into a block including the hash of the previous block, a time stamp and a random number (a nonce) (Pilkington, 2016). From this statement we note one of the main properties of the blockchain technology, namely that it leads to distributed consensus among participating nodes. In this way the blockchain technology is able to remove the need for a trusted third party to guarantee a transaction.

4. THE BITCOIN: SOME MORE DETAILS

Bitcoin is the special case that the digital record represents monetary value. It was the first decentralized public ledger, and has acquired a global status.

We first point out the steps to run the bitcoin network – a special peer-to-peer network – taken from (Nakamoto, 2008) and next provide some details. Nakamoto proposes the following steps:

1) New transactions are broadcast to all nodes.
2) Each node collects new transactions into a block.
3) Each node works on finding a difficult proof-of-work for its block.
4) When a node finds a proof-of-work, it broadcasts the block to all nodes.
5) Nodes accept the block only if all transactions in it are valid and not already spent.
6) Nodes express their acceptance of the block by working on creating the next block in the chain, using the hash of the accepted block as the previous hash.

The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proves that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they will generate the longest chain and outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone. Bitcoin mining is the process of adding transaction records to bitcoin's public ledger of past transactions. This ledger of past transactions is called the blockchain as it is a chain of blocks. The blockchain serves to confirm to the rest of the network that certain transactions have taken place. This approach provides a solution to the double-spending problem.

BITCOIN TRANSACTIONS

To send bitcoins, you need two things: a bitcoin address or wallet (the public key), and a private key because the blockchain includes a public-key encryption scheme similar to the RSA one. A bitcoin address is generated randomly, and is simply a sequence of letters and numbers. The private key is another sequence of letters and numbers, but unlike the bitcoin address, this is kept secret. A transaction is initiated when either, the owner looks up the bitcoin address of the future owner, or, the future owner of the coin sends his/her public key to the original owner, asking him/her for money. Every coin is associated with an address, and a transaction in the crypto-economy is simply a trade of coins from one address to another. Note that there are no physical bitcoins or even digital ones: only records of bitcoin transactions. Another striking feature of the blockchain is that public keys are never directly tied to a real-world identity. Transactions, although traceable, are enabled without disclosing one's identity. This is a major difference with transactions in real-world currencies that, with the exception of (non-traceable) cash transactions, are related to specific economic agents endowed with legal personality, such as banks (Pilkington, 2015).

What does a transaction look like?

If Alice sends some bitcoins to Bob, that transaction will have three pieces of information:

- An input, stating which bitcoin address was used to send the bitcoins to Alice in the first place. This secures the chain of transactions.
- An amount. This is the amount of bitcoins that Alice is sending to Bob. Note that one may send more than one bitcoin and that a bitcoin can be split into 100,000,000 pieces. Each such piece, i.e. 0.00000001 bitcoin, is called a satoshi.
- An output or target address. This is Bob's bitcoin address. Recall that this address is public.

To actually send bitcoins, you moreover need your own private key. When Alice wants to send bitcoins to Bob, she uses her private key to sign a message with the input (the source of the coins), the amount, and the target (Bob's address).

She then sends them from her bitcoin wallet out to the wider bitcoin network and all peers trying to solve blocks collect the transaction records and add them to the block they are working to solve. Miners verify and confirm transactions and get an incentive for doing this because of attached transaction fees.

A transaction in the bitcoin world is final once it is included in the blockchain, thereby becoming simultaneously verifiable by many sources. These fully decentralized blockchains rest on a consensus mechanism of proof-of-work for validation purposes: in the case of bitcoin, the "longest chain – the chain with the most proof-of-work – is considered to be the valid ledger (Swanson, 2015, p.4).

HOW TO MAKE NEW BITCOINS?

We will not discuss how to mine for bitcoins on your own, leaving that to others. But we will explain the main idea. The bitcoin blockchain is a chain of transactional records enriched by a subset of so-called miners who solve difficult computational problems. Miners anonymously compete on the network to solve a mathematical problem, thereby adding the next block to the blockchain. The reward for finding this next block, namely 'newly minted' coins, is sent to the miner's public address. Miners may spend these coins at will, using their private key. However, mining cannot go on forever. When the bitcoin algorithm was created a finite limit on the number of bitcoins that will ever exist was set at 21 million. Currently (January 2018), there are about 18 million and 800,000 bitcoins in circulation. That means that slightly more than two million bitcoins are still to be discovered. New bitcoins must show a proof-of-work to be accepted. This proof-of-work (PoW) is the so-called Hashcash PoW (Pilkington, 2015) proposed by Black (2002). For verifying transactions, and calculating proof-of-work, bitcoin relies on a specific hashing function, called the double SHA256 hashing algorithm, wherein the target is a 256-bit number (a number of the order of 10^{168}). To be accepted by the network the SHA256 hash of a block's header must be lower than or equal to the current target for the block. The lower the target, the more difficult (and processing time consuming) it is to generate a new block. For a block to be valid, it must result in a hash value less than the current target.

The proof-of-work involves randomly searching – as there is no mathematical algorithm – for a value that when hashed with SHA-256, the hash begins with a certain number of zero bits. The average work required is exponential in the number of zero bits required and can be verified by executing a single hash. Searching involves incrementing a nonce in the block until a value is found that gives the block's hash the required zero bits. Once the CPU effort has been expended to make it satisfy the proof-of-work, the block cannot be changed without redoing the work. As later blocks are chained after it, the work to change the block would include redoing all the blocks after it.

To compensate for increasing hardware speed in the real world and varying interest in running nodes over time, the proof-of-work difficulty is determined by a moving average targeting an average number of blocks per hour. If they are generated too fast, the difficulty increases. Similarly, when system-wide mining power increases, so does the difficulty of the computational problems required to mine a new block (Böhme et al., 2015, p. 218). This difficulty level is adjusted to keep the pace with which new blocks are generated constant at roughly one per ten minutes (Dwyer, 2014, p. 5).

Nodes always consider the longest chain to be the correct one and will keep working on extending it. If two nodes broadcast different versions of the next block simultaneously, some nodes may receive one or the other first. In that case, they work on the first one they received, but save the other branch in case it becomes longer. The tie will be broken when the next proof-of-work is found and one branch becomes longer; the nodes that were working on the other branch will then switch to the longer one.

We already know that in the blockchain, bitcoins are registered to bitcoin addresses. Creating a bitcoin address is nothing more than picking a random valid private key and computing the corresponding bitcoin address. This computation can be done in a split second. But the reverse (computing the private key of a given bitcoin address) is mathematically unfeasible and so users can make a bitcoin address public without compromising its corresponding private key. Moreover, the number of valid private keys is so vast that it is extremely unlikely someone will compute a key-pair that is already in use and has funds. The vast number of valid private keys makes it unfeasible that brute force could be used for that. To be able to spend the bitcoins, the owner must know the corresponding private key and digitally sign the transaction. The network verifies the signature using the public key (recall the example of the RSA-system).

