eISSN : 2287-4577
pISSN : 2287-9099
https://www.jistap.org
Vol. 11 No. 2 June 30, 2023

Journal of Information Science Theory and Practice

Journal of Information Science Theory and Practice

JISTaP

01 On the Application of Gamification Elements in Libraries
12 Ontology Development for Cultural Knowledge of Thai-Khmer Textiles
22 Growth and Distribution of Research Journals Across the World
37 Counting Research Publications, Citations, and Topics: A Critical Assessment of the Empirical Basis of Scientometrics and Research Evaluation
67 A Data Mining Approach for a Dynamic Development of an Ontology-Based Statistical Information System
82 Critical Success Factors for the Adoption of Health Management Information Systems in Public Hospitals in Zimbabwe

Vol. 11 | No. 2 | June 30, 2023

KiSTi
www.kisti.re.kr

Indexed/Covered by SCOPUS, LISA, DOAJ, CrossRef, and KCI

JISTaP
Journal of Information Science Theory and Practice

General Information

Aims and Scope

The Journal of Information Science Theory and Practice (*JISTaP*) is an international journal that aims at publishing original studies, review papers and brief communications on information science theory and practice. The journal provides an international forum for practical as well as theoretical research in the interdisciplinary areas of information science, such as information processing and management, knowledge organization, scholarly communication and bibliometrics. JISTaP will be published quarterly, issued on the 30th of March, June, September, and December. JISTaP is indexed in the Scopus, Korea Science Citation Index (KSCI) and KoreaScience by the Korea Institute of Science and Technology Information (KISTI) as well as CrossRef. The full text of this journal is available on the website at http://www.jistap.org

Indexed/Covered by

Scopus LISA DOAJ DIRECTORY OF OPEN ACCESS JOURNALS Crossref Cited-by 한국학술지인용색인 Korea Citation Index

Publisher

Korea Institute of Science and Technology Information
66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea
(T) +82-2-3299-6105
(F) +82-2-3299-6067
E-mail: jistap@kisti.re.kr
URL: http://www.jistap.org

Managing Editor: Seonhee Lee, Eungi Kim

Copy Editor: Ken Eckert

Design & Printing Company: MEDrang Inc.

2nd Floor, Hyoryeong Building, 32 Mugyo-ro, Jung-gu, Seoul 04521, Korea
(T) +82-2-325-2093
(F) +82-2-325-2095
E-mail: info@medrang.co.kr
Homepage: www.medrang.co.kr

Open Access and Creative Commons License Statement

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and au-thors, permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

This paper meets the requirements of KS X ISO 9706, ISO 9706-1994 and ANSI/NISO Z39.48-1992 (Permanence of Paper).

2023 Copyright © Korea Institute of Science and Technology Information

JISTaP
Journal of Information Science Theory and Practice

Editorial Board

Co-Editors-in-Chief

Gary Marchionini
University of North Carolina, USA

Dong-Geun Oh
Keimyung University, Korea

Co-Associate Editors

Kiduk Yang
Youngnam Old Document Archive Center, Korea

Wan-Jong Kim
Korea Institute of Science and Technology Information, Korea

Co-Managing Editors

Seonhee Lee
Korea Institute of Science and Technology Information, Korea

Eungi Kim
Keimyung University, Korea

Editorial Board

Dan Albertson
University at Buffalo, SUNY, USA

Daniel Martínez Ávila
San Paulo State University, Brazil

Beeraka Ramesh Babu
University of Madras, India

Pia Borlund
University of Copenhagen, Denmark

France Bouthillier
McGill University, Canada

Kathleen Burnett
Florida State University, USA

Ina Fourie
University of Pretoria, South Africa

Hamid R. Jamali
Charles Sturt University, Australia

Boryung Ju
Louisiana State University, USA

Thomas Mandl
Universiät Hildesheim, Germany

Lokman I. Meho
American University of Beirut, Lebanon

Jin Cheon Na
Nanyang Technological University, Singapore

Dan O'Connor
Rutgers University, USA

M. P. Satija
Guru Nanak Dev University, India

Ou Shiyan
Nanjing University, China

Paul Solomon
University of South Carolina, USA

Wolfgang G. Stock
Heinrich-Heine-Universität Düsseldorf, Germany

Kulthida Tuamsuk
Khon Kaen University, Thailand

Lihong Zhou
Wuhan University, China

Xiao-Zhi Gao
University of Eastern Finland, Finland

Wole Michael Olatokun
University of Ibadan, Nigeria

Consulting Editors

Reysa Alenzuela
Oriental Institute, ASCR, Czech Republic

Wayne Buente
University of Hawaii, USA

Sujin Butdisuwan
Mahasarakham University, Thailand

Seon Heui Choi
Korea Institute of Science and Technology Information, Korea

M. Krishnamurthy
DRTC, Indian Statistical Institute, India

S.K. Asok Kumar
The Tamil Nadu Dr Ambedkar Law University, India

P. Rajendran
SRM University, India

Tsutomu Shihota
St. Andrews University, Japan

2023 Copyright © Korea Institute of Science and Technology Information

JISTaP
Journal of Information Science Theory and Practice

Table of Contents

Vol. 11 No. 2 June 30, 2023
Journal of Information Science Theory and Practice • http://www.jistap.org

Articles

1 On the Application of Gamification Elements in Libraries
Seong-Kwan Lim

12 Ontology Development for Cultural Knowledge of Thai-Khmer Textiles
Jutamas Promthong, Malee Kabmala, Wirapong Chansanam

22 Growth and Distribution of Research Journals Across the World
Ramesh Pandita, Shivendra Singh

37 Counting Research Publications, Citations, and Topics: A Critical Assessment of the Empirical Basis of Scientometrics and Research Evaluation
Wolfgang G. Stock, Gerhard Reichmann, Isabelle Dorsch, Christian Schlögl

67 A Data Mining Approach for a Dynamic Development of an Ontology-Based Statistical Information System
Mohamed Hachem Kermani, Zizette Boufaida, Amel Lina Bensabbane, Besma Bourezg

82 Critical Success Factors for the Adoption of Health Management Information Systems in Public Hospitals in Zimbabwe
Caleb Manjeese, Indira Padayachee

104 **Call for Paper**

105 **Information for Authors**

2023 Copyright © Korea Institute of Science and Technology Information

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 11(2): 1-11, 2023

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2023.11.2.1

On the Application of Gamification Elements in Libraries

Seong-Kwan Lim*
Educational School Librarianship, Graduate School of Education,
Kyonggi University, Suwon, Korea
E-mail: kglimpro@kyonggi.ac.kr

ABSTRACT

Libraries are still conservative in their approach or practice of providing various services for the unspecified majority of visitors attending these institutions. It is very apparent that libraries are not actively trying to increase the interest and participation of users by applying the various angles or elements associated with gamification. By its very nature, gamification is the application of game-playing elements such as point scoring, peer competition, team work, score tables, and such used to motivate participants and make them more engaged with the subject matter. In areas such as education, marketing, and exercise, the implementation of gamification techniques is actively taking place in order to maximum participation by taking advantage of uncertainties or the competitive nature that many people have with the setting of goals. In this study, four libraries that have applied gamification are analyzed to understand how and to what degree gamification has been applied. Broken down into four different elements, this includes: *Point*, *Reward*, *Leaderboards & Competition*, *Self-expression & Achievement*. By focusing on the results of this analysis, gamification measures that can be applied to other libraries are specifically proposed. The overall results of this study will provide useful guidance and potential plans for libraries seeking to increase the number of users by using gamification to increase user participation and/or satisfaction with additional library services.

Keywords: playing, gamification, library service, case study, Korea, United States of America

Received: February 11, 2022 **Revised:** July 19, 2022
Accepted: July 19, 2022 **Published:** June 30, 2023

***Corresponding Author:** Seong-Kwan Lim
https://orcid.org/0000-0002-9735-1717
E-mail: kglimpro@kyonggi.ac.kr

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© Seong-Kwan Lim, 2023

1. INTRODUCTION

Humans are beings who play. As Bolz (2014) says in *Wer Nicht Spielt, ist Krank* (*Who Doesn't Play, is Sick*):

> Games fascinate because they take us into the paradise of the essential. But what is essential is not what is useful. The playground is a cherished living environment in which everything is done properly. The rules of the game guarantee a good order in which one always knows exactly what to do. And that is why the game world is better than the real world. Games are fascinating because, as a player, you are totally focused and can completely lose yourself in them. (p. 9, German translation)

However, as Bolz (2014) continues, rational society demands human beings who escape from play and participate in productive labor. Therefore, play is regarded as *überflüssiges*, "superfluous" or useless, and those who are deeply immersed in play feel guilty. However, our desire to fall in love with play is neither substituted nor resolved by anything else.

In 1938, Huizinga stated in his book *Home Ludens*, "Playing is not an element of culture, but culture itself has the character of play." (Huizinga, 1955). Therefore, play is older than culture, and it was also seen as the starting point of complex human activities such as language, law, war, philosophy, and art. Huizinga also claimed, "In happier days than now, we once called people 'Homo Sapiens,' but over time, we cannot believe that we are as rational as in the 18th century when we worshiped reason and pursued optimism."

Caillois (2014) then described the core principles of play in *Lemasque et le Vertige* (published in 1958) as Agon, Alea, Mimicry, and Ilinx, explaining them as strong instincts. Caillois' four core principles add two more to Huizinga's suggested principles of competition and motivation as the basic categories of play. Agon determines victory or defeat through competition in fair rules and orders, and Alea can be controlled as a poker game and main chance. Mimicry means creating a fictional self like role playing and then internally accepting and acting on it. Lastly, Ilinx means a pleasant thrill such as rollercoaster riding and air acrobatics.

Therefore, if we look up these principles in our lives, we can see that they are already found everywhere, and that we are already enjoying them every day. A general question arises, though, regarding how much play and game elements can be applied to libraries, especially when the traditional purpose or role of these institutions are to collect and preserve important materials and provide access to them for those needing this service.

Of course, this situation may have occurred because there is an environment where information resources can be easily accessed online, and there are more places where citizens find it much more entertaining and beneficial than via libraries. Therefore, since libraries are organized and operated for the purpose of contributing to reading, research, reference, and hobbies to readers by collecting, organizing, and preserving books and other materials, they will place greater value on activities such as information utilization education. In the end, the operational purpose of the library will not be achieved if it does not receive the proper attention of users.

As previously stated, the purpose of this study is to analyze how and what degree library services gamification can be applied and suggest specific measures for the further application of gamification. Of course, there are implications for libraries implementing gamification techniques and methodology, and these will be discussed and explored in some detail.

2. GAMIFICATION

2.1. Definition of Gamification

Gamification was first used as a concept in 2002 by British engineer Nick Pelling and was officially announced at the Gamification Summit held in January 2011 in San Francisco, California. Since then, this concept has been gradually systematized through interests and applications in various fields.

Although gamification as expressed is a relatively recent concept, the idea behind it has already existed for several decades. Kapp (2012) defines gamification as "using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems." Deterding et al. (2011) said that gamification is the use of game design elements in fields other than games, and Zichermann and Linder (2010) defined it as a process of solving problems by immersing users in using game thinking and techniques. Subsequently, Seaborn and Fels (2015) stated that it is an approach taken to achieve the purpose of motivation, enjoyment, and immersion in a non-game situation.

For the elements of gamification, Bunchball Inc. (2010) has six different types: Reward, Status, Achievement, Self-expression, Competition, and Altruism. Zichermann and

 https://doi.org/10.1633/JISTaP.2023.11.2.1

Cunningham (2011) have the following: Levels, Leaders (1), Leaders (2), Leaders (3), and Border. In addition, Kim (2013) said that it is based on the structure of Rule, Story, Design, and Interaction, and that game mechanics and game components interact to induce user participation and behavior.

Ana Ordás, an expert in games and gamification in Spanish libraries, stated in a chat interview with Lapointe (2017) titled "*Engaging library visitors through gamification*" that

> "Gamification in libraries is the use of elements with games with people – it is to use games and gamification to get closer to the physical and digital spaces. In the end, the significance of gamification in libraries is to put the user in the center of the design of the gamified system, taking into account the user's needs. To do this, you must first focus on preparing an environment where users are encouraged to interact with the library."

2.2. Effects of Gamification

Gamification focuses on "fun theory," where people are willing to do any activity when they have fun, and this provides joy and fun to users so that they can naturally recognize and pay attention to objects (Lee, 2019). Successful gamification must have an attracting theme or element that can immerse users in the beginning, and after that, it must be addictive so as to make them play games periodically. In addition, various devices that can imprint messages or images that the game's design entity wants to convey to the user must be incorporated. Finally, there should be an oral effect that allows users who enjoy games to naturally attract others to the game (Navercast, 2015).

In the end, gamification is not a "real" game. It is a very large feature with the advantage it gives to motivate people to voluntarily participate (through various techniques that induce specific behaviors which satisfy the desire for achievement as well as fun). Currently, gamification techniques are being used in various fields such as marketing by companies around the world, as well as in education, health, and multiple components of public policy.

2.3. Gamification Application Examples in Various Fields or Disciplines

Service cases to which the gamification technique is applied are already known in various fields. Among them, there is *My Starbucks Reward* with Starbucks, *Starbucks Planner*, Apple, and Nike (a healthcare running app service created jointly by the latter two). There are also McDonald's Monopoly partnership, *Duolingo,* which accumulates experience every time users practice and improve their language skills by comparing scores with friends, and *Friendsurance* (the world's first social network service-based insurance recruitment platform).

According to Gartner (2014), a global information technology business research consulting group, at least 70% of the world's 2,000 global companies will introduce one or more gamification marketing programs by 2029 and at least 50% of companies will pursue innovation processes using gamification concepts. Therefore, the list of field applications for gamification will continue to increase, and the types will continue to vary accordingly. The following is a summary of case examples in which gamification concepts and techniques were applied to different disciplines.

2.3.1. Gamification in Marketing: Brandification

Gamify (2022) describes *Brandification* as follows:

> Brandification in its simplest form consists of in-game advertising for real-world products and services. This could look like messages, images, or videos promoting a brand, product, or service within a video game's world.

2.3.2. Gamification in Healthcare: Patient Monitoring and Recovery

Phillips et al. (2019) said that gamification is a pressure to be actively introduced to motivate patients because of studies that have already been conducted on various states. However, this raises the fundamental question of whether it can be used under a doctor's prescription and supervision, as it can lead to other problems, such as creating a game-dependency state for patients. Nevertheless, those with chronic diseases such as cardiac disease, chronological orders of disease, and patients with type 2 diabetes were provided with a gamification app based on clinical scenarios (due to high management costs). The free app *SuperBetter* is highly recommended by psychotherapists because it helps psychological recovery by solving challenges one by one when and as problems occur, such as anxiety, depression, stress, and chronic pain.

2.3.3. Gamification in Education: Classrooms

Encouraging and maintaining students' motivation has been an old challenge in the field of education. Many studies concerning gamification have been conducted to date because gamification has the potential to motivate

students. Dichev and Dicheva (2017) suggested, though, that there is no consistent guideline for game design based on various educational experiences, and that it is necessary to explore the influence of game design elements including game dynamics throughout learning. *Kahoot*, which is currently widely used in schools, is an intuitive and interactive educational program that allows students to submit answers to questions asked by teachers during class in real time.

2.3.4. Gamification in Websites: Gamification Nation

Gamification Nation is a leading company in gamification design consulting in the United Kingdom, Europe, the United States, and Asia, and with corporations such as the sports clothing brand Adidas. In particular, it was selected as the European winner at the Outstanding Gamification Agency Awards in 2017.

2.3.5. Gamification in Influencers: Yu-kai Chou

Yu-kai Chou is an entrepreneur, business consultant, writer, and instructor. As a pioneer in the field of gamification, he gained notoriety and popularity in 2014 for lectures at Technology, Entertainment, and Design, published a book entitled *Actionable Gamification: Beyond Points, Badges, and Leaderboards* in 2017, and won the Gamification Guru of the Year award at the Gamification Europe conference.

According to Yu-kai Chou, gamification focuses on playful engagement, and is increasingly used in marketing and health products to increase incentives to engage with products and systems (Walsh, 2014).

3. LITERATURE REVIEW

Libraries are one of the most important organizations where new ways and innovative approaches. One of the problems with libraries is how librarians can increase users' engagement in library websites and library operations. This question could be answered by use of games and gamification elements in library websites could improve their operation (Bigdeli et al., 2016).

As Sukovic et al. (2011) remarked, "engagement through serious games and plays is proposed as a way of dealing with discrepancies between library traditional roles and contemporary demands, enabling experimentation and exploration of future roles." Actually, library managers now recognize that their instructions should "gamify" services to attract more and more users as well as boost customers' engagement with library websites (Radhakrishnan, 2013, as cited in Bigdeli et al., 2016).

Felker (2014) emphasized that gamification can be a strategy and framework for immersive learning and play, as well-designed games can provide powerful educational experiences that can promote positive interactions between libraries and users. Some libraries are already aware of the possibility of gamification participation, and "Project Velius," an alternative reality game developed by the University of Alabama to participate in library orientation by engaging students in a kind of online mystery scenario, may be an example. Also cited are the following operational areas as obvious candidates for library gamification:

1) Library orientation: A game that can inform users of basic information about how the library operates and what services are available.

2) Information literacy instruction: Games can be used for active experiences of users.

3) Resource usage: Games can give patrons more frequent or longer access to library resources.

4) Reading programs: Games can encourage children to read more to improve their reading interest and reading skills.

Walsh (2014) implemented *Lemontree* at the University of Huddersfield, England, a platform that creates an attractive environment in the library by taking ideas from games and providing a game layer on top of normal library activities. In doing so, he wanted to clarify what kind of relationship there is between them. An online survey was conducted for users registered with *Lemontree*, and as a result, 85% of students answered that gamification was appropriate in education. In addition, 60% of students responded that interaction within *Lemontree* changed their library usage behavior, suggesting that they can accept and easily participate in gamification in a university environment.

Based on the above responses, the gamification application method proposed in this study could be implemented to all primary/secondary, university, and general public libraries. In reality, though, not all libraries will be able to apply gamification immediately. This study is different from previous studies in that it derives gamification elements centering on cases applied in the United States and Korea and suggests ways to apply them. Nonetheless, if the proposed method can be applied to libraries, it will have positive effects in various aspects associated with increased user participation.

4. RESEARCH METHOD

This study began with the following questions:

Q1: In what field(s) is/are gamification applied and what are the specific application methods?

Q2: What are some examples of gamifications applied to library services?

Q3: What gamification measures can be applied to library services and how can they be applied?

Then, in order to achieve the purpose of the study, the following steps were taken.

Step 1: In terms of theory, the concept, effects, and application of gamification were examined, and the results of related previous studies were summarized.

Step 2: Four libraries in the United States (US) and Korea which systematically summarize the process and results of gamification application through literature reviews and Internet information search were selected as targets for analysis. The specific libraries (and games) to be analyzed include the *Mobile Scavenger Hunt* at North Carolina State University Library, and *Summer Game* at Ann Arbor District Library in the US. The *Reading Marathon* at Jeungpyoung County Library, and *Board Game* at Song-gok Girl's High School library in South Korea were also included.

Step 3: Based on the gamification elements organized by previous researchers, the analysis focused on whether these elements were applied to the four library programs selected for the target. Table 1 illustrates the gamification elements organized by previous researchers.

Step 4: From the results and findings gained through case analysis, the gamification plan applicable to all libraries (such as university and public libraries), was proposed focusing on the elements: *Point*, *Reward*, *Leaderboards and Competition*, *Self-expression* and *Achievement*.

Table 1. Gamification elements

Researcher	Gamification Elements
Bunchball Inc. (2010)	Reward, Status, Achievement, Self-expression, Competition, Altruism
Zichermann & Cunningham (2011)	Point, Levels, Leaderboards, Badges, Challenges/Quests, Onboarding, Engagement loops
Kim (2013)	Rule, Story, Design, Interaction

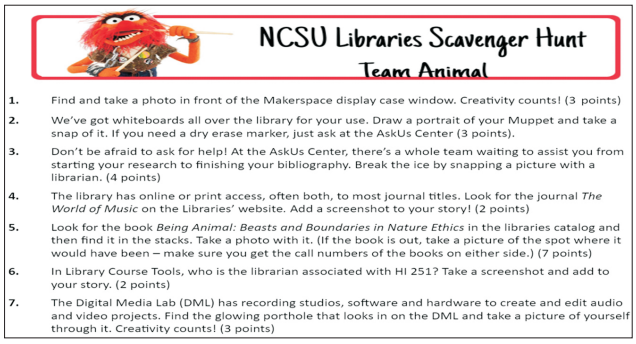
5. CASE ANALYSIS

5.1. Case Analysis 1: Mobile Scavenger Hunt at North Carolina State University Library (Raleigh, NC)

The North Carolina State University Library (NCSU) Library Mobile Scavenger Hunt, developed with feedback from students and instructors during 2020-2021, was aimed at guiding students to libraries, promoting the use of new technologies, and inspiring confidence in using multiple types of materials and media.

Students who participated in the NCSU Library Mobile Scavenger Hunt received instruction on the rules for this activity and on how to use iPads and Snapchat. Afterwards, each team received a clue sheet containing 12 questions about libraries and library services, a library map, and an iPad Air to use for submitting answers. Each team then went to the library for 25 minutes to find answers to the questions and write them on Google Spreadsheets. At the end of the allotted time given, they returned to the starting place, where their responses were reviewed, and the winning team with the most correct answers received a prize.

Based on *Rule*, several students teamed up to *Interact* as *Engagement loops*, and *Leaderboards* were selected to solve *Challenges/Quests* to obtain more *Points* and eventually the winning team was *Rewarded*. Since students have the advantage of being able to explore the library and gain knowledge as if they were game playing, it was helpful in revitalizing library operation of the library. Fig. 1 shows a sample of the scavenger hunt questions posed to participants and Fig. 2 shows the scoresheet template used (North Carolina State University Libraries, 2020).


NCSU Libraries Scavenger Hunt
Team Animal

1. Find and take a photo in front of the Makerspace display case window. Creativity counts! (3 points)
2. We've got whiteboards all over the library for your use. Draw a portrait of your Muppet and take a snap of it. If you need a dry erase marker, just ask at the AskUs Center (3 points).
3. Don't be afraid to ask for help! At the AskUs Center, there's a whole team waiting to assist you from starting your research to finishing your bibliography. Break the ice by snapping a picture with a librarian. (4 points)
4. The library has online or print access, often both, to most journal titles. Look for the journal *The World of Music* on the Libraries' website. Add a screenshot to your story! (2 points)
5. Look for the book *Being Animal: Beasts and Boundaries in Nature Ethics* in the libraries catalog and then find it in the stacks. Take a photo with it. (If the book is out, take a picture of the spot where it would have been – make sure you get the call numbers of the books on either side.) (7 points)
6. In Library Course Tools, who is the librarian associated with HI 251? Take a screenshot and add to your story. (2 points)
7. The Digital Media Lab (DML) has recording studios, software and hardware to create and edit audio and video projects. Find the glowing porthole that looks in on the DML and take a picture of yourself through it. Creativity counts! (3 points)

Fig. 1. Scavenger hunt team animal questions. Reproduced from North Carolina State University Libraries, 2020 with copyright holder's permission.

Team	Clue 1	Clue 2	Clue 3	Clue 4	Clue 5
Question summary ->	Brief reminder of question 1				
Question value ->	2	2	3	3	4
Team 1	Reminder of correct answer to question 1 for team 1				
Team 2	Reminder of correct answer to question 1 for team 2				
Team 3	... etc.				
Team 4					
Team 5					
Team 6					
Team 7					

Fig. 2. Scavenger hunt scoresheet template. Reproduced from North Carolina State University Libraries, 2020 with copyright holder's permission.

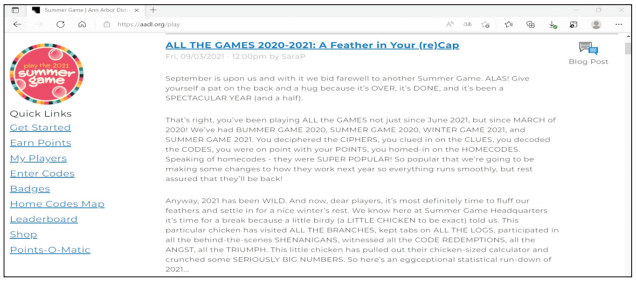


Fig. 3. Introduction to Summer Game. Reproduced from Ann Arbor District Library, 2022 with copyright holder's permission.

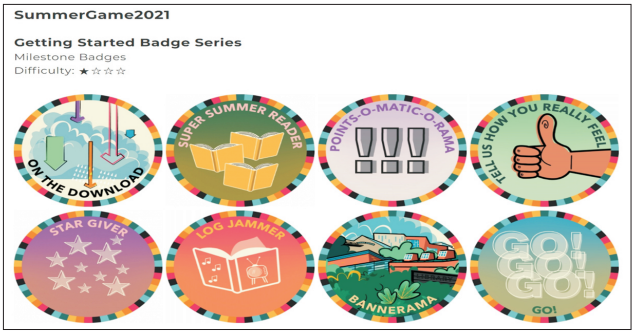


Fig. 4. Summer Game badge series. Reproduced from Ann Arbor District Library, 2022 with copyright holder's permission.

5.2. Case Analysis 2: Summer Game at Ann Arbor, Michigan District Library

Ann Arbor district library began running summer reading programs for children in the 1940s. These programs, evolving over the years as an event for everyone, are divided into three parts or types: *The Summer Reading Game*, *The Summer Game: Explorer Series*, and *The Summer Game: Online Codes*. Children participating in *The Summer Reading Game* have to read ten age-appropriate books, and teenagers and adults also have to read ten books. Children and teenagers who complete the mission will be given 2,000 points and tokens to receive free books, while adults will also be given 2,000 points. *The Summer Game: Explorer Series* is played in "extreme" parks, across all five ADL locations in the city, in attempts to find summer game codes (either participating outside and/or in-person). These codes are considered points. Lastly, *The Summer Game: Online Codes* is played by completing a puzzle by searching secret codes hidden on the Ann Arbor district library Internet site. The winning prize is a special badge given by the library.

Ann Arbor district library's overall *Summer Game* reported that 7,858 players scored 141,708,438 points in 2021, and 266,374 badges were obtained. This result alone shows that the *Summer Game* is an event that many citizens are participating in, and utilizes elements in three ways: *Design*, *Competition* to obtain *Points*; *Badge* in accordance with each *Rule*; and, *Leaderboards* to make for a larger *Reward* factor. Fig. 3 is an introduction page for the Summer Game sponsored by the Ann Arbor District Library and Fig. 4 is part of the badge series that can be obtained by those who participated in those games (Ann Arbor District Library, 2021).

5.3. Case Analysis 3: Reading Marathon at Jeungpyeong County Library

The Jeungpyeong County Library selected Kim Deuk Shin, a poet and reading enthusiast of the Joseon Dynasty, as its representative figure and used his name for various events. The Kim Deuk Shin reading marathon of the Jeungpyeong county library, which started in 2020, is currently celebrating its third anniversary in 2022. In order to participate in the competition, players must join the library as a member, read the appropriate borrowed/lended books, write journal entries in a "reading" diary, or submit certified reading activities on a social network service or blog for final confirmation. Table 2 summarizes the courses of the Kim Deuk Shin Reading Marathon.

In Table 2 the reading marathon course is organized into seven steps. This stage is structured as to raise the level of the game, and pages read are converted into distance traveled and number of volumes/copies. If participants read more than 42,195 pages, which corresponds to a total of 141 books, the course is completed. Thus so, there is a great deal of reading involved with this course.

As may be surmised, it is difficult to complete a whole marathon and there will be many dropouts. If players

Table 2. Kim Deuk Shin's reading marathon course

Conversion	Course name (distance)						
	Turtle (3 km-)	Horse (5 km-)	Rabbit (10 km-)	Tiger (15 km-)	Ostrich (20 km-)	Cheetah (30 km-)	Kim Deuk Shin (42.195 km-)
Page conversion (1 p=1 m)	3,000 p-4,999 p	5,000 p-9,999 p	10,000 p-14,999p	15,000 p-19,999 p	20,000 p-29,999 p	30,000 p-42,194 p	42.195 p-
Converted number of copies (=300 p)	10 copies	17 copies	34 copies	50 copies	67 copies	100 copies	141 copies

p, point.



Fig. 5. Public relations poster for Kim Deuk Shin's reading marathon. Reproduced from Jeungpyeong County Library, 2022, Kim Deuk Shin's reading marathon, with copyright holder's permission.

succeed in the rabbit course, they receive either a puzzle, pop-up book, or toothbrush set. If players succeed in the cheetah section, they receive cold bags or Kim Deuk Shin slippers. Ultimately, if they complete the Kim Deuk Shin course, they participate further (between October and November) to certify their reading activities. In addition, the final finisher will be given the benefit of borrowing from two to four more books.

This activity includes the elements of gamification in that participants can receive *Achievement*, *Rewards* according to the final *Status* and *Point* through the *Self-expression* process. This includes such things as writing diary entries after reading according to *Levels*, which is *Design* in a total of seven stages. Accordingly, this helps increase reading opportunities along with citizens' use of libraries. Fig. 5 is a promotional poster for the 2022 Kim Deuk Shin's reading marathon published in the Jeungpyeong County Library (2022).



Fig. 6. Board game corner of Song-gok Girl's High School library. Reproduced from Lee et al. 2017. Board games in the library. Gyeonggi-do Cyber Library.

5.4. Case Analysis 4: Board Game at Song-gok Girl's High School Library

This is a board game in which at least two people participate according to set rules, using simple tools such as Board, Card, and Tile. Since the early 2000's, various board games have been introduced in Germany and the United States and many of these now-numerous games have been released and can be easily purchased. Board games are also used in various fields such as subject learning, psychological counseling, and personality and career development education, as it is effective in easily understanding specific concepts such as the development of creativity, reasoning, expression, problem-solving, and communication skills.

There are many cases where board games are used as learning tools in Korean school libraries. Song-gok Girl's High School library registers/lends more than 100 board games in the same way as general materials, so that students can borrow them and use them in schools. In addition, an example of using board games in library cooperative classes is the creation of *Dixit* card games in art subjects. The main activity is to divide the *Dixit* cards between students, select one of the cards based on the keyword of the card chosen by the storyteller, infer why

the card was selected, and award scores according to the correct answer or similar reasoning.

Fig. 6 is a board game corner located at the Song-gok Girl's High School library and Fig. 7 is a *Dixit* card drawn by students in their art subject cooperation class (Lee et al., 2017).

5.5. Analysis and Conclusion

According to the results summarized in Table 3, the gamification factor included in all cases was *Point*. Next, *Reward* was included in three cases, and *Rule*, *Leaderboards* and *Design* were found in two places. Lastly, *Interact*, *Engagement Loops*, *Challenges/Quests*, *Competition*, *Badge*, *Self-expression*, *Status*, and *Achievement* were used once each.

6. GAMIFICATION APPLICATION METHOD

Even before the term gamification appeared in general discourse, libraries used related terms or elements. Presently, though, the term gamification has won out and is actively used in various fields, increasing general interest. Therefore, the application plan is to focus on such components as *Point*, *Reward*, *Leaderboards & Competition*, and *Self-expression & Achievement* through case analysis, and this needs to be applied according to different conditions and scenarios.

6.1. Point

A service that provides a customer with a score calculated under a specific condition, according to the purchase amount of various products or the number of visits to a store, is called or designated the term *Mileage*. Customers can use their mileage when they need it, and so they buy many products or visit stores more often to get higher scores. By applying this principle or service, *Point* provides a way or method of converting the number of in-person visits to the library, the number of times a website is accessed online, and the number of book loans, into scored points. In addition to this, accumulated points can be used to pay overdue fees, compensate for lost books or periodicals, and even make use of the library cafeteria.

6.2. Reward

The establishment and operation of libraries already provides sufficient compensation to many users. In addition, awards and products have been provided through various events such as multi-reader awards, reading appreciation contests, and reading quiz contests. Nevertheless, compensation can be a way to increase and maintain users' motivation for using the library. From time to time, it is recommended to proceed with informal things that



Fig. 7. Dixit card made by students. Reproduced from Lee et al. 2017. Board games in the library. Gyeonggi-do Cyber Library.

Table 3. Gamification factor derivation result

Case	Gamification factor
1	Rule, Interact, Engagement loops, Leaderboards, Challenges/Quests, Point, Reward
2	Design, Competition, Point, Badge, Rule, Leaderboards, Reward
3	Design, Levels, Self-expression, Status, Point, Reward, Achievement
4	Point

 https://doi.org/10.1633/JISTaP.2023.11.2.1

are in addition to events that have already been officially announced. As an example of this, markets sell vegetables and fish that are less fresh at lower prices before a deadline, or hold last minute sales on specific products. By applying these same measures, users who visit the library for, say, the 100th time could be allowed twice the set number of book loan outs, or could receive library bags as gifts on a first-come, first-served basis. Their satisfaction and loyalty to the library will likely increase due to these measures taken.

6.3. Leaderboards & Competition

Most users are satisfied with getting what they need when necessary among the various materials available in the library. *Leaderboards* installed in library lobbies can show real-time changes to individual activities, and if players see this every time they enter and exit, the inner competitive sentiment or nature can begin to percolate. Since it is easiest to calculate loan statistics for each user in the library, it would be appropriate to install *Leaderboards* to make or enhance competition between users.

6.4. Self-expression & Achievement

Libraries have materials from many different fields of endeavor such as literature, music, and the arts that users can appreciate and immerse themselves in. It would be nice to provide users with an opportunity to introduce and/or express their appreciation of books they read, such as with book talk sessions or discussion groups. This would create an environment of collective or shared intelligence with these types of interactions. The process will bring a community of participants together through the opportunities made available through outward expression.

7. DISCUSSION

The American Psychological Association established the "Internet Game Disorder" item in the *Diagnostic and Statistical Manual of Mental Disorder-5* revised in 2013, and discussions of this disorder were held at the International Statistical Division in Geneva, Switzerland in May 2019 (72nd General Assembly B Committee). Such a development has sparked a backlash from game developers and gamers around the world. Technology marches on, though, and the continuing development of portable electronic devices such as smartphones and the like has only increased the number of people immersed in games. Without question, games are definitely interesting activities, but it is difficult to set standards for how appropriate they are, especially for those at-risk gamer individuals who have difficulty in self-regulation or who lack discipline.

Those who emphasize the risk of gamification warn that in the case of activities focusing on external compensation, if the latter eventually disappears, behavior disappears while reducing internal motivation. In addition, it is said that if the participants' activities are tracked and scores and results are given, there is a concern regarding the matter of privacy. The library's gamification plans and focus will have to include discussions centering around the questions posed below.

1) To what extent will gamification be applied, and to which service?

Depending on the type of library, such as public or school libraries, the purpose of operation or use may vary, but library orientations, information literature instruction, resource usage, and reading programs, as selected by Felker (2014), can be applied to all libraries. Therefore, it is necessary to discuss in the library community to what extent gamification will be effective in applying to which services according to the type of library.

2) Will the service be effective even when the gamification, which includes external compensation, is stopped?

The library provides users with various services such as the provision of information, education, and recreation. However, with the development of media devices such as smartphones, the social environment has changed where necessary information can be easily found, and the proportion of users who visit the library is gradually decreasing. Therefore, this study proposed to find and apply a gamification method that satisfies all the factors of fun, participation, and external compensation to increase use over the long-term.

8. CONCLUSIONS AND RECOMMENDATIONS

Library managers now recognize that their institutions should gamify services to attract more and more users as well as boost customers' engagement with library websites (Radhakrishnan, 2013). It is necessary to think about how to apply or dovetail traditional services that are loved by many people to the library with the various new elements that appear as by-products of changing times.

Based on the cases of libraries applying gamification,

this study aimed to analyze different gaming aspects, how they were applied, and propose implementation and specific application plans for libraries to increase the number of future users and increase service satisfaction. The measures analyzed to achieve the above research objectives are summarized as follows:

First, the elements of gamification included in the *Mobile Scavenger Hunt* of the North Carolina State University Library are based on *Rule* and several students became part of *Engagement Loops* to solve *Interaction* and *Challenges/Quests* to achieve more *Composition* with other teams.

Second, the elements of gamification included in Ann Arbor district library's *Summer Game* are three *Designed* ways, making *Competition* to obtain *Points* and *Badge* in accordance with each of the *Rules*, and *Leaderboards* to make *Status* larger.

Third, the elements of gamification included in the Kim Deuk-shin's reading marathon Contest of the Jeungpyeong county library, which is *Designed* in a total of seven stages, are *Self-expression* to writing a diary after reading, the *Reward* according to the final *Status* and *Point*, and a sense of *Achievement* felt for all these processes and results.

Fourth, it can be said that a *Board Game*, which is used in curriculum cooperation classes at Song-gok Girl's High School library, contains all elements of gamification, as there are various types and methods. Then, based on the analysis, the gamification plan applied to the library can be implemented as follows.

A) The *Point* element proposes a plan to convert the number of offline visits to the library, access to the homepage online, and the number of book loans, into scores so that they can be used to pay late fees, compensate for lost books, and use the cafeteria in the library.

B) The *Reward* element proposes a strategy to increase loyalty to the library with unexpected satisfaction to users, giving twice the number of books when making a loan on a special day, or giving a library bag as a gift on a first-come, first-served basis.

C) In the *Leaderboards & Competition* element, *Leaderboards* showing loan statistics for each user are installed in the library lobby to propose each other's *Compensation* plan.

D) The *Self-expression & Achievement* element proposes ways to provide users with an opportunity to introduce and express their appreciation of the data they read, and furthermore, to create collective intelligence while interacting with each other.

There is nothing complete or an end-all in any tool or method. Therefore, gamification also needs to be considered in many different aspects that apply to the library. There will be many positive affects if an appropriate plan or roadmap is found. Further research regarding this topic is welcomed to advance the gamification knowledge base from various angles or points of view so that the library can always be fun and interesting, but above all, remain an important institution in our culture and civilization.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Ann Arbor District Library. (2021). *Play the 2021 summer game.* https://aadl.org/play?page=3

Bigdeli, Z., Haidari, G., HajiYakhchali, A., & BasirianJahromi, R. (2016). Gamification in library websites based on motivational theories. *Webology,* 13(1), Article 146. https://www.webology.org/data-cms/articles/20200515050922pma146.pdf

Bolz, N. (2014). *Wer nicht spielt, ist krank: Warum fußball, glücksspiel und social games lebenswichtig für uns sind [Who doesn't play is sick: Why football, gambling, and social games are vital to us].* Redline-Verlag. German.

Bunchball Inc. (2010). *Gamification 101: An introduction to the use of game dynamics to influence behavior.* Bunchball Inc.