If the private key is lost, the bitcoin network will not recognize any other evidence of ownership: the coins are then unusable, and effectively lost. With no central bank backing bitcoins, there is no possible way to recoup losses. Besides the bitcoin other cryptocurrencies have been invented, the so-called altcoins (this naming may sound familiar to informetricians). Besides cryptocurrencies, there exist other applications of the blockchain, among which Ethereum, a blockchain-based platform for smart contracts, is probably the best known.

ANONYMITY (BRITO & CASTILLO, 2013)

Media have given a great deal of attention to the so-called anonymity of the bitcoin. Yet, reality is less simple. On the one hand, bitcoins are like cash in that once Alice gives bitcoins to Bob, she no longer has them and Bob does, and there is no other party that knows their identities. While the public keys for all transactions are recorded in the blockchain, those public keys are, indeed, not tied to anyone's identity. On the other hand, unlike cash, the fact that a transaction took place between two public keys, the time when it happened, the amount that was transferred, and other information is recorded in the blockchain.

If a person's identity were linked to a public key, one could look through the recorded transactions in the blockchain and easily see all transactions associated with that key. For this reason, bitcoin is not anonymous, but pseudonymous at best.

Tying a real-world identity to a bitcoin address is not as difficult as one may imagine. For one thing, a person's identity, such as an IP address, is often recorded when the person makes a bitcoin transaction. Moreover, it is also possible to guess identities simply by looking at the blockchain. Brito and Castillo (2013) mention that in an experiment the identities of 40 percent of bitcoin users were discovered (Androulaki et al., 2012). Moreover, it has been shown several times that studying the bitcoin transaction graph with the appropriate tools can lay bare the financial activities and identities of bitcoin users. We conclude that it is very difficult to stay anonymous in the bitcoin network and pseudonyms tied to transactions recorded in the public ledger can still be identified years after an exchange is made.

Finally, a few words about the bitcoin in the real world. Can it be considered as 'money'? It is not because some newspapers refer to the bitcoin as a currency, that it actually is. "Real" money has three properties: it can be exchanged for something else (something of value, such as a loaf of bread or a house); it can be used to store value (the value of the house you sold can be stored in money) and finally, it has a sufficiently stable value. Although the bitcoin satisfies the first two requirements (at least to some extent), it is the third requirement where things go wrong. We know that even official currencies such as the dollar or the euro are not always completely stable and hence do not reflect a fixed value, yet having a stable value is not at all the case for the bitcoin or any other cryptocurrency. In a sense the bitcoin behaves more like precious metals such as gold, digital gold to be more precise, which is also to a great extent market-dependent (Krugman, 1984).

5. POSSIBLE APPLICATIONS

Nowadays it is more and more realized that algorithms that enable the creation of a blockchain are powerful, disruptive innovations that could transform the delivery of public and private services and enhance productivity through a wide range of applications (Walport, 2016).

Chapron (2017) points to four specific areas in which the blockchain technology could be used: ownership, traceability, incentives and policymaking. We provide some examples, based on Chapron's article.

Proven ownership of fishing or hunting rights or the right to protect animals, such as fish, may prevent selling these rights or denying their existence by corrupt governments. Traceability starts with humans, leading to undeniable birth certificates (they cannot be lost anymore), but, of course, includes tracing physical goods throughout their life cycle. Another example is tracking the origin of green electricity (Fouquet, 2017). Chapron mentions that by using a portable DNA sequencer illegally traded animal or plant parts can be spotted. The blockchain could ensure that conservation and development funding is used as intended (a strong incentive to do so). If insurance money must be paid, e.g. for crop damages, payments can be made with minimal delay, although officials are still needed to assess damages. Scientific advice to cities could be organized along a blockchain framework (Acuto, 2018). Finally, a public, shared and immutable register of assets and transactions can help to hold politicians accountable for their actions.

Traceability and ownership are essential for business enterprises. Not surprisingly large companies such as IBM offer partners a form of private blockchain to track their goods, see https://www.ibm.com/blockchain/. In such blockchains, identities are known and no cryptocurrencies are involved.

RESEARCH AND THE BLOCKCHAIN

In a research context the blockchain could help solve the reproducibility crisis, reduce the power of publishing giants and improve peer review (Van Rossum, 2017). In a 'blockchained' science, performing and communicating science would look very different from what happens nowadays. Indeed: blockchains allow for decentralised, self-regulating data and create a shared infrastructure where all transactions are saved and stored. As scientific information is essentially a large, dynamic body of information related to data that is collaboratively created, altered, used and shared, it lends itself perfectly to the blockchain technology. Working within a blockchain context would mean that whenever researchers create content or interact with it, this action is stored in a single decentralized platform. In this way, everyone has access to the same information. Moreover, in a blockchain for research, critical aspects of scholarly communication such as trust, credit and universal access can be realised and safeguarded. "Blockchained science" would make larger parts of the research cycle open to self-correction, and has therefore the potential to address the reproducibility and credibility crisis (Van Rossum, 2017).

A blockchain could moreover provide a notarization function by allowing scientists to post a text or file with ideas, results or basic data. These time-stamped records would allow researchers to claim to be the origin of a piece of information or of some idea. Such records could potentially replace the function of patent offices. Moreover, researchers would be encouraged to think more freely and share ideas that cannot immediately be placed in contemporary paradigms. Division of labour or specialization would become streamlined: some labs collect the data; others carry out the statistical analysis, etc. This framework could clearly increase the potential for collaboration (Bartling, 2017; Van Rossum, 2017).

Now we turn to the peer review process. A blockchain framework could not only improve reproducibility in general, but would also allow reviewers to do their work more thoroughly as they have more information available to judge originality. Encryption would allow reviews to be validated but in this way they remain anonymous and stored permanently. Moreover, post-publication review in various forms could be integrated easily (Bartling, 2017; Van Rossum, 2017).

DISSEMINATING CONTENT

Van Rossum (2017, p.10) writes:

> *One of the main roles of a publisher is the dissemination of content. After manuscripts are reviewed and accepted by the editorial board, publishers distribute this content to the academic community. Today, this happens largely through online platforms with subscriptions or open access fees as underlying business models. But blockchain holds the promise to change how publishers serve as middlemen in the dissemination process.*

In recent times the possible role of the blockchain in publishing has been investigated predominantly in non-academic publishing, where the move to online has led to a shift in revenue allocation from content creators and publishing companies to hosting companies, social media giants, and advertising intermediates (Van Rossum, 2017). The original business model followed logically from the structure of the Web, which consists of one-way pointers (hyperlinks). Hence, there is no immediate mechanism for allowing small automatic payments for usage. Given this, the only choice for publishers is to impose unfriendly paywalls with expensive forms of payments or to open up content and base their business model on advertising. In a blockchain model for scientific communication this business model could be a thing of the past. Indeed, several applications have been developed that allow for content distribution coupled with micropayments that flow directly to the producers of content (Van Rossum, 2017).

Another interesting potential dimension of the blockchain is digital rights management (Van Rossum, 2017). The coupling of usage to micropayments already makes rights management more straightforward, but digital rights can also relate to more complex aspects like re-use, permissions and royalties that are currently intermediated through large institutions. Here, the combination of a central database with smart contracts could bring huge advantages. Through the blockchain, ownership of content is automatically established, and the use of content and the payment of royalties are executed through smart contracts in which the rights are stored.

A more comprehensive reform of academic endorsement has been proposed in the manifesto ‘Towards Open Science: The Case for a Decentralized Autonomous Endorsement System’, published under a blockchain hash as author name (b8d5ad9d974a44e7e2882f986467f4d3, 2016). The author(s) propose a new academic endorsement system that is not based on current journal publication practices which are argued to be expensive, slow, disregard non-traditional output and negative results, and which give too much power to editors and publishers. Built on the blockchain, the Academic Endorsement System (AES) is based on a new form of currency, named

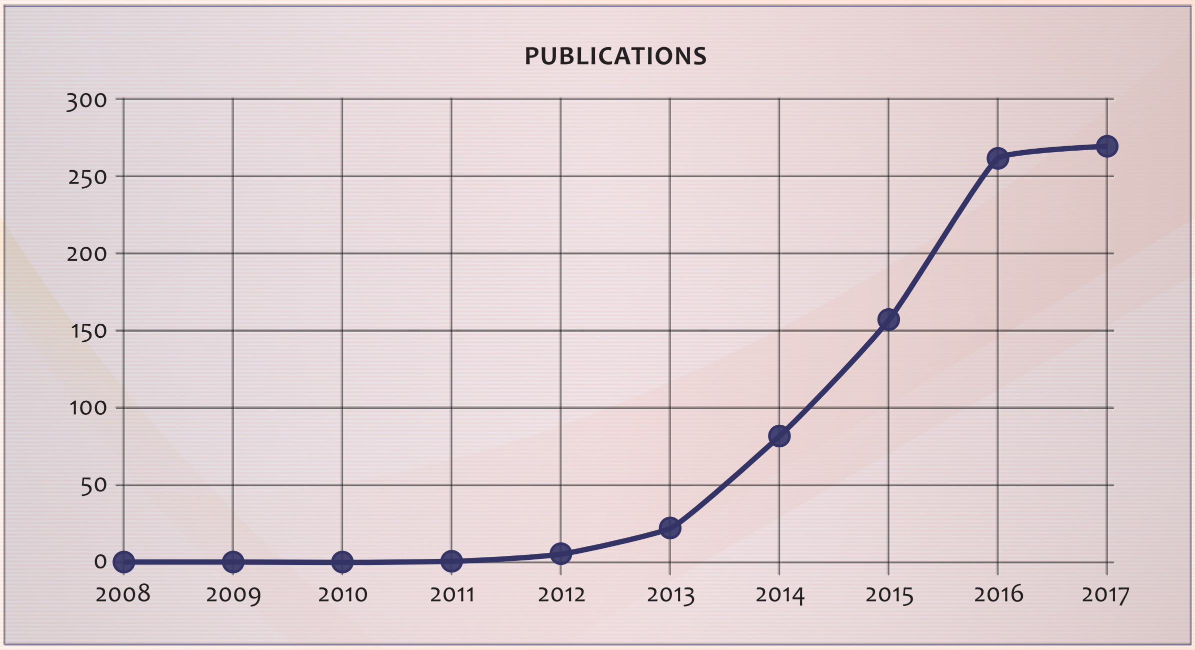


Figure 1. Yearly publications related to blockchain technology (WoS data).

academic endorsement points (AEP), which can be used by scientists to reward scientific work that is worthy of endorsement. Researchers whose output has been endorsed to a high degree will have a larger influence in the community. Any kind of research output could be endorsed including blog posts, data sets, software etc.

6. A BIBLIOMETRIC ANALYSIS

As we are interested to find out how academic authors have reacted to recent developments related to the blockchain we performed the following query in the Web of Science (WoS), on January 4, 2018.

TS = (bitcoin* OR blockchain* OR cryptocurrenc* OR ethereum OR "block chain algorithm*" OR "block chain technolog* " OR ("block chain" AND bitcoin*)).

We did not try to fully cover all aspects of blockchains or cryptocurrency but are convinced that we were able to capture the main ones, at least those included in the WoS. We also note that the query TS="block chain" on its own gave many false positives. The final query resulted in 800 publications (3 had 2018 as publication date and are not shown on Fig.1) with an average number of 1.6 citations and an h-index of 17. Yearly publications (all types) are shown in Figure 1. Taking into account that the year 2017 is not complete, this table suggests an exponential increase. As Nakamoto wrote his article on the bitcoin in 2008, we start the time axis in the year 2008.

Table 1 shows the ten countries with the most publications. For simplicity we used whole counts (if a publications is written by authors with addresses in three different countries, then each country receives a score). Moreover, we follow the WoS in considering England, Scotland, Wales and Northern Ireland as four different regions. We moreover compared rankings obtained from our blockchain query with the ranking based on a query for all publications in the research area of Computer Science. Results are shown in Table 1.

Some countries such as the USA, England and Switzerland are, relatively speaking, more interested in blockchain technology than in computer science in general, while the opposite holds for China, Spain and Japan.

Table 2 shows publication types and the number published for each of them. Not surprisingly, proceedings papers lead the rankings. Also the relative high number of editorials and news items catch the eye.

There are no institutes with a high number of WoS publications on blockchain technology. Cornell (USA) leads with 16 publications. When it comes to *Sources Titles,* the Lecture Notes in Computer Science (107 records) lead by far. Similarly, for *Research Areas,* it is no surprise that computer science leads (see Table 3). The fact that areas in the social sciences and humanities occupy ranks 7 to 10 may be somewhat of a surprise (but numbers are small). Given the legal implications of the existence of the bitcoin, the position of Government Law may be less of a surprise.

We also search for the most cited articles, but found that none of the publications on blockchain technology included in the WoS is highly cited. This is illustrated in Table 4.