Caillois, R. (2014). *Les jeux et les hommes: Le masque et le vertige.* Gallimard. French.

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education,* 14, 9. https://doi.org/10.1186/s41239-017-0042-5

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September 28-30). From game design elements to gamefulness: Defining "gamification". In A. Lugmayr, H. Franssila, C. Safran, & I. Hammouda (Eds.), *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)* (pp. 9-15). ACM.

Gamify. (2022). *What is gamification?* https://www.gamify.

com/what-is-gamification

Gartner. (2014). *Gartner's gamification predictions for 2020.* https://www.growthengineering.co.uk/future-of-gamification-gartner/

Huizinga, J. (1955). *Homo ludens: A study of the play-element in culture.* Beacon Press. (Original work published 1938)

Jeungpyeong County Library. (2022). *Kim Deuk Shin's reading marathon.* https://lib.jp.go.kr/front/index.php?g_page=culture& m_page=culture04

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education.* Pfeiffer.

Kim, H. T. (2013). *Gamification marketing.* Youngjin.com.

Lapointe, M. (2017). *Engaging library visitors through gamification – Interview with Ana Ordás. https://princh.com/blog-engaging-library-visitors-through-gamification-interview-ana-ordas/#.ZDbnSLrP3b0*

Lee, D. J., Kim, S. H., Shin, S. Y., Oh, K. S., Jung, Y. H., & Han, H. J. (2017). *Board games in the library*. Gyeonggi-do Cyber Library.

Lee, Y. H. (2019). *A study on the effects of fun and flow through gamification on memorable positive tourism experience.* (Doctoral dissertation). Dong-Eui University, Busan, Korea.

Navercast. (2015). *Fun is the trend, gamification wind blowing in Hong Kong.* https://terms.naver.com/entry.naver?docId=3579278&cid=58784&categoryId=58787

North Carolina State University Libraries. (2020). *Mobile scavenger hunt: A mobile, game-based way to introduce students to the library.* https://www.lib.ncsu.edu/projects/mobile-scavenger-hunt

Phillips, E. G., Jr, Nabhan, C., & Feinberg, B. A. (2019). The gamification of healthcare: emergence of the digital practitioner? *American Journal of Managed Care,* 25(1), 13-15.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies,* 74, 14-31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Sukovic, S., Litting, D., & England, A. (2011). Playing with the future: Library engagement and change. *Australian Academic & Research Libraries,* 42(2), 70-87. https://doi.org/10.1080/00048623.2011.10722215

Walsh, A. (2014). The potential for using gamification in academic libraries in order to increase student engagement and achievement. *Nordic Journal of Information Literacy in Higher Education,* 6(1), 39-51. https://doi.org/10.15845/noril.v6i1.214

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps.* O'Reilly Media.

Zichermann, G., & Linder, J. (2010). *Game-based marketing: Inspire customer loyalty through rewards, challenges, and contests.* Wiley.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 11(2): 12-21, 2023

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2023.11.2.2

Ontology Development for Cultural Knowledge of Thai-Khmer Textiles

Jutamas Promthong
Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kean University, Khon Kean, Thailand
E-mail: jutamat.pt@bru.ac.th

Malee Kabmala
Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kean University, Khon Kean, Thailand
E-mail: malee_ka@kku.ac.th

Wirapong Chansanam*
Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kean University, Khon Kean, Thailand
E-mail: wirach@kku.ac.th

ABSTRACT

This study aims to develop ontologies regarding cultural knowledge of Thai-Khmer textiles by applying the Knowledge Engineering Methodology to build upon the ontologies. The process includes 1) generating the ontologies' objectives, 2) building ontologies, and 3) evaluating the ontologies. The researchers used OntOlogies Pitfall Scanner (OOPS!) to minimize defects and asked two experts to evaluate the ontologies' design. Protégé was used to design the ontologies, and WIDOCO was used to present the ontologies through the World Wide Web. It was found that the developed ontology consists of two classes, 16 sub-classes, and 16 relationships. The ontologies assessment found that there were seven items to fix according to the OOPS! software. Apart from the assessment program, the experts mentioned that all five aspects were suitable; namely, the ontology design was evaluated at 4.51 (Likert), the process of identifying scopes of definitions and objectives of development was 4.61, the applications and guidelines for further development was 4.58, the process of forming classes was 4.53, and the process of generating class's properties was 4.50.

Keywords: ontologies of cultural knowledge in Thai-Khmer textiles, ontologies, culture in Thai-Khmer textiles, OntOlogies Pitfall Scanner

Received: August 31, 2022 **Revised:** January 25, 2023
Accepted: January 31, 2023 **Published:** June 30, 2023

***Corresponding Author:** Wirapong Chansanam
https://orcid.org/0000-0001-5546-8485
E-mail: wirach@kku.ac.th

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© Jutamas Promthong, Malee Kabmala, Wirapong Chansanam, 2023

1. INTRODUCTION

Recently, each country has been trying to support every sector in terms of applying cultural knowledge as a foundation for national improvement, since cultural knowledge relates to values, practices, and innovations to correspond to the needs of communities. Furthermore, cultural knowledge reflects the environment and ways of life (Wei Choo, 2000). Therefore, the cultural knowledge of each community tells stories related to culture in order to demonstrate the significance of the community's culture. This was also the foundation for countries' product and service development and innovation (Howkins, 2002; Iamkhajornchai & Manmart, 2018).

Textiles can portray a culture's core values, which is a significant capital to build on existing knowledge. In the past, textiles were considered one of the economic indicators which could be sold or traded with other goods or money. Also, textiles are counted as cultural capital since they are bound with traditions, rituals, and ways of life (Pogodina et al., 2021). For example, people use textiles as offerings in rituals or make textiles as ethnic costumes. More than this, culture in textiles is bound with social values, especially regarding women and textiles, and it reflects patience, handicraft, and creativity.

Additionally, ways of culture in textiles from birth to death can demonstrate capital in religions and traditions since beliefs regarding religions and traditions are attached in every step of the weaving process. Textiles describe their legacies in cultural knowledge, identity, local wisdom, and weaving skills (Sakamoto, 2014). Therefore, textiles are not only for daily use or aesthetics, but they are resources in the culture of ethnic groups which illustrate contexts in history and culture (Martins et al., 2013) to develop products further because all knowledge of textiles lies within each fabric: preparing threads, colors, patterns, weaving tools, weaving process, weaving techniques, and decorations. These elements are in each fabric of the Thai-Khmer ethnic group as reflected in ways of life, rituals, traditions, and beliefs.

Cultural knowledge in each area is different due to distinctive geography, climate, natural surroundings, needs, and ethnic groups. The abovementioned concepts lead to adaptations and unique ways of life, beliefs, culture, traditions, tools, and needs. When ethnic groups migrated, they often brought their culture with them. Therefore, we can distinguish the uniqueness of each ethnic group. For instance, in Lawa costume, women often wear dark blue or black sleeveless shirts called *Pok Tia* or *Koh* in Thai. For the So ethnic group, women wear *ikat* or cotton sarongs with black button-up shirts decorated with red threads and a breast cloth.

Nevertheless, married women do not decorate their shirts with red threads. Instead, Thai-Khmer women wear sarongs which are *Hol*, *Umprom*, *Unloonseem*, and *Sakoo*, and shirts. Hence, cultural knowledge in textiles can identify specific characteristics of each ethnic group.

In accordance with the National Strategy on Diverse Tourism (2018-2037) and the Northeast Region Development Plan, Thai-Khmer ethnic groups' woven textiles will be marketed globally (2018-2022). The textiles are to be a part of cultural tourism and souvenir sales. Thai-Khmer textiles have a number of distinctive characteristics (Office of the National Economic and Social Development Council, 2019; Office of Small and Medium Enterprises Promotion, 2017), including 1) premium-quality yarns for a piece of delicate and lustrous fabric; 2) silkworm farming, a superior form of natural farming; 3) only Thai Khmer-made woven fabrics use the wrap-and-weft *ikat* weaving techniques; 4) dyeing the hues red, yellow, black, green, and blue under the influence of the Khmer culture; 5) using natural dyes; and 6) *Hol*, *Kanaw*, *Umprom*, *Unloonseem*, *Labek*, *Samore*, *Chanotluek*, and *Sakoo* are original patterns (Boonpok, 2019).

Cultural knowledge of Thai-Khmer textiles could be applied to teaching, learning, and researching the weaving process and textile product developments. In accordance with studies by Arayaphan and Panommit (2019) and Iamkhajornchai and Manmart (2018), it was found that issues in cultural knowledge application were 1) knowledge classifications; even cultural knowledge of textiles is specific, as it was classified as common knowledge and placed in several categories. Consequently, users could no longer retrieve the cultural knowledge; 2) existing keywords; they do not cover all the content of cultural knowledge of textiles; and 3) keyword settings; each textile word is specific and called differently in each area. Moreover, some words in Thai share the same sounds or same forms which might cause misspellings. As a result, users might take time to retrieve and get inaccurate meanings. Therefore, a quality tool for retrieving cultural knowledge of textiles containing functional features, e.g., specific classifications and accurate keywords for global, local, and tribal use, should be developed for users to access the cultural knowledge of Thai-Khmer textiles faster.

Ontology is one of the helpful retrieval tools. It contains practical features for Thai-Khmer textile cultural knowledge such as setting representative words, specific

properties, and relationship attributes, and developing relationships by using classes to show clear connections between classes, properties of each class, and instances. Employing classes in an ontology could provide accurate settings and share knowledge sets among the system. It also reduces redundancies in word uses (Uschold & Gruninger, 1996) and helps users retrieve and access this cultural knowledge effectively.

As a result, the development of an ontology in the cultural knowledge of Thai-Khmer textiles could help 1) set keywords representing standard concepts of cultural knowledge of Thai-Khmer textiles, 2) enhance a process's automatic clear methodology for managing disputes and improving content, enabling a common namespace for Thai-Khmer textile terms that facilitates querying and understanding the relationships between concepts in a particular cultural knowledge domain, and 3) reduce redundancy and improve accuracy. Additionally, this ontology will be an alternative for the textile category to widely share cultural knowledge of Thai-Khmer textiles.

2. RESEARCH METHODOLOGY

In general, modeling is the process of creating a representation of a system or concept to understand it better and make predictions about it. Many different methodologies can be used for modeling, depending on the nature of the system or concept being studied and the research goals. Some common methodologies used in academic research include: 1) mathematical modeling, which involves using mathematical equations and simulations to represent a system and make predictions about it; 2) statistical modeling, which involves using statistical techniques to analyze data and build a model representing relationships within the data; 3) simulation modeling: This involves creating a computer program that simulates the behavior of a system over time; 4) conceptual modeling: This involves creating a diagram or other visual representation of a concept or system to better understand its structure and relationships; 5) prototyping: This involves creating a physical or digital prototype of a product or system in order to test and refine it; 6) case study: This involves in-depth analysis of a specific case or situation in order to understand it in more depth; 7) experimental design: This involves manipulating variables and measuring the results in order to understand the relationships between variables and to test hypotheses; and 8) grounded theory: This involves using inductive reasoning to develop a theory based on data collected through observations or interviews.

It is important to choose the most appropriate methodology for the research question being studied and to clearly describe the methodology used in the research in order to allow other researchers to replicate or build upon the work. Therefore, in this study, we chose the conceptual modeling methodology for constructing an ontology. Also, research and development research was employed, and there were three phases as follows:

Phase 1. Investigating applications of cultural knowledge of Thai-Khmer textiles. This was done to investigate how the target groups (university lecturers, scholars, weavers, entrepreneurs, and university students in Buriram and Surin provinces) applied cultural knowledge to gain insightful information regarding applications of the cultural knowledge of Thai-Khmer textiles, issues in applying the cultural knowledge of Thai-Khmer textiles, and suggestions to develop a quality tool for retrieving cultural knowledge of Thai-Khmer textiles.

Phase 2. Studies on structures of cultural knowledge in Thai-Khmer textiles. This phase was done to analyze data from four resources: Thai Journal Online, Thailand Library Integrated System, Thailand Culture Research, and Digital Research Information Center. After that, the structures were designed by applying the cultural frameworks of United Nations Educational Scientific and Cultural Organization (2003) and criteria for selecting the cultural heritage of the nation (Department of Cultural Promotion [Thailand], 2017). Then, the data was classified by classes, sub-classes, and divisions (Broughton, 2006; Kumar, 1992); moreover, the properties of cultural knowledge in Thai-Khmer textiles were analyzed by the Facet Classification Approach of Spiteri (1998).

Phase 3. Ontologies design of cultural knowledge in Thai-Khmer textiles. This was accomplished to develop a structure of cultural knowledge of Thai-Khmer textiles from the analyzed content, which was 1) textiles and 2) cultural knowledge of Thai-Khmer textiles and building an ontology. The researchers employed the ontology development frameworks of Noy and McGuinness (2001) and Uschold and King (1995) to get a body of ontology and evaluation. Processes in building ontologies were 1) generating objectives of ontologies, 2) building ontologies, 2.1) forming classes, sub-classes, and classifying class relationships, 2.2) setting properties, 2.3) creating keywords or data representatives, and 3) preparing two types of ontologies evaluations using OntOlogies Pitfall Scanner (OOPS!) (Poveda-Villalón, 2021) with six experts in ontologies and Thai-Khmer textiles. The tool for building

 https://doi.org/10.1633/JISTaP.2023.11.2.2

ontologies was Protégé version 5.5 (Stanford University, 2021) and the tool for disseminating ontologies through World Wide Web was WIDOCO (Garijo, 2017), as shown in Fig. 1.

3. IMPLEMENTATION

The researchers built ontologies regarding cultural knowledge in Thai-Khmer textiles by using two classes: textiles and cultural knowledge in the textiles. This ontologies development applied the methodologies of Uschold and King (1995) and Noy and Mcguinness (2001) in the form of 1) generating objectives of ontologies, 2) building ontologies, 2.1) setting classes, sub-classes, and classifying class relationships, 2.2) setting properties, and 2.3) creating keywords or data representatives, and 3) preparing two types of ontologies evaluations using OOPS! (Poveda-Villalón, 2021) with three experts in ontologies and three experts in Thai-Khmer textiles.

3.1. Generating Objectives of Ontologies

As a result of insufficient data and key terms regarding standard words, local words, cultural knowledge of textile words, and specific ethnic words, the ontology was developed and aimed at providing ease to users in retrieving and accessing data on cultural knowledge of Thai-Khmer textiles. Further, users would like databases to have more key terms regarding standard words, local words, cultural knowledge in textile words, and specific ethnic words. For example, when users searched for "Ka new fabric," "Ka neaw fabric," "Hang Ka Ronk silk," and "Kuab fabric," they got the same search result. To solve this problem, the researchers formed objectives in building ontologies regarding the cultural knowledge in Thai-Khmer textiles, as follows:

1) Provide easy access to cultural knowledge in Thai-Khmer textiles with related keywords.

2) Provide distinctive scopes of cultural knowledge in Thai-Khmer textiles with definitions, properties, characteristics, sample data, and relationships to ease users' access and retrieve needed information.

After generating objectives, the researchers formed structures of the ontologies, including two main groups: textiles and culture in textiles. Then, these two groups were developed into a major domain called cultural knowledge in Thai-Khmer textiles, and finally, they were classified into ontologies of cultural knowledge in Thai-Khmer textiles.

3.2. Building Ontologies of Cultural Knowledge in Thai-Khmer Textiles

3.2.1. Forming Classes, Sub-Classes, and Relationships

The ontologies of cultural knowledge in Thai-Khmer textiles consist of two main classes and sixteen sub-classes:

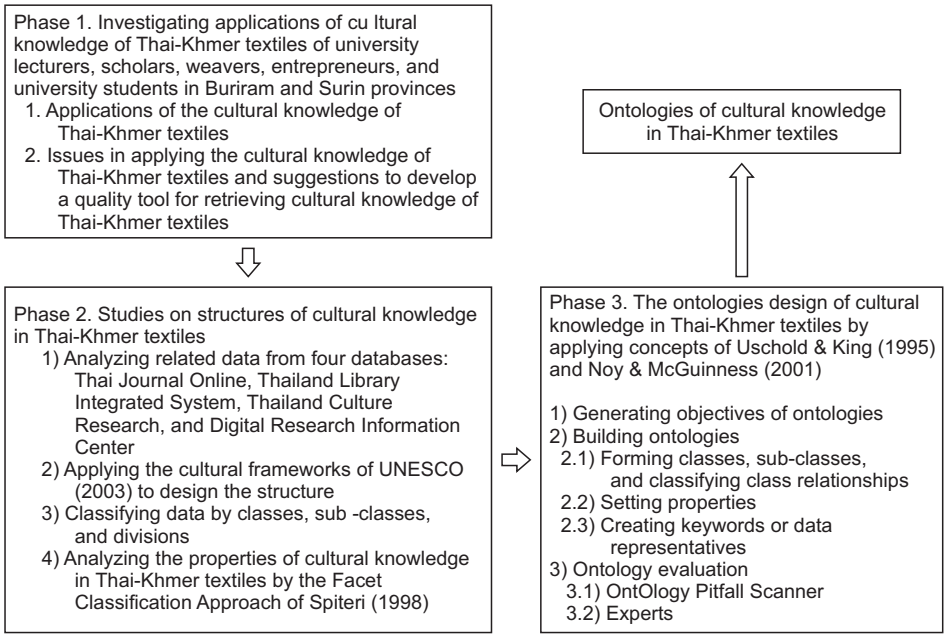


Fig. 1. Research conceptual framework. UNESCO, United Nations Educational Scientific and Cultural Organization.

namely 1) the textile class, including silkworm, mulberry, diseases and pest, silk thread, color, mordant, pattern, weaving tool, weaving process, weaving techniques, decoration, and fabric, and 2) culture in textiles, including belief, clothes, ritual, and tradition, as shown in Fig. 2.

Forming relationships between classes in the ontologies of cultural knowledge in Thai-Khmer textiles by object properties can be classified as sixteen types of relationships, including MadeFrom, hasRelationship, hasColor, hasDyeAuxiliaries, DyeAuxiliariesOf, ToolOf, hasTool, UseFor, BeliefOf, hasBelief, ProcessOf, hasProcess, DeseaseOf, hasDesease, hasCloth, and ClothOf.

3.2.2. Forming Properties

The researchers investigated two existing ontologies, which were Prea-wa silk ontology, consisting of four classes: 1) production, 2) textile products, 3) textile types, and 4) patterns (Singthongchai et al., 2012); and Lanna textile ontology, consisting of four classes: 1) weaving materials, 2) weaving tools, 3) raw materials for textile production, 4) textile production process, and 5) weaving techniques (Arayaphan & Panommit, 2019), with the finding that the similar classes were patterns, production process, textile types, and textile products. Further, the researchers found that these two ontologies did not include cultural knowledge of textiles; therefore, classes of patterns, production processes, and textiles were applied along with cultural knowledge of textiles to the developed ontology.

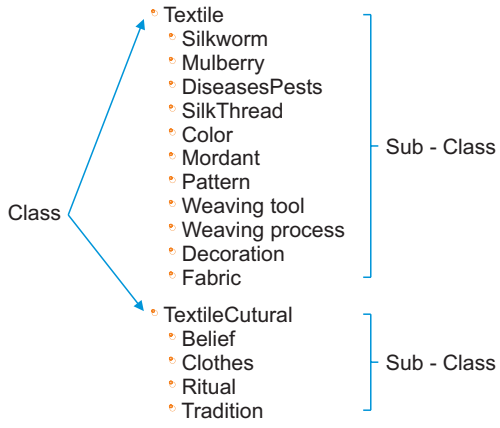


Fig. 2. Classes and sub-classes of ontologies of cultural knowledge in Thai-Khmer textiles.

3.2.3. Forming Keywords or Data Representatives

There are two class properties: object properties and datatype properties, which describe the properties of information access between classes through meaning. There were 69 class properties of the developed ontology consisting of 1) textile class, including 58 properties; e.g., the Disease and pest class includes two types of properties: object properties (Name, Cause, Period, Symptom, and Prevention) and datatype properties (String); and 2) culture of textiles, consisting of 11 properties; e.g., the Ritual class includes two types of properties: object properties (Name, Ritual nature, and Description) and datatype properties (String). Setting class properties helps picture cultural knowledge of Thai-Khmer textiles vividly. There are three data properties, including Name, Period, and Process. The Name property is a string data type used to contain any name class in the ontology. Period property is a string data type used for the beginning to the end of the moment. Process property is a string data type used for the procedure of class in the ontology. Examples of class and property definitions are in Tables 1 and 2.

3.2.4. Creating Individuals or Instances

The researchers examined cultural knowledge in Thai-Khmer textiles and scopes of knowledge structures to form individuals or instances by inserting data according to class properties and attributes. 493 instances included 1) textile class for 422 instances and 2) culture in textiles class for 71 instances. For more details of the whole instance or individual, please access online: http://reg-rdi.com/OntologyTextile/index-en.php#home.

Table 1. Example of ritual class: definitions and class properties

Ritual	Using textiles in rituals of the Thai-Khmer ethnic group inherited from generation to generation			
Object properties	Datatype Property	Description	Class relationships	Types of relationship
Name	String	Name of a ritual	Belief	hasBelief
			Cloth	hasCloth
Ritual nature	String	Type of a ritual		
Description	String	Explanation of textile applications		

 https://doi.org/10.1633/JISTaP.2023.11.2.2

Table 2. Example of sub-classes and instance of ritual class

Class	Sub-class	Sub-class	Description	Instance
Ritual			Using textiles in rituals of the Thai-Khmer ethnic group inherited from generation to generation	
	StatusChangeRitual		Rituals related to changes from one status to one another, e.g. changes from birth to illness and death	
		Birth ritual	Rituals related to birth	
			Rituals that summon guardians of newborns	Johm kroo gamnert ritual
			Ritual of shaving newborn hair	Shaving newborn hair ritual
		Curing ritual	Rituals related to curing	
			Rituals to treat illnesses	Jol mamuad ritual
			Rituals related to summoning the souls of ill people	Soul summoning ritual

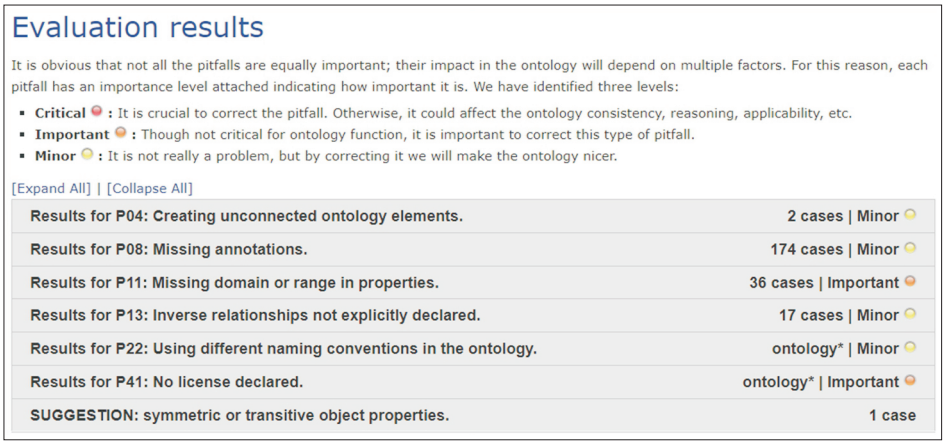


Fig. 3. Evaluation results from OntOlogies Pitfall Scanner.

3.3. Evaluations of the Ontologies of Cultural Knowledge in Thai-Khmer Textiles

The researchers employed two types of evaluation: 1) OOPS!, which detects defects regarding the structure, function, and usability. The application provides three recommendations: critical issues which the developers have to fix immediately, since defects affect the structure and might cause damages; important issues which the developers can fix later, since defects might not cause significant damages; and minor issues which the developers can fix, since defects might cause small damages to classes or relationships. The application results can help developers build effective ontologies (Chansanam et al., 2021; Poveda-Villalón et al., 2014).

Moreover, the application suggested 40 fixed issues and 2) expert evaluation consisting of two groups of experts who were purposively selected as three experts in ontologies and three experts in Thai-Khmer textiles. They were asked to evaluate five aspects of ontologies building processes: the process of identifying scopes of definitions and objectives of development, generating classes, forming properties of classes, setting instances of classes, and applications and guidelines for further development. The evaluation results are interpreted as average scores and then summarized with the descriptive analysis, as shown in Fig. 3.

3.4. WIDOCO-Presenting the Ontologies through World Wide Web: WWW with WIDOCO

The researchers brought WIDOCO to present the developed ontologies: 1) selecting a template to be the Ontologies Web Languages file, 2) uploading metadata, and 3) uploading the textile data, as shown in Fig. 4. Once the preparation is finished, all files of cultural knowledge

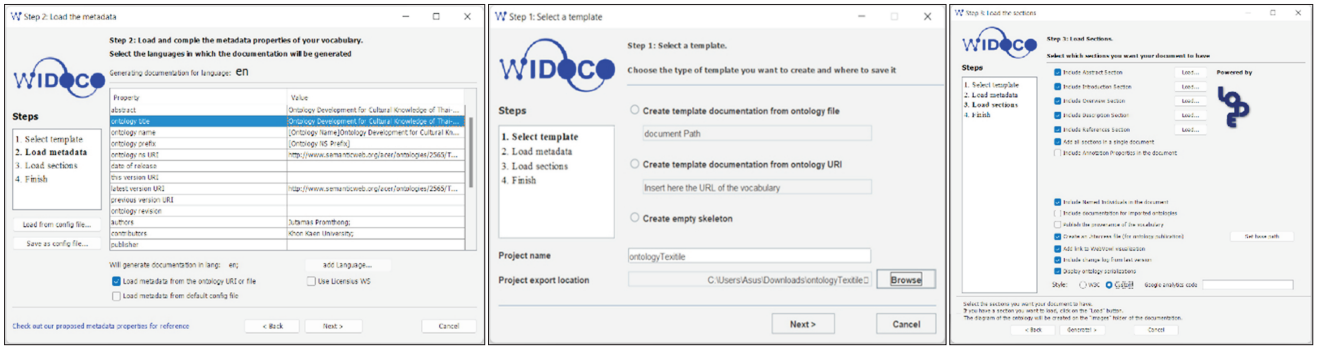


Fig. 4. Steps in creating hypertext markup language file on WIDOCO.

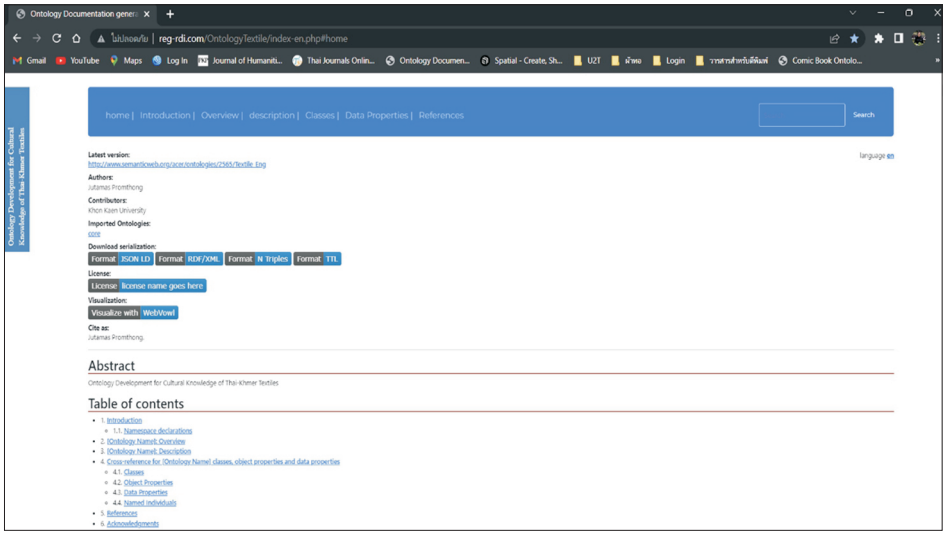


Fig. 5. Screenshot of the ontologies of cultural knowledge in Thai-Khmer textiles through World Wide Web (http://reg-rdi.com/OntologyTextile/index-en.php#home).

in Thai-Khmer textiles will be uploaded and presented at http://reg-rdi.com/Texttile1/index-en.php, as shown in Fig. 5.

4. RESULT AND DISCUSSION

4.1. Result

4.1.1. The Classes of Cultural Knowledge Thai-Khmer Textiles Ontology

The ontologies of cultural knowledge in Thai-Khmer textiles consist of two classes, which are 1) textiles and 2) cultural knowledge in Thai-Khmer textiles. For the first class, there were 12 sub-classes: 1.1) Silkworm sub-class, consisting of 7 divisions, 3 meaning properties, 9 class properties, and 17 instances; 1.2) Mulberry sub-class, consisting of 5 divisions, 5 class relationships, 5 class properties, and 19 instances; 1.3) Diseases and pests sub-class, consisting of 2 divisions, 2 class relationships, 9 class properties, and 16 instances; 1.4) Silk threads sub-class, consisting of 2 class relationships, 4 class properties, and 3 instances; 1.5) coloring sub-class, consisting of 9 divisions, 3 class relationships, 6 class properties, and 52 instances; 1.6) Mordant sub-class, consisting of 2 divisions, 1 class relationshi, 2 class properties, and 23 instances; 1.7) Pattern sub-class, consisting of 12 divisions, 1 class relationships, 3 class properties, and 80 instances; 1.8) Weaving tools sub-class, consisting of 12 divisions, 2 class relationships, 4 class properties, and 59 instances; 1.9) Weaving processes sub-class, consisting of 3 divisions, 2 class relationships, 4 class properties, and 10 instances; 1.10) Weaving techniques sub-class, consisting of 4 divisions, 3 class relationships, 4 class properties, and 12 instances; 1.11) Decoration techniques sub-class, consisting of 4 class relationships and 6 instances; and 1.12) Fabric sub-class, consisting of 4 divisions, 3 class relationships, and 58 instances.

The second main class consists of 4 sub-classes, which

 https://doi.org/10.1633/JISTaP.2023.11.2.2

are 2.1) Dressing sub-class, consisting of 6 divisions, 1 class relationship, 3 class properties, and 30 instances; 2.2) Rituals sub-class, consisting of 19 divisions, 1 class relationship, 2 class properties, and 36 instances; 2.3) Traditions sub-class, consisting of 7 divisions, 1 class relationship, 2 class properties, and 12 instances; and 2.4) Beliefs sub-class, consisting of 7 divisions, 1 class relationship, 2 class properties, and 37 instances.

4.1.2. Evaluations on The Ontologies of Cultural knowledge in Thai-Khmer Textiles

The results from the OOPS! application show seven issues to fix: 1) two issues of P04 Creating unconnected ontologies elements; 2) 173 issues of P08 Missing annotations in sub-class; 3) 36 issues of P11 Missing domain or range in properties; 4) 17 issues of P13 Missing inverse relationships; 5) one issue of P22 Using different naming criteria in the ontologies; 6) one issue of P41 No license declared; and 7) a suggestion of symmetric or transitive object properties.

Further, evaluation results from the experts revealed that the overall ontologies development was suitable at 4.51 on a 5-point Likert scale (Likert, 1967). However, the process of identifying scopes of definitions and objectives of development was 4.61, the applications and guidelines for further development was 4.58, and the process of forming classes or concepts was 4.53. The least suitable process was forming the class's instance (4.33), as shown in Fig. 6.

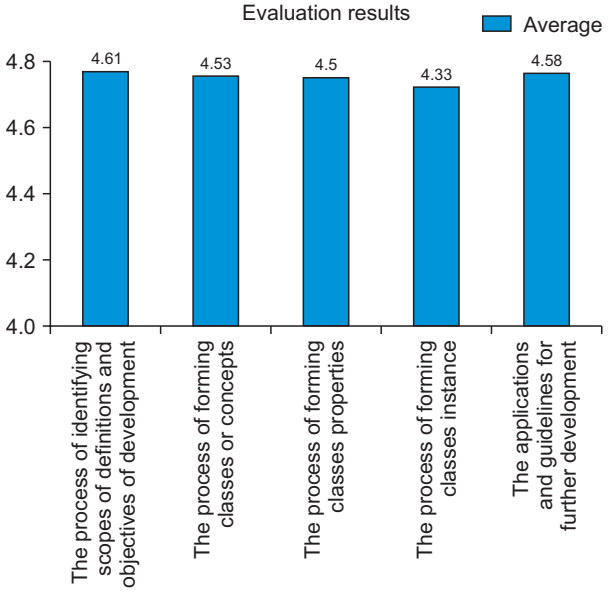


Fig. 6. Evaluation results from the experts (5 pt. Likert).

4.2. Discussion

The cultural knowledge in Thai-Khmer textiles is distinctive and unique. According to the results, each class links to the other; for example, the pattern class relates to classes of weaving techniques, fabrics, and beliefs. The knowledge is a legacy, handed on from generation to generation. In addition, each generation has further developed patterns and weaving techniques to have various patterns of textiles. This goes along with Gruber's (1995) finding that ontologies are basic knowledge formed by having specific domains and that it is important to reuse and add more values. Besides this, the present paper's results revealed that the ontologies of cultural knowledge in Thai-Khmer textiles are specific, including plants used in the dying thread process, mordants, weaving tools, patterns, traditions, and rituals. Additionally, the ontologies show relationships among classes by meaning and keywords. In regard to Browne and Jermey (2004) and Taylor (2004), ontology is a process of generating a controlled vocabulary to explain objects and their relationships by terms, synonyms, properties, and instances; then creating relations by linking meaning with terms.

The ontologies of cultural knowledge in Thai-Khmer textiles are formal representations of a set of concepts within a domain and the relationships between those concepts. When built and used in a search or reasoning system, ontologies can provide several benefits:

1) Improved accuracy: By providing a clear and standardized structure for representing and organizing information, ontologies can help ensure that search results are more accurate and relevant.

2) Enhanced search functionality: Ontologies can enable more sophisticated search functionality, such as searching for conceptually related terms or narrowing down results based on specific relationships between concepts.

3) Better integration of diverse data sources: Ontologies can help integrate information from multiple sources by providing a shared vocabulary and structure for organizing that information.

4) Greater consistency in reasoning: Ontologies can support automated reasoning by providing a clear set of rules for inferring new information based on existing knowledge. This can lead to more consistent and reliable conclusions.

5) Increased reuse of knowledge: Ontologies can facilitate the reuse of knowledge by providing a clear and standardized representation of concepts and relationships

that can be accessed and utilized by different systems and applications.

5. CONCLUSIONS

The ontologies of cultural knowledge in Thai-Khmer textiles are considered domain ontologies that store specific knowledge regarding cultural knowledge in Thai-Khmer textiles. The ontologies development was a top-down process accomplished by 1) generating classes in a domain of cultural knowledge in Thai-Khmer textiles from general definitions; 2) forming main classes, textiles and cultural knowledge in Thai-Khmer textiles; 3) forming sub-classes such as setting textiles class and adding sub-classes such as silkworm, mulberry, diseases, silk threads, color, mordant, weaving tools, weaving process, weaving technique, decoration, and fabrics; and 4) managing class relationships, setting main class and sub-class properties, and generating vocabulary or data representatives.

The evaluation results from the application and the experts revealed that the ontologies are suitable for further development in knowledge retrieval regarding the meaning and to further build cultural knowledge in other ethnic groups. Moreover, these ontologies could be further developed as a recommendation system of cultural knowledge to guide users regarding Thai-Khmer textiles.

In sum, using the ontologies of cultural knowledge in Thai-Khmer textiles can improve search and reasoning capabilities by providing a more structured and standardized representation of knowledge and enabling more sophisticated search and reasoning functionality.

ACKNOWLEDGMENTS

This article was a part of a graduate research grant supported by the National Research Council of Thailand for the year 2020.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Arayaphan, W., & Panommit, P. (2019). Ontology development for knowledge of lanna textiles. *Journal of Human Sciences,* 20(2), 133-170. https://so03.tci-thaijo.org/index.php/JHUMANS/article/view/213167

Boonpok, D. (2019). [Identity and symbols of the ethnic Khmer in Thailand's silk weaving cloths]. *Faculty of Humanities and Social Sciences Thepsatri Rajabhat University Journal*, 10(1), 91-102. Thai. https://so01.tci-thaijo.org/index.php/truhusocjo/article/view/210012

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings,* 58(1/2), 49-72. https://doi.org/10.1108/00012530610648671

Browne, G., & Jermey, J. (2004). *Website indexing: Enhancing access to information within websites* (2nd ed.). Auslib Press.

Chansanam, W., Tuamsuk, K., Kwiecien, K., Sutthiprapa, K., & Supnithi, T. (2021). Thai tattoo wisdom's representation of knowledge by ontology. *Informatics,* 8(1), 3. https://doi.org/10.3390/informatics8010003

Department of Cultural Promotion (Thailand). (2017). *Criteria for selecting the cultural heritage of the nation 2017*. https://www2.m-culture.go.th/satun/ewt_dl_link.php?nid=1370

Garijo, D. (2017, October 21-25). WIDOCO: A wizard for documenting ontologies. *Proceedings of the 16th International Semantic Web Conference (ISWC 2017)* (pp. 94-102). Springer.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies,* 43(5-6), 907-928. https://doi.org/10.1006/ijhc.1995.1081

Howkins, J. (2002). *The creative economy: How people make money from ideas.* Penguin Books.

Iamkhajornchai, P., & Manmart, L. (2018). Knowledge organization system and access of Thai cultural knowledge. *Journal of Information Science,* 36(1), 1-40. https://so03.tci-thaijo.org/index.php/jiskku/article/view/104955

Kumar, K. (1992). *Theory of classification*. Vikas Publishing House.

Likert, R. (1967). The method of constructing an attitude scale. In M. Fishbein (Ed.), *Readings in attitude theory and measurement* (pp. 90-95). Wiley.

Martins, L., Helmer, S., & Arnold, D. Y. (2013). Exploring weaving structures in the andes: Reflections on the creation of a digital archive. *Visual Resources,* 29(1-2), 59-71. https://doi.org/10.1080/01973762.2013.761121

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology.* Stanford University.

Office of the National Economic and Social Development Council. (2019). *[Northeastern development plan 2017-2022]* (review ed.). Office of the National Economic and

Social Development Council. Thai.

Office of Small and Medium Enterprises Promotion (OSMEP). (2017). *20-Year national strategy (2017-2036).* OSMEP.