Not surprisingly, the most-cited articles about the blockchain deal with the bitcoin. In particular they discuss the so-called Silk Road, an online black market, launched in 2011, for selling drugs and other illegal goods. As part of the dark web it was operated in such a way that users were able to browse it anonymously. Selling and buying were conducted with bitcoins. Yet it was shut down by the FBI in October 2013. It then re-emerged as Silk Road 2.0, but was again shut down by the FBI and Europol on 6 November 2014. A new version, Silk Road 3.0, went offline in 2017 due to loss of funds.

Although these articles are generally poorly cited, Nakamoto's paper (not in the Web of Science) is much more cited. We found 403 citations (in the WoS), mostly recorded for *Nakamoto S. bitcoin Peer to Peer* (and some variations), but some also for *Nakamoto S., consulted, freely available, technical report* or *working paper*.

Nakamoto's article received 2312 citations according to Google Scholar; Pilkington's 2015 contribution received 89 citations and The "Silk Road" article by Van Hout and Bingham (the most cited one on the WoS) received 99 citations in Google Scholar.

Table 1. Countries with the most publications on blockchain technology, compared with computer science in general (period: 2009-2017).

Countries	# Publications on blockchain	Rank: blockchain	Rank: computer science
USA	233	1	2
ENGLAND	94	2	6
PEOPLES R CHINA	68	3	1
GERMANY	45	4	4
AUSTRALIA	40	5	13
FRANCE	32	6	5
SWITZERLAND	29	7	18
CANADA	28	8	9
ITALY	27	8	10
SOUTH KOREA	21	10	11
INDIA	19	11	8
SPAIN	19	11	3
JAPAN	17	13	7

Table 2. Types of publications on blockchain technology

Type of publication	Number of publications
PROCEEDINGS PAPER	370
ARTICLE	334
EDITORIAL MATERIAL	51
NEWS ITEM	21
BOOK REVIEW	11
LETTER	8
REVIEW	8
CORRECTION	6

Table 3. Number of publications on blockchain technology per research area

Research Areas	# publ.
1 COMPUTER SCIENCE	412
2 BUSINESS ECONOMICS	152
3 ENGINEERING	113
4 SCIENCE TECHNOLOGY OTHER TOPICS	58
5 TELECOMMUNICATIONS	58
6 GOVERNMENT LAW	44
7 INFORMATION SCIENCE LIBRARY SCIENCE	14
8 SOCIAL SCIENCES OTHER TOPICS	13
9 PHILOSOPHY	9
10 ARTS HUMANITIES OTHER TOPICS	8

© International Society for Scientometrics and Informetrics

Table 4. Most-cited articles related to blockchain technology in the WoS (PY stands for Publication Year)

Authors	Title	Source	PY	# Cit.
Van Hout, Marie Claire; Bingham, Tim	'Silk Road', the virtual drug marketplace: A single case study of user experiences	INTERNATIONAL JOURNAL OF DRUG POLICY	2013	45
Miers, Ian; Garman, Christina; Green, Matthew; Rubin, Aviel D.	Zerocoin: Anonymous distributed E-cash from bitcoin	2013 IEEE SYMPOSIUM ON SECURITY AND PRIVACY	2013	40
Kristoufek, Ladislav	BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era	SCIENTIFIC REPORTS	2013	38
Van Hout, Marie Claire; Bingham, Tim	Responsible vendors, intelligent consumers: Silk Road, the online revolution in drug trading	INTERNATIONAL JOURNAL OF DRUG POLICY	2014	36
Boehme, Rainer; Christin, Nicolas; Edelman, Benjamin; Moore, Tyler	Bitcoin: Economics, Technology, and Governance	JOURNAL OF ECONOMIC PERSPECTIVES	2015	33

7. DISCUSSION AND CONCLUSION

In conclusion we recall that the blockchain technology involves a network of computers and records a digital object's existence.

Advantages and disadvantages of the blockchain technology

1) Advantages

Blockchain technology replaces a system based on trust by one of mathematically defined and mechanically enforceable rules. The bitcoin solved the double-spending-problem. We further recall that Chapron (2017) and Van Rossum (2017) noted several areas in which the use of blockchain technology is advantageous. Some of these are related to science, the way science is performed and how its results are distributed.

2) Disadvantages

However, Chapron (2017) also mentioned that, when it comes to the bitcoin, this technology is estimated to consume about 10.4 terawatt hours (TWh) a year, which is almost twice the amount used by Google (5.7 TWh). Of course, most of the so-called 'trusted' third parties such as banks and governments, also consume large amounts of electricity and are expected to oppose this new technology as it would make their privileged role in society largely or completely superfluous. Notwithstanding organized crime syndicates, whose Silk Road experiment did not turn out very well, also those used to act in 'grey zones' will probably not immediately embrace a system that makes 'everything' traceable. Although, for instance, laundering money through bitcoin is possible, this may be seen as more risky than using a more traditional method.

This leads to the question: is blockchain technology the solution for all problems? The answer is clearly no. By its nature this technology is not efficient: one registration takes much more time (Chapron mentions that the bitcoin can only manage seven transactions per second) than when registration is done by one – trusted – party. Moreover, nowadays transactions become slower and slower.

We already mentioned the loss of a private key in the context of bitcoins. Similar losses for contracts or ownership (your house for example) or in the context of inheritances are catastrophic and as far as we know, no good solutions exist for the moment.

We are interested to see updated and expanded versions of our elementary bibliometric analysis related to blockchain technology.

ACKNOWLEDGEMENT

It is a pleasure to thank my children Brendan, Cynthia and Sandra for suggestions and advice in the context of software engineering, banking and economics.

REFERENCES

Acuto, M. (2018). Global science for city policy. *Science*, 359(6372), 165-166.

b8d5ad9d974a44e7e2882f986467f4d3 (2016). Towards Open Science: The case for a decentralized autonomous academic endorsement system. https://zenodo.org/record/60054

Androulaki, E., Karame, G.O., Roeschlin, M., Scherer, T. & Capkun, S. (2012). Evaluating User Privacy in Bitcoin. In: IACR Cryptology ePrint Archive 596 (2012), http://fc13.ifca.ai/proc/1-3.pdf.

Bartling, S. (2017). Blockchain for Science and Knowledge Creation. Available at: http://www.blockchainforscience.com/2017/02/23/blockchain-for-open-science-the-living-document/

Böhme, R., Christin, N., Edelman, B, & Moore, T. (2015). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives*, 29(2): 213-38, DOI: 10.1257/jep.29.2.213

Brito, J. & Castillo, A. (2013). *Bitcoin. A primer for Policymakers.* Arlington (VA): Mercatus Center.