Pogodina, I. V., Mammadov, S. N., & Avdeev, D. A. (2021). [Textile products as a reflection of religious culture: Retrospective and trends]. *Известия высших учебных заведений. Технология текстильной промышленности,* 2(392), 140-143. Russian. https://doi.org/10.47367/0021-3497_2021_2_140

Poveda-Villalón, M. (2021). *OOPS! (OntOlogy Pitfall Scanner!).* https://oops.linkeddata.es

Poveda-Villalón, M., Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2014). OOPS! (Ontology Pitfall Scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems,* 10(2), 7-34. https://doi.org/10.4018/ijswis.2014040102

Sakamoto, M. (2014). [The change of ethnic relations between Khmu and Tai Lue: A study of textiles cultures in Muang Ngoen, Sayabury province, Lao PDR]. *Journal of Fine Arts,* 5(1), 67-124. Thai. https://so02.tci-thaijo.org/index.php/finearts Journal/article/view/77237

Singthongchai, J., Naenudorn, E., Kittidachanupap, N., Rotkanok, D., & Niwattanakul, S. (2012, May 25-27). Development for Prae-wa silk knowledge base using ontology. In A. Editor & B. Editor (Eds.), *Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)* (pp. 93-97). IEEE.

Spiteri, L. (1998). A simplified model for facet analysis: Ranganathan 101. *Canadian Journal of Information and Library Science,* 23(1/2), 1-30. https://www.researchgate.net/publication/260328294_A_simplified_model_for_facet_analysis_Ranganathan_101

Stanford University. (2021). *Protégé 5.5.0.* https://protege.stanford.edu/

Taylor, A. G. (2004). *The organization of information.* Libraries Unlimited.

United Nations Educational Scientific and Cultural Organization (UNESCO). (2003). *2003 Convention for the safeguarding of the intangible cultural heritage.* UNESCO.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93-136. https://doi.org/10.1017/S0269888900007797

Uschold, M., & King, M. (1995). *Towards a methodology for building ontologies.* University of Edinburgh.

Wei Choo, C. (2000). Working with knowledge: How information professionals help organisations manage what they know. *Library Management,* 21(8), 395-403. https://doi.org/10.1108/01435120010342770

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 11(2): 22-36, 2023

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2023.11.2.3

Growth and Distribution of Research Journals Across the World

Ramesh Pandita*
Central Library, Baba Ghulam Shah Badshah University, Rajouri, India
E-mail: rameshpandita90@gmail.com

Shivendra Singh
Central Library, Guru Gobind Singh Medical College and Hospital, Baba Farid University of Health Sciences, Farikdot, India
E-mail: shiv.mail@gmail.com

ABSTRACT

This study assesses the growth and distribution of research journals indexed in Scopus from 2000 through 2019, using SCImago data for journal and country ranking. A total of 15,484 journals were indexed in Scopus up till 1999, and from 2000 to 2019, 11,037 journals were indexed in Scopus at an average of 581.85 journals per year at an average annual corresponding growth of 2.74%. Of the total journals indexed in Scopus, 41.61% were indexed during the last two decades. Europe is the leading continent publishing nearly 60% of the total world journals, while Africa publishes around 1% of journals, the lowest among all the continents. The United Kingdom, United States (US), China, Brazil, Australia, and Egypt are the leading countries publishing the maximum number of journals from their respective continents. Of the total journals indexed across the world, nearly one-fourth (6,425, 24.23%) are published in the US. Still more, of the 109 journal publishing countries across the world, nearly 90% of journals are published in twenty leading journal publishing countries. Among the 27 major Scopus recognized subject disciplines, the majority (1,606, 6.06%) of journals are published in the field of medicine. The growth of research journals and the growth of research results has moved at a uniform pace, and any asymmetry between the two is bound to result either in the introduction of predatory and dubious journals or in the production of sub-standard research results.

Keywords: journal growth, journal distribution, Scopus, world, SCImago, scientometrics

Received: November 14, 2022 **Revised:** January 19, 2023
Accepted: January 31, 2023 **Published:** June 30, 2023

***Corresponding Author:** Ramesh Pandita
https://orcid.org/0000-0001-8568-412X
E-mail: rameshpandita90@gmail.com

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© Ramesh Pandita, Shivendra Singh, 2023

1. INTRODUCTION

Researchers across the world are always keen to publish their research results in quality research journals that have greater visibility and acceptance among information seekers. The research journals indexed in Scopus and Web of Science (WoS) are by and large considered as qualitative and authoritative sources of information, and enjoy both popularity and acceptance across the world for maintaining a very high indexing standard for research journals desirous to be indexed. Given this fact, research journals indexed in any of these two popular indexes are considered both qualitative and authoritative, whereby the genuineness of such journals is not generally questioned by researchers while submitting their research results for publication. Researchers publishing their research results in journals not indexed in Scopus or WoS are generally considered substandard, low quality, fake, dubious, and even of a predatory nature. The question remains, are mainstream recognized research journals available in sufficient numbers to publish all the research results produced across the world? This is more a race between demand and supply, whereby the demand and availability of quality and authentic research journals to publish all the research results produced by the world scientific community is met or not. Any existing gap between the quality journals published and the number of research results published in them somewhere offers a breeding ground for predatory, dubious, sub-standard, and poor-quality research journals.

The research results publishing platform is as important as the research itself. What is more a paradox is that a journal attains recognition by publishing quality research results over the years, and thereafter it is the very journal which becomes the parameter in deciding research quality; furthermore, the irony is that the scientific community too upholds this notion that research papers published in a particular mainstream recognized journal shall be more qualitative. The need is to understand that a journal is not the reason for quality research, but quality research is the reason for the quality journal. In absence of a good publishing platform, it is difficult to ascertain and validate the research results. Research journals are the globally accepted medium to publish research results produced, still more when the journal publishing market is flooded with predatory, dubious, sub-standard, and low-quality journals; thereon the only choice left with the scientific and research community is to trust the journals indexed in Scopus and Web of Science, or for that matter, journals indexed in any other popular index for being both qualitative and authoritative. Accordingly, in the undergoing study, an attempt has been made to reflect the growth and distribution of journals indexed in Scopus during the last two decades, viz. for the period 2000 through 2019.

2. PROBLEM STATEMENT

The research results produced across the world and the number of quality research journals available for publishing all such research results is quite disproportionate. The requirement of quality journals is more than available, hence the research results produced are beyond the publishing capacity of the available quality journals; as a result, researchers seek alternate routes to publish their research results, which mostly lead to predatory, dubious, sub-standard, poor, and low-quality research journals. Still more, researchers spend a good amount of time in ascertaining whether a particular journal is indexed in a specific index or not, so apart from relevance, the indexing of a journal plays an important part in submitting a manuscript for publication. Given this fact, it becomes quite imperative to ascertain the growth and distribution of quality research journals across the world on the similar lines as the quantification of research growth at subject level is done. Accordingly, the present study assesses the growth, distribution, and quantum of quality research journals indexed in Scopus and launched globally on an annual basis.

3. OBJECTIVES OF THE STUDY

The present study has been undertaken with the following objectives.

- To assess the growth of research journals at the global level based on indexing of journals in Scopus during the last two decades, viz., the year 2000 through the year 2019.
- To compute the distribution and growth of Scopus-indexed research journals at the continental level.
- To find out the average annual corresponding growth of journals both at the continental and world level, along with the average number of journals introduced each year.
- To identify the leading journal publishing countries at the world level along with the leading journal publishing countries at the continental level.
- To work out the distribution and growth of journals

across the major subject disciplines as identified in Scopus alongside some other leading sub-subject disciplines.

4. REVIEW OF LITERATURE

Quantification of research output of individuals, institutions, countries, continents, subject disciplines, etc., is very common among the research and scientific community of the world. The quantification of research journals rather than the platforms available where researchers may publish their research results is quite uncommon. Meadows (1993) did some pioneering work to assess the growth of research journals at the global level and found around 10,000 journals published during 1951, and the number reached around 71,000 journals during 1987, while Price (1963) was of the view that the number of journals would reach around one million towards the end of the twentieth century. Ulrich's periodical directory is known for its comprehensiveness in covering almost all research journals published across the length and breadth of the world, available both in CD-ROM and online form (Mabe, 2003). Around one lakh research journals are indexed in Ulrich's periodical directory, including both open and toll-access journals published all across the world (ProQuest, 2021).

The fact of the matter is that research journals are as important as the research itself, because it is these recognized mainstream journals that not only help convey the research results to its end users but also are the authority to uphold and validate the research results (Mesa Melgarejo & Galindo Huertas, 2011). In a study Laakso et al. (2011) analyzed the growth of research journals and research articles during the period 2000 to 2009 and found that research articles on average grew at 30% annually, while the research journals on average grew at 18%. For the timely publication of research results and for sustainable and balanced research growth, introduction of new research journals is equally important (Pandita et al., 2017). The introduction of new journals or the growth of journals depends upon several factors, foremost being the quantum of research work undertaken in any given subject discipline (Singh, 1978). It is being observed that chemistry is one of the leading subject disciplines in which research results are published in far greater numbers across the world; consequently, manifold growth is being observed in the introduction of new research journals in the field of chemistry. By the year 1910, chemical abstract covered 435 research journals in the field of chemistry, and this number grew up to 14,000 journals by 1975-1976 (Dutta, 1968; Singh, 1978). Similarly, research journals in the field of library and information sciences were found to double in every 13.8 years (Bottle & Efthimiadis, 2016).

In another study, while assessing the growth of journals, articles, and authors in the field of malaria for the period 1955 through 2005, Rao and Srivastava (2010) observed that the number of journals, articles, and authors in the malaria field grew annually at 5.31%, 7.38%, and 10.06%, respectively. While analyzing the growth of research journals in India indexed in Scopus and the growth of research articles published by Indian authors in Scopus indexed journals, Singh and Pandita (2017) found that between 2002 to 2010 research journals in India grew at 3.29% annually, while research articles grew at 5.44% annually, which also means the growth of research articles in India during the period of study remained 65.34% greater than the growth of research journals. In a similar study, Mabe (2003) computed the growth of journals in the twentieth century during three different periods. The first period lasted between 1900 through 1940 and the researcher found that journals during this period grew by 3.23% annually; during the second phase, which lasted between 1945 through 1976, the journals grew at 4.35% annually and during the third phase, which lasted from 1977 to 2000, journals grew at 3.26% in number annually.

There are other similar studies which were undertaken to assess the growth of research journals at the local level. The growth of journals at the local level may correspond to country level, subject level, regional level, or even the continental level. In a similar study, while assessing the growth of journals indexed in WoS from Los Angeles (LA), the researcher observed that journals from Latin American and the Caribbean increased from 69 in 2006 to 248 in 2009; however, the researcher owed this manifold growth to change in the editorial policy of WoS which resulted in the inclusion of more existing journals in WoS rather than the introduction of new journals (Collazo-Reyes, 2014). Accordingly, Pandita et al. (2017) found that 15,631 journals were introduced in India from 2005 through 2014, but only 3% of them were indexed in Scopus while the majority of them were non-indexed and grew annually in quantity by 31.44%. This also indicates that around 97% of journals introduced failed to get recognized and listed in quality indexes. Given this fact, Sen and Lakshmi (1992), while finding only a few journals in science and technology from India among Science Citation Index (SCI) journals, observed that slight improvements in the quality of publishing can help many Indian research journals to find a place among SCI journals.

 https://doi.org/10.1633/JISTaP.2023.11.2.3

The introduction of new journals is important to cater to both present and future research results publishing requirements. Besides this, every journal has its limitations of being confined to certain subject disciplines, having a limited scope, covering research articles of local interest generally, and more. Accordingly, all such journals published in countries like India, Australia, Israel, and Canada have been dubbed as journals with an "island effect" (Arunachalam & Markanday, 1981). Factually, it takes time for a journal to mature over the years and then get indexed in any quality index. Indexing, apart from giving recognition to a research journal as a quality journal, also gives it more visibility among the scientific community. The low visibility of a journal among its professionals has been referred to as "cloistering culture" (Mesa Melgarejo & Galindo Huertas, 2011), but the fact of the matter is that each journal enjoys a certain degree of reputation and popularity among its readers, which generally depends upon the degree of standard each journal follows while publishing research results (Seglen, 1997). Still more, when the journal publishing market is flooded with predatory and dubious journals, it becomes imperative for all the quality and genuine journal publishers to index their journals in quality indexes. Indexing status is one important criterion for a journal to establish itself as a recognized mainstream research journal (Ayeni & Adetoro, 2017).

5. MATERIALS AND METHODS

This study is based on secondary data retrieved from the SCImago journal and country rankings on July 30, 2021, and can be accessed at www.scimago.com (SCImago, 2021). The data upon retrieval was in semi-structured form, and as such was structured keeping in view the objectives of the study. While structuring the data, trade journals, conferences, and proceedings and book series were excluded from the study, and only journals were considered for the present study. It is pertinent to mention that SCImago data is based on a Scopus data source maintained by Elsevier. It is a well-known fact that it takes at least anything between four to five years for a journal to mature and be indexed in Scopus or WoS, and accordingly, the growth and distribution of journals has been computed based on their indexing in Scopus. To avoid any controversy or ambiguity around the concept of the introduction of new research journals, we have considered all such journals from the date they have been indexed in Scopus. The scope of the study is global, and to make it more authentic and relevant, it was desired to analyze the data for a longer period. Accordingly, the undergoing study has been undertaken on the data lasting for the period 2000 through 2019.

6. RESULTS

To structure the data and to perform mathematical expressions like addition, subtraction, multiplication, and drawing percentages, Microsoft Excel 2010 (Microsoft, Redmond, WA, USA) has been used. The percentage at most of the places has been drawn up to two decimal places and has been rounded off to a hundred percent figure. The formula used to perform mathematical expressions has been reflected in each column.

The primary objective of the present study is to reflect the growth of research journals across the world during the last two decades; as such the data reflected in Table 1 lasts for the period year 2000 through 2019. It is pertinent to mention that up till the year 1999, 15,484 research journals were indexed in Scopus, and accordingly, the corresponding growth and percentage corresponding to the growth of research journals for the year 2000 have been computed. A total of 11,037 research journals were indexed in Scopus at an average of 551.85 research journals each year, constituting around 2.74% of the average annual corresponding growth. This also means that approximately 550 new research journals are launched each year across the world. Nevertheless, it takes anything between three to five years for a new research journal to get indexed in a database like Scopus, but for our study and to avoid any controversy or ambiguity around the introduction of new research journals we have considered all journals from the date they have been indexed in Scopus. The highest number (1,009, 9.14%) of research journals were indexed during the year 2011 and the lowest number (154, 1.40%) were indexed during the year 2000. On the date of data retrieval, a total of 26,521 research journals were indexed in Scopus, and of these 11,037 (41.61%) were indexed during the last two decades.

It is quite interesting to know that up till 1999, 15,484 (58.39%) journals were indexed in Scopus. It can also be inferred that post-year 2000, Scopus has somewhat raised the quality bar to index newly introduced research journals, which has resulted in indexing a lesser number of research journals post-2000. On the other hand, post-2000, exponential growth was observed in the launch of new research journals, which in turn should have resulted in the indexing of a far greater number of research journals,

Table 1. Year wise growth of research journals indexed

Serial no.	Year (A)	Journals indexed (B)	Corresponding growth (C)	Percentage corresponding growth (D)[a)]	Percentage of C (E)[b)]	Percentage of cumulative growth (F)[c)]
n	1999	15,484	-	-	-	-
01	2000	15,638	154	0.99	1.40	58.96
02	2001	15,971	333	2.13	3.02	60.22
03	2002	16,723	752	4.71	6.81	63.06
04	2003	16,900	177	1.06	1.60	63.72
05	2004	17,214	314	1.86	2.84	64.91
06	2005	17,624	410	2.38	3.71	66.45
07	2006	18,151	527	2.99	4.77	68.44
08	2007	18,673	522	2.88	4.73	70.41
09	2008	19,623	950	5.09	8.61	73.99
10	2009	20,622	999	5.09	9.05	77.76
11	2010	21,462	840	4.07	7.61	80.92
12	2011	22,471	1,009	4.70	9.14	84.73
13	2012	23,105	634	2.82	5.74	87.12
14	2013	23,642	537	2.32	4.87	89.14
15	2014	24,252	610	2.58	5.53	91.44
16	2015	24,718	466	1.92	4.22	93.20
17	2016	25,117	399	1.61	3.62	94.71
18	2017	25,460	343	1.37	3.11	96.00
19	2018	26,257	797	3.13	7.22	99.00
20	2019	26,521	264	1.01	2.39	100.00
Total (average)			11,037 (551.85)	54.71 (2.74)		

a) $\frac{Bn\text{-}B(n\text{-}1)}{B(n\text{-}1)} \times 100$. b) $\frac{C}{Total\ of\ C} \times 100$. c) $\frac{B}{Total\ of\ B} \times 100$.

but the case is not so. Also, the recent exponential growth in research journals has raised much hue and cry among the research and scientific community across the world, whereby many predatory publishers and predatory journals have surfaced. Given this fact, it has become more important for indexes like Scopus or WoS to opt for more stringent measures to index a research journal. A good number of predatory research journals were previously found to be indexed in indexes like Scopus and WoS; thereby it somewhere doubles the responsibility of all such agencies to be careful with such predatory publishers while indexing new research journals. If the research journals indexed in Scopus continue to grow at the same pace, then by 2030 there will be around 36,000 journals indexed in Scopus.

It is equally desirable and important to assess the growth of research journals across different continents of the world. Accordingly, Table 2 depicts the scenario of the number of research journals introduced across different continents during the last two decades and indexed in Scopus. From the data it is evident that Europe is the leading continent, publishing as of date 14,944 research journals, distantly followed by North America and Asia, publishing 6,709 and 3,308 research journals, respectively. As reflected in Table 1, on average around 550 research journals are being introduced each year across the world

 https://doi.org/10.1633/JISTaP.2023.11.2.3

Table 2. Year wise journals indexed across continents

Serial no.	Year	Africa			Asia			Australia			Europe			North America			South America			Total		
		Jrls	CG	CG (%)	Jrls	CG	CG (%)	Jrls	CG	CG (%)	Jrls	CG	CG (%)	Jrls	CG	CG (%)	Jrls	CG	CG (%)	Jrls	CG	CG (%)
	1999	79	-	-	1,336	-	-	177	-	-	8,498	-	-	5,135	-	-	256	-	-	15,481	-	-
1	2000	76	-3	-3.80	1,371	35	2.62	169	-8	-4.52	8,611	113	1.33	5,154	19	0.37	257	1	0.39	15,638	157	1.01
2	2001	79	3	3.95	1,451	80	5.84	166	-3	-1.78	8,838	227	2.64	5,174	20	0.39	263	6	2.33	15,971	333	2.13
3	2002	83	4	5.06	1,488	37	2.55	175	9	5.42	9,228	390	4.41	5,467	293	5.66	282	19	7.22	16,723	752	4.71
4	2003	88	5	6.02	1,531	43	2.89	167	-8	-4.57	9,380	152	1.65	5,451	-16	-0.29	283	1	0.35	16,900	177	1.06
5	2004	93	5	5.68	1,576	45	2.94	166	-1	-0.60	9,607	227	2.42	5,485	34	0.62	287	4	1.41	17,214	314	1.86
6	2005	97	4	4.30	1,670	94	5.96	173	7	4.22	9,865	258	2.69	5,527	42	0.77	292	5	1.74	17,624	410	2.38
7	2006	101	4	4.12	1,810	140	8.38	187	14	8.09	10,128	263	2.67	5,573	46	0.83	352	60	20.55	18,151	527	2.99
8	2007	104	3	2.97	1,869	59	3.26	195	8	4.28	10,463	335	3.31	5,618	45	0.81	424	72	20.45	18,673	522	2.88
9	2008	131	27	25.96	2,048	179	9.58	236	41	21.03	10,964	501	4.79	5,709	91	1.62	535	111	26.18	19,623	950	5.09
10	2009	157	26	19.85	2,292	244	11.91	292	56	23.73	11,370	406	3.70	5,925	216	3.78	586	51	9.53	20,622	999	5.09
11	2010	193	36	22.93	2,446	154	6.72	331	39	13.36	11,781	411	3.61	6,087	162	2.73	624	38	6.48	21,462	840	4.07
12	2011	211	18	9.33	2,625	179	7.32	350	19	5.74	12,399	618	5.25	6,225	138	2.27	661	37	5.93	22,471	1,009	4.70
13	2012	221	10	4.74	2,730	105	4.00	353	3	0.86	12,809	410	3.31	6,278	53	0.85	714	53	8.02	23,105	634	2.82
14	2013	239	18	8.14	2,796	66	2.42	352	-1	-0.28	13,144	335	2.62	6,374	96	1.53	737	23	3.22	23,642	537	2.32
15	2014	254	15	6.28	2,870	74	2.65	361	9	2.56	13,542	398	3.03	6,463	89	1.40	762	25	3.39	24,252	610	2.58
16	2015	252	-2	-0.79	2,934	64	2.23	368	7	1.94	13,884	342	2.53	6,501	38	0.59	779	17	2.23	24,718	466	1.92
17	2016	251	-1	-0.40	3,032	98	3.34	366	-2	-0.54	14,124	240	1.73	6,539	38	0.58	805	26	3.34	25,117	399	1.61
18	2017	254	3	1.20	3,085	53	1.75	361	-5	-1.37	14,334	210	1.49	6,589	50	0.76	837	32	3.98	25,460	343	1.37
19	2018	265	11	4.33	3,247	162	5.25	377	16	4.43	14,777	443	3.09	6,711	122	1.85	880	43	5.14	26,257	797	3.13
20	2019	273	8	3.02	3,308	61	1.88	379	2	0.53	14,944	167	1.13	6,709	-2	-0.03	908	28	3.18	26,521	264	1.01
	Avg (AACG)	6.65 (9.70)			4.67 (98.60)			4.13 (10.10)			2.87 (322.30)			1.35 (78.70)			6.75 (32.60)			2.74 (552.00)		

Jrl, journals; CG, corresponding growth; Avg, average; AACG, average annual corresponding growth.

at an average annual growth of 2.74%. Similarly, at the continental level, Europe is the largest publisher of research journals, which on average introduces 322.30 new research journals each year at an average annual growth of 2.87%. This also means that of the total new research journals introduced across the world, nearly 60% of journals are introduced alone from Europe, and the average annual growth of European research journals is almost on par with the average global journal growth, viz. 2.87%. Asia is the second leading continent, which on average introduces 98.60 new research journals each year, followed by North America with an average introduction of 78.70 research journals, while South America on average introduces 32.40, Australia 10.10, and Africa 9.70 research journals each year. The growth of research journals in terms of percentage is highest across South America at 6.75% followed by Africa at 6.65%, Asia at 4.67%, Australia at 4.13, Europe at 2.87%, and North America at 1.35%, which is even below the average global journal growth rate. It is needless to mention that the higher percentage growth of research journals across continents like South America and Africa is due to the lower number of journals published across these continents, while the introduction of every new research journal across these continents immediately inflates the average annual percentage growth of research journals when compared to the average annual percentage growth rate of journals at the global level.

The growth of research journals in Asia, Europe, and South America was consistent and showed growth over the past two decades, while the growth of research journals across Africa and Australia remained inconsistent, recording mixed corresponding growth. North America by and large recorded consistent corresponding growth in the introduction of new research journals. However, during the years 2003 and 2019 North America recorded negative corresponding growth.

The highest annual corresponding growth rate of newly introduced research journals recorded at the global level remained at 1,009 research journals during the year 2011, while at the continental level, the highest annual corresponding growth levels of newly introduced research journals were recorded for Africa (36, 2010), Asia (244, 2009), Australia (56, 2009), Europe (618, 2011), North America (293, 2002), and South America (111, 2008). Almost all the continents recorded the highest annual corresponding growth in journals during different years; however, the annual corresponding growth of research journals introduced across Europe can be seen as having a significant impact on the overall global journal growth levels.

Of the total 26,521 research journals indexed in Scopus across the world, the majority (14,944, 56.35%) of research journals are published alone across Europe, reflecting the dominance of Europe in publishing research journals across the world. Even putting together, the research journals published across the remaining five continents, the rate is lesser by 6.35% than the number of research journals published alone across Europe. This dominance of Europe in publishing research journals at such a large scale can be seen for several reasons, the foremost being that it is the home of the world's leading academic and research institutions producing quality research results. The other leading continents publishing research journals are North America, publishing 6,709 (25.30%) research journals, and Asia, publishing 3,308 (12.47%) of the total research journals. Together, Europe, North America, and Asia are publishing 24,961 (94.12%) of world research journals, while Africa, Australia, and South America together are publishing a meager 5.88% of research journals. Africa is publishing a mere 273 journals, constituting 1.03% of the total global research journals, the lowest among all the continents (Table 3).

There are around 195 recognized countries all over the world; however, as per the data retrieved from SCImago, there are 218 countries around the world that are actively undertaking research. Of these, 109 (50%) countries are publishing research journals, while the remaining countries depend on these 109 countries for publishing their research results. Each research country is supposed to have its own publishing channels as well, not only to publish the research results of their local researchers but also to provide a publishing platform to other global researchers, which is also imperative to have balanced and sustainable growth. Even if we take 50% of countries as a standard measure for publishing research journals in a region or in a continent, then Africa and South America are the two continents that have inadequate research results publishing platforms. But the fact of the matter is that even 50% of countries publishing journals is not sufficient to fulfill the actual publishing requirement of a region or the world. Contrary to the other five continents, 81% of countries across Europe are publishing their research journals, which reflects the availability of a balanced and wholesome publishing platform for researchers at both the local and global level.

While assessing the leading journal publishing countries at the continental level, it emerged that Egypt is the leading country from Africa, publishing the highest

 https://doi.org/10.1633/JISTaP.2023.11.2.3

Table 3. Journals published across continents

Serial no.	Name of continent (A)	No. of journals published (B)	Percentage of B (C)[a)]	No. of journal publishing countries (D)	Total research countries (E)	Leading journal publishing country at continental level		Percentage share of G from B (H)[b)]
						Name of Country (F)	No. of Journals (G)	
01	Africa	273	1.03	17	48	Egypt	130	47.61
02	Asia	3,308	12.47	31	59	China	665	20.10
03	Australia	379	1.43	03	06	Australia	242	63.85
04	Europe	14,944	56.35	43	53	United Kingdom	6,337	42.40
05	North America	6,709	25.30	02	04	United States	6,425	95.76
06	South America	908	3.42	13	48	Brazil	401	44.16
	Total	26,521	100.0	109	218		14,200	53.54

a) $\frac{B}{Total\ of\ B} \times 100$. b) $\frac{G}{B} \times 100$.

number (130) of journals, constituting 47.61% of the total research journals published across Africa. Similarly, 20.10% of the total research journals published across Asia are published in China, 63.85% of the total research journals published across Oceania are published in Australia, 42.40% of the total research journals published across Europe are published across North America, 95.76% of journals published across North America are published in the United States, and 44.16% of the research journals published across South America are published in Brazil. These six leading journal publishing countries from their respective continents together publish 53.54% of the total research journals globally.

Except in Asia, more than 40% of research journals in each continent are published in one country. This again somewhere reflects that the concentration and dominance of journal publishing in each continent is almost limited to one country. This inequitable and disproportionate publication of research journals across countries at the continental level is bound to be the reason for inequitable and unsustainable growth across countries in each continent.

The dominance of Europe also gets reflected by the number of countries actively publishing research journals across Europe. Of the 109 countries actively publishing research journals across the world, 43 countries come from Europe alone, constituting 39.44% of the total nation-states publishing research journals at the global level. Still more, Europe is a continent with 53 nation-states, which means that over 81% of the countries across Europe are publishing their own research journals. This also somewhat signifies the fact that equitable distribution of resources and services is the sole reason for the proportionate and sustainable growth and development of a country or a region, and so this holds true of Europe, an evenly developed continent in the world. Asia despite being the largest continent in the world both in terms of population and the number of countries, only 31 countries publish research journals, constituting 28.44% of the total countries publishing research journals in the world. Together, 67.88% of countries publishing research journals across the world come from Europe and Asia alone.

Australia and North America are the two continents from where the lowest number of countries, three (2.75%) and two (1.83%), publish research journals, respectively. It is needless to mention that Australia and North America are largely known as single nation continents, and as such these figures cannot be counted as low or less for any reason.

The production of journals around the world is disproportionate. More than half of all journals published worldwide are based in Europe, while continents like Africa, Australia, and South America appear to publish for participation's sake. The research journals published across Africa, Australia, and South America appear even insufficient to publish the research results produced from these continents. Understandably, Europe is one of the most developed continents in the world and is also the home of the world's leading academic, scientific, technical, and research institutions; as such, it being the home of most of the research journals of the world is not that surprising. But what concerns the most is that somewhere this over-

and-above dominance of Europe in publishing research journals is bound to create an imbalance in the research result publishing world, and this concentration of journal publishing in Europe may lead to monopolizing publishing research results in days to come.

It is highly desirable that journal publishing be promoted across Africa, Australia, and South America, but what is more important is that promoting journal publishing should not come at the cost of compromising the quality of research journals. Institutions or publishers across these continents can seek the necessary expertise, support, and help from the institutions and publishers known for publishing quality research journals, even be they from other parts of the world. Equitable and proportionate publication of research journals across continents is bound to promote equitable and sustainable growth across continents, regions, and countries all over the world. There is a need to let quality academics, research, and scientific investigations move outside the few concentrated pockets of the world to lesser-known regions and countries to extend equal opportunities to underprivileged ones.

The distribution of research journals published across different continents reflects a huge imbalance, somewhat reflecting the research prowess of different continents in producing and publishing research results.

Of the total research journals indexed during the last two decades across the world, the majority of 6,425 (24.23%) journals were indexed alone from the United

Table 4. Twenty leading journal publishing countries across the world

Serial no.	Name of country (A)	Continent (B)	Journals indexed as of 2019 (C)	Percentage share (D)[a)]
1	United States	North America	6,425	24.23
2	United Kingdom	Europe	6,337	23.89
3	Netherlands	Europe	1,771	6.68
4	Germany	Europe	1,521	5.74
5	China	Asia	665	2.51
6	Switzerland	Europe	660	2.49
7	Spain	Europe	655	2.47
8	France	Europe	584	2.20
9	Russian Fed	Asia	571	2.15
10	India	Asia	564	2.13
11	Italy	Europe	553	2.09
12	Japan	Asia	499	1.88
13	Poland	Europe	459	1.73
14	Brazil	South America	401	1.51
15	Canada	North America	284	1.07
16	South Korea	Asia	286	1.08
17	Turkey	Europe	245	0.92
18	Australia	Australia	242	0.91
19	Iran	Asia	214	0.81
20	Czech Republic	Europe	208	0.78
21-109	Rest of the world		3,377	12.73
	Total (average)		26,521 (1,263)	100.0 (4.76)

a) $\frac{C}{Total\ of\ C} \times 100$.

 https://doi.org/10.1633/JISTaP.2023.11.2.3

States (US), constituting around one-fourth of the total research journals indexed across the world. The US is closely followed by the United Kingdom (UK) by indexing as many as 6,337 (23.89%) research journals, which again comprises around one-fourth of total journals indexed across the world. Around half of the research journals indexed across the world during the last two decades are from the US and the UK. Similarly, of the total research journals published across the world, 88.91% are published across the twenty leading tabulated countries, while the remaining 11.09% of journals were indexed from 89 countries of the world (Table 4).

A country that does not have its own recognized channels or mediums to communicate its research results can be considered as a country lagging in wholesome research activity. After all, the purpose of research gets fulfilled the moment the end results of research reach their end users for practical use, and these research results are delivered through the recognized channels, what we know as research journals. The need is to promote a journal publishing culture among all such countries which are dependent on other countries to publish their own research results. All this, apart from providing a platform for local and budding researchers to publish their research results, will

Table 5. Twenty leading subject disciplines

Serial no.	Name of the subject (A)	Journals Indexed as of 2019 (B)	Percentage share of B (C)[a)]
1	Medicine	1,606	6.06
2	Education	734	2.77
3	History	687	2.59
4	Literature and literary theory	608	2.29
5	Cultural studies	606	2.28
6	Law	518	1.95
7	Language and linguistics	457	1.72
8	Engineering	413	1.56
9	Animal science and zoology	350	1.32
10	Agronomy and crop science	332	1.25
11	Business and international management	328	1.24
12	Cardiology and cardiovascular medicine	323	1.22
13	Electrical and electronic engineering	310	1.17
14	Economics and econometrics	298	1.12
15	Applied mathematics	291	1.10
16	Agricultural and biological sciences	261	0.98
17	Anthropology	244	0.92
18	Earth and planetary sciences	241	0.91
19	Communication	234	0.88
20	Geography, planning, and development	226	0.85
21-297	Rest of the subject areas	17,454	65.81
	Total	26,521	99.99

The sum of the percentage is not 100% because of rounding.

a) $\frac{B}{Total\ of\ B} \times 100$.

Table 6. Decadal growth of research journals across main subject disciplines

Serial no.	Subject (A)	2000 (B)	2010 (C)	CG (D)[a]	CG of D (%) (E)[b]	2019 (F)	CG (G)[c]	CG of G (%) (H)[d]	Avg CG (I)[e]	Avg CG of I (%) (J)[f]
01	Agriculture and biological sciences	1,246	1,830	584	46.87	2,236	406	22.19	495	34.53
02	Arts and humanities	1,399	2,776	1377	98.43	4,257	1,481	53.35	1,429	75.89
03	Biochemistry, genetics, and molecular biology	1,300	1,829	529	40.69	2,130	301	16.46	415	28.57
04	Business, management, and accounting	654	1,118	464	70.95	1,463	345	30.86	404.5	50.90
05	Chemical engineering	437	500	63	14.42	637	137	27.40	100	20.91
06	Chemistry	645	771	126	19.53	870	99	12.84	112.5	16.18
07	Computer sciences	730	1,296	566	77.53	1,717	421	32.48	493.5	55.00
08	Decision sciences	172	305	133	77.33	430	125	40.98	129	59.15
09	Dentistry	108	174	66	61.11	211	37	21.26	51.5	41.18
10	Earth and planetary Sciences	979	1,022	43	4.39	1,193	171	16.73	107	10.56
11	Economics, econometrics, and finance	438	831	393	89.73	1,130	299	35.98	346	62.85
12	Energy	244	332	88	36.07	477	145	43.67	116.5	39.87
13	Engineering	1,671	2,217	546	32.68	2,956	739	33.33	642.5	33.00
14	Environmental science	939	1,122	183	19.49	1,535	413	36.81	298	28.15
15	Health professions	336	490	154	45.83	583	93	18.98	123.5	32.40
16	Immunology and microbiology	384	521	137	35.68	591	70	13.44	103.5	24.56
17	Material science	792	982	190	23.99	1,235	253	25.76	221.5	24.87
18	Mathematics	744	1,199	455	61.16	1,569	370	30.86	412.5	46.01
19	Medicine	5,863	6,949	1,086	18.52	7,401	452	6.50	769	12.51
20	Multidisciplinary	67	111	44	65.67	143	32	28.83	38	47.25
21	Neuroscience	341	500	159	46.63	592	92	18.40	125.5	32.51
22	Nursing	384	623	239	62.24	641	18	2.89	128.5	32.56
23	Pharmacology, toxicology, and pharmaceutics	489	715	226	46.22	748	33	4.62	129.5	25.42
24	Physics and Astronomy	768	959	191	24.87	1,113	154	16.06	172.5	20.46
25	Psychology	780	1,049	269	34.49	1,289	240	22.88	254.5	28.68
26	Social sciences	2,816	4,629	1,813	64.38	6,779	2,150	46.45	1,981.5	55.41
27	Veterinary sciences	146	209	63	43.15	256	47	22.49	55	32.82
	Total (Avg)	24,872 (921.18)	35,059 (1,298.4)	10,187 (377.2)	1,262.50 (46.74)	44,182 (1,636.37)	9,123 (456.15)	682.50 (25.27)	9,655 (482.75)	972.27 (36.01)

Avg, average; CG, corresponding growth.
[a] $C\text{-}B$. [b] $\frac{D}{B}\times 100$. [c] $F\text{-}C$. [d] $\frac{G}{C}\times 100$. [e] $\frac{D+G}{2}$. [f] $\frac{E+H}{2}$.

 https://doi.org/10.1633/JISTaP.2023.11.2.3

also help in promoting equitable research activities and communicating and availing research benefits across the world uniformly for the sustainable growth and development of all nation-states.

To assess the growth of research journals at the subject discipline level, it emerged that a maximum of 1,606 journals were indexed in the field of medicine, constituting 6.06% of the total journals introduced during the last two decades across all subject disciplines in the world. Medicine is distantly followed by education and cultural studies with the indexing of 734 and 606 research journals each, respectively. A minimum of 226 research journals were introduced in the subject area of geography, planning, and development, constituting around 0.85% of the total journals indexed during the last two decades. In all, 297 subject areas were identified in which 26,521 research journals are published, of which 9,067 research journals are published across the twenty leading tabulated subject disciplines, constituting 34.14% of the total research journals indexed, while 17,454 (65.81%) research journals were indexed across the remaining 277 subject disciplines. On average, 88.38 research journals are published within each individual identified subject discipline (Table 5).

The actual number of journals indexed as reflected in Tables 1 to 2 is 26,521, while the number of journals reflected in columns B, C, and F of Table 6 vary considerably, solely because of the scope of the journals which most of the time can be found falling under different subject heads, as such resulting in variation in the actual number of journals indexed.

As per SCImago data, there are 27 main subject disciplines under which all 26,521 journals are indexed under 297 major and minor subject disciplines as reflected in Table 5. The data reflected in Table 6 shows the growth of research journals across 27 major subject disciplines at the decadal level. However, a considerable difference can be found in the total number of journals indexed across 27 major subject disciplines, and this difference is mainly because a good number of journals are multidisciplinary and a few others have scopes beyond a specific subject discipline; as such they stand indexed under different subject disciplines, hence there is a considerable difference in the actual number of journals indexed and those indicated in Table 6.

While analyzing the growth of research journals at the subject level during the first decade of the twenty-first century, it emerged that arts and humanities recorded the maximum 1,377 (98.43%) growth rate in research journals, followed by economics, econometrics, and finance with 89.73%, and computer sciences, which recorded 77.53%. The minimum 4.39% growth rate of research journals was recorded in the subject discipline earth and planetary sciences. During the second decade of the century, again the maximum growth of research journals was recorded in arts and humanities, registering a growth of 1,481 (53.35%) research journals. Arts and humanities is the only subject discipline that recorded more than 50% growth, while the rest of the subject disciplines recorded less than 50% growth. Only 18 new research journals were indexed in nursing in the second decade of the century across the world, thereby registering 2.89% of decadal corresponding growth, the lowest among all the subject disciplines. On average research journals in each subject discipline during the first decade of the century grew at 46.74%, while during the second decade of the twenty-first century, the research journals in each subject discipline on average grew at 25.27%.

From 2000 through 2019, research journals in the field of arts and humanities grew by 75.89%, the highest among the 27 subject disciplines listed in Table 6. Arts and humanities are followed by economics, econometrics, and finance with 62.85% of journals, and decision sciences with 59.15%. However, a minimum of 10.56%, of research journals were indexed in the field of earth and planetary sciences. Overall, research journals in each subject discipline during the period of study on average grew at 36% each decade.

7. DISCUSSION

With the manifold growth of research journals across the world, the existence of journal publishers has become questionable, especially when the journal publishing market is flooded with fake, sub-standard, and predatory research journals; with this result, the research journals indexed in quality indexes like Scopus and WoS are also being seen with suspicion regarding their genuineness and authenticity. It is quite interesting to note that research journals over the years indexed in Scopus have grown in number at a very moderate and steady pace, viz., 2.74% per annum. The steady growth of these research journals also confirms that journal indexing parameters and standard practices followed over the years by Scopus are intact and have not been diluted. In fact, in post-year 2000 the process of indexing research journals in Scopus has become more rigorous, with the result that of the total journals indexed in Scopus only 40% of journals have been added during the last two decades. Post-2011, a steady

decline can be observed in the indexing of new research journals in Scopus, and this may be primarily because of the manifold growth in predatory and sub-standard journals, with the result that leading indexing agencies want to doubly ensure the legitimacy and genuineness of a research journal before its indexing.

The credibility of indexes like Scopus and WoS may become questionable if the scenario of indexing dubious, predatory, or sub-standard research journals continues. It is a well-known fact that a newly launched research journal must prove its credibility and seriousness in publishing research results in the early years of its publication and thereafter it qualifies for indexing in Scopus or WoS or any other quality index. Since most researchers presume that everything indexed in Scopus or WoS is credible, as such they do not bother to look at or ascertain the credibility of a journal at their level, with the result that indexing service providers have more responsibility to ensure that no dubious research journal is indexed. There is a far greater need to raise the standard of indexing a research journal, whereby all the dubious, sub-standard, and predatory journals may be kept at bay.

The scenario of journal publishing at the continental level is entirely different, which cannot be considered ideal for holistic, uniform, and sustainable research growth across the world. Europe as a continent is publishing around 57% of the total global research journals, which speaks about the tilt in the global journal publishing industry of not only being concentrated in Europe, but also create an imbalance which may over the years monopolize the research results publishing industry of the world. Europe is far ahead of the other five continents even if put together in publishing research results, and this scenario has not changed over the years. Of the newly introduced journals across the world, nearly 60% are launched in Europe alone, with the result that the facts and figures of journal publishing in Europe move almost neck and neck with that of overall global journal publishing figures, thereby confirming the dominance of Europe over the rest of the world in publishing research journals. The global journal publishing market is primarily dominated by Europe, which is not good for the overall health and wholesome growth of research practices across the world. The only way out to overcome this European dominance in publishing research journals is by improving the quality of research journals published across the rest of the world in general and in the continents like Africa, Australia, and South America in particular, who together publish less than 6% research journals of the world. This small percentage of contribution from these continents needs to be improved considerably. It would be ideal if the publishers from Europe might come forward and extend support to journal publishers from these continents in improving the quality of their research journals. The world can grow and prosper only when there is uniform growth across all the regions of the world. The irony is that research journals published across continents like Africa, Australia, and South America are concentrated in countries like Egypt, Australia, and Brazil respectively, and as such need to be pushed and promoted beyond these countries. The journal publishing scenario across other continents is almost similar, whereby countries like China, the UK, and the US dominate journal publishing in their respective continents. These six countries together publish nearly 54% of the total global research journals, which again is alarming, whereby participation of the rest of the countries from each continent must be promoted.

Publishing research journals and publishing research articles are altogether two different aspects, and there are countries in the world that do not publish even a single Scopus-indexed journal, but on the other hand, most of these countries publish their research results in Scopus-indexed journals. Accordingly, at the global level of the 218 recognized research countries, only 109 countries are actively publishing research journals, constituting less than 50% of countries of the world, while of the 109 countries actively publishing research journals, 43 countries come alone from Europe. This also signifies that 50% of the countries of the world completely depend on other countries to publish their research results, which by no means can be considered good for the research health of the world.

The dominance of journal publishing is equally prevalent at the country level. The US and the UK alone publish around one-fourth of global research journals each, together constituting around 50% of the research journals published across the world. This dominance of individual countries should be seen as more dangerous than the dominance of an individual continent vis-à-vis Europe. The concentration of publishing research journals limited to a few countries somewhat questions the research prowess of the rest of the world and indirectly signifies that if not all, nearly 50% of research results produced in any part of the world are to seek the route of the UK and the US journal publishing industry to validate their results. Here again, the need is to promote the publishing of research journals across different countries of the world for uniform and sustainable growth. Still more, of the 109

 https://doi.org/10.1633/JISTaP.2023.11.2.3

journal publishing countries a meager 11% of journals are published across 89 countries, which is far less than the desired levels of journal publishing, both in terms of number and percentage.

The volume of research undertaken in any given subject discipline at the global level has a direct bearing on the number of research journals published in any subject field, and the practice proceeds according to demand and supply. The greater the number of research results produced in any subject discipline; the larger will be the platform required to publish all such research results. Accordingly, the growth of research journals can be seen as an indicator to assess the research output in any subject discipline. The results depict that medicine is one of the leading research areas, covering more than 6% of core research journals of the world, the highest among all the subject disciplines. However, when taking together all the allied subject disciplines of medicine and other inter-disciplinary journals, the number of journals published in the field of medicine then rises to around 7,400, which constitutes around 27.90% of the total journals published. The same is the case with other subject disciplines where the number of core journals published in that subject discipline may be less, but while aggregating other interdisciplinary and allied subjects, the number becomes far greater. There are some subject areas where the required number of journals available for publishing research results may be less, with the result that the journal packing density (Singh & Pandita, 2017) of research journals available in that particular subject may be higher than the average packing density of research journals. The availability of an adequate platform to publish research results is as important as the research itself, while the absence of the same or the availability of an inadequate platform can hamper research growth in that subject discipline.

While assessing the growth of subject disciplines at the decadal level, it emerged that arts and humanities recorded the highest growth in the introduction of new journals. The higher growth of research journals is also an indicator of the fact that the volume of research work undertaken in the field of arts and humanities has also increased manifold. Besides this, subject disciplines being interdisciplinary as such encourages the growth of newer research journals.

8. CONCLUSION

A shortage of quality research journals in any subject discipline can badly hamper the growth of that subject discipline, and the obligation to have adequate and quality journals in that very subject discipline rests with the existing scholarly community of that subject discipline. The research journals published across continents like Africa, Australia, and South America are not sufficient even to cater to the research results publishing requirements of their local researchers, hence there is a need to improve the journal publishing scenario across these continents among all the major subject disciplines. The countries, regions, or for that matter continents lesser known for publishing quality research journals are not generally preferred by researchers from other regions for publishing their research results. The journal publishers from these continents have to more or less compete with the publishers from other continents both in terms of quality and quantity. Similarly, the number of research journals published across different subject disciplines is far lesser than desired. The concentration of journal publishing being limited to a few continents or a few countries in a continent equally acts as an impediment to the wholesome and sustainable growth of the entire region or the continent. There is a far greater need to have a proportionate distribution of journal publishing across different regions of the world. Every country undertaking research should publish its quality research journals and by no means should it completely depend on other countries for publishing their research results.

The introduction of new quality research journals is as important as publishing new research results or research concepts. The introduction of new journals becomes even more important for the fact that an existing lot of journals published across the world is not sufficient to publish all the research results produced in a time-bound manner. It is imperative to have sustainable growth and distribution of quality research journals across the world in each subject discipline. It is very important to have evenly distributed quality research journals across all the continents, countries, and subject disciplines, and more so the researchers from all continents, countries, and subject disciplines should have equal opportunity to publish their research results in any quality journal all over the world.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Arunachalam, S., & Markanday, S. (1981). Science in the

middle-level countries: A bibliometric analysis of scientific journals of Australia, Canada, India and Israel. *Journal of Information Science,* 3(1), 13-26. https://doi.org/10.1177/016555158100300103

Ayeni, P. O., & Adetoro, N. (2017). Growth of predatory open access journals: Implication for quality assurance in library and information science research. *Library Hi Tech News,* 34(1), 17-22. https://doi.org/10.1108/LHTN-10-2016-0046

Bottle, R. T., & Efthimiadis, E. N. (1984). Library and information science literature: Authorship and growth patterns. *Journal of Information Science,* 9(3), 107-116. https://doi.org/10.1177/016555158400900304

Collazo-Reyes, F. (2014). Growth of the number of indexed journals of Latin America and the Caribbean: The effect on the impact of each country. *Scientometrics,* 98(1), 197-209 https://doi.org/10.1007/s11192-013-1036-2

Dutta, S. (1968). Frequency symbol for periodical publications. *Annals of Library and Information Studies,* 15(1), 35-36. https://nopr.niscpr.res.in/handle/123456789/28315

Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B. C., & Hedlund, T. (2011). The development of open access journal publishing from 1993 to 2009. *PLoS One,* 6(6), e20961. https://doi.org/10.1371/journal.pone.0020961

Mabe, M. (2003). The growth and number of journals. *Serials: The Journal for the Serials Community,* 16(2), 191-197. https://doi.org/10.1629/16191

Meadows, A. J. (1993). Too much of a good thing? Quality versus quantity. In H. M. Woodward, & S. Pilling (Eds.), *The international serials industry* (pp. 23-44). Gower.

Mesa Melgarejo, L., & Galindo Huertas, S. (2011). [Characterization of periodical publications on nursing in Colombia, visible in the Internet]. *Avances en Enfermería,* 29(1), 159-168. Spanish. http://www.scielo.org.co/scielo.php?pid=S0121-45002011000100016&script=sci_abstract

Pandita, R., Koul, M., & Singh, S. (2017). Growth of research journals in India during last decade (2005-2014): An overview. *Collection Building,* 36(4), 143-154. https://doi.org/10.1108/CB-02-2017-0006

Price, D. J. D. S. (1963). *Little science, big science.* Columbia University Press.

ProQuest. (2021). *Ulrichsweb™.* https://about.proquest.com/en/products-services/Ulrichsweb.

Rao, I. K. R., & Srivastava, D. (2010). Growth of journals, articles and authors in malaria research. *Journal of Informetrics,* 4(3), 249-256. https://doi.org/10.1016/j.joi.2009.12.003

Scimago. (2021). *SJR: Scimago journal & country rank.* https://www.scimagojr.com/

Seglen, P. O. (1997). Citations and journal impact factors: Questionable indicators of research quality. *Allergy,* 52(11), 1050-1056. https://doi.org/10.1111/j.1398-9995.1997.tb00175.x

Sen, B. K., & Lakshmi, V. V. (1992). Indian periodicals in the science citation index. *Scientometrics,* 23(2), 291-318 https://doi.org/10.1007/BF02017043

Singh, M. (1978). Studies of chemical literature and changes in the ranking of periodicals by citation analysis of data for 1967-76. *Annals of Library and Information Studies,* 25(1-4), 55-61. https://nopr.niscair.res.in/handle/123456789/28105

Singh, S., & Pandita, R. (2017). Indexing of Indian journals in SCImago citation database: A decadal study (2005-2014). *DESIDOC Journal of Library & Information Technology,* 37(1), 51-57. https://doi.org/10.14429/djlit.37.1.10343

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 11(2): 37-66, 2023

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2023.11.2.4

Counting Research Publications, Citations, and Topics: A Critical Assessment of the Empirical Basis of Scientometrics and Research Evaluation

Wolfgang G. Stock*
Depearment of Information Science, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
E-mail: wolfgang.stock@hhu.de

Gerhard Reichmann
Department of Operations and Information Systems, University of Graz, Graz, Austria
E-mail: gerhard.reichmann@uni-graz.at

Isabelle Dorsch
ZBW – Leibniz Information Centre for Economics, Kiel, Germany
E-mail: i.dorsch@zbw-online.eu

Christian Schlögl
Department of Operations and Information Systems, University of Graz, Graz, Austria
E-mail: christian.schloegl@uni-graz.at

ABSTRACT

Scientometrics and research evaluation describe and analyze research publications when conducting publication, citation, and topic analyses. However, what exactly is a (scientific, academic, scholarly or research) publication? This article demonstrates that there are many problems when it comes to looking in detail at quantitative publication analyses, citation analyses, altmetric analyses, and topic analyses. When is a document a publication and when is it not? We discuss authorship and contribution, formally and informally published documents, as well as documents in between (preprints, research data) and the characteristics of references, citations, and topics. What is a research publication? Is there a commonly accepted criterion for distinguishing between research and non-research? How complete and unbiased are data sources for research publications and sources for altmetrics? What is one research publication? What is the unit of a publication that causes us to count it as "1?" In this regard, we report problems related to multi-author publications and their counting, weighted document types, the unit and weighting of citations and references, the unit of topics, and counting problems—not only at the article and individual researcher level (micro-level), but also at the meso-level (e.g., institutions) and macro-level (e.g., countries). Our results suggest that scientometric counting units are not reliable and clear. Many scientometric and research evaluation studies must therefore be used with the utmost caution.

Keywords: scientometrics, research publication, counting unit, reference, citation, topic

Received: October 12, 2022 **Revised:** February 18, 2023
Accepted: February 21, 2023 **Published:** June 30, 2023

***Corresponding Author:** Wolfgang G. Stock
https://orcid.org/0000-0003-2697-3225
E-mail: wolfgang.stock@hhu.de

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© Wolfgang G. Stock, Gerhard Reichmann, Isabelle Dorsch, Christian Schlögl, 2023

1. INTRODUCTION

Scientometrics is a research area which focuses on formal scholarly communications (Borgman & Furner, 2002; Leydesdorff, 1995) and on likewise informal communications with the research environment, which include internal informal communication amongst researchers and external communication between researchers and the general public. "Communication" means that there are sustainable documents such as journal articles, papers in proceedings, patents, books, and postings in social media services, but not ephemeral documents such as oral communications. Formal scholarly communication happens via research documents, which deal with research topics and nearly always contain references. Informal communications with the research environment comprise all kinds of mentions in social media, in the newspapers, or in other mass media, some of them studied by altmetrics (Thelwall et al., 2013). Scientometric results play a role in the evaluation of research units at different levels in relation to single researchers, institutions, journals, disciplines, or even countries, in the analysis of age- and gender-related science studies (Larivière et al., 2013; Ni et al., 2021) as well as in research policy and management (Mingers & Leydesdorff, 2015), too. We use the term "scientometrics" as a narrower term of "informetrics." While scientometrics studies research information, "informetrics" is the study of all quantitative aspects of all kinds of information (Dorsch & Haustein, 2023; Stock & Weber, 2006; Tague-Sutcliffe, 1992).

According to Leydesdorff (1995, p. 4), the science of science in general has three dimensions: scientists, cognitions, and texts. The subjects of scientometrics are the multiple relationships between scientists and their texts, i.e., the published research information. The remaining disciplines of the science of science are the sociology of knowledge, studying the relationships between scientists and cognitions, and the theories of information and communication, which analyze the relationships between cognitions and texts. In scientometrics, the "methodological priority of textual data" dominates (Leydesdorff, 1995, p. 57).

Scientometrics includes quantitative studies on all kinds of research. Particularly, quantitative research studies call for a solid empirical observation base. Each research field has its own technical terms which directly relate to observation objects and their properties, and sentences reflecting protocols of propositions with those technical terms (Carnap, 1931; 1932; Neurath, 1932). Important elementary concepts of scientometrics are "research publication" and "word in a research document," or, alternatively, "index term" in a record of a bibliographic information service, "reference" / "citation," and "mention in the research environment." A protocol sentence may be "X has 350 publications as of 2023-01-01." Scientometricians *measure* counts of publications, words, index terms, references, citations, and mentions, as these objects are directly observable. In scientometrics, numbers of publications are an *indicator* for research output, words in research documents or index terms in bibliographic records are an indicator for research topics, and numbers of citations are an indicator for research impact, but not for research quality, as this is not measurable at all (Glänzel, 2008); and, finally, in altmetrics, the numbers of mentions in the research environment, for instance in research blogs, in microblogs, in social media services, or in newspapers, can be interpreted as an indicator of research interestingness or attention the research receives. A sentence such as "X's research productivity is higher than average" would be a typical proposition applying a research indicator. We should have in mind that research output or research productivity, which is output in relation to input (Stock et al., 2023), topics, impact, and interestingness are not directly observable but only mediated by measures and deviations from measures, for instance, publications per year or co-citations.

References are hints for information flows from previous research towards a research publication; citations are hints for information flows towards further research; however, citations are references in the citing publications. As research publications contain words and references and so indirectly also citations, an important empirical observation base of scientometrics is publications. Here an issue arises: "What is a scientific, academic, scholarly, or research publication?" This issue leads to three research questions (Reichmann, 2020; Stock, 2000; 2001):

1. What is a publication? When is a research document a publication and when is it not? What are references and citations? And what is a topic?
2. What is a research publication? Are there criteria for demarcation of research from non-research? What are appropriate data sources for research publications?
3. What is one research publication? What is the unit of a publication inducing us to count it as "1?" And, of course, additionally, what are *one* reference, one citation, and one topic?

 https://doi.org/10.1633/JISTaP.2023.11.2.4

2. METHODS

This paper intends to answer these research questions by means of a critical review of the state of the art in scientometrics and research evaluation, i.e., we collected literature and tried to generate new knowledge about the nature of the properties of research publications as the base unit for many scientometric studies.

The reviewed literature was found through searches in bibliographic information services (Web of Science [WoS], Scopus, Google Scholar, and Dimensions) and on SpringerLink, as Springer is the publishing house of the journal *Scientometrics*, applying both the citation pearls growing strategy and the building block strategy (Stock & Stock, 2013, pp. 261-263) wherever possible. Additionally, we followed the links backwards via references and forwards via citations of found articles as a kind of snowball sampling. We collected relevant hits from the direct database results and from their references and citations; additionally, we found some hits from browsing and serendipity, thus applying a berrypicking strategy (Bates, 1989) in the end.

We searched separately for every topic, as a global search strategy failed. For, e.g., "criterion of demarcation" we initially looked for "demarcation criterion" and hoped to find some very relevant "citation pearls." In the next step, we checked the found pearls for further search arguments such as, for instance, "pseudo-science," "science and non-science," or "problem of demarcation" and put them together in the sense of the building block strategy to: "(science OR 'pseudo-science' OR 'non-science') AND ('demarcation criterion' OR 'criterion of demarcation' OR 'problem of demarcation')." In the last step, we browsed through the references and citations.

This paper is not a scientometric research study and, in particular, it does not provide a quantitative analysis. It is a paper on fundamental *problems of scientometrics*, i.e., on problems of the quantitative research approach in the science of science. Our focus is on the main research object of scientometrics: the research publication, including its authors, topics, references, citations, and mentions in social media sources, but also the associated questions and problems. Are quantitative scientometric studies really free from problems with the correct collection of empirical data, the use of counting methods and indicators, and methodological problems? If there were any fundamental problems with scientometrics, would its application in research evaluation really be helpful? Or is it dangerous to rely solely on scientometric results in research description, analysis, evaluation, and policy?

It is not the purpose of this study to mention all relevant literature, but to draw a comprehensive picture of research publications as counting units of scientometrics and—first and foremost—on all related problems. The main contribution of our paper is a critical assessment of the empirical foundations of scientometrics: Does scientometrics really have empirical evidence? Is scientometrics a sound scientific discipline? Or are there so many problems with data, indicators, and research practices that scientometrics and research evaluation as scientific disciplines should be fundamentally reconsidered?

We can reveal one of this study's main results already now: There are many serious problems. Many scientometricians are aware of these issues, but researchers from other knowledge domains conducting scientometric studies may not be. Scientometric publications—conducted by scientometric laypeople—are easy to find in many research areas. Therefore, this article is not only addressed to students and early career researchers to understand important topics of scientometrics and to act very carefully with problematic data and indicators in their own scientometric research, but also to senior researchers to revisit and solve the major issues—if this is ever possible—in the evolving field of scholarly communications.

3. WHAT IS A PUBLICATION?

In this paragraph, we discuss authorship as well as formally and informally published documents in general and, more specific, published documents in research. Of course, every research publication is a publication, but not all publications are research publications. With the emergence of social media, the concept of "publication" has changed. Tokar (2012) distinguishes between publications also by and on researchers in such social media as, e.g., Facebook, Twitter, Reddit, Instagram, or YouTube and academic or scholarly publications in "classical" online as well as offline media, e.g., in journals or proceedings, documents of intellectual property, or books. Also, publications in mass media such as newspapers, TV shows, or on the radio are publications. Documents in "classical" research media are genuine formal research publications, while publications in social media and mass media are parts of the communication with the research environment and so are—from the viewpoint of the research system—informally published.

A crucial question for scientometrics is: What exactly is a research publication? And what is it not? For the decision for or against the consideration that a document is

a research publication, Dorsch et al. (2018) propose two rules, namely the rule of formal publishing and the rule of research content. Of course, there is an additional at a first glance rather self-evident norm: In the publication's byline, the name of the author is stated.

3.1. Publication and Authorship

The "self-evident" criterion of authorship is not always clear as there exists rather unfair or unethical behavior by some researchers (Flanagin et al., 1998; Teixeira da Silva & Dobránszki, 2016): There are "ghost writers," i.e., authors who are not mentioned in the byline but who contributed to the document, and there is honorary authorship, i.e., persons who did not contribute to the document but are stated in the byline. In scientometrics, however, we have no possibilities for identifying ghost writers and honorary authors in an individual document. Therefore, we can only mention this problem but are not able to show a way to solve it. With regard to honorary authorship, the number of publications per year for individual researchers could be looked at. If this number is unrealistically high, this could be an indication of the existence of honorary authorship.

In multi-authored papers, the individual authors may have different roles. The Contributor Roles Taxonomy provides a standardized set of 14 research contributions such as, e.g., conceptualization, data curation, funding acquisition, or writing the original draft, and nowadays we find many journals indicating those contributor roles (Larivière et al., 2021). Of course, we have to presuppose that the given information is correct and unbiased. Taken seriously, we have to shift away from authorship and to move to contributorship instead (Holcombe, 2019).

3.2. Formally and Informally Published Documents

Authors publish documents or contribute to documents. But what is a document? In a first approximation, we follow Briet (1951/2006, p. 10) in defining a "document" as "any concrete or symbolic indexical sign [*indice*], preserved or recorded toward the ends of representing, of reconstituting, or of providing a physical or intellectual phenomenon". It is intuitively clear that all manners of texts are documents. But for Briet, non-textual objects may also be documents. The antelope in a zoological garden she described has become the paradigm for the additional processing of non-textual documents in documentation. An antelope in a zoo is a document and has nothing in common with a publication, but an artifact of engineering, say, a bridge or a software program, does (Heumüller et al., 2020; Jansz & le Pair, 1991). Following Buckland (1997), objects are "documents" if they meet the following four criteria: materiality: they are physically—including the digital form—present; intentionality: there is a motivation behind them, and they carry meaning; development: they are created; and perception: they are described and perceived as documents. This broad definition of a "document" includes textual and non-textual, and digital and non-digital resources. Every text, every data set, every "piece" of multimedia on the web, every artwork, and every engineering artifact is a document.

What is a formally published document? These are all documents from publishing houses, i.e., books, newspapers as well as journals, all intellectual property documents, grey literature, which are publications not provided by commercial publishers like university publications, working papers, or company newspapers (Auger, 1998), and all audio-visual materials from broadcasting companies.

All documents which are not publicly accessible have to be excluded. Such documents are, for instance, letters and other types of personal correspondence unless the correspondence is published, and confidential documents from companies and institutions including expert opinions and technical reports to be kept under lock and key.

Many web documents including all social media documents with user-generated content are informally published. Here we find the empirical basis for many studies in altmetrics. In "outbreak science" (Johansson et al., 2018), especially during the Coronavirus disease 2019 (COVID-19) pandemic, many research papers, including preprints, were described and discussed both in social media and newspapers (Massarani & Neves, 2022) and the general amount of preprints for both COVID-19 and non-COVID-19 related papers seemed to have increased (Aviv-Reuven & Rosenfeld, 2021). During COVID-19, the role of preprints further advanced and so the science communication landscape changed as well (Fraser et al., 2021; Watson, 2022). The status of preprints in the research system seems to alter. Since 2013, the number of preprint server is increasing (Chiarelli et al., 2019; Tennant et al., 2018). Accordingly, the following questions arise: What is the position of preprints? To which extend can they be considered as research publication? Based on our definition of formal and informally published documents, they would be positioned somewhere in-between, as they in most cases might undergo peer review. In general, there exist different definitions of preprints within the literature, concerning their genre, position over time, versioning, ac-

cessibility, responsibility, and value (Chiarelli et al., 2019). However, it is always possible that an author publishes any kind of paper with any kind of content on a preprint server without any quality control (Añazco et al., 2021).

Also, research data are provided by specialized services, which were simply uploaded. However, there are formally published documents which are based on the data. Hence, both preprints and research data can be considered as documents in between formal and informal publishing.

Most publications are written in a discipline-specific publication culture. This "culture" includes the preference for different document types, for example, journal articles in the natural sciences, proceedings papers in computer science, or books in the humanities, as well as differences in the mean length of reference lists and in the mean number of citations received per paper. Additionally, some document types such as, for instance, review articles tend to get more citations than others. Field-specific publication and citation cultures, different document types, but also field-specific differences in informally published sources require normalization in order to make figures from various research disciplines, document types, or social media mentions comparable (Haustein, 2012, pp. 274-289). Normalization is usually based on the respective mean values of a field, document type, and social media service or on a combination of the three aspects. An important problem is the granularity of the definition of the field as it may be defined at various levels, from small research areas to broad academic disciplines (Zitt et al., 2005).

Following library science, a published document is a unit comprising two areas: intellectual or artistic content and physical (also digital) entities, whereby the terms "work" and "expression" are related to the content and "manifestation" and "item" to the physical entity: "The entities defined as *work* (a distinct intellectual or artistic creation) and *expression* (the intellectual or artistic realization of a *work*) reflect intellectual or artistic content. The entities defined as *manifestation* (the physical embodiment of an *expression* of a *work*) and *item* (a single exemplar of a *manifestation*), on the other hand, reflect physical form" (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1998, p. 12). Works, i.e., authors' creations, can thus have one or more forms of expression (concrete realizations, including translations or illustrations), which are physically reflected in a manifestation such as, e.g., one printing of a book, and the single items of this manifestation. For scientometrics, it makes a difference to count works, expressions (say, a work in English and the same in German), or manifestations (the same expression in different journals or books or in different editions). Even the items are of interest for scientometrics, be it as the number of printed copies or downloads of a book edition, the circulation of a journal, or the download numbers of an article. In cataloging, there is a cut-off point between the same work such as, e.g., a translation or an edition in multiple sources, and a new work, for instance, the work's summary or a free translation (Tillett, 2001). There are some special problems concerning manifestations. One problem is concerned with gen-identical documents, i.e., documents with slightly changing content over time. These documents can be found in loose-leaf binders such as with many legal documents, or on websites. How can we count manifestations with several parts? And how can we count documents with shared parts, e.g., the same text body but a new title (Stock & Stock, 2013, pp. 572-573), or even the same document in multiple manifestations (Bar-Ilan, 2006)? Should such duplicate or redundant publications (Lundberg, 1993) be skipped in scientometric analyses? A related problem is the "salami slicing" of research results into "least publishable units" (Broad, 1981). If such units are redundant and aiming for a quick gain of publication and citation numbers, it is indeed an issue for the research system and for scientometrics as well. However, different small salami slices of the same research project can be a beneficial mechanism in the research system when it facilitates data management and rapid publication, among others (Refinetti, 1990), of course, always under the premise of being truly necessary. Additionally, some journals strongly encourage short papers (Roth, 1981). An open question is how to decide whether a paper is small but original, or redundant or even a duplicate.

The scientometric literature on works, expressions, etc. and on non-textual documents is rather limited. What we additionally miss are precise descriptions of the concepts of "formally published documents," "informally published documents," and "documents in between" like preprints or research data, as the units of many or even of all scientometric studies.

3.3. What is a Citation? What is a Reference?

Almost always publications do not stand in isolation, but have relations to other publications. With regard to textual publications in contrast to works of tonal art, pictures, or movies this is described as "intertextuality" (Rauter, 2005). Intertextuality is expressed by quotations or references. Important for scientometrics are the refer-

ences, which are bibliographic descriptions of publications authors have read and used while writing their papers. References are set by the authors themselves, or—for example, in patents—by others (in patents, by patent examiners). References can be found in the text, in footnotes, end notes, in a list of prior art documents in patents, or in a bibliography. References thus show information flows towards a publication or concerning patents towards a patent examination process. At the same time, we can see in the reverse direction how reputation is awarded to the cited document. From the citer's perspective, an end note, an entry in the bibliography, etc. is a reference, whereas from the citee's perspective, it is a citation.

The act of setting references is not unproblematic. It has been known for decades that authors do not cite all influences, that there is unspecific self-citing, and that some references are meant negatively (MacRoberts & MacRoberts, 1989), including pejorative, disagreeing, condemning, and other negative denotations. Additionally, some researchers concert reciprocal citations, thus forming a citation cartel (Poppema, 2020). For MacRoberts and MacRoberts (2018), citation analysis is a mismeasure of science. The starting position for both reference analysis and citation analysis is also bad, since one cannot trust the authors as to whether they actually cite everything they have read and used and whether they do not cite anything else.

If citations are used as an indicator for research impact, it should also be mentioned that there is a difference between research and practitioner journals. The former usually have an orientation on other researchers who write articles and potentially cite the read papers. The latter have an orientation on practitioners who also read, but do not always write their own papers and therefore cannot cite the read material so often. As a consequence, practitioners are less regarded in citation analyses (Scanlan, 1988; Schlögl & Stock, 2008).

3.4. What is a Topic?

Publications are about something; they are dealing with topics. Topics of research publications must be identified so that they can be captured in scientometric topic analyses (Stock, 1990). In information science, we speak of a publication's "aboutness" (Maron, 1977; Stock & Stock, 2013, pp. 519-523). Ingwersen (2002) differentiates between indexer aboutness (the indexer's interpretation of the aboutness in a bibliographic record) and author aboutness (the content as it is) among others. It is possible but very difficult to describe the author aboutness in a quantitative way; here, scientometricians have to work with the words (including the references) in the publication and have to prepare more informative indicators such as, for instance, word counts or co-word analyses. It is also a possible path of topic analysis to restrict to the publications' titles and authors' keywords, if they really reflect all discussed topics. Furthermore, scientometricians can work with citation indexing, as a citation, for Garfield (1979, p. 3), may represent "a precise, unambiguous representation of a subject."

The easier way is to work with the indexer aboutness, i.e., the topics in their representations through knowledge organization methods (Haustein, 2012, p. 78), i.e., nomenclatures, classification systems, thesauri, and the text-word method. In this way, the aboutness is captured by index terms. If we apply knowledge representation systems for scientometric topic analysis, we have to deal with the respective method (e.g., classifications have notations which are built without natural languages, while thesauri work with natural language descriptors), with the concrete tool (e.g., the International Patent Classification with all its rules or the Medical Subject Headings [MeSH] with their sets of descriptors and qualifiers), and, finally, with the tool's indexing rules and the concrete work of an indexer.

4. WHAT IS A RESEARCH PUBLICATION?

"Research" covers all activities in the natural sciences, the engineering sciences, the social sciences, and the humanities. However, depending on the sources, the definitions of research differ. The Organisation for Economic Co-operation and Development (OECD) (2015), for instance, defines research very broadly in terms of progress and new applications, while, according to Popper (1962), it is necessary for scientific results to be falsifiable. In this section, we describe the criterion of demarcation of science in order to find a boundary between research and non-research. Then we discuss data sources of research publications including bibliographic information services as well as personal and institutional repositories. Finally, we describe advantages and shortcomings of altmetrics, which analyze research-related publications in social media.

4.1. Criterion of Demarcation

In the philosophy of science, authors discuss criteria for demarcation between science and non-science or pseudo-science. For Carnap (1928, p. 255), *Wissenschaft* formulates questions which are answerable in principle.

 https://doi.org/10.1633/JISTaP.2023.11.2.4

The answers are formed through scientific concepts, and the propositions are basically true or false. Please note that the German term *Wissenschaft* includes natural sciences, social sciences, and humanities. Scientific sentences are protocol sentences reflecting observable facts; Carnap's (1931) prototype of a scientific language is the language of physics. Following Neurath (1932) such a protocol sentence is defined through justifications. If a justification fails, the sentence can be deleted. Therefore, *Wissenschaft* is always on the go. Popper (1935, p. 15) reformulates Neurath's definition: An empirical-scientific system must be able to fail because of experience; so, Popper (1962) calls only falsifiable, testable, or refutable propositions scientific. However, for Kuhn (1970) there are problems when we pay attention to scientific development. Sometimes there is a radical change in the development of a discipline, a so-called "scientific revolution," and we meet supporters of the old view and at the same time supporters of the new one. For both groups, believing in their own "normal science," led by an accepted "paradigm," the other group performs non-science and they themselves do good science. Lakatos (1974) calls Popper's attempt "naïve," but also declines Kuhn's theory of science, as there is only a demarcation criterion from the view of a special normal science and no general criterion; in fact, we cannot speak about progress in science, if we agree with Kuhn's approach (Fasce, 2017). If we—with Feyerabend (1975)—accept that in science "anything goes," then there will be no demarcation criterion at all. For Lakatos (1974), progressive research programs outperform degenerating ones. A research program is progressive if its theory leads to the discovery of hitherto unknown facts. Now, progress is the demarcation criterion; progress leading to new applications, too (Quay, 1974). Also, for OECD (2015)'s Frascati Handbook, new knowledge and new applications determine science. According to the OECD (2015, p. 45), research and development activities in natural sciences, social sciences, humanities and the arts, and engineering sciences must be novel, creative, uncertain, systematic, and transferable and/or reproducible.

Fernandez-Beanato (2020, p. 375) sums up: "The problem of demarcating science from nonscience remains unsolved." As all empirical sciences and also all formal sciences (e.g., mathematics) can be falsified, they are science for Popper (1935); and religion (Raman, 2001), metaphysics, astrology (Thagard, 1978), art, or fiction books are not. Following the philosophical ideas of Popper (1935; 1962), Carnap (1928; 1931; 1932), and Neurath (1932), psychoanalysis is pseudo-science, because there is no empirically observable data base; for Grünbaum (1977), it is science as it is successfully applicable. Parapsychology may be falsified, so it is science (Mousseau, 2003); so is—against common sense—homeopathy (Fasce, 2017, p. 463). The philosophy of science clearly shows the problems of demarcation. But it has no satisfying answer for the use of scientometrics. Perhaps it is better to formulate acceptable criteria for pseudo-science (e.g., Fasce, 2017, p. 476 or Hansson, 2013, pp. 70-71). These are Fasce's (2017) criteria: Pseudo-science is presented as scientific knowledge, but it refers to entities outside the domain of science, applies a deficient methodology, or is not supported by evidence. However, this definition makes use of the here undefined terms of science, scientific knowledge, and evidence. If we are going to define these concepts we must start from the beginning.

In scientometrics, it would be very problematic to check demarcation for every single research publication. Since scientometric studies often deal with large amounts of data, such a check would not even be feasible. In line with Dorsch (2017) one can propose to be guided by the scientific, academic, or scholarly character of the publishing source and not to consider all those documents which are not published in such media. Pragmatic criteria for a research publication include the existence of an abstract (however, these are not available in every case, e.g., not in editorials); the affiliation of the authors, e.g., universities, research companies, etc. (however, these might be not at hand for independent scholars); the existence of references, if applicable; and the use of a citation style. It seems to be questionable and worthy of discussion to add an institutionalized peer review process by the publishing source as a criterion of a research publication, as the quality of peer review is questionable (Henderson, 2010; Lee et al., 2013). Furthermore, not all research journals perform peer review, and, vice versa, some sources promise peer review and fail to run it satisfactorily. Hence, peer review was seen as a "flawed process" (Smith, 2006, p. 182), but we have to mention that there is no convincing alternative.

One could assume that so-called predatory journals, i.e., online journals without serious peer-review, solely maximizing profit (Cobey et al., 2018), are not scholarly sources. There are different lists of predatory journals (e.g., Beall's List or Cabells' Predatory Reports; Chen, 2019) and also white lists of non-predatory journals, but they do not match (Strinzel et al., 2019). It seems not to be possible to delete articles in predatory journals or proceedings of predatory conferences, as there is no clear cut distinction between predatory sources as pseudo-science and

non-predatory sources as science. It is possible to find even highly cited quality research articles in such sources (Moussa, 2021).

A special problem for scientometrics is retracted papers (Bar-Ilan & Halevi, 2017), i.e., articles which were formally published but later on taken back by the authors or by the publishing source, e.g., due to ethical misconduct, fake data, or false reports. Do retracted papers count as publications for the authors? And do citations to such documents count as equivalent to citations of non-retracted papers? A special problem seems to be the attention paid to retracted papers on social media, as for some retracted articles their attention increases. When papers written by famous authors or published in famous journals "are retracted for misconduct, their altmetric attention score increases, representing the public's interest in scandals, so to speak" (Shema et al., 2019).

If a researcher or a company asks for patent protection, the document is a patent application, an "A Document," at first. But if the inventor does not apply for granting the patent or if the patent authority refuses to grant it, is it a research publication in this case? It seems necessary only to consider granted patents, "B Documents," as research publications (Archambault, 2002). Besides such technical property rights as patents and utility models there are also non-technical property rights such as designs and trademarks. Non-technical property documents, however, as well as utility models in many countries, do not pass through a testing process at the patent office as there is only pure registration without examination. For non-technical intellectual property documents, it seems that there is only a marginal scientific research background, but a strong background in arts (designs) as well as marketing and advertising (trademarks); so it is an open question to include design and trademark documents in scientometric research or not (Cavaller, 2009; Stock, 2001).

In the case of rankings, the research publications used are often limited to single types of publications, for instance, to journal articles (Röbken, 2011). The reasons for this can be limited data availability, e.g., almost only papers in journals are recorded in some sub-databases of WoS. An additional reason can be that certain research rankings only focus on journal articles. As a consequence, only a small fraction of research publications is considered. Besides journal papers, there are monographs, papers in edited books, papers in conference proceedings, patents, and other articles (all other types, such as editorials or reviews) according to Reichmann et al. (2022).