Chapron, G. (2017). The environment needs cryptogovernance. *Nature*, 545(7655), 403-405.

Cryptography: https://en.wikipedia.org/wiki/Cryptography

Dwyer, G. (2014). The Economics of Bitcoin and Similar Private Digital Currencies. July 8. dx.doi.org/10.2139/ssrn.2434628

Fouquet, R. (2017). From knowledge comes power. *Nature*, 551(7686), S141.

Krugman, P. R. (1984). The international role of the dollar: theory and prospect. In: (John F. O. Bilson and Richard C. Marston, Eds.) *Exchange rate theory and practice* (pp. 261-278). University of Chicago Press.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved from http://bitcoin.org/bitcoin.pdf

Pilkington, M. (2015). Blockchain Technology: Principles and Applications (September 18, 2015). *Research Handbook on Digital Transformations*, edited by F. Xavier Olleros and Majlinda Zhegu. Edward Elgar, 2016. Available at SSRN: https://ssrn.com/abstract=2662660

Pilkington, M. (2016). Bitcoin through the Lenses of Complexity Theory: Some Non-Orthodox Implications for Economic Theorizing. *Handbook of the Geographies of Money and Finance.* Martin, R.; Pollard.J. (Eds.). Edward Elgar: Cheltenham

Rivest, R.L., Shamir, A. & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2), 120-126.

Swanson, T. (2015). Consensus-as-a-service: a brief report on the emergence of permissioned, distributed ledger systems. Working paper: 6 April 2015. Retrieved from http://www.ofnumbers.com/wp-content/uploads/2015/04/Permissioned-distributed-ledgers.pdf

Van Rossum, J. (2017). Blockchain for research. Digital Science Report. DOI: https://doi.org/10.6084/m9.figshare.5607778

Walport, M. (2016). *Distributed Ledger Technology: beyond blockchain.* A Report by the UK Government Chief Scientific Adviser.

CROSSREF AS A NEW SOURCE OF CITATION DATA: A COMPARISON WITH WEB OF SCIENCE AND SCOPUS

NEES JAN VAN ECK
CWTS,
Leiden University,
The Netherlands
ecknjpvan@cwts.leidenuniv.nl

LUDO WALTMAN
CWTS,
Leiden University,
The Netherlands
waltmanlr@cwts.leidenuniv.nl

VINCENT LARIVIÈRE
Université de Montréal,
Canada
vincent.lariviere@umontreal.ca

CASSIDY R. SUGIMOTO
Indiana University Bloomington,
USA
sugimoto@indiana.edu

Abstract: Thanks to the Initiative for Open Citations, a large number of references in the scholarly literature have become openly available through Crossref. In addition, there are also many references that have been deposited in Crossref but that have not (yet) been made openly available. To better understand the value of Crossref as a new source of citation data, we compare the citation data in Crossref with the corresponding data in Web of Science and Scopus. We show that more than three-quarters of the references in WoS and more than two-thirds of the Scopus references can be found in Crossref, with about half of these references being openly available. However, we also find that many publications have been deposited in Crossref without their references, resulting in millions of missing references in Crossref.

INTRODUCTION

The Initiative for Open Citations (I4OC)[1] encourages scholarly publishers to make the references found in their journals and books openly available through Crossref (Shotton, 2018). With a few exceptions (most notably the American Chemical Society, Elsevier, IEEE, and Wolters Kluwer Health), almost all large publishers support the initiative. So far, this support has resulted in approximately half of all references deposited in Crossref being openly available, yielding about half a billion open references (Shotton, 2017).

I4OC has attracted widespread attention. The initiative is of particular importance for the scientometric community. Thanks to I4OC, Crossref has the potential to become an openly available source of citation data covering a large share of all scholarly litera-

[1] See https://i4oc.org/.

ture. I4OC has been endorsed by CWTS[2] and the International Society for Scientometrics and Informetrics[3]. In December 2017, an open letter from the scientometric community[4] was published calling for publishers to open their references. This letter has already been signed by more than 300 individuals.

At present, scientometricians typically obtain citation data from Web of Science (WoS) and Scopus, two proprietary data sources. In this paper, we provide empirical insights into the value of Crossref as a new source of citation data. We compare Crossref with WoS and Scopus, focusing on the citation data that is available in the different data sources. Our analysis will show that more than three-quarters of the references in WoS and more than two-thirds of the Scopus references can be found in Crossref, with about half of these references being openly available. On the other hand, it will also be shown that millions of references are missing in Crossref. These references occur in publications that have been deposited in Crossref without their references.

The statistics presented in this paper are based on WoS and Scopus data provided to CWTS in September 2017 and May 2017, respectively. The Crossref data was downloaded through the Crossref API in August 2017. For WoS, we consider the Science Citation Index Expanded, the Social Sciences Citation Index, the Arts & Humanities Citation Index, and the Conference Proceedings Citation Index. Other citation indices included in WoS, in particular the Emerging Sources Citation Index and the Book Citation Index, are not taken into account, as we do not have access to them.

MATCHING CROSSREF WITH WEB OF SCIENCE AND SCOPUS

To compare Crossref with WoS and Scopus, we matched publications using Digital Object Identifiers (DOIs). However, as we will see, this matching approach is not perfect. Every publication in Crossref has a DOI, but only a selection of the publications in WoS and Scopus have such an identifier. Furthermore, not all publications with a DOI in WoS and Scopus have a matching DOI in Crossref.

We begin by analyzing the extent to which WoS and Scopus provide DOIs, and in particular DOIs that can be used to match publications with Crossref. We consider publications in WoS and Scopus from the period 2012–2016. Recent publications are more likely to have DOIs in WoS and Scopus, and our focus is therefore on publications from these recent years. Table 1 provides statistics both for WoS and for Scopus. Statistics are presented for all document types (which includes proceedings papers, letters, editorials, book reviews, etc. in addition to research and review articles) and exclusively for research and review articles. We note that the total number of publications in Crossref in the period 2012–2016 equals 19.1 million, which is substantially more than the 11.9 and 13.9 million publications reported in Table 1 for WoS and Scopus, respectively.

Table 1. Number of publications in WoS and Scopus, with a breakdown based on whether a publication has a DOI and whether it has a Crossref match (in millions; period 2012–2016).

	All document types		Research and review articles	
	WoS	Scopus	WoS	Scopus
All publications	11.9 (100.0%)	13.9 (100.0%)	7.6 (100.0%)	9.9 (100.0%)
Publications with DOI	8.3 (69.6%)	11.3 (80.9%)	6.8 (90.2%)	8.3 (83.8%)
Publications with Crossref match	8.2 (68.3%)	10.7 (76.9%)	6.7 (88.9%)	7.9 (79.7%)

As shown in Table 1, 68.3% and 76.9% of the publications in WoS and Scopus have a DOI match with Crossref. Focusing on re-

[2] See www.cwts.nl/news?article=n-r2r244.