As well, this section ends rather unsatisfactorily, as we did not find clear criteria to draw the line between research documents as objects of scientometrics and pseudo-science or even non-science publications to be omitted or marked as problematic in scientometric studies. But it should be clear that a set of demarcation criteria would be very helpful for scientometrics. Perhaps this article helps to initialize studies concerning demarcation in scientometrics.

4.2. Data Sources of Research Publications: How to Find Research Publications?

Scientometricians can use personal publication lists of individual researchers, research databases from research institutions (e.g., universities), discipline-specific information services (e.g., Chemical Abstracts Service for chemistry, Inspec for physics, Medline for medicine, or Philosopher's Index for philosophy), or multi-disciplinary citation databases in order to collect a comprehensive set of publications. Frequently used data sources are the WoS (Birkle et al., 2020), Scopus (Baas et al., 2020), and less frequently applied are Google Scholar (Aguillo, 2012) and (since the end of 2021 inactive) Microsoft Academic (Wang et al., 2020), and the recently established Dimensions (Herzog et al., 2020) as well as OpenAlex, providing "pinnacles" like the ancient Library of Alexandria with access via API (Priem et al., 2022). There are additional sources which are specialized on research data and their citations such as WoS's Data Citation Index (Robinson-García et al., 2016), DataCite (Ninkov et al., 2021), or Overton to track policy citations (Szomszor & Adie, 2022). Besides this, there exist several repositories, especially utilized as preprint publication data source; well-established ones include arXiv (McKiernan, 2000), medRxiv (Strcic et al., 2022), bioRxiv (Fraser et al., 2020; 2022), Research Square (Riegelman, 2022), and Zenodo (Peters et al., 2017). Additionally, there are discipline-specific preprint services such as Social Science Research Network (SSRN) and SocArXiv for the social sciences.

Concerning the number of research publications, Google Scholar retrieves the most documents, followed by the former Microsoft Academic, Scopus, Dimensions, and the WoS (Martín-Martín et al., 2021). However, all mentioned information services are incomplete when compared with researchers' personal publication lists (Dorsch et al., 2018; Hilbert et al., 2015). For WoS, Tüür-Fröhlich (2016) could show that there are many misspellings and typos, "mutations" of author names, and data field confusions, all leading to erroneous scientometric results. Concerning the literature of technical intellectual property

rights, i.e., patents and utility models, the information services of Questel and of Derwent are complete for many countries and therefore allow for comprehensive quantitative studies (Stock & Stock, 2006).

In comparison to the other information services, Google Scholar is challenging when advanced searching is required. Since it does not support data downloads, it is difficult to use as a scientometric data source. Additionally, it lacks quality control and clear indexing guidelines (Halevi et al., 2017). For WoS, it is known that is has a bias in disciplines (Mongeon & Paul-Hus, 2016). There is a clear bias in languages in favor of English for WoS (Vera-Baceta et al., 2019), which is also true to a lesser extent for Scopus (Albarillo, 2014). As WoS has many sub-databases and as research institutions commonly subscribe to WoS only in parts (mainly Science Citation Index Expanded [SCIE], Social Sciences Citation Index [SSCI], Arts & Humanities Citation Index [A&HCI], and Emerging Sources Citation Index [ESCI], i.e., sources which only cover journal articles), research disciplines publishing also in proceedings (as does computer science), books (as in many humanities), and patents (as does technology) are systematically discriminated against (Stock, 2021). Besides the WoS Core Collection, we find regional databases for the respective research output from Russia (Russian Science Citation Index; Kassian & Melikhova, 2019), China (Chinese Science Citation Index; Jin & Wang, 1999), Korea (KCI Korean Journal Database; Ko et al., 2011), Arabia (Arabic Citation Index; El-Ouahi, 2022), and South America (SciELO; Packer, 2000), which minimize WoS's language bias. But the scientific institutions and their libraries must subscribe to all sub-databases, which cannot always be realized because of budget reasons. About half of all research evaluations using WoS data did not report the concrete sub-databases (Liu, 2019) and similarly 40% of studies on the h-index withheld information on the analyzed sub-databases (Hu et al., 2020).

Some scientometric studies are grounded on data from an information service (e.g., WoS), which are further processed in an in-house database inside an institution, and which are as a consequence not available for researchers outside the institution (e.g., the study of Bornmann et al., 2015). Hence, an intersubjective examination and replication of such results by independent researchers is not possible, leading Vílchez-Román (2017) to press for general replicability in scientometrics to ensure—paraphrasing Popper (1935)—falsifiability of the results.

There are huge problems with the data sources on all scientometric levels (Rousseau et al., 2018). For Pranckutė (2021, p. 48), there are serious biases and limitations for both WoS and Scopus. At the micro-level (of individual researchers), Dorsch (2017) found massive differences in publication counts between information services, and Stock (2021) was able to present similarly high differences between different library subscriptions of the sub-databases of WoS. Concerning the macro-level (here, universities) Huang et al. (2020) reported on different rankings of the evaluation units depending on the data source. For van Raan (2005, p. 140) many university rankings must be considered as "quasi-evaluations," which is "absolutely unacceptable."

Dorsch et al. (2018) also used personal, institutional, or national publication lists as data sources for quantitative science studies. In their case studies, the personal publication lists of authors in scientometrics are much more complete than the research results on Google Scholar, Scopus, and WoS. However, these publication lists are not complete in an ideal way ("truebounded" in the terminology of Dorsch et al. (2018). In many cases, they are slightly "underbounded," i.e., there are missing items, and in a few cases, they are "overbounded," i.e., there are documents not having been formally published or not being research publications. In comparison to the truebounded publication lists, personal publication lists in their study consider about 93% of all publications on average, Google Scholar 88%, Scopus 63%, and WoS Core Collection (only SSCI, SSCI, A&HCI, and ESCI) 49% accordingly. Not every author maintains a current and correct publication list. In such cases, it might be possible to consider institutional repositories (Marsh, 2015) or even data collections on a national level (Sīle et al., 2018). Following Dorsch (2017) we must take into account that the entries in the personal, institutional, or national bibliographies will not be in a standardized format, so that further processing is laborious.

However, it can also happen that in the context of rankings, the objectively best, i.e., most complete data source is not used, although it would be accessible, but rather the one the use of which most likely leads to the desired results. For example, instead of publication lists, the WoS could be used to favor researchers who mainly publish in English-language journals. However, if a research evaluation is generally limited to the publication type "papers in journals," the differences depending on the data source used are much smaller. And if there is another restriction to English-language papers, these differences decrease again (Reichmann & Schlögl, 2021). In general, the use of international databases such as WoS or Scopus

implicitly restricts the research publications retrieved to mainly those being published in English (Schlögl, 2013).

With regard to high-quality professional information services (e.g., WoS and Scopus) one may argue that they only include "quality papers" being published in "quality journals." However, according to Chavarro et al. (2018), the inclusion of a journal in WoS or Scopus is no sound criterion for the quality of this journal.

Apart from the good situation with patent information, the data sources for other scientometric studies are either relatively complete (as are Google Scholar and some personal, institutional, or national repositories) but poorly suited for quantitative studies; or they are well-fitted for scientometrics (as are WoS, Scopus, or Dimensions), but lack a thorough database. Indeed, many scientometric studies rely on data from WoS and Scopus. For Fröhlich (1999, p. 27), they are "measuring what is easily measurable," and Mikhailov (2021, p. 4) states that "the problem of an adequate and objective assessment of scientific activity and publication activity is still rather far from being solved."

4.3. Altmetrics

More than two decades ago, the first social media platforms were launched. Soon after social media also became popular in the world of research, they became used for communication between researchers and between researchers and the general public. This subarea of social media metrics is called "altmetrics," a term which was originally coined by Priem (2010).

In the following, we will concentrate on basic problems of altmetrics, namely the diversity of data and the diversity of data sources of altmetrics (Holmberg, 2015) and on the reasonableness of altmetrics indicators. Following García-Villar (2021, p. 1), "altmetrics measure the digital attention received by an article using multiple online sources." With the upcoming of the Internet and digitalization, new ways of informal communication between researchers as well as between researchers and the general public emerged. Besides popular-science descriptions of research results in newspaper or on TV (Albrecht et al.,1994), now, for instance, numbers of views, downloads, discussions in blogs or microblogs, mentions in social networking services, or recommendations can be such new characteristics of research publications (Lin & Fenner, 2013). Download numbers of individual research papers (Schlögl et al., 2014) and contributions in social media such as, e.g., blogs, microblogs, posts, as well as in mass media (e.g., newspapers) especially extended the repertoire of data sources and research indicators. In their first generation, the analysis of these aspects is summarized as "webometrics" (Almind & Ingwersen, 1997), followed by "altmetrics" (Priem, 2010; 2014) as the second generation of web-based alternative indicators and methods, especially from the social web (Thelwall, 2021).

Altmetrics describes the field of study and the metrics themselves. It is broadly defined as the "study and use of scholarly impact measures based on activity in online tools and environments" (Priem, 2014, p. 266). Since its emergence, the concept itself, the setup of an altmetrics framework, and the methodologies, advantages, challenges, and main issues of altmetrics are discussed. First and foremost, the question is, what do altmetrics really measure? However, this is not the only aspect in question over the last ten years. Priem (2014) already lists a lack of theory, ease of gaming, and bias as concerns. Sugimoto et al. (2017) summarize the conceptualization and classification of altmetrics as well as data collection and methodological limitations in their literature review as challenges. Erdt et al. (2016) show in a publication analysis of altmetrics that aspects like coverage, country biases, data source validity, usage, and limitations of altmetrics were already discussed in research literature in the beginning of altmetrics in 2011. Later normalization issues (2014-2015) and detecting gaming as well as spamming were found as further topics. Coverage of altmetrics and cross-validation studies were the main research topics on altmetrics between 2011 and 2015 (Erdt et al., 2016). Still now, Thelwall (2021, p. 1302) identifies such common issues as "selecting indicators to match goals; aggregating them in a way sensitive to field and publication year differences; largely avoiding them in formal evaluations; understanding that they reflect a biased fraction of the activity of interest; and understanding the type of impact reflected rather than interpreting them at face value."

Altmetrics are heterogeneous in terms of their characteristics of measurements, the availability of the data, and the ease of data collection (Haustein, 2016; Sugimoto et al., 2017). Depending on the audience, "views of what kind of impact matters and the context in which it should be presented" (Liu & Adie, 2013, p. 31) vary. This overall heterogeneity enables altmetric applications in a variety of different ways and cherry-picking from available data and tools (Liu & Adie, 2013). However, this also makes it difficult to define the impact of the measurements. If one conducts altmetric studies, extensive data cleansing is necessary at any rate (Bar-Ilan, 2019).

If not collected directly from the respective data

 https://doi.org/10.1633/JISTaP.2023.11.2.4

source, altmetrics data can be obtained from different data aggregators. Established and frequently used aggregators are Altmetric.com, Plum Analytics, ImpactStory (formerly Total-Impact), and PLOS Article-Level Metrics. Similar to raw data from different sources, these processed data differ (Zahedi & Costas, 2018). In their comparison of the four aggregators, Zahedi and Costas (2018) argue that most data issues are related to methodological decisions in data collection, data aggregation and reporting, updating, and other technical choices. With these inequalities, it is still of utmost importance to understand how differences can cause variations within altmetrics analyses (Zahedi & Costas, 2018). Likewise, the underlying social media platforms affect altmetric indicators in general (Sugimoto et al., 2017). Potential data quality issues can happen at the level of data providers, data aggregators, and users. Furthermore, for providers like Twitter or Facebook, data quality is not a priority as they are not targeted at academia (Haustein, 2016). When altmetrics cover dynamic publications such as, e.g., Wikipedia entries, such texts may be in constant flux. Not uncommon in social media are deleted publications such as, e.g., tweets on Twitter or posts on social networking services. Moreover, hyperlinks included in tweets or posts may not work sometime after their publication. Deleted or modified documents on social media services are a serious problem for the replicability of studies in altmetrics. Nonetheless, Twitter is one of the most popular sources of altmetrics, its data are easily accessible, and the data could be used as early indicators of an article's possible impact (see, e.g., Haustein, 2019).

There are many different sources for altmetrics with different possible measures. One may describe Mendeley readers, Wikipedia mentions, microblogs on Twitter, posts in weblogs, news in Reddit, posts on Facebook, and videos in YouTube, among others. All these measures capture different aspects of communication, and there is not or not always a high correlation between the single measures (Meschede & Siebenlist, 2018). From the perspective of the researcher or person who is using altmetric data, there is the risk of making application errors. Following Thelwall (2021), using a "basket of indicators," or "a single hybrid altmetrics," without a theoretical or problem-based reason instead of using indicators tailored to the research framework and goals is not uncommon. Likewise, the usage of sparse indicators for too small datasets is not recommended. Just as with conventional scientometrics, ignoring field differences or publication year differences would affect the analysis. For example, the comparison of two fields where one field is much more engaged and discussed within social media would require field normalized indicators. Calculating arithmetic means or using Pearson correlations to validate altmetrics indicators without log-transformation or normalization is also not recommended, as altmetrics, like webometrics and citation counts, are highly skewed. It is furthermore suggested not to over-interpret correlation values. In general, altmetric indicators are also biased and only "reflect a very small fraction of impacts of the type that they are relevant to" (Thelwall, 2021, p. 1308). For more general indicators like Twitter citations, it is important to attribute meaning with, e.g., content analysis of citation sources, surveys, or interviews instead of "equating the type of impact that they might reflect with the type of impact that they do reflect" (Thelwall, 2021, p. 1308). Another aspect is to analyze indicators separately instead of using hybrid indicators (e.g., Altmetric Attention Score) that aggregate several altmetrics into a single score, as the latter have different interpretations. Altmetrics data is in constant flux, so not reporting on the date of data collection and how the data were collected would further add against interpretability. It is also important to conduct an effective literature review to keep updated about issues and directions in altmetrics. There is always the possibility that a social media service applied for altmetrics such as, for instance, Twitter, will be shut down or no longer used by researchers. In terms of formal research evaluations, altmetric indicators should be avoided (Thelwall, 2021).

Altmetric measures provide both further characteristics of formal research publications besides citations, for instance, number of views, downloads, or recommendations, and further research-related informal publications in social media as well as in newspapers, TV, or radio. Overall, altmetrics share many characteristics, and thus problems, with citations so that methodological solutions already exist and can be imported (Sugimoto et al., 2017). Altmetrics add to the bouquet of a research publication and can contribute to gathering more information on different societal impacts of research (Thelwall, 2021). Herb (2016) objects that altmetrics refrain from every serious theoretical foundation; further—for Herb (2016, p. 406)—it is unclear which properties of which objects with which operationalizations altmetrics really measure. There is a huge problem of altmetrics: Altmetric measures capture different aspects of the informal internal as well as external research communications—the problem is how to merge all these aspects as there is no standard (Holmberg, 2015). Perhaps there should be no merging of different data from different data sources at all. While such single

measures may be indeed okay, there are serious problems with indicators derived from those measures. Especially, aggregated indicators are more or less senseless, as we do not know what they exactly tell us.

5. WHAT IS ONE RESEARCH PUBLICATION?

In this section, we discuss the counting units of scientometrics. What causes us to count "1" for a research publication, reference, citation, or topic? Counting problems can arise from multi-author publications as well as from weightings for publications. We also discuss the units of references and citations and of topics. Finally, we deal with counting problems on the meso- and macro-level of scientometrics.

5.1. Multi-author Publications and Their Counting

If a research publication has one author, then this publication counts as "1" for the researcher. With multi-author publications, counting problems start (Gauffriau, 2017), as counting methods are decisive for all measures of authors' productivity and for rankings which are based upon publication and citation studies (Gauffriau & Larsen, 2005). What is the share of each author? Egghe and Rousseau (1990, pp. 275-276) differentiate between straight counting (considering only first authors and ignoring all co-authors), unit counting or whole counting (for all co-authors the publication counts "1"), and fractional counting (1/n in the context of n co-authors). The formula 2/(n+1) has advantages in terms of a balanced consideration of co-authorship in relation to single authorship (Schneijderberg, 2018, p. 343). It takes into account the fact that co-authorship requires additional work due to the necessary coordination. However, the fact that the sum of the author values is greater than one in this variant is a decisive flaw.

There are further weighting methods considering different values for different kinds of contribution following the Contributor Roles Taxonomy (Rahman et al., 2017). Other approaches consider the past roles of the co-authors such as, for instance, senior versus junior authors (Boxenbaum et al., 1987), the co-authors' h-indices (Egghe et al., 2013), or the co-authors' numbers of publications and received citations (Shen & Barabasi, 2014). All in all, Gauffriau (2017; 2021) found 32 different counting methods for an author's share of a multi-authored research publication. Among them are problematic practices as well, like the authors' h-index.

Straight counting of only the first author is unfair towards all authors who are not the first author. If we want to weight contributorship fairly, we need weighting schemes for all kinds of contributions and aggregation rules for authors with different contribution types. However, a practical problem arrives as no information service provides contributor data for now. The consideration of the authors' histories with, e.g., the amount of citations, their position in the institution, or their h-indices, is in particular unfair towards younger researchers.

Anyway, unit counting and fractional counting are fairer methods, which do not discriminate against any of the co-authors. However, when we aggregate data from the individual researcher level to higher levels, unit counting may lead to wrong results. If we assume that two authors from the same institution co-publish an article and count "1" for each author, this article would count "2" for the institution. As co-authorship of researchers from the same institution, city, country, etc. is a matter of fact, the adding of the single values would result in too high values at the aggregate level. This mistake can be prevented by relating whole counting directly to the aggregating level (Reichmann & Schlögl, 2021).

Even when applying fractional counting, it must be considered that proportional authorship according to the formula 1/n does not reflect the "true" contribution of every author; it is a kind of leveling down of each co-author. An information science article with, for instance, two authors counts 1/2 for each; a high-energy physics article with 250 authors (which is thoroughly common there) counts 1/250 for each. This makes every comparison between disciplines with different publication cultures nearly impossible. A solution can be a field-specific normalization of co-author values (Waltman & van Eck, 2019). However, such a normalization is a difficult task, as there are no realistic normalization factors for single disciplines, as one must consider not only the research area, but also the kind of papers, such as, e.g., a research paper versus review article. Furthermore, the underlying distributions are usually skewed, leading to problems of the calculation of arithmetic means or median values (Kostoff & Martinez, 2005).

Many research journals ask for research data if the article reports on empirical findings. If the research data, say, the applied questionnaire or sheets of numerical data, is published in a digital appendix of the paper, the unit of article and data counts "1." But what should be counted if the research data is published in another source such as, e.g., on Zenodo? Should scientometricians ignore the publication of the research data when counting an author's publications? A possible solution would be a separate

 https://doi.org/10.1633/JISTaP.2023.11.2.4

counting of research articles and research data. The same course of action can also be applied for the calculation of the impact. Accordingly, the citations and mentions in social media on those research data could also be observed apart (Peters et al., 2016). As information services for research data do not check the data's quality, they must be regarded as informally published. However, since they are connected with their formally published research paper, they are in between formally and informally published documents.

It is clear that different counting procedures "can have a dramatic effect on the results" (Egghe & Rousseau, 1990, p. 274). In scientometrics, unit counting and fractional counting are often applied; however, both approaches lead to different values (for the country-level see Aksnes et al., 2012; Gauffriau et al., 2007; Huang et al., 2011). Of course, rankings derived from different counting approaches cannot be compared (Gauffriau & Larsen, 2005). Additionally, there are scientometric studies which do not even give detailed information on the applied counting approach (Gauffriau et al., 2008). For Osório (2018, p. 2161) it is impossible to name a general counting method able "to satisfy no advantageous merging and no advantageous splitting simultaneously." Consequently, the problem of the unit of publication for multi-authored research publications remains far away from a satisfying solution.

5.2. Weighting Publications

A book of, say, 800 pages by two authors has with fractional counting a value of 0.5 for each of them; a two-page review of this book by one author counts 1.0. Is this fair? A citation of this monograph in a long article on this work in, for instance, *Science*, counts 1.0; and a citation without substantial text in a rarely read regional journal counts 1.0, too. Is this fair? If the answers for those questions are "no," scientometric studies have to work with weighting factors for research publications and citations.

How can we weight different publication types? In a rather arbitrary way, some authors have introduced weighting factors, for instance, 50 for a monograph, 10 for an article, and 1 for a book review (e.g., Finkenstaedt, 1986). These weighting factors are often based on the average size, which is why monographs usually receive the highest number of points. Papers in journals sometimes receive a higher number of points than those in edited works (Reichmann et al., 2022). However, such document-specific weighting approaches were not able to find broad application, because all introduced weighting factors are more or less employed arbitrarily.

Should a book be weighted with its length (Verleysen & Engels, 2018), with the number of published printed copies (Stock, 2001, p. 21), or with download numbers? Is a book with different chapters one publication, or is every chapter also one publication each? If an information service such as, for instance, WoS in its Book Citation Index, indexes chapter-wise, every chapter counts as a publication. The same procedure is usually applied in the case of edited books. The question remains whether the mere editing should also count as a publication for the editor.

If the number of published pages is used for taking into account the size of a publication, there can be strong distortions due to different formats, font sizes, and layouts. It would be better to count the size in words or characters, but such data is only available for digital publications. The size of publications can also be used as a filter criterion, for example, by only considering publications with a certain minimum size in a research evaluation (Reichmann et al., 2022). There are no international standards for the optimal length of a publication, which will certainly depend heavily on the type and medium of a publication, but scientific contributions, especially those in journals and edited works, should have a certain minimum length in order to be regarded as full-fledged. However, if there were, for example, a minimum size of five printed pages, many articles in highly respected journals such as *Nature* or *Science* would no longer be considered.

Since the papers by Seglen (1992; 1997) it is known that an article's weighting with the impact factor of the publishing journal is misleading, since impact factor is a time-dependent measure for a journal and not for individual articles. Furthermore, if we sort articles in a journal by their citation counts, we can find a highly left-skewed distribution with few heavily cited papers and a long tail of articles being much lesser or even not cited at all. As the impact factor is a quotient of the number of citations to a journal in a given year Y and the number of source papers in the same journal in the years Y-1 and Y-2 (Garfield, 1979), journals which publish primarily up-to-date results (i.e., from the last two years), and therefore are cited soon afterwards, are heavily favored in contrast to journals with longer dissemination periods such as, for instance, in disciplines like history or classical philology. Quantitative descriptions of journals such as the impact factor can lead to a kind of halo effect, i.e., the link of a current evaluation to previous judgments, which is a cognitive bias (Migheli & Ramello, 2021). The impact factor depends on only a few highly cited articles which are generalized to all papers, so the highly cited papers generate the halo effect for

the less-cited ones. Accordingly, the meaning of an impact factor is "systematically biased" (Migheli & Ramello, 2021, p. 1942). The application of the journal impact factor on individual papers was called a "mortal sin" (van Noorden, 2010, p. 865, citing a statement by van Raan); however—and here the application of pseudo-scientometric results becomes problematic—one can find appointment committees in universities working with such misleading indicators (Martyn, 2005).

Similarly, it is not without problems to use discipline-related journal rankings like, for example, JOURQUAL 3 for the field of business administration (Schrader & Hennig-Thurau, 2009) for the weighting of articles. However, this approach also ignores the fact that it is hardly possible to infer the quality of a single paper from the quality of the journal in which it is published.

As already mentioned, granted patents can also be counted as research and development publications. A weighting factor for patents would be the extension of the respective patent family, i.e., the number of countries for which the invention is granted (Neuhäusler & Frietsch, 2013). Additionally, one can weigh the patents by the economic power of the granting country (Faust, 1992), so a patent granted in the United States weights more than one granted in, for instance, Romania or Austria.

To sum up, we did not find hints for weighting factors for research publications, which are generally acknowledged by the scientometrics community as all factors represent more or less arbitrary decisions.

5.3. The Unit of References and Citations and Weighting Citations

A similar problem as determining the unit and weighting of publications comes with the unit and weighting of references and citations. What is one reference and the respective citation? It is common practice in scientometric studies and in citation databases to count every different reference in a publication as a citation of the mentioned publication, so we count the fact of information dissemination from one publication to another. But we do not count the number of references to a cited publication in the text or in footnotes with "loc. cit.," "ibid.," etc. (Stock, 1985). Let us assume that in a hypothetical paper there are two references XYZ, 2021 and ABC, 2020. In the paper's body, ABC, 2020 is mentioned ten times, but XYZ, 2021 only once in a short sentence. Intuitively, we would count ABC, 2020 ten times and XYZ, 2021 once. However, every citation database counts only "1." For such in-text citations, Rauter (2005) and later Zhao et al. (2017) introduced frequency-weighted citation analysis, counting not only "10" in our hypothetical case for ABC, but also the sentiment of the citation (supportive, negative, etc.). Concerning negative citations, Farid (2021, p. 50) states that they have to be excluded for further calculations. However, there is a large practical problem: how can we determine a citation's sentiment? Should every author mark every citation in their article as "positive," "negative," or "neutral," as Rauter (2005) suggests? Or should scientometrics work with automatic sentiment analysis in the sentences in the article's body, where the reference is found (Aljuaid et al., 2021)? Additionally, one could locate the citation in the article, for instance, in the introduction, methods, results, or discussion section (Ding et al., 2014).

How can we weight citations? Self-citations play a special role in citation analyses, as researchers can inflate their citation rate through massive self-citations. It is not useful to count self-citations with "0," as it may be very informative that an author uses former insights in a later publication. Additionally, self-citations may signal continuity and competence in a particular research field. Schubert et al. (2006, p. 505) suggest weights for author self-citations following the kind of self-citation. Accordingly, a self-citation link between a citing and a cited work is stronger if a single-authored paper is cited in a later single-authored paper by the same author than between two multi-authored papers where the citation is made by only one joint co-author (Schubert et al., 2006, p. 505).

Citations of co-authored publications can be counted "1" for each co-author (whole citation counting) or—using analog to fractional counting of publications—1/n for every co-author (fractional citation counting) (Leydesdorff & Shin, 2011).

As not all citing sources are equally important, one can weight a citation by the importance of the citing articles or by the importance of the citing source. It does not seem to be a good idea to work with the journal impact factor of a citing article as this is the same dead end as in the case of the weighting of an individual article with the journal's impact factor. But it is possible to weight a citation with the importance of the citing articles. Here, network analysis comes into play. Many scientometric studies work with PageRank (Brin & Page, 1998) or similar eigenvector-based measures of the citing articles (e.g., Del Corso & Romani, 2009; Ma et al., 2008), which are, however, very time-dependent.

In a review on citation weighting solutions, Cai et al. (2019, p. 474) state that "there isn't a perfect weighting approach since it's hard to prove that one indicator out-

performs another." Additionally, the weighting values of citations (and, not to forget, the absolute numbers of citations) change starkly in the course of time. There is always a cold-start problem in the citation history of a research publication, as in the beginning there is no citation: The publication has to be read, and the citing publication has to be written and formally published—a process which takes months or even some years.

5.4. The Unit of Topics

How can scientometricians determine what a topic is? What is the unit of a topic? And do weighted topics exist? As there are different methods of the representation of a research publication's aboutness, there are also different answers to our questions. We will discuss scientometric topic analysis based on title indexing, text statistics, citation indexing, author keywords, index terms (nomenclatures, thesauri, and classifications), text-word method, and content analysis. The methods of title indexing, text statistics, citation indexing, and author keywords are able to capture the author aboutness, while index terms, text-word method, and content analysis describe the indexer aboutness of a publication.

Topic identification, which is based upon text statistics (including the statistics of title words), works with identified single terms and phrases or alternatively with n-grams in the text (Leydesdorff & Welbers, 2011). To create relationships between terms, phrases, or n-grams, in many cases co-word analysis and co-word mapping are applied. One may weigh the importance of terms, etc. following the frequency of occurrences in both the single publications and in a set of documents (e.g., a hit list of an information service). Further steps of processing may be calculations of term similarity or a factor analysis of the terms. The main problems of text statistics are different publication languages, different language usages of the authors, linguistic challenges such as, e.g., lemmatization or stemming, phrase detection, name recognition, the selection of "important" topics, the choice of the data processing methods such as, for instance, similarity calculation, the creation of topic clusters, and, finally, the presentation of the clusters in an expressive image. An additional problem for title indexing is that there are only few terms for processing. However, an advantage of text statistics for the task of topic identification is that all steps can be automatized.

For Garfield (1979, p. 3), citations represent concepts and so topics. However, with new citations of the same publication the represented concept and topic may change, as the citing papers come from different subject areas. In addition to the analysis of direct citations (Garfield et al., 2006), scientometricians can apply co-citation and bibliographic coupling for topic identification (Kleminski et al., 2022). Honka et al. (2015) describe a combination of citation indexing and title indexing. In their topical impact analysis, they analyze the title terms of the citing publications of all the literature of a research institution. Sjögårde and Didegah (2022) found associations between citations and topics insofar as publications covering fast-growing topics are more cited compared to publications in slow-growing or declining topics. However, we have to recall that if we apply citation analysis for topic analysis, we inherit all discussed problems of citation analysis.

Keywords are often given by the authors themselves, but information services like WoS also generate so-called "KeyWords Plus" automatically, which are single words or phrases that frequently appear in the titles of an article's references besides the author keywords (Haustein, 2012, pp. 90-91). Both types of keywords include hints to research topics (Li, 2018) and complement title indexing. But author keywords depend heavily on the author's language usage and WoS's KeyWords Plus depends on the author's selection of cited papers (if there are references at all).

Index terms are found in bibliographic records of library holdings or discipline-specific information services such as Medline or PubMed (medicine) or Derwent (patents). Nomenclatures (Stock & Stock, 2013, pp. 635-646) or thesauri (Stock & Stock, 2013, pp. 675-696) such as the MeSH reflect the topics of the sources of the respective service. For Shu et al. (2021, p. 448), MeSH or the controlled vocabulary of similar nomenclatures and thesauri are well suitable for mapping science using subject headings. Concerning classification systems (Stock & Stock, 2013, pp. 647-674), the International Patent Classification (IPC) is often used for topic analyses. Scientometric—or better, patentometric—studies apply both the notations of the IPC and the co-occurrences of those notations (da Silveira Bueno et al., 2018). WoS offers with its Subject Categories a very rough classification system for the topics of covered journals (Haustein, 2012, pp. 94-101), which is indeed applied in scientometric topic analyses, mainly for high-level aggregates such as cities and regions (Altvater-Mackensen et al., 2005). As journals in WoS may be assigned to different classes, this has also a certain effect on topic identification (Jesenko & Schlögl, 2021).

Content analysis works with the classification method, for which the (as a rule: few) classes are built with respect

to the set of publications which should be analyzed. The class construction is both inductive (derived from the publications to be analyzed) and deductive (derived from other research literature). The construction of the classes and the sorting of the publications in the classes is a complex procedure, which is also sometimes used in scientometrics (Järvelin & Vakkari, 2022).

While nomenclatures, thesauri, and classification systems work with a standardized controlled terminology, the text-word method (Stock & Stock, 2013, pp. 735-743) only allows words from the concrete publication as index terms, making this indexing method as low-interpretative as possible. For Henrichs (1994; see also Stock, 1984), the text-word method is ideally suited for contextual topic analyses and for the history of ideas.

For topic analyses, there are ways to weight the topics in a research paper. If one applies MeSH, one can work with a binary weight, i.e., major and non-major topics. The text-word method as a form of syntactical indexing weights every index term in the interval between greater 0 and 100 while calculating the importance of the term in the text (Stock & Stock, 2013, pp. 766-767). If an information service works with coordinating indexing, every index term forms a topic unit and weights "1."

Indeed, there are many approaches to describe topics in scientometrics. But there is no widely accepted "golden path" to scientometric topic analysis. As the concept of "aboutness" itself is rather blurred, so are topic analyses. However, if we work in the framework of one method such as with classification systems or with thesauri, and with exactly one tool such as, e.g., IPC or MeSH, there should be reliable results for the analyzed topic domain (in our examples, patent literature or medicine). Cross-frontier comparisons between domains, methods, and concrete tools are not possible unless there are semantic crosswalks between different vocabularies. Sometimes mapping is applied, where the concepts of different tools, for instance of different thesauri, are set in relation (Doerr, 2001). It is also possible to create a crosswalk between different methods, e.g., between the Standard Thesaurus Economy and the classification system of Nomenclature statistique des activités économiques dans la Communauté Européenne (NACE) (Stock & Stock, 2013, p. 724). Nevertheless, such crosswalks are scarce and their applications in scientometrics are nearly unknown outside of patent analyses.

All in all, the scientometric description, analysis, and evaluation of topics of research publications is still a great challenge, as there are no broadly accepted methods and also no accepted indicators.

5.5. Counting Problems on the Meso- and Macro-levels

Scientometrics on the meso- and macro-levels can apply size-dependent indicators (irrespective of the size of a research unit) or size-independent indicators (considering the size of a unit of research) (Waltman et al., 2016). It seems to be unfair to compare different research institutions (scientometric meso-level) or even countries (macro-level) through their absolute numbers of research publications and citations, since this favors large institutions or countries. Nevertheless, this often happens, for example, in the case of university rankings (Olcay & Bulu, 2017). For all studies on the meso- and macro-level which use whole counting it must be taken into account that co-authors from the same institution or the same country are not counted multiple times. A possible solution might be to relate the fractional counts to the total number of relevant researchers (as did, e.g., Abramo & D'Angelo, 2014 and Akbash et al., 2021) or—better—to the relevant full-time equivalent numbers (Kutlača et al., 2015, p. 250; Stock et al., 2023) for the whole analysis period (Docampo & Bessoule, 2019).

In contrast to evaluations of single researchers, scientometric studies on the meso- and macro-level have, first of all, the challenge of capturing the exact number of staff members in the institution or country in the considered time frame (Toutkoushian et al., 2003) as the "underlying production capability" (Thelwall & Fairclough, 2017, p. 1142). In this context, it must also be clarified whose research performance should be relevant for the research evaluation. The answer to this question seems simple at first glance, namely that of the evaluated units, i.e., in the case of a research evaluation at the institute level, that of the compared institutes (Rousseau et al., 2018). If the research performance is to be recorded as comprehensively as possible, all researchers, i.e., all scientific employees, must be taken into account. However, it must be considered that the research performance varies strongly between research staff. Pre-doctoral researchers, for instance, sometimes have hardly any publications because they often need several years to familiarize themselves with research. Therefore, it could make sense only to consider professors and post-doctoral researchers as representatives of a research institution. There is an additional problem with the acquisition of the researchers' personal data in particular for longer time periods (Reichmann & Schlögl, 2022), as such data are not always and everywhere

 https://doi.org/10.1633/JISTaP.2023.11.2.4

freely accessible, and the access may be restricted due to national data protection acts.

6. KEY FINDINGS AND OUTLOOK

The main contribution of this article is a compilation of problems in scientometrics and research evaluation when it comes to the exact determination of the counting units of these knowledge fields. There is only a little completely new in the paper as most issues can be found in the literature, but the concise synopsis reveals several serious problems.

6.1. Main Results: A Dozen Problem Areas

The key findings of our paper can be subsumed as follows: Problems with scientometric analyses already start with basic assumptions that a publication is (1) formally or informally published, that it (2) contains research content, and that (3) in the publication's byline, the names of the authors are stated. The concepts of (4) references and citations and of (5) topics and their descriptions are also not free of fundamental problems. Coming to the aspect of "research" in research publications, we found three problem areas: the criterion of demarcation (extension of 2), (6) the data sources of formally published publications and their problems, and (7) the data sources of informally published publications and their problems. Finally, there are challenges concerning the counting unit of research publications, references or citations, and topics. What exactly are 1 publication, 1 reference, 1 citation, and 1 topic? Identified problem areas here are (8) multi-author publications, (9) weighted publications, (10) the unit of references and citations, (11) the unit of topics, and (12) additional counting problems on the meso- and macro-levels. These problem areas are numbered according to their first appearance in this paper. This sub-section also provides the answers to our three research questions.

6.1.1. What are a publication, a reference or citation, and a topic?

(1) Formal publishing happens through classical communication channels including journals, proceedings, and patents; informal publishing uses newspapers as well as other mass media sources and sources of the social web such as Mendeley or Twitter. The classical formal channels have been studied by scientometrics for decades, while the observation of the digital informal channels was added as "altmetrics" with the upcoming of the social web with its user-created content. Articles on preprint servers and published research data cannot be regarded as formally published as there is no gatekeeping process, but they are also not informally published as they are connected with their respective formal publications, if they exist at all. So they are "in between" formal and informal publishing.

Publications are—in the sense of Briet (1951/2006) and later Buckland (1997)—documents. However, documents are not only papers, but sometimes also artifacts as buildings as documents of engineering or software programs and information systems as documents of computer science, which are not captured by scientometric methods at all. Published non-artifact documents may be analyzed as works, expressions, manifestations, or items, but scientometrics makes these important distinctions only rarely. Since there are different field-specific cultures concerning publications, citations, and mentions in newspapers and social media, scientometric and altmetric analyses have to consider these differences, e.g., by normalization; however, we miss broadly accepted methods.

(2) Research content calls for the criterion of demarcation between research and non-research (see below!).

(3) Authorship is partly indeterminable as there are ghost writers and also honorary authors. As not all authors contribute equally to a publication, it could be wiser to work with contributorship instead of authorship. But also, in this case it is not determinable whether the declarations of contributorship are correct. Moreover, there is a practical problem as no bibliographic information service contains such data.

(4) References are bibliographic descriptions of documents used in publications and so an expression of intertextuality. For an individual publication, intertextuality is an information flow from an older publication to the actual one as shown by the reference; in the opposite direction intertextuality is considered as a citation and is regarded as reputation for the cited publication. The usefulness of reference and citation analyses clearly depends on the veracity of the references given. And this veracity is not or not always prevalent, as authors cite publications which were not used in the actual paper, and they do not cite all influences. Furthermore, contributions by practitioners in research journals are less regarded in citation studies since they do not cite to the same degree as academics.

(5) Topics are descriptions of the publication's aboutness. The author aboutness can be analyzed by means of text-statistical methods with all the problems of author-specific language use. The indexer aboutness can be identified by applying the index terms from discipline-specific databases. Here, the method (e.g., classification or

thesaurus), the vocabulary of the concrete tool (e.g., IPC or MeSH), the tool's indexing rules, and the work of an indexer play crucial roles. All the steps in the process of the description of topics are prone to failure.

6.1.2. What is a research publication?

(2 in detail) In the philosophy of science, the problem of defining "research" in contrast to non-research or pseudo-research is discussed as the criterion of demarcation, but this discussion did not find an accepted solution. In scientometrics, we were not able to find serious discussions on the demarcation criterion at all. A criterion for scientometrics could be to bank on the claimed or supposed research character of the publishing sources. For instance, scientometricians could rely on journals or proceedings meeting basic requirements to ensure quality standards like, for instance, peer review. But peer review may work badly, and there is the additional issue of predatory journals and conferences. It is also unclear how to handle papers on preprint servers, which are not published on a classical channel, and how to count published research data.