[3] See www.issi-society.org/blog/posts/2017/september/issi-supports-i4oc/.

[4] See www.issi-society.org/open-citations-letter/.

search and review articles, these figures increase to 88.9% for WoS and 79.7% for Scopus. This demonstrates that a relatively large share of the publications not classified as research or review article in WoS lack a DOI.

Matching based on DOIs involves various difficulties. When a publication does not have a DOI in WoS or Scopus, there are two possibilities. Either the publication truly does not have a DOI or it does have a DOI, but the DOI is missing in WoS or Scopus. Based on a manual examination of a small sample of publications, we estimate that about 75% of the publications without a DOI in WoS or Scopus truly do not have a DOI. The other 25% do have a DOI, but the DOI is missing in WoS or Scopus.

Duplicate DOIs also cause problems in matching. DOIs are assumed to be unique. One would not expect to have multiple publications with the same DOI. However, duplicate DOIs can be found both in WoS and in Scopus. The problem is particularly sizeable in Scopus. In the period 2012–2016, there are 161,446 duplicate DOIs in Scopus, some of them assigned to more than 100 publications. There are 8,087 duplicate DOIs in WoS in the same period.

In addition to Crossref, there are also other organizations that register DOIs. This causes another complication in matching. As shown in Table 1, of all publications in Scopus, 4.0% cannot be matched with Crossref even though they do have a DOI. We performed a manual examination of a small sample of these publications. In about half of the cases, a DOI was registered not with Crossref but with another organization, such as the China National Knowledge Infrastructure (CNKI). In other cases, DOIs in Scopus are incorrect (i.e., different from DOIs reported on publishers' websites), DOIs were never registered, or registration was not yet completed when the Crossref data was downloaded. As can be seen in Table 1, the number of publications in WoS with a non-matching DOI is relatively limited.

COMPARING CITATION DATA IN CROSSREF WITH WEB OF SCIENCE AND SCOPUS

How many of the references in WoS and Scopus are also available in Crossref? As discussed above, matching Crossref with

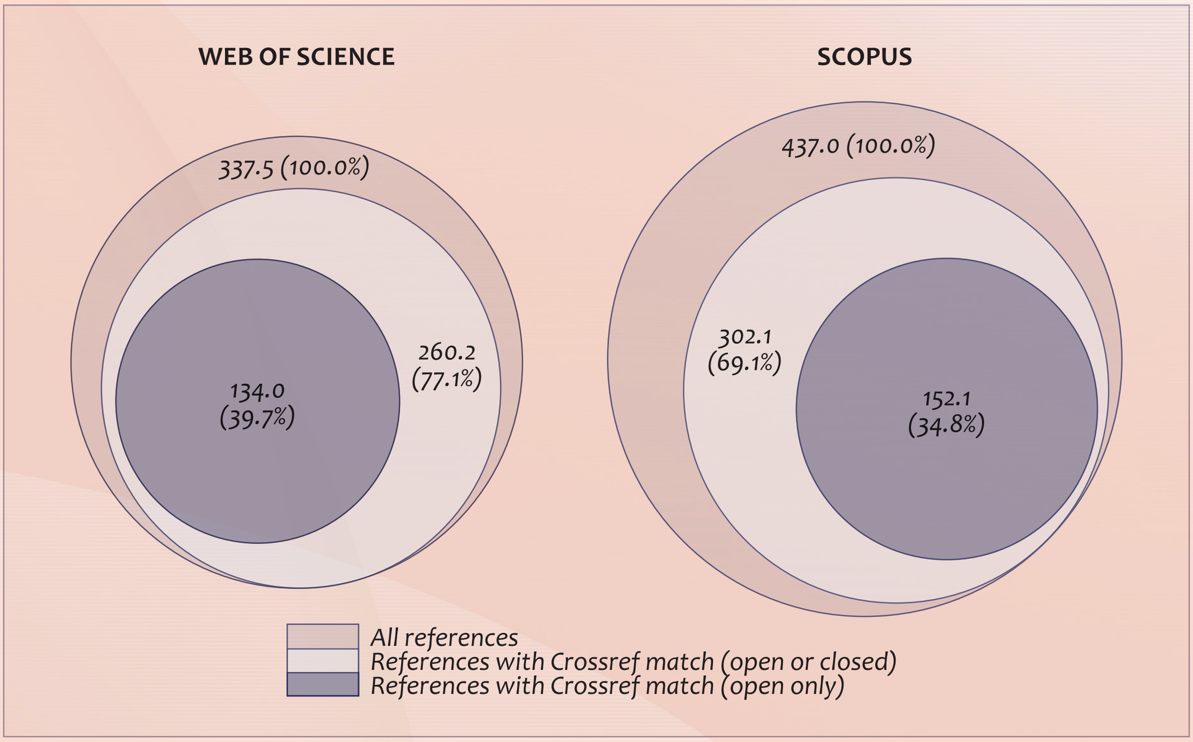


Figure 1. Number of references in WoS and Scopus, with a breakdown based on whether a reference has a Crossref match and whether it is open or closed in Crossref (in millions; period 2012–2016).

WoS and Scopus involves various challenges, and we therefore cannot give a precise answer to this question. However, by matching publications based on DOIs, an approximate lower bound can be provided for the number of references in WoS and Scopus that can also be found in Crossref.

Figure 1 shows how many of the references in WoS and Scopus publications from the period 2012–2016 have a matching reference in Crossref. In addition, the figure also shows how many of the matching references are openly available in Crossref. A reference in WoS or Scopus is considered to have a matching reference in Crossref if the citing publication has a DOI match with a publication in Crossref that also has references. All citing publications in WoS and Scopus are taken into account, irrespective of their document type.

As shown in Figure 1, 77.1% of the references in WoS have a matching reference in Crossref, but only 39.7% of the references in WoS have a matching reference that is openly available. For Scopus these statistics are somewhat lower, 69.1% and 34.8%, respectively. It needs to be emphasized that these results are likely to underestimate the true overlap in terms of references between WoS and Scopus on the one hand and Crossref on the other hand. Because of missing and incorrect DOIs in WoS and Scopus, our matching of Crossref with WoS and Scopus is incomplete, leading to an underestimation of the overlap between the different data sources. Both for WoS and for Scopus, Figure 1 shows that slightly more than half of all references with a Crossref match are openly available. This is in line with overall statistics reported for Crossref, where about 50% of all references are found to be open (Shotton, 2017).