(6) There are many data sources which can be used for scientometric studies including multi-disciplinary information services such as, e.g., WoS, Scopus, Dimensions, or Google Scholar, and discipline-specific databases as Medline for medicine. All named sources have problems concerning completeness. WoS consists of diverse subdatabases, which are differently subscribed by research libraries. Some sources do not satisfactorily support scientometric analyses (such as Google Scholar). Alternatively or additionally to information services, scientometrics can apply the data of personal, institutional, or national publication lists. However, these data sources are also not fully complete and the data are not in a standardized format.

(7) Some of the social media sources, for instance, tweets in microblogging services, mentions in Wikipedia, references in Mendeley, posts in other sources with user-created content, download numbers, and—not to forget—articles in newspapers or mentions in TV shows—may give supplementary scientometrics information. However, documents on social media may be deleted later or included links may lead to nowhere, so replications of studies in altmetrics are partly not possible. One may interpret measures of altmetrics as indicators of research-internal and research-external interestingness, but a theoretical justification is still pending.

There are some online sources which are easily accessible and the data of which are well processable, e.g., WoS, Scopus, and Dimensions for formally published documents and Twitter for informally published documents, which are frequently used in scientometrics and altmetrics. None the less, scientometricians should always use those data sources which are best suited for their scientometric problems to be investigated (e.g., researchers' personal publication lists) instead of easy-to-edit sources.

6.1.3. What are one research publication, one reference or one citation, and one topic?

(8) Scientometricians may apply different counting methods for determining the contribution of authors to an individual article. Often used are unit counting (all authors count 1) and fractional counting (e.g., counting 1/n for every co-author given n names in the byline). Different counting methods can have a huge effect on the findings.

(9) It is evident that a book with twelve chapters cannot be counted as 1, whereas the twelve chapters, published as articles, would count for 12. However, there are no valid and reliable weighting methods so far. Many approaches are arbitrary (as the weighting of publications by document type) or simply wrong (e.g., applying the journal impact factor to individual papers). For problems no. 8 and 9, we were not able to find an accepted gold standard in scientometrics.

In scientometrics, (10) a reference is considered as an information flow from an earlier document to the citing one, while not considering the extent of the information flows. If an article references another paper, for instance, 20 times and another one only once in a short note, both count—against evidence—1. Citations of multi-authored publications are normally counted 1 for every cited author; but fractional citation counting (e.g., 1/n for every co-author) would represent a much fairer solution. We also found approaches for weighted citations as, e.g., using the PageRanks of the citing articles, but no perfect solution.

In order to capture (11) research topics for quantitative scientometric descriptions and analyses, there are indeed many paths including text-statistical approaches and the application of index terms from bibliographical information services. We found options for binary topic weighting (for instance, major versus non-major topics) and interval-scaled weighting (by applying a database working with syntactical indexing). However, we failed to identify a widely accepted method for topic description and for topic weighting.

On the meso- and macro-levels of scientometrics and especially of research evaluation, special problems arise

 https://doi.org/10.1633/JISTaP.2023.11.2.4

(12): Who is a member of an institution? Should research evaluation consider the institution's size? If yes, do we count faculty or all staff? Should full-time equivalents be used instead of head counts? Is it possible to get these numbers at all due to data protection regulations?

6.2. Counting Problems of Scientometrics in an Overview

In Fig. 1, we summarized important problems regarding the empirical basis of scientometrics and research evaluation. Additionally, we categorized the problems into three areas. Firstly, there are problems of the research system such as, for instance, authorship, redundant publications, and veracity of references. These problems are caused by the research system, and scientometrics is not able to solve them alone. The second problem category includes the areas of data collection and determination of the counting units of the publications, references, citations, and topics such as the handling of multi-authored publications and their citations, normalization, the weighting of publications and citations, and the methods for capturing topics. We miss clear definitions of formally published documents, informal published documents, and documents in between such as preprints and research data. The last category includes problems with scholarly indicators. As there is a gap between measures and indicators, the interpretation of the indicators is essential. Measures, e.g., counts of citations, are safe terrain, apart from the problems listed above and apart from measurement errors. Indicators refer to not directly observable constructs. For instance, counts of citations are regarded as indicators of impact in the research community, which is a construct. They allow for interpretations and are therefore open for uncertainties. The inference from an



Fig. 1. Problem areas with regard to the empirical basis of scientometrics and research evaluation.

indicator of a thing to the thing itself is called "reification," which should be avoided (Stock, 1994, p. 24; Weingart & Winterhager, 1984, p. 26). Therefore, "indicators built from patents and articles cannot be perfect" (Narin et al., 1994, p. 69). Besides the interpretation of measures and the mentioned gap between scientometric and altmetric measures and their indicators, there are the problems concerning size-dependent and size-independent indicators and the problem areas of derived indicators.

Quantitative scientometric studies are therefore by no means unproblematic. Though problems in the research system are outside of scientometrics, these issues are reflected in scientometric studies. The problems increase when it comes to the correct data collection including the counting of research publications, citations, and topics. Finally, there are problems with scientometric indicators and their interpretation. Indeed, it seems dangerous to rely only on scientometric indicators for the description, analysis, evaluation, and policy of research. Additionally, all mentioned problems may lead to malpractices especially in research evaluation, when evaluators are driven by research-external interests and prefer measures or indicators which primarily meet their interests.

6.3. Outlook

We concentrated on the counting unit of scientometrics and research evaluation, namely the research publication, its measures, and its indicators. When analyzing the foundations of derived indicators, we would probably find further problems. What are the strengths or weaknesses of, for instance, the impact factor, the h-index, or the altmetrics score? What do they really measure? Are researchers aware of scientometric and altmetric indicators (Keng et al., 2022)? How do scientists appraise the indicators' importance for themselves and for their knowledge field (Kamrani et al., 2020; 2021)? Do the scientists have a sufficient level of metrics literacies to be able to understand all measures and indicators and their problems (Dorsch et al., 2021)? Of course, there are several articles on these topics. However, there is room left for further reviews and for detailed criticism about the basics of scientometrics.

Every scientometric study should take into account the problem areas mentioned in this article as much as possible. One should be aware, however, that compromises always have to be made in such analyses. Nevertheless, quantitative scientometric studies usually have a certain benefit and added value. However, it is absolutely necessary to point out the limitations of such a study and, ideally, to rethink the scientometric results by conducting an additional qualitative analysis. We have learned from Leydesdorff (1995) that the science of science is broader than scientometrics as it also comprises the sociology of knowledge and theories of information and communication. For Leydesdorff (1995, p. 10), there is room "for the use of qualitative data and dynamic analysis" in addition to purely quantitative scientometrics. Therefore, scientometricians should overcome the divide between quantitative and qualitative studies in the science of science (Cambrosio et al., 2020; Kang & Evans, 2020; Leydesdorff et al., 2020). Qualitative methods in science studies cover, for instance, interviews with researchers, content analysis of research publications, ethnographic field research (e.g., observations of researchers at work), surveying researchers (also with open questions), and grounded theory (in order to arrive at theories on research communication) (see from another knowledge domain: Gremm et al., 2018, pp. 75-78).

Furthermore, the adoption of theoretical approaches from communications science (e.g., the Uses and Gratifications Theory or the Lasswell Formula of Communication; see Zimmer et al., 2018) is helpful for the analysis of scholarly communication in scientometrics, too. With the help of such approaches, we may analyze, for instance, in the context of the Lasswell Formula (Lasswell, 1948), the way from research papers' authors (who?) to the intended target group (to whom?), considering also the content (says what?), the paths of transmission (in which channels?), and the readers' reactions, e.g., applying or citing the paper (with what effect?). Or we may analyze the research communication from the opposite direction: Why do people read, apply, and cite a research paper? For the Uses and Gratifications Theory by Katz et al. (1973-1974), there are seven aspects in the audience's research usage: (1) the social and psychological origins of (2) the audience members' needs, which generate (3) expectations of (4) the research literature or other sources, which lead to (5) different patterns of research exposure, resulting in (6) need gratifications and (7) other consequences (including unintended ones). In this manner, scientometrics will not only count and interpret publications, topics, and citations, but the entire process of scholarly communication.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

 https://doi.org/10.1633/JISTaP.2023.11.2.4

REFERENCES

Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics,* 101(2), 1129-1144. https://doi.org/10.1007/s11192-014-1269-8

Aguillo, I. F. (2012). Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics,* 91(2), 343-351. https://doi.org/10.1007/s11192-011-0582-8

Akbash, K. S., Pasichnyk, N. O., & Rizhniak, R. Y. (2021). Analysis of key factors of influence on scientometric indicators of higher educational institutions of Ukraine. *International Journal of Educational Development,* 81, 102330. https://doi.org/10.1016/j.ijedudev.2020.102330

Aksnes, D. W., Schneider, J. W., & Gunnarsson, M. (2012). Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalised counting methods. *Journal of Informetrics,* 6(1), 36-43. https://doi.org/10.1016/j.joi.2011.08.002

Albarillo, F. (2014). Language in social science databases: English versus non-English articles in JSTOR and Scopus, *Behavioral & Social Sciences Librarian,* 33(2), 77-90. https://doi.org/10.1080/01639269.2014.904693

Albrecht, K., Frost, M., & Handtke, U. (1994, November 2-4). Informetrische Vermessung eines Forschungsinstituts. In W. Rauch, F. Strohmeier, H. Hiller, & C. Schlögl (Eds.), *Mehrwert von information – Professionalisierung der Informationsarbeit: Proceedings des 4. Internationalen Symposiums für Informationswissenschaft (ISI '94)* (pp. 151-163). Universitätsverlag Konstanz. German.

Aljuaid, H., Iftikhar, R., Ahmad, S., Asif, M., & Tanvir Afzal, M. (2021). Important citation identification using sentiment analysis of in-text citations. *Telematics and Informatics,* 56, 101492. https://doi.org/10.1016/j.tele.2020.101492

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to 'webometrics'. *Journal of Documentation,* 53(4), 404-426. https://doi.org/10.1108/EUM0000000007205

Altvater-Mackensen, N., Balicki, G., Bestakowa, L., Bocatius, B., Braun, J., Brehmer, L., Brune, V., Eigemeier, K., Erdem, F., Fritscher, R., Jacobs, A., Klingsporn, B., Kosinski, M., Kuntze, J., Lee, J. R., Osterhage, A., Probost, M., Risch, T., Schmitt, T., . . . Werner, K. (2005). Science and technology in the region: The output of regional science and technology, its strengths and its leading institutions. *Scientometrics,* 63(3), 463-529. https://doi.org/10.1007/s11192-005-0225-z

Añazco, D., Nicolalde, B., Espinosa, I., Camacho, J., Mushtaq, M., Gimenez, J., & Teran, E. (2021). Publication rate and citation counts for preprints released during the COVID-19 pandemic: the good, the bad and the ugly. *PeerJ,* 9, e10927. https://doi.org/10.7717/peerj.10927

Archambault, É. (2002). Methods of using patents in cross-country comparisons. *Scientometrics,* 54(1), 15-30. https://doi.org/10.1023/A:1015654903218

Auger, P. (1998). *Information sources in grey literature: Guides to information sources* (4th ed.). De Gruyter Saur.

Aviv-Reuven, S., & Rosenfeld, A. (2021). Publication patterns' changes due to the COVID-19 pandemic: A longitudinal and short-term scientometric analysis. *Scientometrics,* 126(8), 6761-6784. https://doi.org/10.1007/s11192-021-04059-x

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies,* 1(1), 377-386. https://doi.org/10.1162/qss_a_00019

Bar-Ilan, J. (2006). An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes. *Information Processing & Management,* 42(6), 1553-1566. https://doi.org/10.1016/j.ipm.2006.03.019

Bar-Ilan, J. (2019). Data collection from the web for informetric purposes. In W Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 781-800). Springer.

Bar-Ilan, J., & Halevi, G. (2017). Post retraction citations in context: A case study. *Scientometrics,* 113(1), 547-565. https://doi.org/10.1007/s11192-017-2242-0

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review,* 13(5), 407-424. https://doi.org/10.1108/eb024320

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies,* 1(1), 363-376. https://doi.org/10.1162/qss_a_00018

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology,* 36(1), 2-72. https://doi.org/10.1002/aris.1440360102

Bornmann, L., Wagner, C., & Leydesdorff, L. (2015). BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. *Journal of the Association for Information Science and Technology,* 66(7), 1507-1513. https://doi.org/10.1002/asi.23333

Boxenbaum, H., Pivinski, F., & Ruberg, S. J. (1987). Publication rates of pharmaceutical scientists: Application of the waring distribution. *Drug Metabolism Reviews,* 18(4), 553-571. https://doi.org/10.3109/03602538708994132

Briet, S. (2006). *What is documentation?* (R. E. Day, L. Martinet, & H. G. B. Anghelescu, Trans.). Scarecrow Press.

(Original work published 1951)

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems,* 30(1-7), 107-117. https://doi.org/10.1016/S0169-7552(98)00110-X

Broad, W. J. (1981). The publishing game: Getting more for less. *Science,* 211(4487), 1137-1139. https://doi.org/10.1126/science.7008199

Buckland, M. K. (1997). What is a "document"? *Journal of the American Society for Information Science,* 48(9), 804-809. https://doi.org/10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V

Cai, L., Tian, J., Liu, J., Bai, X., Lee, I., Kong, X., & Xia, F. (2019). Scholarly impact assessment: A survey of citation weighting solutions. *Scientometrics,* 118(2), 453-478. https://doi.org/10.1007/s11192-018-2973-6

Cambrosio, A., Cointet, J. P., & Abdo, A. H. (2020). Beyond networks: Aligning qualitative and computational science studies. *Quantitative Science Studies,* 1(3), 1017-1024. https://doi.org/10.1162/qss_a_00055

Carnap, R. (1928). *Der logische Aufbau der Welt.* Meiner. German.

Carnap, R. (1931). Die physikalische sprache als Universalsprache der Wissenschaft. *Erkenntnis,* 2(1), 432-465. German. https://doi.org/10.1007/BF02028172

Carnap, R. (1932). Über Protokollsätze. *Erkenntnis,* 3(1), 215-228. https://doi.org/10.1007/BF01886421

Cavaller, V. (2009). Scientometrics and patent bibliometrics in RUL analysis: A new approach to valuation of intangible assets. *Vine,* 39(1), 80-91. https://doi.org/10.1108/03055720910962461

Chavarro, D., Ràfols, I., & Tang, P. (2018). To what extent is inclusion in the Web of Science an indicator of journal 'quality'? *Research Evaluation,* 27(2), 106-118. https://doi.org/10.1093/reseval/rvy001

Chen, X. (2019). Beall's list and Cabell's blacklist: A comparison of two lists of predatory OA journals. *Serials Review,* 45(4), 219-226. https://doi.org/10.1080/00987913.2019.1694810

Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and scholarly communication: Adoption, practices, drivers and barriers. *F1000Research,* 8, 971. https://doi.org/10.12688/f1000research.19619.2

Cobey, K. D., Lalu, M. M., Skidmore, B., Ahmadzai, N., Grudniewicz, A., & Moher, D. (2018). What is a predatory journal? A scoping review. *F1000Research,* 7, 1001. https://doi.org/10.12688/f1000research.15256.2

da Silveira Bueno, C., da Silveira, J. M. F. J., Buainain, A. M., & Poz, M. E. S. D. (2018). Applying an IPC network to identify the bioenergy technological frontier. *Revista Brasileira De Inovação,* 17(2), 259-286. https://doi.org/10.20396/rbi.v17i2.8652020

Del Corso, G. M., & Romani, F. (2009). Versatile weighting strategies for a citation-based research evaluation model. *Bulletin of the Belgian Mathematical Society - Simon Stevin,* 16(4), 723-743. https://doi.org/10.36045/bbms/1257776244

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology,* 65(9), 1820-1833. https://doi.org/10.1002/asi.23256

Doerr, M. (2001). Semantic problems of thesaurus mapping. *Journal of Digital Information,* 1(8), 1-15. https://www.backbonethesaurus.eu/sites/default/files/Semantic-problems-of-thesaurus-mapping(2001).pdf

Docampo, D., & Bessoule, J. J. (2019). A new approach to the analysis and evaluation of the research output of countries and institutions. *Scientometrics,* 119(2), 1207-1225. https://doi.org/10.1007/s11192-019-03089-w

Dorsch, I. (2017). Relative visibility of authors' publications in different information services. *Scientometrics,* 112(2), 917-925. https://doi.org/10.1007/s11192-017-2416-9

Dorsch, I., Askeridis, J., & Stock, W. (2018). Truebounded, overbounded, or underbounded? Scientists' personal publication lists versus lists generated through bibliographic information services. *Publications,* 6(1), 7. https://doi.org/10.3390/publications6010007

Dorsch, I., & Haustein, S. (2023). Bibliometrie. In R. Kuhlen, D. Lewandowski, W. Semar, & C. Womser-Hacker (Eds.), *Grundlagen der Informationswissenschaft* (pp. 271-280). De Gruyter Saur.

Dorsch, I., Jeffrey, A., Ebrahimzadeh, S., Maggio, L. A., & Haustein, S. (2021, July 12-15). Metrics literacies: On the state of the art of multimedia scholarly metrics education. In W. Glänzel, S. Heeffer, P. S. Chi, & R. Rousseau (Eds.), *Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI2021)* (pp. 1465-1466). ISSI.

Egghe, L., Guns, R., & Rousseau, R. (2013). Measuring co-authors' contribution to an article's visibility. *Scientometrics,* 95(1), 55-67. https://doi.org/10.1007/s11192-012-0832-4

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Elsevier.

El-Ouahi, J. (2022). The Arabic Citation Index - Toward a better understanding of Arab scientific literature. *Research Square.* https://doi.org/10.21203/rs.3.rs-1917781/v2

Erdt, M., Nagarajan, A., Sin, S. C. J., & Theng, Y. L. (2016). Altmetrics: An analysis of the state-of-the-art in measuring

research impact on social media. *Scientometrics,* 109(2), 1117-1166. https://doi.org/10.1007/s11192-016-2077-0

Farid, F. O. (2021). On a new index for research assessment. *Journal of Information Science Theory and Practice,* 9(3), 56-75. https://doi.org/10.1633/JISTaP.2021.9.3.5

Fasce, A. (2017). What do we mean when we speak of pseudoscience? The development of a demarcation criterion based on the analysis of twenty-one previous attempts. *Disputatio. Philosophical Research Bulletin,* 6(7), 459-488. https://doi.org/10.5281/zenodo.1433737

Faust, K. (1992, March 19-20). Patentanmeldungen als Indikator von technologischen Entwicklungen – Ergebnisse und Erfahrungen mit der IFO Patentstatistik. In A. Editor & B. Editor (Eds.), *Proceedings of the 1st EPO-IFO-Workshop: Results and Methods of Economic Patent Research* (pp. 155-172). European Patent Office. German.

Fernandez-Beanato, D. (2020). The multicriterial approach to the problem of demarcation. *Journal for General Philosophy of Science,* 51(3), 375-390. https://doi.org/10.1007/s10838-020-09507-5

Feyerabend, P. (1975). Against method: Outline of an anarchistic theory of knowledge. New Left Books.

Finkenstaedt, T. (1986). Forschungsmessung in den Geisteswissenschaften. In R. Fisch, & H. D. Daniel (Eds.), *Messung und Förderung von Forschungsleistung* (pp. 51-60). Universitätsverlag Konstanz. German.

Flanagin, A., Carey, L. A., Fontanarosa, P. B., Phillips, S. G., Pace, B. P., Lundberg, G. D., & Rennie, D. (1998). Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *Journal of the American Medical Association,* 280(3), 222-224. https://doi.org/10.1001/jama.280.3.222

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biology,* 19(4), e3000959. https://doi.org/10.1371/journal.pbio.3000959

Fraser, N., Mayr, P., & Peters, I. (2022). Motivations, concerns and selection biases when posting preprints: A survey of bioRxiv authors. *PLoS One,* 17(11), e0274441. https://doi.org/10.1371/journal.pone.0274441

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies,* 1(2), 618-638. https://doi.org/10.1162/qss_a_00043

Fröhlich, G. (1999). Das messen des leicht meßbaren: Output-indikatoren, impact-maße: Artefakte der szientometrie? In J. Becker, & W. Göhring (Eds.), *Kommunikation statt markt: Zu einer alternativen theorie der informationsgesellschaft* (pp. 27-38). GMD - Forschungszentrum Informationstechnik.

García-Villar, C. (2021). A critical review on altmetrics: Can we measure the social impact factor? *Insights Imaging,* 12, 92. https://doi.org/10.1186/s13244-021-01033-2

Garfield, E. (1979). *Citation indexing - Its theory and application in science, technology, and humanities.* Wiley.

Garfield, E., Paris, S. W., & Stock, W. G. (2006). HistCite™: A software tool for informetric analysis of citation linkage. *NFD Information-Wissenschaft und Praxis,* 57(8), 391-400. https://www.isi.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Sprache_und_Information/Informationswissenschaft/Dateien/Wolfgang_G._Stock/Daten_vor_2011/1166781999eugenegarf.pdf

Gauffriau, M. (2017). A categorization of arguments for counting methods for publication and citation indicators. *Journal of Informetrics,* 11(3), 672-684. https://doi.org/10.1016/j.joi.2017.05.009

Gauffriau, M. (2021). Counting methods introduced into the bibliometric research literature 1970–2018: A review. *Quantitative Science Studies,* 2(3), 932-975. https://doi.org/10.1162/qss_a_00141

Gauffriau, M., & Larsen, P. O. (2005). Counting methods are decisive for rankings based on publication and citation studies. *Scientometrics,* 64(1), 85-93. https://doi.org/10.1007/s11192-005-0239-6

Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics,* 73(2), 175-214. https://doi.org/10.1007/s11192-007-1800-2

Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A., & von Ins, M. (2008). Comparisons of results of publication counting using different methods. *Scientometrics,* 77(1), 147-176 (2008). https://doi.org/10.1007/s11192-007-1934-2

Glänzel, W. (2008). Seven myths in bibliometrics about facts and fiction in quantitative science studies. *COLLNET Journal of Scientometrics and Information Management,* 2(1), 9-17. https://doi.org/10.1080/09737766.2008.10700836

Grünbaum, A. (1977). Is psychoanalysis a pseudo-science? Karl Popper versus Sigmund Freud. *Zeitschrift Für Philosophische Forschung,* 31(3), 333-353. http://www.jstor.org/stable/20482837

Gremm, J., Barth, J., Fietkiewicz, K. J., & Stock, W. G. (2018). *Transitioning towards a knowledge society: Qatar as a case study.* Springer.

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of Informetrics,* 11(3), 823-834. https://doi.

org/10.1016/j.joi.2017.06.005

Hansson, S. O. (2013). Defining pseudoscience and science. In M. Pigliucci, & M. Boudry (Eds.), *Philosophy of pseudoscience: Reconsidering the demarcation problem* (pp. 61-78). University of Chicago Press.

Haustein, S. (2012). *Multidimensional journal evaluation: Analyzing scientific periodicals beyond the impact factor.* De Gruyter Saur.

Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics,* 108(1), 413-423. https://doi.org/10.1007/s11192-016-1910-9

Haustein, S. (2019). Scholarly Twitter metrics. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 729-760). Springer.

Henderson, M. (2010). Problems with peer review. *British Medical Journal,* 340, c1409. https://doi.org/10.1136/bmj.c1409

Henrichs, N. (1994). Begriffswandel in Datenbanken: Kontextuelle Inhaltsanalyse für Disambiguierung und ideengeschichtliche Analyse. In H. Best, B. Endres-Niggemeyer, M. Herfurth, & H. P. Ohly (Eds.), *Informations- und Wissensverarbeitung in den Sozialwissenschaften* (pp. 225-239). VS Verlag für Sozialwissenschaften. German.

Herb, U. (2016). Altmetrics zwischen Revolution und Dienstleistung : Eine methodische und konzeptionelle Kritik. In H. Staubmann (Ed.), *Soziologie in Österreich – Internationale Verflechtungen* (pp. 387-410). Innsbruck University Press. German.

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies,* 1(1), 387-395. https://doi.org/10.1162/qss_a_00020

Heumüller, R., Nielebock, S., Krüger, J., & Ortmeier, F. (2020). Publish or perish, but do not forget your software artifacts. *Empirical Software Engineering,* 25(6), 4585-4616. https://doi.org/10.1007/s10664-020-09851-6

Hilbert, F., Barth, J., Gremm, J., Gros, D., Haiter, J., Henkel, M., Reinhardt, W., & Stock, W. G. (2015). Coverage of academic citation databases compared with coverage of scientific social media: Personal publication lists as calibration parameters. *Online Information Review,* 39(2), 255-264. https://doi.org/10.1108/OIR-07-2014-0159

Holcombe, A. (2019). Farewell authors, hello contributors. *Nature,* 571(7764), 147. https://doi.org/10.1038/d41586-019-02084-8

Holmberg, K. (2015). *Altmetrics for information professionals: Past, present and future.* Chandos Publishing.

Honka, A., Orszullok, L., Dorsch, I., & Frommelius, N. (2015, May 19-21). Topical impact analysis - A new informetric indicator for the assessment of a scientific institution. In F. Pehar, C. Schlögl, & C. Wolff (Eds.), *Proceedings of the 14th International Symposium on Information Science (ISI 2015): Re:inventing Information Science in the Networked Society* (pp. 410-421). Verlag Werner Hülsbusch.

Hu, G., Wang, L., Ni, R., & Liu, W. (2020). Which h-index? An exploration within the Web of Science. *Scientometrics,* 123(3), 1225-1233. https://doi.org/10.1007/s11192-020-03425-5

Huang, C. K., Neylon, C., Kenworthy, C. B., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies,* 1(2), 445-478. https://doi.org/10.1162/qss_a_00031

Huang, M. H., Lin, C. S., & Chen, D. Z. (2011). Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact. *Journal of the American Society for Information Science and Technology,* 62(12), 2427-2436. https://doi.org/10.1002/asi.21625

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional requirements for bibliographic records.* K. G. Saur.

Ingwersen, P. (2002, July 21-25). Cognitive perspectives of document representation. In H. Bruce, R. Fidel, P. Ingwersen, & P. Vakkari (Eds.), *Proceedings of the 4th International Conference on Conceptions of Library and Information Science (CoLIS 4)* (pp. 285-300). Libraries Unlimited.

Järvelin, K., & Vakkari, P. (2022). LIS research across 50 years: Content analysis of journal articles. *Journal of Documentation,* 78(7), 65-88. https://doi.org/10.1108/JD-03-2021-0062

Jansz, C. N. M., & le Pair, C. (1991). Die bibliometrische Unsichtbarkeit des technischen Fortschritts. In P. Weingart, R. Sehringer, & M. Winterhager (Eds.), *Indikatoren der Wissenschaft und Technik: Theorie, Methoden, Anwendungen* (pp. 209-223). Campus-Verlag. German.

Jesenko, B., & Schlögl, C. (2021). The effect of web of science subject categories on clustering: The case of data-driven methods in business and economic sciences. *Scientometrics,* 126(8), 6785-6801. https://doi.org/10.1007/s11192-021-04060-4

Jin, B., & Wang, B. (1999). Chinese science citation database: Its construction and application. *Scientometrics,* 45(2), 325-332. https://doi.org/10.1007/BF02458440

Johansson, M. A., Reich, N. G., Meyers, L. A., & Lipsitch, M. (2018). Preprints: An underutilized mechanism to accelerate outbreak science. *PLoS Medicine,* 15(4), e1002549.

https://doi.org/10.1371/journal.pmed.1002549

Kang, D., & Evans, J. (2020). Against method: Exploding the boundary between qualitative and quantitative studies of science. *Quantitative Science Studies,* 1(3), 930-944. https://doi.org/10.1162/qss_a_00056

Kamrani, P., Dorsch, I., & Stock, W. G. (2020). Publikationen, Zitationen und H-index im Meinungsbild Deutscher Universitätsprofessoren. *Beiträge zur Hochschulforschung,* 42(3), 78-98. German.

Kamrani, P., Dorsch, I., & Stock, W. G. (2021). Do researchers know what the h-index is? And how do they estimate its importance? *Scientometrics,* 126(7), 5489-5508. https://doi.org/10.1007/s11192-021-03968-1

Katz, E., Blumler, J. G., & Gurevitch, M. (1973-1974). Uses and gratifications research. *Public Opinion Quarterly,* 37(4), 509-523. https://doi.org/10.1086/268109

Kassian, A., & Melikhova, L. (2019). Russian Science Citation Index on the WoS platform: A critical assessment. *Journal of Documentation,* 75(5), 1162-1168. https://doi.org/10.1108/JD-02-2019-0033

Keng, S. T., Ooi, P. B., Pang, K. Y., & Phan, C. W. (2022). Awareness, familiarity and usage of traditional research metrics and altmetrics among academics in Malaysia. *Malaysian Journal of Library and Information Science,* 27(3), 1-19. https://doi.org/10.22452/mjlis.vol27no3.1

Kleminski, R., Kazienko, P., & Kajdanowicz, T. (2022). Analysis of direct citation, co-citation and bibliographic coupling in scientific topic identification. *Journal of Information Science,* 48(3), 349-373. https://doi.org/10.1177/0165551520962775

Ko, Y. M., Cho, S. R., & Park, Y. S. (2011). A study on the optimization of KCI-based index (Kor-Factor) in evaluating Korean journals. *Scientometrics,* 88(1), 61-71. https://doi.org/10.1007/s11192-011-0384-z

Kostoff, R. N., & Martinez, W. L. (2005). Is citation normalization realistic? *Journal of Information Science,* 31(1), 57-61. https://doi.org/10.1177/0165551505049260

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

Kutlača, D., Babić, D., Živković, L., & Štrbac, D. (2015). Analysis of quantitative and qualitative indicators of SEE countries scientific output. *Scientometrics,* 102(1), 247-265. https://doi.org/10.1007/s11192-014-1290-y

Lakatos, I. (1974). Science and pseudoscience. *Conceptus,* 8, 5-9.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature,* 504(7479), 211-213. https://doi.org/10.1038/504211a

Larivière, V., Pontille, D., & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies,* 2(1), 111-128. doi: https://doi.org/10.1162/qss_a_00097

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37-51). Harper & Brothers.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology,* 64(1), 2-17. https://doi.org/10.1002/asi.22784

Leydesdorff, L. (1995). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications.* DSWO Press.

Leydesdorff, L., Ràfols, I., & Milojević, S. (2020). Bridging the divide between qualitative and quantitative science studies. *Quantitative Science Studies,* 1(3), 918-926. https://doi.org/10.1162/qss_e_00061

Leydesdorff, L., & Shin, J. C. (2011). How to evaluate universities in terms of their relative citation impacts: Fractional counting of citations and the normalization of differences among disciplines. *Journal of the American Society for Information Science and Technology,* 62(6), 1146-1155. https://doi.org/10.1002/asi.21511

Leydesdorff, L., & Welbers, K. (2011). The semantic mapping of words and co-words in contexts. *Journal of Informetrics,* 5(3), 469-475. https://doi.org/10.1016/j.joi.2011.01.008

Li, M. (2018). Classifying and ranking topic terms based on a novel approach: Role differentiation of author keywords. *Scientometrics,* 116(1), 77-100. https://doi.org/10.1007/s11192-018-2741-7

Lin, J., & Fenner, M. (2013). Altmetrics in evolution: Defining & redefining the ontology of article-level metrics. *Information Standards Quarterly,* 25(2), 20-26. http://doi.org/10.3789/isqv25no2.2013.04

Liu, J., & Adie, E. (2013). Five challenges in altmetrics: A toolmaker's perspective. *Bulletin of the American Society for Information Science and Technology,* 39(4), 31-34. https://doi.org/10.1002/bult.2013.1720390410

Liu, W. (2019). The data source of this study is Web of Science Core Collection? Not enough. *Scientometrics,* 121(3), 1815-1824. https://doi.org/10.1007/s11192-019-03238-1

Lundberg, G. D. (1993). Statement by the International Committee of Medical Journal Editors on duplicate or redundant publication. *Journal of the American Medical Association,* 270(20), 2495. https://doi.org/10.1001/jama.1993.03510200101040

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing & Management,*

44(2), 800-810. https://doi.org/10.1016/j.ipm.2007.06.006

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science,* 40(5), 342-349. https://doi.org/10.1002/(SICI)1097-4571(198909)40:5%3C342::AID-ASI7%3E3.0.CO;2-U

MacRoberts, M. H., & MacRoberts, B. R. (2018). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology,* 69(3), 474-482. https://doi.org/10.1002/asi.23970

Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science,* 28(1), 38-43. https://doi.org/10.1002/asi.4630280107

Marsh, R. M. (2015). The role of institutional repositories in developing the communication of scholarly research. *OCLC Systems & Services: International Digital Library Perspectives,* 31(4), 163-195. https://doi.org/10.1108/OCLC-04-2014-0022

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics,* 126(1), 871-906. https://doi.org/10.1007/s11192-020-03690-4

McKiernan, G. (2000). arXiv.org: The Los Alamos National Laboratory e-print server. *International Journal on Grey Literature,* 1(3), 127-138. https://doi.org/10.1108/14666180010345564

Martyn, C. (2005). The malign influence of impact factors. *Journal of the Royal Society of Medicine,* 98(12), 556. https://doi.org/10.1177/014107680509801221

Massarani, L., & Neves, L. F. F. (2022). Reporting COVID-19 preprints: Fast science in newspapers in the United States, the United Kingdom and Brazil. *Ciencia & Saude Coletiva,* 27(3), 957-968. https://doi.org/10.1590/1413-81232022273.20512021

Meschede, C., & Siebenlist, T. (2018). Cross-metric compatability and inconsistencies of altmetrics. *Scientometrics,* 115(1), 283-297. https://doi.org/10.1007/s11192-018-2674-1

Migheli, M., & Ramello, G. B. (2021). The unbearable lightness of scientometric indices. *Managerial and Decision Economics,* 42(8), 1933-1944. https://doi.org/10.1002/mde.3486

Mikhailov, O. V. (2021). Introduction from the guest editor of special issue "Modern problems of scientometric assessment of publication activity". *Publications,* 9(2), 19. https://doi.org/10.3390/publications9020019

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research,* 246(1), 1-19. https://doi.org/10.1016/j.ejor.2015.04.002

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics,* 106(1), 213-228. https://doi.org/10.1007/s11192-015-1765-5

Moussa, S. (2021). Citation contagion: A citation analysis of selected predatory marketing journals. *Scientometrics,* 126(1), 485-506. https://doi.org/10.1007/s11192-020-03729-6

Mousseau, M. C. (2003). Parapsychology: Science or pseudoscience? *Journal of Scientific Exploration,* 17(2), 271-282. https://www.scientificexploration.org/docs/17/jse_17_2_mousseau.pdf

Narin, F., Olivastro, D., & Stevens, K. A. (1994). Bibliometrics/theory, practice and problems. *Evaluation Review,* 18(1), 65-76. https://doi.org/10.1177/0193841X9401800107

Neuhäusler, P., & Frietsch, R. (2013). Patent families as macro level patent value indicators: Applying weights to account for market differences. *Scientometrics,* 96(1), 27-49. https://doi.org/10.1007/s11192-012-0870-y

Neurath, O. (1932). Protokollsätze. *Erkenntnis,* 3(1), 204-214. German. https://doi.org/10.1007/BF01886420

Ni, C., Smith, E., Yuan, H., Larivière, V., & Sugimoto, C. R. (2021). The gendered nature of authorship. *Science Advances,* 7(36), eabe4639. https://www.science.org/doi/10.1126/sciadv.abe4639

Ninkov, A., Gregory, K., Peters, I., & Haustein, S. (2021, April 30). Datasets on DataCite - An initial bibliometric investigation. *Proceeding of the 18th International Conference on Scientometrics & Informetrics (ISSI 2021)* (pp. 1-6). Zenodo.

Organisation for Economic Co-operation and Development (OECD). (2015). *Frascati manual 2015: Guidelines for collecting and reporting data on research and experimental development.* OECD.

Olcay, G. A., & Bulu, M. (2017). Is measuring the knowledge creation of universities possible?: A review of university rankings. *Technological Forecasting and Social Change,* 123, 153-160. https://doi.org/10.1016/j.techfore.2016.03.029

Osório, A. (2018). On the impossibility of a perfect counting method to allocate the credits of multi-authored publications. *Scientometrics,* 116(3), 2161-2173. https://doi.org/10.1007/s11192-018-2815-6

Packer, A. L. (2020). The pasts, presents, and futures of SciELO. In M. P. Eve, & J. Gray (Eds.), *Reassembling scholarly communications: Histories, infrastructures, and global politics of open access* (pp. 297-313). MIT Press.

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. (2016). Research data explored: An extended analysis of citations and altmetrics. *Scientometrics,* 107(2), 723-744.

https://doi.org/10.1007/s11192-016-1887-4

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. I. (2017). Zenodo in the spotlight of traditional and new metrics. *Frontiers in Research Metrics and Analytics,* 2, 13. https://doi.org/10.3389/frma.2017.00013

Poppema, S. (2020). Citations are the currency of science: Then there are the counterfeiters. *The Scientist,* 34(12), 14-15. https://www.the-scientist.com/critic-at-large/citations-are-the-currency-of-science-68204

Popper, K. R. (1935). *Logik der Forschung: Zur Erkenntnistheorie der Modernen Naturwissenschaft.* Julius Springer Verlag. German.

Popper, K. R. (1962). *Conjectures and refutations: The growth of scientific knowledge.* Basic Books.

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. *Publications,* 9(1), 12. https://doi.org/10.3390/publications9010012

Priem, J. (2010). *I like the term #articlelevelmetrics, but it fails to imply *diversity* of measures. Lately, I'm liking #altmetrics.* https://twitter.com/jasonpriem/status/25844968813

Priem, J. (2014). Altmetrics. In B. Cronin, & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 263-287). MIT Press.

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv. https://doi.org/10.48550/arXiv.2205.01833*

Quay, P. M. (1974). Progress as a demarcation criterion for the sciences. *Philosophy of Science,* 41(2), 154-170. http://www.jstor.org/stable/186865

Rahman, M. T., Regenstein, J. M., Kassim, N. L. A., & Haque, N. (2017). The need to quantify authors' relative intellectual contributions in a multi-author paper. *Journal of Informetrics,* 11(1), 275-281. https://doi.org/10.1016/j.joi.2017.01.002

Raman, V. V. (2001). Science and religion: Some demarcation criteria. *Zygon,* 36(3), 541-556. https://doi.org/10.1111/0591-2385.00380

Rauter, J. (2005, September 13-16). Wissenschaftliche Forschung und ihre Evaluation: Expansive Zitationsanalyse auf deskriptiv-intertextueller Basis. In E. Pipp (Ed.), *Proceedings of the 11th Austrian Online Information Meeting, 12th Austrian Documentary Day (ODOK '05)* (pp. 160-184). Wolfgang Neugebauer Verlag. German.

Refinetti, R. (1990). In defense of the least publishable unit. *FASEB Journal,* 4(1), 128-129. https://doi.org/10.1096/fasebj.4.1.2295373

Reichmann, G. (2020). Was zählt eine Publikation? Möglichkeiten zur Durchführung einer Forschungsevaluierung im Bereich der Betriebswirtschaftslehre. In I. Dorsch, K. Fietkiewicz, A. Ilhan, C. Meschede, & T. Siebenlist (Eds.), *Facetten von Wolf Stock und ihre Bedeutung für die Informationswissenschaft: Festschrift zu Ehren von Wolfgang G. Stock, Schriften zur Informationswissenschaft* (pp. 15-34). Verlag Werner Hülsbusch. German.

Reichmann, G., & Schlögl, C. (2021). Möglichkeiten zur Steuerung der Ergebnisse einer Forschungsevaluation: Ein Vergleich der Forschungsleistung zweier informationswissenschaftlicher Institute. *Information - Wissenschaft & Praxis,* 72(4), 212-220. German. https://doi.org/10.1515/iwp-2021-2148

Reichmann, G., & Schlögl, C. (2022). On the possibilities of presenting the research performance of an institute over a long period of time: The case of the Institute of Information Science at the University of Graz in Austria. *Scientometrics,* 127(6), 3193-3223. https://doi.org/10.1007/s11192-022-04377-8

Reichmann, G., Schlögl, C., Stock, W. G., & Dorsch, I. (2022). Forschungsevaluation auf Institutsebene – Der Einfluss der gewählten Methodik auf die Ergebnisse. *Beiträge zur Hochschulforschung,* 44(1), 74-97. German. https://www.bzh.bayern.de/fileadmin/user_upload/Publikationen/Beitraege_zur_Hochschulforschung/2022/2022-1-Reichmann-Schloegl-Stock-Dorsch.pdf

Riegelman, A. (2022). Research square. *Charleston Advisor,* 24(2), 43-45. https://doi.org/10.5260/chara.24.2.43

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology,* 67(12), 2964-2975. https://doi.org/10.1002/asi.23529

Roth, J. (1981). Least publishable unit. *Science,* 212(4494), 494. https://doi.org/10.1126/science.212.4494.494.b

Röbken, H. (2011). Forschungsproduktivität von Wissenschaftlern und Wissenschaftlerinnen – Eine empirische Analyse von Publikationsaktivitäten vor und nach der Berufung. *Beiträge zur Hochschulforschung,* 33(3), 62-81. German. https://www.bzh.bayern.de/fileadmin/news_import/3-2011-roebken_01.pdf

Rousseau, R., & Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers.* Chandos.

Scanlan, B. D. (1988). Coverage by *current contents* and the validity of impact factors. *Serials Librarian,* 13(2-3), 57-66. https://doi.org/10.1300/J123v13n02_07

Schlögl, C. (2013). Internationale Sichtbarkeit der europäischen und insbesondere der deutschsprachigen Informationswissenschaft [International visibility of the European informa-

tion science, especially in German language]. *Information - Wissenschaft & Praxis,* 64(1), 1-8. German. https://doi.org/10.1515/iwp-2013-0001

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics,* 101(2), 1113-1128. https://doi.org/10.1007/s11192-014-1365-9

Schlögl, C., & Stock, W. G. (2008). Practitioners and academics as authors and readers: The case of LIS journals. *Journal of Documentation,* 64(5), 643-666. https://doi.org/10.1108/00220410810899691

Schneijderberg, C. (2018). *Promovieren in den Sozialwissenschaften: Eine sozialisationstheoretische Erschließung des Forschungsfeldes Promotion.* Springer. German.

Schrader, U., & Hennig-Thurau, T. (2009). VHB-JOURQUAL2: Method, results, and implications of the German Academic Association for Business Research's journal ranking. *BuR Business Research Journal,* 2(2), 180-204. https://ssrn.com/abstract=1568651

Schubert, A., Glänzel, W., & Thijs, B. (2006). The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics,* 67(3), 503-514. https://doi.org/10.1556/Scient.67.2006.3.11

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science,* 43(9), 628-638. https://doi.org/10.1002/(SICI)1097-4571(199210)43:9%3C628::AID-ASI5%3E3.0.CO;2-0

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ,* 314(7079), 498-502. https://doi.org/10.1136/bmj.314.7079.497

Shema, H., Hahn, O., Mazarakis, A., & Peters, I. (2019). Retractions from altmetric and bibliometric perspectives. *Information - Wissenschaft & Praxis,* 70(2-3), 98-110. https://doi.org/10.1515/iwp-2019-2006

Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences,* 111(34), 12325-12330. https://doi.org/10.1073/pnas.1401992111

Shu, F., Qiu, J., & Larivière, V. (2021). Mapping the biomedical sciences using Medical Subject Headings: A comparison between MeSH co-assignments and MeSH citation pairs. *Journal of the Medical Library Association,* 109(3), 441-449. https://doi.org/10.5195/jmla.2021.1173

Sīle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., Dušková, M., Faurbæk, L., Holl, A., Kulczycki, E., Macan, B., Nelhans, G., Petr, M., Pisk, M., Soós, S., Stojanovski, J., Stone, A., Šušol, J., & Teitelbaum, R. (2018). Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey. *Research Evaluation,* 27(4), 310-322. https://doi.org/10.1093/reseval/rvy016

Sjögårde, P., & Didegah, F. (2022). The association between topic growth and citation impact of research publications. *Scientometrics,* 127(4), 1903-1921. https://doi.org/10.1007/s11192-022-04293-x

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine,* 99(4), 178-182. https://doi.org/10.1177/014107680609900414

Stock, M., & Stock, W. G. (2006). Intellectual property information: A comparative analysis of main information providers. *Journal of the American Society for Information Science and Technology,* 57(13), 1794-1803. https://doi.org/10.1002/asi.20498

Stock, W. G. (1984). Informetrische Untersuchungsmethoden auf der Grundlage der Textwortmethode. *International Classification,* 11(1984), 151-157. German. https://www.phil-fak.uni-duesseldorf.de/publikationen/publikationen/publikationen-1990-und-frueher/019-informetrische-untersuchungsmethoden-auf-der-grundlage-der-textwortmethode/

Stock, W. G. (1985). Die Bedeutung der Zitatenanalyse für die Wissenschaftsforschung. *Zeitschrift für Allgemeine Wissenschaftstheorie,* 16(2), 304-314. German. https://doi.org/10.1007/BF01803677

Stock, W. G. (1990). Themenanalytische informetrische Methoden: Vorbemerkungen zur Methode. In M. Stock, & W. G. Stock (Eds.), *Psychologie und Philosophie der Grazer Schule* (pp. 7-31). Rodopi. German.

Stock, W. G. (1994). *Wissenschaftsevaluation: Die Bewertung wissenschaftlicher Forschung und Lehre.* IFO Institut für Wirtschaftsforschung. German.

Stock, W. G. (2000). Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In K. Fuchs-Kittowski, H. Laitko, H, Parthey, & W. Umstätter (Eds.), *Wissenschaft und Digitale Bibliothek: Wissenschaftsforschung Jahrbuch 1998* (pp. 239-282). Gesellschaft für Wissenschaftsforschung. German.

Stock, W. G. (2001). *Publikation und Zitat: Die problematische Basis empirischer Wissenschaftsforschung* (Band 29). Fachhochschule Köln. German.

Stock, W. G. (2021). Die Teildatenbanken von Web of Science und deren Subskriptionen von Wissenschaftseinrichtungen: Auswirkungen auf Forschungsevaluation und Sichtbarkeit von Forschenden. *Zeitschrift für Bibliothekswesen und Bibliographie,* 68(3), 152-157. German. https://doi.org/10.3196/186429502068331

Stock, W. G., Dorsch, I., Reichmann, G., & Schlögl, C. (2023). Labor productivity, labor impact, and co-authorship of research institutions: Publications and citations per full-time equivalents. *Scientometrics,* 128(1), 363-377. https://doi.org/10.1007/s11192-022-04582-5

Stock, W. G., & Weber, S. (2006). Facets of informetrics. *Information - Wissenschaft & Praxis,* 57(8), 385-389. https://www.ling.hhu.de/fileadmin/redaktion/Fakultaeten/Philosophische_Fakultaet/Sprache_und_Information/Informationswissenschaft/Dateien/Wolfgang_G._Stock/Daten_vor_2011/1166781846sonjaweber.pdf

Stock, W. G., & Stock, M. (2013). *Handbook of information science.* De Gruyter Saur.

Strcic, J., Civljak, A., Glozinic, T., Pacheco, R. L., Brkovic, T., & Puljak, L. (2022). Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv. *Scientometrics,* 127(5), 2791-2802. https://doi.org/10.1007/s11192-022-04346-1

Strinzel, M., Severin, A., Milzow, K., & Egger, M. (2019). Blacklists and whitelists to tackle predatory publishing: A cross-sectional comparison and thematic analysis. *mBio,* 10(3), e00411-e00419. https://doi.org/10.1128/mBio.00411-19

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology,* 68(9), 2037-2062. https://doi.org/10.1002/asi.23833

Szomszor, M., & Adie, E. (2022). Overton: A bibliometric database of policy document citations. *Quantitative Science Studies,* 3(3), 624-650. https://doi.org/10.1162/qss_a_00204

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing & Management,* 28(1), 1-3. https://doi.org/10.1016/0306-4573(92)90087-G

Teixeira da Silva, J. A., & Dobránszki, J. (2016). Multiple authorship in scientific manuscripts: Ethical challenges, ghost and guest/gift authorship, and the cultural/disciplinary perspective. *Science and Engineering Ethics,* 22(5), 1457-1472. https://doi.org/10.1007/s11948-015-9716-3

Tennant, J., Bauin, S., James, S., & Kant, J. (2018). The evolving preprint landscape: Introductory report for the knowledge exchange working group on preprints. *MetaArXiv.* https://doi.org/10.31222/osf.io/796tu

Thagard, P. R. (1978). Why astrology is a pseudoscience. *Proceedings of the Biennial Meeting of the Philosophy of Science Association,* 1978, 223-234. http://www.jstor.org/stable/192639

Thelwall, M. (2021). Measuring societal impacts of research with altmetrics? Common problems and mistakes. *Journal of Economic Surveys,* 35(5), 1302-1314. https://doi.org/10.1111/joes.12381

Thelwall, M., & Fairclough, R. (2017). The research production of nations and departments: A statistical model for the share of publications. *Journal of Informetrics,* 11(4), 1142-1157. https://doi.org/10.1016/j.joi.2017.10.001

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS One,* 8(5), e64841. https://doi.org/10.1371/journal.pone.0064841

Tillett, B. B. (2001). Bibliographic relationships. In C. A. Bean, & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 19-35). Springer.

Tokar, A. (2012). Semantic change of the publication-concept? In A. Tokar, M. Beurskens, S. Keuneke, M. Mahrt, I. Peters, C. Puschmann, T. van Treeck, & K. Weller (Eds.), *Science and the Internet* (pp. 193-205). Düsseldorf University Press.

Toutkoushian, R. K., Porter, S. R., Danielson, C., & Hollis, P. R. (2003). Using publications counts to measure an institution's research productivity. *Research in Higher Education,* 44(2), 121-148. https://doi.org/10.1023/A:1022070227966

Tüür-Fröhlich, T. (2016). *The non-trivial effects of trivial errors in scientific communication and evaluation.* Verlag Werner Hülsbusch.

Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature,* 465(7300), 864-866. https://doi.org/10.1038/465864a

van Raan, A. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics,* 62(1), 133-143. https://doi.org/10.1007/s11192-005-0008-6

Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics,* 121(3), 1803-1813. https://doi.org/10.1007/s11192-019-03264-z

Verleysen, F. T., & Engels, T. C. E. (2018). How arbitrary are the weights assigned to books in performance-based research funding? An empirical assessment of the weight and size of monographs in Flanders. *ASLIB Journal of Information Management,* 70(6), 660-672. https://doi.org/10.1108/AJIM-05-2018-0110

Vílchez-Román, C. (2017). BRICS' scientific excellence and the search for relevance and replicability. *Journal of the Association for Information Science and Technology,* 68(2), 539. https://doi.org/10.1002/asi.23654

Waltman, L., & van Eck, N. J. (2019). Field normalization of scientometric indicators. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 281-300). Springer.

Waltman, L., van Eck, N. J., Visser, M., & Wouters, P. (2016). The elephant in the room: The problem of quantify-

ing productivity in evaluative scientometrics. *Journal of Informetrics,* 10(2), 671-674. https://doi.org/10.1016/j.joi.2015.12.008

Wang, K., Shen, Z., Huang, C., Wu, C. H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies,* 1(1), 396-413. https://doi.org/10.1162/qss_a_00021

Watson, C. (2022). Rise of the preprint: How rapid data sharing during COVID-19 has changed science forever. *Nature Medicine,* 28(1), 2-5. https://doi.org/10.1038/s41591-021-01654-6

Weingart, P., & Winterhager, M. (1984). *Die Vermessung der Forschung. Theorie und Praxis der Wissenschaftsindikatoren.* Campus. German.

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLoS One,* 13(5), e0197326. https://doi.org/10.1371/journal.pone.0197326

Zhao, D., Cappello, A., & Johnston, L. (2017). Functions of uni- and multi-citations: Implications for weighted citation analysis. *Journal of Data and Information Science,* 2(1), 51-69. https://doi.org/10.1515/jdis-2017-0003

Zimmer, F., Scheibe, K., & Stock, W. G. (2018, July 15-20). A model for information behavior research on social live streaming services (SLSSs). In G. Meiselwitz (Ed.), *Proceedings of the 10th International Conference, SCSM 2018: Social Computing and Social Media. Technologies and Analytics* (pp. 429-448). Springer.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics,* 63(2), 373-401. https://doi.org/10.1007/s11192-005-0218-y

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 11(2): 67-81, 2023

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2023.11.2.5

A Data Mining Approach for a Dynamic Development of an Ontology-Based Statistical Information System

Mohamed Hachem Kermani*
National Polytechnic School of Constantine-Malek Bennabi, Constantine, Algeria
Laboratoire d'Informatique Repartie (LIRE) Laboratory, Constantine, Algeria
E-mail: hachem.kermani@enp-constantine.dz

Zizette Boufaida
University of Constantine 2 - Abdelhamid Mehri, Constantine, Algeria
Laboratoire d'Informatique Repartie (LIRE) Laboratory, Constantine, Algeria
E-mail: zizette.boufaida@univ-constantine2.dz

Amel Lina Bensabbane
University of Constantine 2 - Abdelhamid Mehri, Constantine, Algeria
Laboratoire d'Informatique Repartie (LIRE) Laboratory, Constantine, Algeria
E-mail: amel.bensabbane@univ-constantine2.dz

Besma Bourezg
University of Constantine 2 - Abdelhamid Mehri, Constantine, Algeria
Laboratoire d'Informatique Repartie (LIRE) Laboratory, Constantine, Algeria
E-mail: besma.bourezg@univ-constantine2.dz

ABSTRACT

This paper presents a dynamic development of an ontology-based statistical information system supporting the collection, storage, processing, analysis, and the presentation of statistical knowledge at the national scale. To accomplish this, we propose a data mining technique to dynamically collect data relating to citizens from publicly available data sources; the collected data will then be structured, classified, categorized, and integrated into an ontology. Moreover, an intelligent platform is proposed in order to generate quantitative and qualitative statistical information based on the knowledge stored in the ontology. The main aims of our proposed system are to digitize administrative tasks and to provide reliable statistical information to governmental, economic, and social actors. The authorities will use the ontology-based statistical information system for strategic decision-making as it easily collects, produces, analyzes, and provides both quantitative and qualitative knowledge that will help to improve the administration and management of national political, social, and economic life.

Keywords: statistical information, data mining, dynamic development, ontology-based information system, statistical information system, strategic decision-making

Received: May 1, 2022 **Revised:** March 3, 2023
Accepted: March 9, 2023 **Published:** June 30, 2023

***Corresponding Author:** Mohamed Hachem Kermani
https://orcid.org/0000-0002-2315-6951
E-mail: hachem.kermani@enp-constantine.dz

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© Mohamed Hachem Kermani, Zizette Boufaida, Amel Lina Bensabbane, Besma Bourezg, 2023

1. INTRODUCTION

The development of ontology-based information systems is argued when applied to the analysis, conceptual modeling, design, and re-engineering of complex information systems. Ontologies serve as a theoretical foundation for the conceptual model, which is used to design and develop the entire information system. Ontologies enable specialists, software agents, systems, and services to share a common understanding of a domain. Ontologies have also influenced research and development in other areas of computational intelligence, resulting in a plethora of hybrid information systems. They have introduced a new way of thinking about, researching, and developing information systems, which, in conjunction with other technologies, has resulted in the emergence and development of intelligent information systems, such as statistical information systems.

Indeed, a system that provides statistical information aids in the preparation, execution, and assessment of decisions and actions pertaining to a complete or partial system, such as a community and its inhabitants, establishments and their operations, or a company and its staff, customers, and vendors with their activities. The statistical information system is intended to acquire, retain, handle, scrutinize, display, and obtain information. These responsibilities are comparable to those of data mining, which refers to the examination of a vast pre-existing database to create new insights. A variety of data mining techniques have been formulated and utilized in recent times, including association, classification, clustering, decision tree, prediction, and Neural Networks (Kermani & Boufaida, 2022).

Obviously, these methodologies are interconnected with the operation of ontologies, and it can be challenging to differentiate the boundary between data mining and ontology. According to Plirdpring and Ruangrajitpakorn (2022), an ontology is a precise and explicit account of the concepts within a specific domain, including the properties and attributes associated with each concept. The collection of individual instances of these concepts with an ontology constitutes a knowledge base. This research presents a data mining strategy for the real-time construction of an ontology-based statistical information system that arranges, sorts, and categorizes information concerning citizens, enabling government, economic, and social authorities to access accurate statistical data.

To accomplish this, we must 1) dynamically collect (i.e. automatically and continuously) all data pertaining to citizens from publicly available data sources (i.e. civil status files, government databases, etc.); 2) structure, classify, and categorize knowledge about citizens before dynamically integrating it into the ontology; and 3) develop an intelligent ontology exploitation platform that provides authorities with quantitative and qualitative information about the country's social and economic situation. The rest of this paper is organized according to the following. Section 2 provides an overview of research that is related to our approach. Section 3 presents our proposal, which is a data mining approach for the dynamic development of an ontology-based statistical information system. Section 4 presents a software application and experimentation of the proposed approach. Section 5 presents a discussion. Finally, Section 6 concludes the paper and suggests directions for future research.

2. RELATED WORK

In our proposed approach, we investigate data mining techniques, ontologies, and statistical information systems development. These aspects have been the subject of numerous works in the literature, which we will discuss below.

2.1. Statistics Data Mining Approaches

The discipline of data mining encompasses computer science and statistics, and seeks to intelligently extract information from data sets and transform it into a comprehensible format for further use. Various statistical data mining methods have been proposed for this purpose, such as the approach described in a study by Baek et al. (2018). This study used electronic health records to analyze event logs of patients admitted to a South Korean university hospital between January and December 2013, with the aim of identifying factors affecting the length of hospital stays. The authors aimed to develop a system that could assist hospitals in managing inpatient admissions more effectively.

Furthermore, in a study conducted by Fernandes et al. (2019), a data mining approach was used to predict the academic performance of public school students in the capital city of Brazil. The authors initially performed a descriptive statistical analysis to gain insights from the data, which resulted in the creation of two datasets. The first dataset consisted of variables obtained before the start of the school year, while the second contained academic variables collected two months after the semester began. The authors developed classification models based on Gradient Boosting Machine for each dataset, to predict the academic outcomes of student performance at the end of the

 https://doi.org/10.1633/JISTaP.2023.11.2.5

school year. The study revealed that 'grades' and 'absences' were the most important attributes for predicting end-of-year academic outcomes, but demographic attributes such as 'neighborhood,' 'school,' and 'age' also showed potential as indicators of academic success or failure.

In the same context, another study conducted by Adekitan and Noma-Osaghae (2019) introduced a data mining technique to predict the academic performance of first-year university students. The research examined the correlation between the academic records of students admitted into Covenant University in Nigeria and their performance during their first year using predictive data mining and regression models. The data mining model was applied on KNIME and Orange platforms, and the accuracy of the prediction was verified using regression analysis. The study recorded maximum accuracies of 50.23% and 51.9%, respectively, with R2 values of 0.207 and 0.232, indicating that students' cognitive entry requirements do not entirely explain their performance during the first year.

Besides this, a research paper authored by Jatav (2018) presented an algorithm for medical diagnosis using predictive data mining. The study focused on predicting diabetes, kidney, and liver disease using a vast number of input attributes. To evaluate the performance of the algorithm, data mining classification techniques such as support vector machine (SVM) and random forest (RF) were used on respective databases for each disease. The study compared the performance of these techniques using metrics such as precision, recall, accuracy, f-measure, and time. The experimental results showed an accuracy of 99.35%, 99.37%, and 99.14% on diabetes, kidney disease, and liver disease, respectively, and concluded that the proposed algorithm was created using the SVM and RF algorithms.

2.2. Approaches for Developing Ontology-Based Information Systems

According to Sowa (2011), all software systems have an implicit or explicit ontology, which means they work with a knowledge base that can be represented using ontologies. Indeed, Nawi et al. (2021) suggest that designing an information system based on a formal ontology is possible, and many works on ontology-based information systems have been proposed in recent years. For instance, Luković et al. (2019) introduced an ontology-based module for the information system ScolioMedIS that uses ontologies for 3D digital diagnosis of adolescent scoliosis. The researchers followed four steps, namely specification, conceptualization, formalization, and implementation, to develop the OBR-Scolio ontology and the ontology-based module of the ScolioMedIS. They used the Protégé-OWL API to create, edit, delete, and query the ontology. The module was tested on datasets of 20 female and 15 male patients with AIS, and it automatically generated statistical indicators about the frequency and characteristics of spinal curvatures based on the Lenke classification system and Lenke scoliosis types. The results were compared with the analysis of 315 observed patients performed using traditional radiation techniques.

Furthermore, a study by Comas Rodríguez et al. (2019) presented a data management model called OntoSIGOBE that is applied to the Geographic Information System SIGOBE, which is a part of the Business Management System of the Cuban Electric Union (SIGE). This model utilizes a developed domain ontology and a query answering process based on case-based reasoning technique to achieve semantic interoperability between heterogeneous data sources and the query answering process. The results of the study showed that OntoSIGOBE provides flexible and integrated data access in SIGOBE and is highly satisfactory to end users. Similarly, Ledvinka et al. (2019) presented an ontology-based information system for aviation safety data integration that was developed in cooperation with the Civil Aviation Authority of the Czech Republic. The study described the system's design, core technologies, and achieved functionalities, and compared it to other available solutions in the domain. The results showed that by linking the features to ontologies, the system achieved many desired characteristics of a safety information system, providing a systematic and sustainable way to improve aviation safety.

Besides this, Choi and Choi (2019) put forward a security context reasoning approach based on ontology for power IoT-Cloud security service, which presented a suitable framework for power IoT-Cloud environment security service. Furthermore, an effective security mechanism was implemented and tested using attack context scenarios based on a smart meter, a critical component of power systems. Inference rules were created for each attack stage to identify attack paths that exploited smart meter vulnerabilities, and the results showed a high level of attack detection. While most literature focuses on data mining and ontology-based information systems for statistical purposes, we propose a data mining approach for a dynamic ontology-based statistical information system to provide reliable quantitative and qualitative information on citizens to governmental, economic, and social author-

ities, thereby facilitating a comprehensive understanding of any country's socio-economic situation.

3. THE PROPOSED APPROACH

Statistics is the science of learning from data by collecting and analyzing information in order to effectively present the results. Statistics is an important part of how we make scientific discoveries, make data-driven decisions, and make predictions. Statistics are useful when used to improve management and decision making. Therefore, recent research has focused on employing computational techniques from statistics, data mining, and information theory to automatically collect and analyze massive amounts of data. Our proposed data mining approach for the dynamic development of an ontology-based statistical information system (Fig. 1) is part of these research efforts to provide governmental, economic, and social authorities with reliable statistical data in order to improve management and decision making. The proposed approach is based on the Model View Controller (MVC) architecture, with the *Controller* representing a data mining algorithm that dynamically (i.e. automatically and continuously) collects all data pertaining to citizens from publicly available data sources (i.e. civil status files, government databases, etc.), and then structures, classifies, and categorizes these knowledge into the ontology. The *Model* is a national ontology that integrates all knowledge about citizens, and the *View* is an intelligent ontology exploitation platform that provides authorities with quantitative and qualitative data about the country's social and economic situation.

Our approach to developing a statistical information system based on an ontology uses a data mining method. This method is based on the MVC pattern, an architectural framework that breaks down an application into three major conceptual components: the Model, the View, and the Controller. While originally designed for desktop graphical user interface applications, this pattern gained popularity with web apps and is now widely supported across common languages (Kermani et al., 2021). Our approach follows this pattern by dividing each component according to its adherence to the MVC pattern. The Model, which serves as the central component of our system, is the national ontology. The View, which represents the output of knowledge, is an intelligent ontology exploitation platform. The Controller is a data mining algorithm that dynamically collects, structures, and categorizes citizen knowledge into an ontology. More details on each of the three components of our system are presented in the following sections.

3.1. The Data Mining Algorithm

In the MVC pattern, the controller plays a vital role as it oversees the decision-making process, system logic, and serves as a mediator between the model and the view. In our suggested MVC framework, the data mining algorithm fulfills the controller's duties. It is the critical element of our system that gathers, arranges, and groups all citizen information into the ontology.

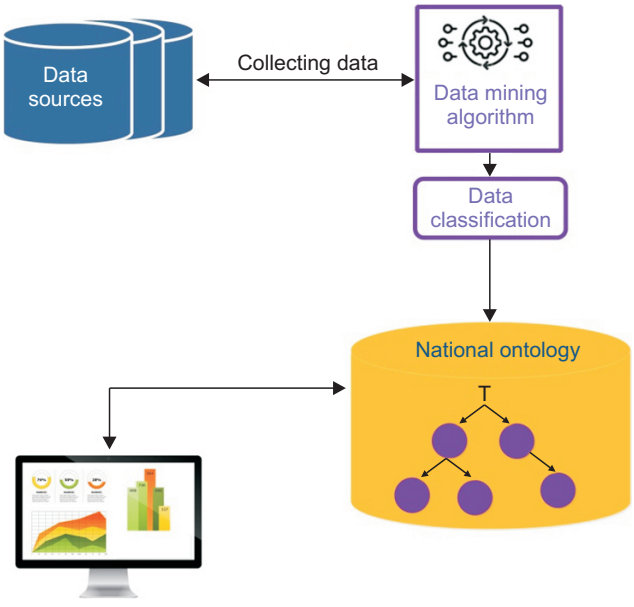


Fig. 1. The ontology-based statistical information system.

The data mining algorithm

Input: CSF *// Civil Statutes Files.*
Input: ADB *// All Data Bases.*
Begin
ID = 1
While (ID ≠ 0) **Do**
Locate (ID, CSF) *// Select ID from Civil Statutes Files.*
Recuperate Data ( ) *// Retrieve all data related to the selected ID.*
Search (ID, CSF) *// Searching ID in CSF where ID is a foreign key.*
If (ID.Found = = 'True') **Then**
Recuperate DFK ( ) *// Collect all data where ID is a Foreign Key From CSF.*
End
Data Cleaning ( ) *// Correcting errors in the recovered data.*
Create Generic Classes ( ) *// Categorize the collected data.*
Search (ID, ADB) *// Searching in ADB where ID is a Foreign Key.*
If (ID.Found = = 'True') **Then**
Recuperate ADB ( ) *// Recuperate data from All Data Bases.*
Data Cleaning ( )
Create R Classes ( ) *// Categorize related data.*
End
Formulate C R ( ) *// Formulate Concepts and Relations.*
Formalize C ( ) *// Define concepts with a formal language.*
Integrate C ( ) *// Insert concepts in the ontology.*
ID = ID + 1
End
End

 https://doi.org/10.1633/JISTaP.2023.11.2.5

The data mining algorithm is comprised of three steps: public data collection, data classification, and knowledge integration, as described in Fig. 2. This proposed algorithm is a complete process in that it builds the national ontology dynamically from publicly available data sources.

3.1.1. Step 1: Public Data Collection

In this step, the data mining algorithm connects to publicly available data sources (such as civil status files, government databases, and so on) in order to dynamically collect (i.e. automatically and continuously) all data pertaining to citizens. After uploading the data sources, the algorithm starts by successively locating citizen national identification number (ID) in the civil status files in order to retrieve all information (i.e. name, date of birth, place of birth, nationality, etc.) related to the selected ID as well as all ID information in relation with the located ID. The algorithm then searches for the citizen ID in all data sources (e.g., social security, financial institution, tax administration, etc.) in order to collect all data related to the located citizen ID. Once all data has been recovered, the data mining algorithm performs data cleaning, which consists of detecting and correcting (or removing) errors in the recovered data. This process involves the following sub-steps:

- Removing duplicate or irrelevant data: While we combine data sets from multiple sources, we may encounter duplicate data. Therefore, the algorithm removes unwanted findings from the recovered dataset, such as duplicate or irrelevant data.
- Fixing structural errors: Unusual naming conventions, typos, and incorrect capitalization will be corrected to avoid mislabeled categories or classes.
- Handling missing data: Missing values must not be ignored. Therefore, the algorithm will insert missing values based on other data. For example: when the first name value is missing, we insert it from another source where the ID is similar.

At the end of the data cleaning process, we will acquire coherent collected data (Fig. 3).

The public data collection is an ongoing process that involves collecting continuously and successively all data pertaining to citizens. Then, in the next algorithm step, the collected data will be classified and categorized automatically.

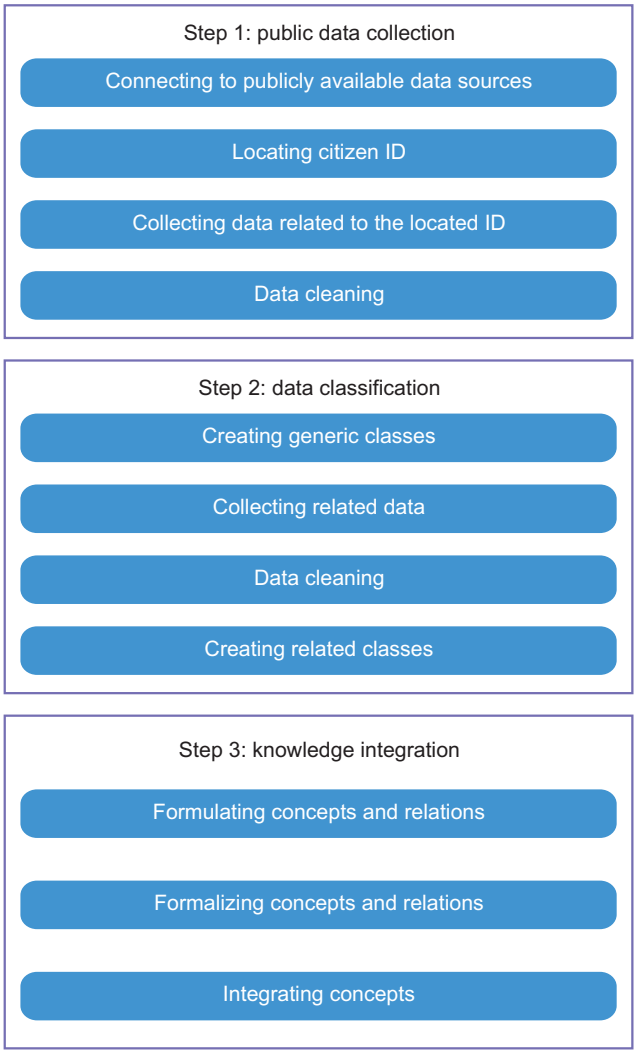


Fig. 2. The steps of the data mining algorithm. ID, national identification number.

3.1.2. Step 2: Data Classification

This step aims to categorize the collected data using a data mining classification technique. First, after retrieving all information related to the selected citizen ID, the data mining algorithm creates generic classes that represent the data source from which we retrieve information. First and foremost, since we have acquired the data pertaining to the citizen ID from the civil status file, the algorithm will create the generic class 'Citizens,' and then creates a sub-class of 'Citizens' with the retrieved information (i.e. name, date of birth, place of birth, nationality, etc.) as attributes. In addition, more 'Citizens' sub-classes of related IDs (i.e. foreign keys) will be created and relations between all sub-classes will be generated, as described in Fig. 4.

Moreover, and based on the searching results of the

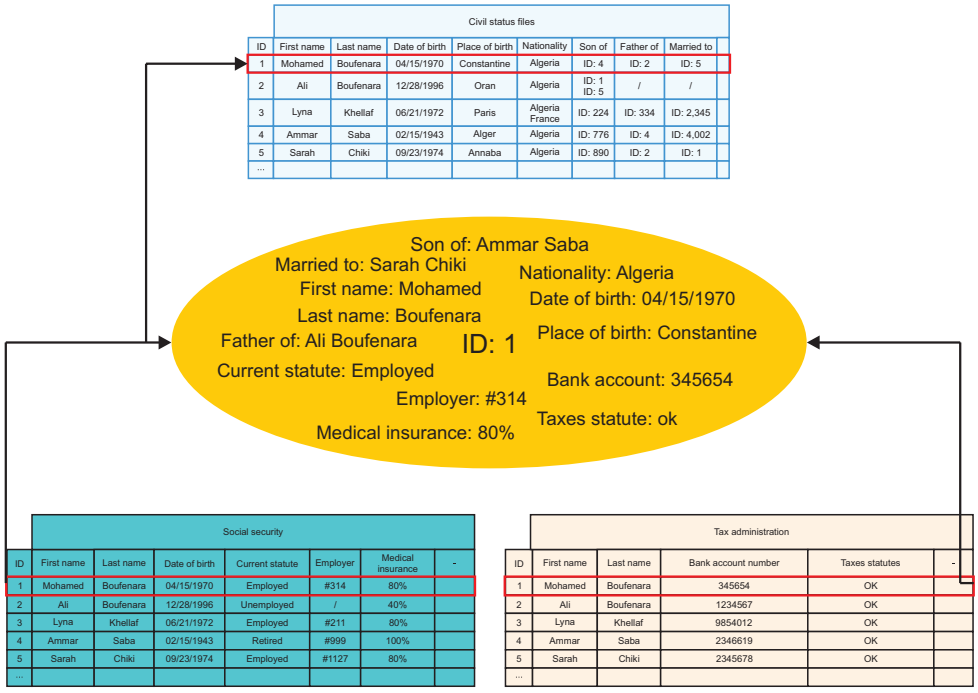


Fig. 3. The Public data collection. ID, national identification number.

citizen ID in all data sources (e.g., social security, financial institution, tax administration, etc.), the data mining algorithm will create classes of related information (Fig. 5).

Once all data has been classified and categorized, the data mining algorithm proceeds to the final step, which is to integrate the categorized knowledge, allowing the dynamic development of the national ontology.

3.1.3. Step 3: Knowledge Integration

The main task of this step is to automatically build the national ontology by structuring and integrating the retrieved data. The collected information is first formulated as ontological concepts by the Data Mining algorithm, with each class created in Step 2 converted into a concept and each correlation between classes converted into a relation. Then, using the syntax of the SHIQ description logic (Baader et al., 2017), the data mining algorithm formalizes concepts with a formal and operational language as described below:

1) Concept representation
ID1:
(∀ Part of.Citizens) ⋂ (≥ 1Married to .ID5 ⋂ ≤
1Married to .ID5) ⋂ (∃ Son of.ID4) ⋂ (∃ Father of.ID2) ⋂ (≥
1Works in .314 ⋂ ≤ 1Works in .314) ⋂ (≥ 1 have an account˙.Bank1 ⋂ ≤
1 have an account˙.Bank1) ⋂ (≥ 1 Pay taxes .T ax administration ⋂ ≤
1 Pay taxes .T ax administration) ⋂ (≥
1Have insurance coverage .Social security ⋂ ≤
1Have insurance coverage .Social security)

2) Relation representation
Married to (ID1, ID5)
Son of (ID1, ID4)
Father of (ID1, ID2)
Works in (ID1, 314)
Have an account (ID1, Bank1)
Pay taxes (ID1, Tax administration)
Have insurance coverage (ID1, Social security)

Following concept formalization, we use the classification algorithm described below to perform concept integration:

Classification algorithm

1: **Input: X** *// The concept to integrate.*
2: **Begin**
3: Sps := **SPS(X)** ; *// Search the super concepts of X*
4: Spg := **SPG(X)** ; *// Search the sub concepts of X*
5: **If** ((Sps ∩ Spg) = ∅) **Then** *// X is a new concept to integrate.*
6: **For each** (S ∈ Sps) **Do**
7: **Add** (X ⊆ S) ; *// Adding the subsumption link X ⊆ S.*
8: **End**
9: **For each** (G ∈ Spg) **Do**
10: **Add** (G ⊆ X) ; *// Adding the subsumption link G ⊆ X.*
11: **End**
12: **For each** (S ∈ Sps) **Do**
13: **For each** (G ∈ Spg) **Do**
14: Remove (G ⊆ S) ; *// Removing the subsumption link G ⊆ S.*
15: **End**
16: **End**
17: **Else** *// X is an existing concept.*
18: **Return** (nil);
19: **End if**
20: **End**

 https://doi.org/10.1633/JISTaP.2023.11.2.5

Civil status files								
ID	First name	Last name	Date of birth	Place of birth	Nationality	Son of	Father of	Married to
1	Mohamed	Boufenara	04/15/1970	Constantine	Algeria	ID: 4	ID: 2	ID: 5
2	Ali	Boufenara	12/28/1996	Oran	Algeria	ID: 1 ID: 5	/	/
3	Lyna	Khellaf	06/21/1972	Paris	Algeria France	ID: 224	ID: 334	ID: 2,345
4	Ammar	Saba	02/15/1943	Alger	Algeria	ID: 776	ID: 4	ID: 4,002
5	Sarah	Chiki	09/23/1974	Annaba	Algeria	ID: 890	ID: 2	ID: 1

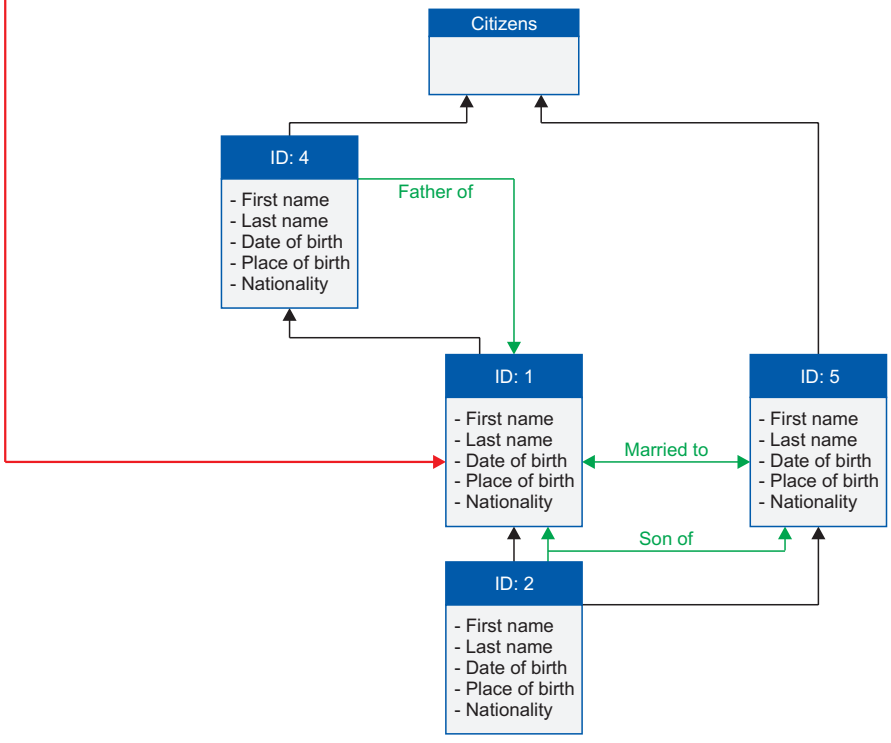


Fig. 4. Citizen classes creation. ID: national identification number.