We note that the total number of references in Crossref in the period 2012–2016 is 339.2 million, counting both open and closed references. Incidentally, this is very close to the 337.5 million references in WoS reported in Figure 1. With 437.0 million references, Scopus provides the largest number of references. In fact, since there is an overlap of about 300 million references between Crossref and Scopus, almost 90% of the references in Crossref are also available in Scopus. Hence, in terms of references, the content of Crossref that is unique relative to Scopus is fairly small. Relative to WoS, the

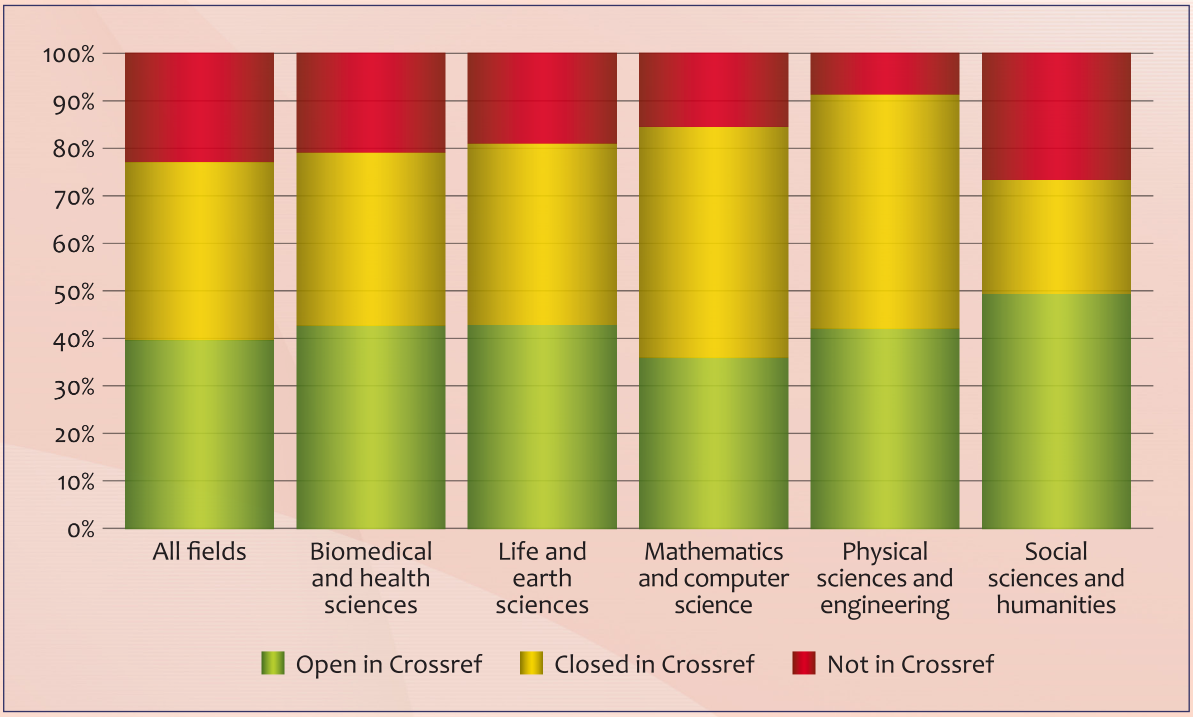


Figure 2. Breakdown for the percentage of references in WoS that (i) have a Crossref match and are openly available, (ii) have a Crossref match but are not openly available, and (iii) do not have a Crossref match (period 2012–2016).

unique content of Crossref is more substantial. Of the 339.2 million references in Crossref, about 75% can also be found in WoS.

The statistics presented in Figure 1 can be further broken down by field. Figure 2 shows such a breakdown for WoS. Five main fields are distinguished, following the definitions used in the CWTS Leiden Ranking. Substantial differences between fields can be observed. For instance, in the physical sciences and engineering, more than 90% of the WoS references have a match with Crossref, while this is the case for less than 75% of the WoS references in the social sciences and humanities. Nevertheless, almost half of the WoS references in the social sciences and humanities are openly available in Crossref. In mathematics and computer science, just around 35% of the WoS references are open in Crossref.

MISSING REFERENCES IN CROSSREF

Discussions about the Initiative for Open Citations (I4OC) have focused mostly on references that have been submitted to Crossref. So far, little attention has been paid to references that are missing in Crossref. These are references that have not been deposited in Crossref, even though the publications in which they occur have been deposited. It is clear that missing references may significantly reduce the value of Crossref as a source of citation data. To a certain extent, the issue of missing references may also explain why some references in WoS and Scopus do not have a match with Crossref, as discussed in the previous section.

For each publication in Crossref in the period 2012–2016, we tried to find a publication in WoS or Scopus with the same DOI. We then identified publications that do not have references in Crossref while they do have references in WoS or Scopus. By counting the number of references in these publications in WoS or Scopus, a lower bound is obtained for the number of references that are missing in Crossref. The results are presented in Table 2. (We note that, because of duplicate DOIs in WoS or Scopus, a publication in Crossref may sometimes have multiple matching publications in WoS or Scopus. We then used the publication with the largest number of references.)

Table 2. Number of references in Crossref, and number of missing references, based on comparisons with WoS and Scopus (in millions; period 2012–2016).

References in Crossref, both open and closed	339.2
Missing references in Crossref, based on a comparison with WoS	34.2
Missing references in Crossref, based on a comparison with Scopus	64.5

Table 2 makes clear that the number of missing references in Crossref is substantial. The comparison with Scopus shows that at least 64.5 million references are missing in Crossref. If publishers take the initiative to deposit these references in Crossref, the number of references in Crossref will increase by 64.5M / 339.2M = 19.0%. Combining this with the statistics presented in Figure 1, it follows that the share of references in Scopus with a Crossref match will increase from 69.1% to (302.1M + 64.5M) / 437.0M = 83.9%. For WoS, there will be an increase from 77.1% to (260.2M + 34.2M) / 337.5M = 87.2%.

Figure 3 shows the top 15 publishers with the largest number of missing references. The numbers reported in this figure are based on a comparison with Scopus, which yields more comprehensive statistics on missing references than a WoS-based comparison. Publishers that support I4OC are presented in yellow in the figure, while those that do not support the initiative are presented in red. Interestingly, various publishers have a large number of missing references, even though they support I4OC. These publishers make the references they deposit in Crossref openly available, but a sizeable share of their references have not been deposited in Crossref at all.

As can be seen in Figure 3, the most significant example of such a publisher is

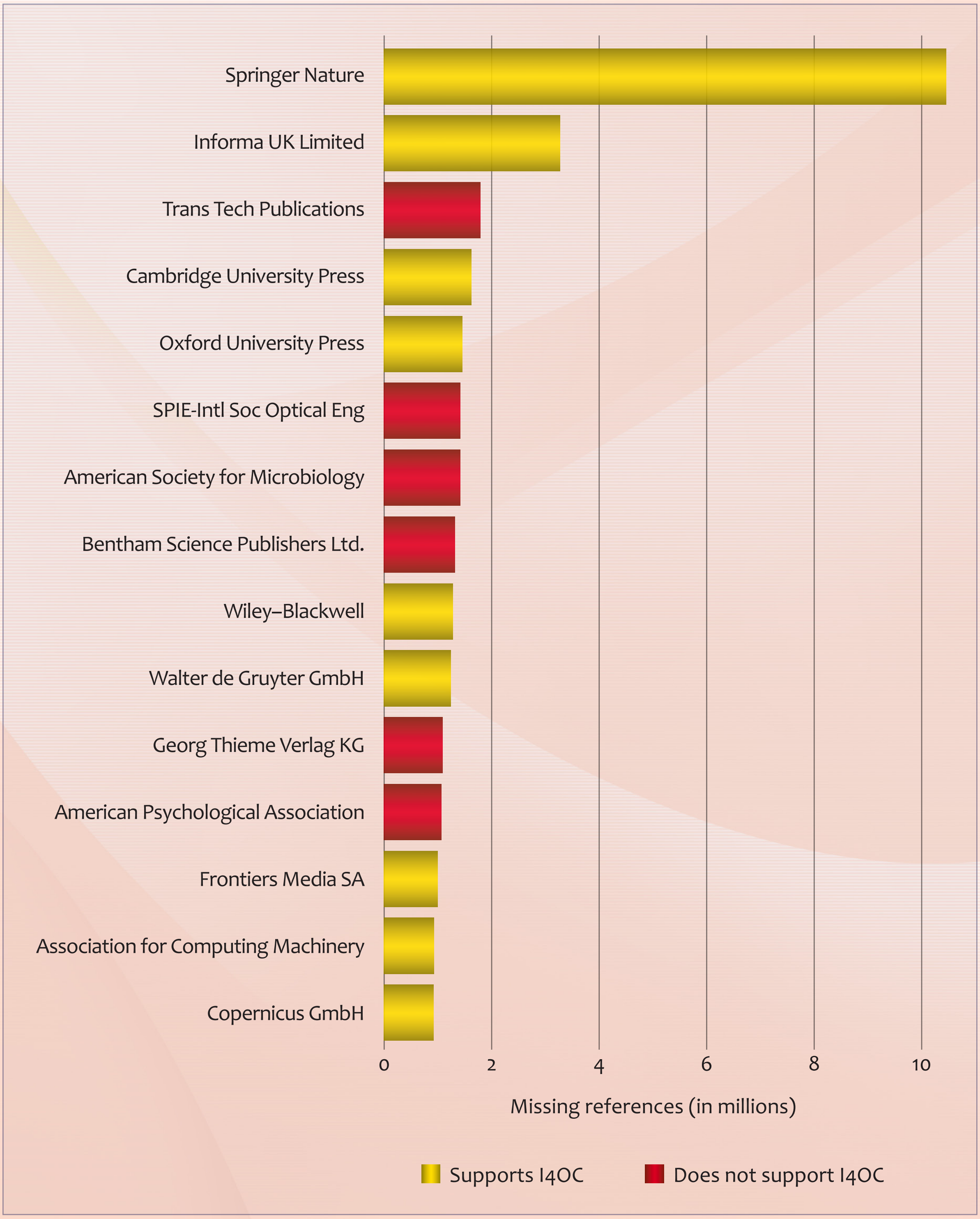


Figure 3. Top 15 publishers with the largest number of missing references in Crossref, based on a comparison with Scopus (in millions; period 2012–2016).

Springer Nature, with more than 10 million references that are missing in Crossref. A more detailed examination of the missing references of Springer Nature revealed that these are mostly references in books and book chapters. For journal articles published by Springer Nature, the number of missing references is much more limited. We are currently in contact with Springer Nature about their missing references in Crossref. Springer Nature informed us that they are investigating why so many of their references are missing, and they assured us that it is their intention to make these references available.

CONCLUSIONS

A large share of the scholarly literature indexed in WoS and Scopus is also available in Crossref. For recent years, 68% of the WoS publications and 77% of the Scopus publications can be matched with Crossref using DOIs as a crosswalking mechanism. These figures are likely to underestimate the true overlap between the data sources, since matching based on DOIs presents several difficulties, such as missing, incorrect, and duplicate DOIs. To improve matching, publishers and data providers need to work together to offer more comprehensive and more accurate DOI data.

The coverage of references is a critical concern for scientometricians. For recent years, more than three-quarters of the references in WoS and more than two-thirds of the Scopus references can be found in Crossref. Slightly more than half of the matched references are openly available. Our analysis also demonstrates that millions of references are missing in Crossref. These missing references occur in publications that have been deposited in Crossref without their references. We estimate that the deposit of missing references in Crossref would increase the share of matched references to 87% for WoS and 84% for Scopus. In order to create a comprehensive source of citation data, publishers must not only open their deposited references, but also attend to missing references.

Several next steps are needed to take full advantage of the infrastructure offered by Crossref. Many references are either closed or missing in Crossref. We therefore call for publishers to deposit their references in Crossref and to make them openly available. Moreover, the quality of reference data in Crossref can be improved. For instance, there is no standardized format for author names, and DOIs appear to be missing for a significant share of the references. Also, quite a lot of references are incomplete, with missing data for some or even all elements of a reference (e.g., author name, journal title, publication year, etc.). Publishers should work together with Crossref to improve the quality of reference data. Finally, we urge scientometricians to perform more in-depth studies of the data available in Crossref, to investigate possible systematic differences between references that are open and closed (e.g., in terms of geography, language, and research area), and to assess the suitability of Crossref data for different types of scientometric analyses.

ACKNOWLEDGEMENT

We thank David Shotton for his comments on an earlier version of this paper. This paper has also appeared as a blog post (see www.cwts.nl/blog?article=n-r2s234).

REFERENCES

Shotton, D. (2017, November 24). Milestone for I4OC – open references at Crossref exceed 50% [Blog post]. Retrieved from https://opencitations.wordpress.com/2017/11/24/milestone-for-i4oc-open-references-at-crossref-exceed-50/

Shotton, D. (2018). Funders should mandate open citations. *Nature, 553,* 129.