The knowledge integration process will enable the ontology to be constructed automatically and continuously. The developed ontology and its features will be presented in the following section.

3.2. The National Ontology

An ontology is, in its most basic form, a table of categories in which each type of entity is represented by a node in a hierarchical tree. We rely on the national ontology developed by the data mining algorithm to achieve our goal of providing authorities with quantitative and qualitative data about any country's social and economic situation. The national ontology serves as a reference knowledge base that can be used to provide reliable statistical knowledge to governmental, economic, and social actors. It includes concepts (type definitions) that are data descriptors for citizen information, as well as the relationships between these concepts. The following are the key features of our national ontology:

- A hierarchical classification of concepts (classes) from general to specific.
- A list of attributes for each class.
- A set of relations between classes to link concepts.

In the national ontology, various generic classes serve to define intricate concepts like 'citizens,' 'employer,' 'social security,' 'banks,' and 'tax administration.' These classes are combined using data correlation, as illustrated in Fig. 6.

The ontology tree will be restructured based on the classification algorithm presented above each time the data mining algorithm collects, categorizes, and integrates concepts. New concepts will be structured and integrated into the ontology as subclasses of the generic classes.

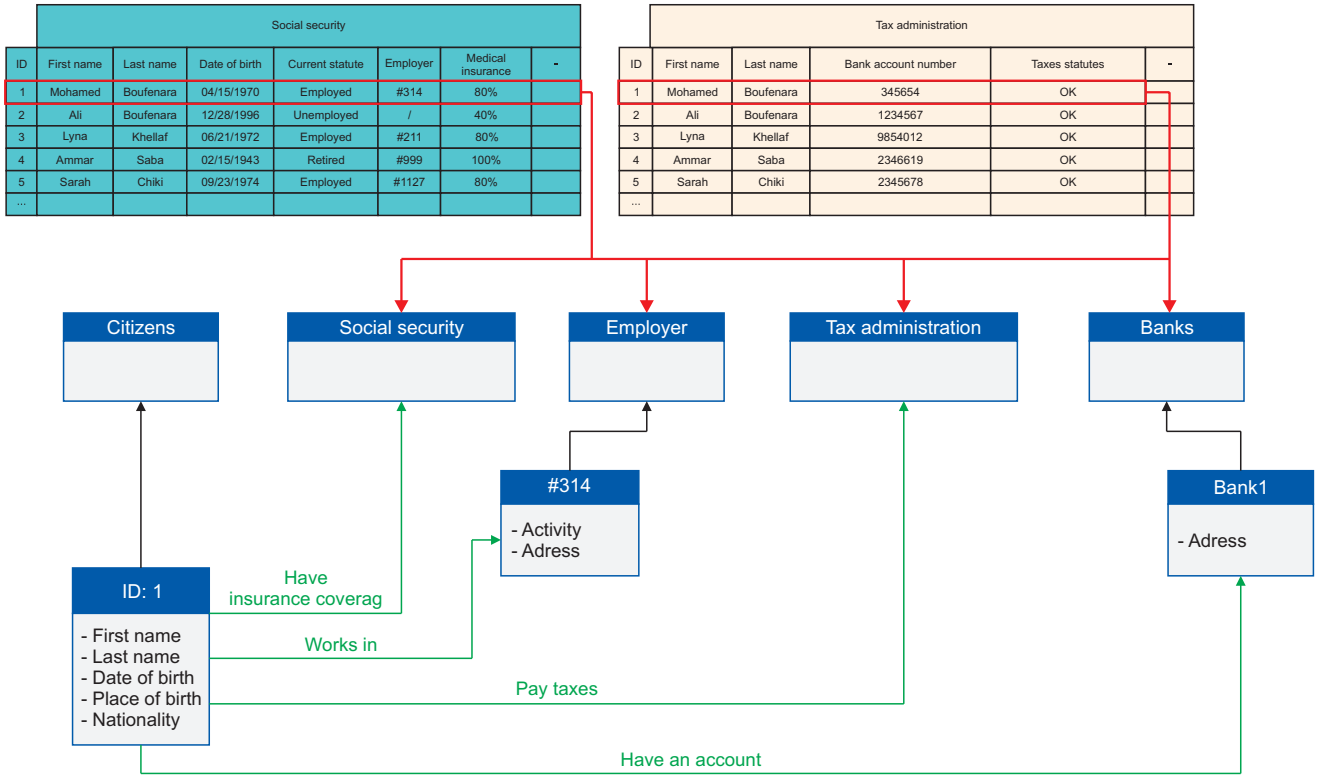


Fig. 5. Classes structuring. ID, national identification number.

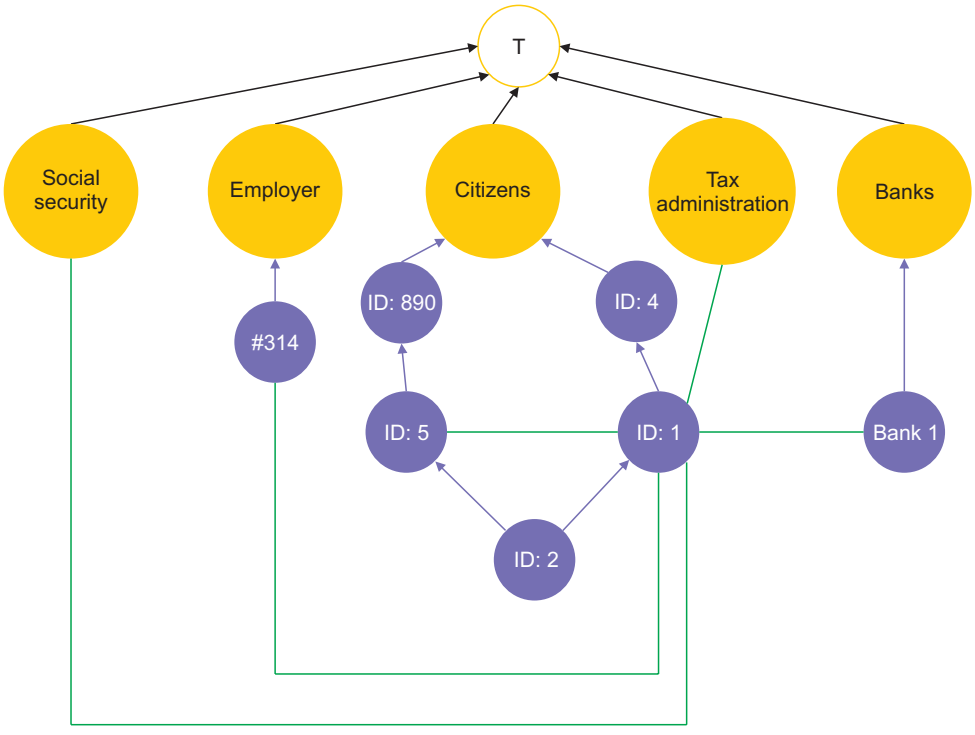


Fig. 6. The ontology concepts tree. ID, national identification number.

 https://doi.org/10.1633/JISTaP.2023.11.2.5

The national ontology is being created dynamically (i.e. automatically and continuously) and will be used as a reference knowledge base to provide reliable statistical knowledge to governmental, economic, and social actors. Therefore, we propose an intelligent exploitation platform to enable the use of the developed national ontology.

3.3. The Intelligent Ontology Exploitation Platform

Our proposed MVC architecture presents the view as an ontology exploitation platform that utilizes artificial intelligence for autonomy, information exchange, and cooperative management. The intelligent platform provides two features that allow users to interact with the national ontology. The first feature enables citizens and government administrators to explore and consult available knowledge, facilitating the digitization of administrative tasks and citizens' daily lives. The second feature continuously offers quantitative and qualitative information about the country's socioeconomic situation to governmental, economic, and social actors, as illustrated in Fig. 7.

Access to the platform is granted to government administrators based on their profile. For example, police services can only access legal, criminal, and civil status information about citizens. On the other hand, tax services have access to all information pertaining to citizens' earnings. in addition, governmental, economic, and social actors have privileged access to the statistical information provided by the platform. These quantitative and qualitative knowledge assets about the country's socioeconomic situation are continuously produced, analyzed, and provided by a platform module interacting with the national ontology and applying both quantitative and qualitative statistical techniques.

Since the national ontology consists of concepts that serve as data descriptors for citizen information, as well as the relationships between these concepts, the platform module will consider the national ontology as the population and will perform quantitative and qualitative analysis on it. To begin, the platform module utilizes different types of variables, including discrete and continuous quantitative variables and nominal and ordinal qualitative variables. These variables are detailed in Table 1.

The platform module can also use multiple variables at the same time. These variables can be qualitative, quantitative, or a mix of both. For example, the module combines multiple statistical variables to determine the number of citizens earning at least the minimum wage, the number of unemployed women, the number of low-income citizens, and so on. To acquire such statistical knowledge, the platform module constantly queries the national ontology using the SPARQL querying language, which is a standardized language that is similar to SQL but is used for Resource Description Format (RDF) graphs.

The national ontology is an OWL ontology in RDF that can be queried as an RDF graph using the SPARQL query language. SPARQL provides a means to query diverse data sources by allowing the expression of both mandatory and optional graph patterns, as well as their combinations. Additionally, it includes aggregate functions that can be used to select and return calculated values based on groups of query results. SPARQL can generate statistical data by using common aggregate functions such as: COUNT, SUM,

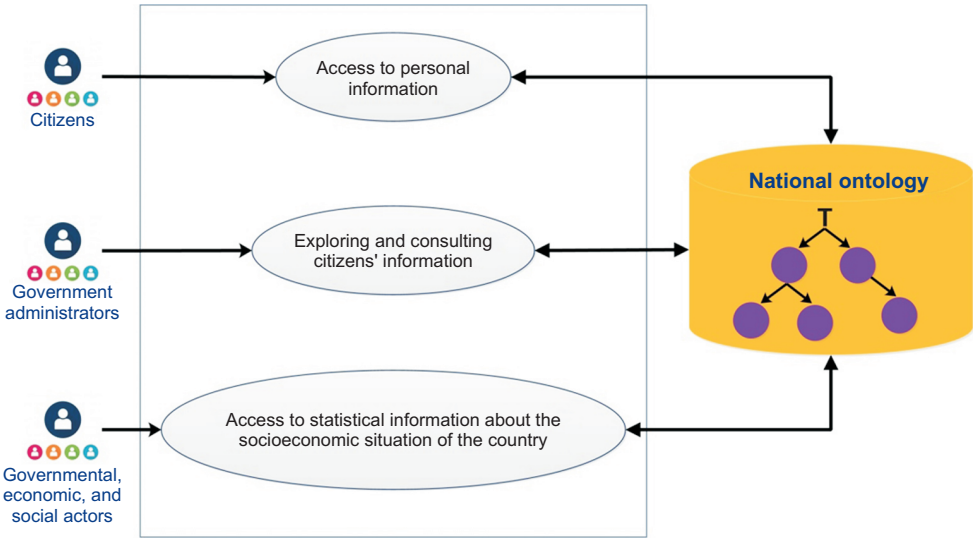


Fig. 7. Use cases of the intelligent platform.

Table 1. Quantitative and qualitative variables

Quantitative		Qualitative	
Discrete	Continuous	Nominal	Ordinal
All numerical values that can be counted	All numerical values that can be measured	Labels that describe groups of observations without logical order	Labels that describe groups of observations with a logical order
Examples: A country's total population A city's total population Total number of voters Total number of citizens with medical coverage	Examples: Citizens' payroll Citizens' tax amount Tax evasion amount	Examples: Employed citizens Unemployed citizens Married, unmarried	Examples: Level of tax evasion (high>medium>low) Level of citizens without medical coverage (high>medium>low)

MIN, MAX, AVG, DISTINCT, HAVING, and GROUP BY. The platform module employs these aggregate functions to generate reliable statistical knowledge, as illustrated in the following examples:

1) To determine a country's total population, the platform constantly queries the national ontology by counting instances belonging to the "Citizens" class.

```
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?total population)
```

2) To determine the total population of a city, the national ontology will be queried by counting instances of the "Citizens" class with the given city as an attribute.

```
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?CTP)
WHERE {
?citizens:city 'Paris' .
```

3) The platform module queries the national ontology for the number of unemployed women by counting instances of the "Citizens" class with female sex and unemployed status as attributes.

```
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?UW)
WHERE {
?citizens:sex 'female' ?citizens:status 'unemployed.'
```

Following the SPARQL querying of the national ontology, graphical representations will be charted to provide reliable and understandable statistical knowledge, allowing for better management and decision making.

4. SOFTWARE APPLICATION AND EXPERIMENT

We developed a software application to simulate and to experiment with our proposed data mining approach for the dynamic development of an ontology-based statistical information system. We performed a case study on Algerian citizens, using publicly Algerian data sources (i.e. civil status files, government databases, etc.).

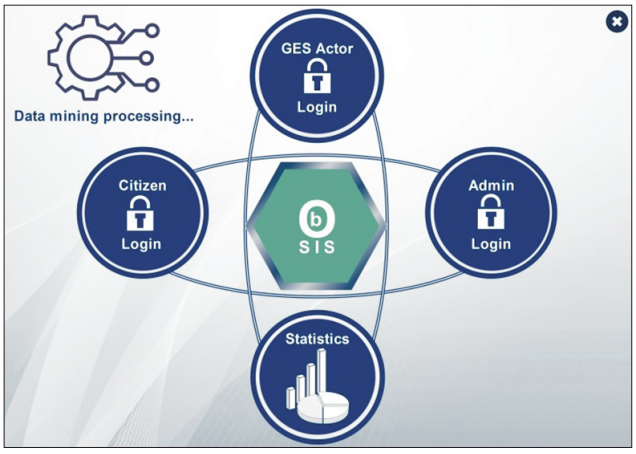


Fig. 8. The system interface. GES, governmental, economic, and social actors; OBSIS, ontology-based statistical information system.

The software application was designed as a back and front interface (Fig. 8) that enables the dynamic ontology development as well as the use of the national ontology.

The interface back-end represents the behavior of the data mining algorithm, which automatically and continuously collects, classifies, and categorizes citizens' knowledge into the ontology, as described in Fig. 9.

The proposed intelligent ontology exploitation platform (Fig. 10) serves as the interface's front end, enabling citizens, government administrators, and governmental, economic, and social actors to use the national ontology.

Citizens can log in to their personal profiles to manage their daily paperwork, applications for employment, hous-

 https://doi.org/10.1633/JISTaP.2023.11.2.5

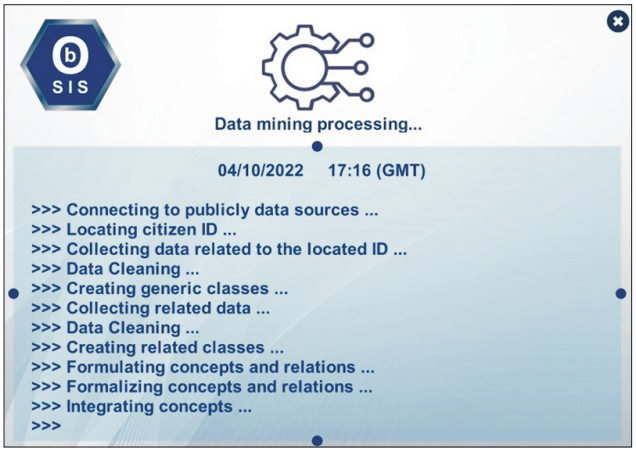


Fig. 9. The interface back-end. GMT, greenwich mean time; ID, national identification number.



Fig. 11. Citizen profile. OBSIS, ontology-based statistical information system; ID, national identification number.



Fig. 10. The interface front-end. GES, governmental, economic, and social actors; OBSIS, ontology-based statistical information system.



Fig. 12. Admin access. OBSIS, ontology-based statistical information system; ID, national identification number.

ing, and assistance, and so on (Fig. 11).

Moreover, administrators representing governmental services such as health and social care, justice, taxes, police, and so on can interact with "the national ontology" by exploring or consulting available knowledge, which will aid in the digitization of administrative and governmental tasks. Fig. 12 depicts a scenario in which the police services use the platform to obtain information about a citizen.

In addition to and as illustrated in Fig. 13, governmental, economic, and social actors have privileged access to the platform's statistical information on the country's socioeconomic situation.

To experiment with our proposed approach, we performed a case study on Algerian citizens, using publicly Algerian data sources in order to constantly develop a national ontology. Statistical knowledge assets about Algeria's socioeconomic situation are continuously produced, analyzed, and provided by SPARQL querying this ontology and applying both quantitative and qualitative statistical variables. Besides this, graphical representations will be charted to provide reliable and understandable statistical knowledge, allowing for better management and decision making, as described below:

- The average amount of unemployed citizens.

```
SELECT (AVG(?UC) AS ?avg-Unemployed )
WHERE {
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?UC)
WHERE { ?citizens:status 'unemployed' }
}
```

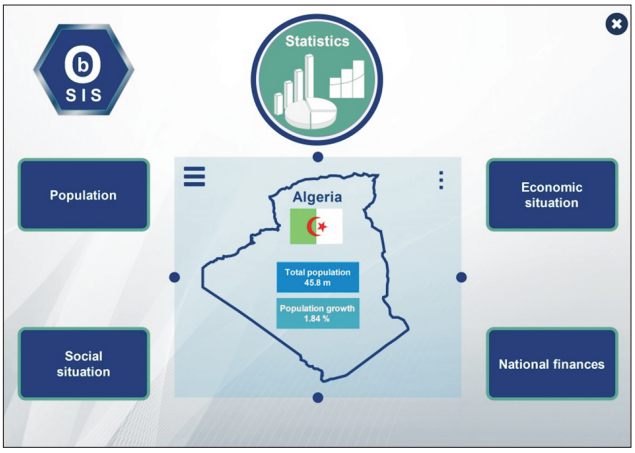


Fig. 13. Access to statistics. OBSIS, ontology-based statistical information system.

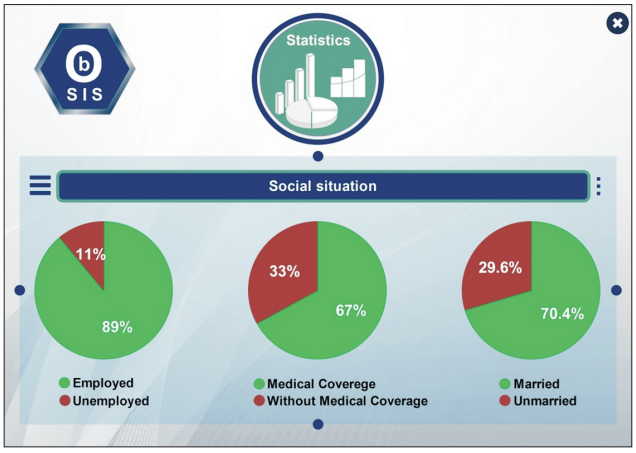


Fig. 14. Social situation parameters. OBSIS, ontology-based statistical information system.

- The average amount of unmarried citizens.

```
SELECT (AVG(?MW) AS ?avg-Min-wage )
WHERE {
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?MW)
WHERE {
?citizens:Marital-status 'unmarried' }
}
```

- The average amount of citizens without medical coverage.

```
SELECT (AVG(?WMI) AS ?avg-Without-M-Insurance )
WHERE {
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?WMI)
WHERE { FILTER (?medical insurance = 0) }
}
```

This ontology querying allows the platform to display graphical representations of some social situation parameters, as shown in Fig. 14.

Furthermore, the following SPARQL queries can be used to obtain some economic situation parameters:

- The proportion of citizens earning the minimum wage.

```
SELECT (AVG(?MW) AS ?avg-Min-wage )
WHERE {
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?MW)
WHERE { FILTER (?salary = 20000) } }
```

- The percentage of citizens earning the average salary.

```
SELECT (AVG(?AS) AS ?avg-med-wage )
WHERE {
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?AS)
WHERE { FILTER (AVG(?salary) } }
```

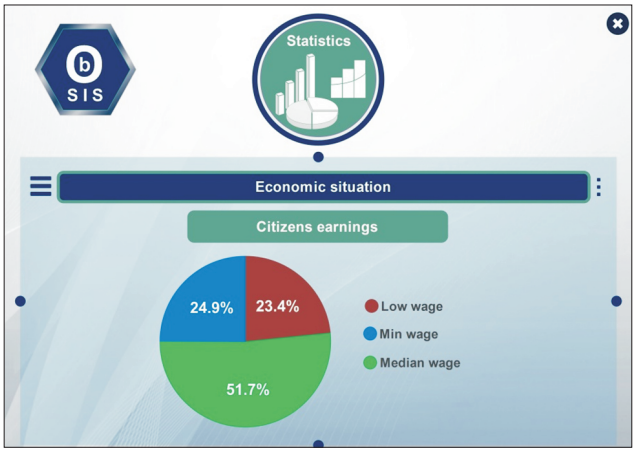


Fig. 15. Economic situation parameters. OBSIS, ontology-based statistical information system.

- The average amount of low-income citizens.

```
SELECT (AVG(?LIC) AS ?low-income )
WHERE {
SELECT DISTINCT ?Citizens (COUNT(?Citizens) as ?LIC)
WHERE { FILTER (?salary ¡ 15000) } }
```

Based on the outcomes of these queries, graphical representations of the economic situation will be charted, as shown in Fig. 15.

More SPARQL queries will be applied in order to constantly obtain reliable quantitative and qualitative knowledge about the country's socioeconomic situation. This produced statistical knowledge will be used by governmental, economic, and social actors for strategic decision-making, administration, and management of the national

 https://doi.org/10.1633/JISTaP.2023.11.2.5

political, social, and economic life.

5. DISCUSSION

Statistical data are used in many sectors, including science, government, manufacturing, population, psychology, banking, and financial markets. Statistics is the science that allows data to be identified, collected, organized, interpreted, and presented. Data can be either qualitative or quantitative. Statistics facilitate information-based decision-making. Indeed, since they collect, store, transform, and provide statistical data, both statistical Information Systems and Data Mining methods have played important roles in improving decision-making. Several statistical information systems and statistical data mining approaches have been developed as a result of these challenges. In Table 2, we compare some of these approaches by Bacchi et al. (2022), Improta et al. (2021), Ngugi (2021), Ngugi et al. (2021), and Osi et al. (2020) with our own method.

Our proposed approach employs a data mining technique to dynamically develop an ontology-based statistical information system. In comparison to some existing solutions that develop static statistical information systems, our developed system allows for the automatic and continuous collection, storage, processing, analysis, and presentation of statistical data at the national scale. To reach this goal, a classification data mining technique was developed to automatically collect, classify, and integrate citizen data into a national ontology. Then, an intelligent platform was developed to generate quantitative and qualitative statistical information based on the knowledge stored in the ontology in order to provide instant reliable knowledge about the country's socioeconomic situation. The provided statistical information will be used by governmental, economic, and social actors for strategic decision-making, administration, and management of national political, social, and economic life. In addition, as illustrated by the software application and experiment, the proposed approach has successfully digitized administrative and governmental tasks, thereby making citizens' lives easier. Moreover, the proposed system has successfully provided reliable and understandable statistical knowledge with graphical representations, allowing for better management and decision making.

6. CONCLUSION

In the age of big data, statistical knowledge is a valuable resource in decision-making. Statistics are becoming increasingly important in decision-making as an increasing amount of data is collected and converted into usable information and actionable steps. As a result, recent

Table 2. A Comparison of Some Existing Approaches

Approaches	Statistical data mining approaches			Developing statistical information systems	
	Classification	Clustering	Prediction	Static	Dynamic
Adekitan and Noma-Osaghae, 2019	X	X	X		
Luković et al., 2019				X	
Baek et al., 2018	X		X		
Ledvinka et al., 2019				X	
Fernandes et al., 2019			X		
Comas Rodríguez et al., 2019				X	
Jatav, 2018			X		
Choi and Choi, 2019				X	
Improta et al., 2021			X		
Ngugi, 2021				X	
Bacchi et al., 2022			X		
Ngugi et al., 2021				X	
Osi et al., 2020	X		X		
Our approach	X				X

research has focused on using computational techniques from statistics, data mining, and information theory to automatically collect and analyze massive amounts of data, thereby improving the availability of information on the state and evolution of a social or economic situation. Our proposed data mining approach for the dynamic development of an ontology-based statistical information system is part of these research efforts to provide governmental, economic, and social authorities with reliable statistical information in order to improve management and decision making. Unlike some existing statistical information solutions, the proposed ontology-based statistical information system enabled the national scale automatic and continuous collection, storage, processing, analysis, and presentation of statistical data. We proposed an MVC-inspired approach for developing our system, with the Controller representing a data mining algorithm that dynamically (i.e. automatically and continuously) collects all data pertaining to citizens from publicly available data sources (i.e. civil status files, government databases, etc.), and then structures, classifies, and categorizes this knowledge into the ontology. The Model is a national ontology that incorporates all citizen knowledge, and the View is an intelligent ontology exploitation platform that provides authorities with quantitative and qualitative data about the country's social and economic situation. To evaluate the proposed approach, a case study on Algerian citizens was conducted, using publicly available Algerian data sources to continuously develop a national ontology. SPARQL queried the developed ontology and applied both quantitative and qualitative statistical variables to continuously produce, analyze, and provide statistical knowledge about Algeria's socioeconomic situation. Furthermore, in addition to demonstrating the feasibility of our approach, the software application with the modeled system have demonstrated additional benefits such as the digitization of administrative and governmental tasks, thereby making citizens' lives easier, and providing reliable and understandable statistical knowledge with graphical representations, allowing governmental, economic, and social actors to better manage and make decisions.

ACKNOWLEDGEMENTS

The authors acknowledge support from the General Directorate of Scientific Research and Technological Development (DGRSDT), Ministry of Higher Education and Scientific Research, Algeria.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies,* 24(2), 1527-1543. https://doi.org/10.1007/s10639-018-9839-7

Baader, F., Horrocks, I., Lutz, C., & Sattler, U. (2017). *An introduction to description logic.* Cambridge University Press.

Bacchi, S., Tan, Y., Oakden-Rayner, L., Jannes, J., Kleinig, T., & Koblar, S. (2022). Machine learning in the prediction of medical inpatient length of stay. *Internal Medicine Journal,* 52(2), 176-185. https://doi.org/10.1111/imj.14962

Baek, H., Cho, M., Kim, S., Hwang, H., Song, M., & Yoo, S. (2018). Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. *PLoS One,* 13(4), e0195901. https://doi.org/10.1371/journal.pone.0195901

Choi, C., & Choi, J. (2019). Ontology-based security context reasoning for power IoT-cloud security service. *IEEE Access,* 7, 110510-110517. https://doi.org/10.1109/ACCESS.2019.2933859

Comas Rodríguez, R., Simón-Cuevas, A., Sánchez Fleitas, N., & García Lorenzo, M. M. (2019, October 28-31). An ontology-based data management model applied to a real information system. In J. Nummenmaa, F. Pérez-González, B. Domenech-Lega, J. Vaunat, & F. O. Fernández-Peña (Eds.), *Proceedings of the Conference on Computer Science, Electronics and Industrial Engineering (CSEI 2019)* (pp. 36-50). Springer.

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research,* 94, 335-343. https://doi.org/10.1016/j.jbusres.2018.02.012

Improta, G., Colella, Y., Rossi, G., Borrelli, A., Russo, G., & Triassi, M. (2021, October 29-31). Use of machine learning to predict abandonment rates in an emergency department. *Proceedings of the 2021 10th International Conference on Bioinformatics and Biomedical Science (ICBBS '21)* (pp. 153-156). ACM.

Jatav, S. (2018). An algorithm for predictive data mining approach in medical diagnosis. *International Journal of Com-*

puter Science & Information Technology, 10(1). https://ssrn.com/abstract=3633801

Kermani, M. H., & Boufaida, Z. (2022). I 3D3P: An intelligent 3D protein prediction platform. *Proceedings of International Conference on Applied Innovation in IT,* 10(1), 37-42. https://doi.org/10.25673/76930

Kermani, M. H., Boufaida, Z., Benredjem, S., & Saker, A. N. (2021). An MVC-inspired approach for an intelligent annotation of a protein ontology: IA-PrOnto. *International Journal of Computer Information Systems and Industrial Management Applications,* 13, 308-318. http://www.mirlabs.org/ijcisim/regular_papers_2021/IJCISIM_29.pdf

Ledvinka, M., Lališ, A., & Křemen, P. (2019). Toward data-driven safety: An ontology-based information system. *Journal of Aerospace Information Systems,* 16(1), 22-36. https://doi.org/10.2514/1.I010622

Luković, V., Ćuković, S., Milošević, D., & Devedžić, G. (2019). An ontology-based module of the information system ScolioMedIS for 3D digital diagnosis of adolescent scoliosis. *Computer Methods and Programs in Biomedicine,* 178, 247-263. https://doi.org/10.1016/j.cmpb.2019.06.027

Nawi, R. M., Noah, S. A. M., & Zakaria, L. Q. (2021). Issues and challenges in the extraction and mapping of linked open data resources with recommender systems datasets. *Journal of Information Science Theory and Practice,* 9(2), 66-82. https://doi.org/10.1633/JISTaP.2021.9.2.5

Ngugi, P., Babic, A., & Were, M. C. (2021). A multivariate statistical evaluation of actual use of electronic health record systems implementations in Kenya. *PLoS One,* 16(9), e0256799. https://doi.org/10.1371/journal.pone.0256799

Ngugi, P. N. (2021). *A systematic method for evaluating implementations of electronic medical records systems in low- and medium-income countries.* (Doctoral dissertation). University of Bergen, Bergen, Norway.

Osi, A. A., Abdu, M., Muhammad, U., Ibrahim, A., Isma'il, L. A., Suleiman, A. A., Abdulkadir, H. S., Sada, S. S., Dikko, H. G., & Ringim, M. Z. (2020). A classification approach for predicting COVID-19 patient's survival outcome with machine learning techniques. *medRxiv.* https://doi.org/10.1101/2020.08.02.20129767

Plirdpring, P., & Ruangrajitpakorn, T. (2022). Using ontology to represent cultural aspects of local products for supporting local community enterprise in Thailand. *Journal of Information Science Theory and Practice,* 10(1), 45-58. https://doi.org/10.1633/JISTaP.2022.10.1.4

Sowa, J. F. (2011). *Serious semantics, serious ontologies, panel.* Paper presented at the Semantic Technology Conference (SEMTEC 2011), San Francisco, CA, USA.

JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 11(2): 82-103, 2023

eISSN : 2287-4577 pISSN : 2287-9099

https://doi.org/10.1633/JISTaP.2023.11.2.6

Critical Success Factors for the Adoption of Health Management Information Systems in Public Hospitals in Zimbabwe

Caleb Manjeese*
School of Management, IT and Governance, University of KwaZulu-Natal, Durban, South Africa
E-mail: cmanjeese83@gmail.com

Indira Padayachee
School of Management, IT and Governance, University of KwaZulu-Natal, Durban, South Africa
E-mail: padayacheei@ukzn.ac.za

ABSTRACT

The Zimbabwean healthcare sector faces huge challenges due to increased demands for improved services for a growing number of patients with fewer resources. The use of information and communications technologies, prevalent in many industries, but lacking in Zimbabwean healthcare, could increase productivity and innovation. The adoption of health management information systems (HMISs) can lead to improved patient safety and high-level patient care. These technologies can change delivery methods to be more patient focused by utilising integrated models and allowing for a continuum of care across healthcare providers. However, implementation of these technologies in the health care sector remains low. The purpose of this study is to demonstrate the advantages to be attained by using HMISs in healthcare delivery and to ascertain the factors that influence the uptake of such systems in the public healthcare sector. A conceptual model, extending the technology, organization, and environment framework by means of other adoption models, underpins the study of adoption behavior. A mixed method methodology was used to conduct the study. For the quantitative approach, questionnaires were used to allow for regression analysis. For the qualitative approach, thematic analysis was used to analyse interview data. The results showed that the critical success factors (namely, relative advantage, availability, complexity, compatibility, trialability, observability, management support, information and communication technology expertise, communication processes, government regulation, infrastructure support, organizational readiness, industry and competitive support, external support, perceived ease of use, perceived usefulness, attitude, and intention to use) influenced adoption of HMISs in public hospitals in Zimbabwe.

Keywords: health management information systems, technological factors, organizational factors, environmental factors, individual factors, technology, organization, and environment framework

Received: May 24, 2022 **Revised:** March 14, 2023
Accepted: March 15, 2023 **Published:** June 30, 2023

***Corresponding Author:** Caleb Manjeese
https://orcid.org/0000-0003-0338-7063
E-mail: cmanjeese83@gmail.com

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

© Caleb Manjeese, Indira Padayachee, 2023

1. INTRODUCTION

Information and communication technologies (ICTs) have been applied in various industries for the collection, editing, transmission, and storage of information to improve service delivery. They aid organizational members in communicating, cooperating, and collaborating freely by removing the challenges of distance and time (Bacigalupe, 2011). This trend also occurs in health delivery institutions as they search for ways to improve service delivery, with one such innovation being the health management information system (HMIS). The main goal of incorporating information technology (IT) into the healthcare industry is improving patient care procedures already in existence and introducing new and innovative capabilities by removing the disadvantages of limited opening times, queues, paperwork, travel, and other challenges inherent in the old means of service provision. The use of HMISs has the potential to improve patient outcomes, while at the same time lowering costs.

1.1. Health Management Information System

An HMIS is designed to manage healthcare data by collecting, storing, manipulating, and transmitting patient data, by helping with hospital operational management and by supporting healthcare policy decisions at community, district, provincial, and national levels (Ariani et al., 2017). HMISs are critical in unifying communications through fully integrated audio-visual data communications and electronic systems. The architecture of an HMIS comprises three key levels of use and aggregation, namely, the patient-carrying level, the territory level, and the central government level. The patient level carries information about a patient's medical history, whilst the territory and government levels carry aggregated information about the medical staff, departments, operations, and management policies, as well as other decision-making related information (Asemahagn, 2017).

A good HMIS should provide relevant information that shows where service is lacking and should highlight the intervention needed in such areas. It should also allow access to reliable, timely, and authorised data by all relevant people, including partners, for implementation of different programs that will improve the healthcare (Chatterjee, 2015). Information is considered an indispensable operational asset necessary for effective management and development of health services; hence the need for an HMIS to support management at all levels (Alwan et al., 2015). The result of using an HMIS will be quality, cost effective, and efficient healthcare services for all (Asemahagn, 2017).

The adoption of HMISs in Zimbabwe would provide horizontal and vertical linkages of health systems such that patient data can be shared across all health systems with the aim of providing improved healthcare to patients. It should encompass all the aspects of healthcare, such as patient data storage, management of healthcare, and allocation of resources. Typically, an HMIS comprises various subsystems such as billing, intensive care unit systems, outpatients, maternity, drug inventory, duty allocations, and many others and should be accessible throughout the whole healthcare system.

1.2. Public Health System in Zimbabwe

In Zimbabwe, the largest proportion of healthcare services is accessed through the public health system and supplemented by mission hospitals and sometimes by non-governmental organizations. The health delivery system in Zimbabwe is divided into a four-level structure: primary, secondary, tertiary, and quaternary. The primary level, known as primary healthcare, drives most of the healthcare programmes such as the essential drugs programme, expanded immunisation, communicable disease control, curative and essential preventive services, food production and nutrition education, health education, and maternal and child services (Ministry of Health and Child Welfare, 2009).

According to official studies such as the Zimbabwe Maternal and Perinatal Mortality study of 2007, the Zimbabwe Demographic and Health Survey of 2005/2006, and the Multiple Indicator Monitoring Survey of 2009, the Zimbabwean populace is succumbing to diseases that can easily be prevented or treated, such as diarrhoea, Human immunodeficiency virus (HIV) and tuberculosis (TB)-related illnesses, malaria, malnutrition, and injuries, and in most cases, deaths are recorded in unusually high numbers. The management and treatment of patients suffering from such illnesses and other chronic diseases takes place over a lengthy period and is carried out by different healthcare providers in different settings. To achieve higher quality and safety in the delivery of healthcare, information should be presented in a structured manner, ready for use in clinical decision making, which can be achieved with the implementation of HMISs that support online access to health records, secure transmission of medical information between systems, and shared longitudinal electronic health records (RTI International, 2014).

The current system from facility to district health in-

formation is paper based, becoming computer based on provincial health information to national health information (United Nations Development Programme [UNDP], 2014b), and as such there is little support for accessibility to health consumer information. Attempts at computerisation in Zimbabwe have seen the development of various systems targeted at specific programmes, diseases, or areas. The systems are not integrated, nor do they cover the entire healthcare continuum of patients, thus allowing for a national picture of healthcare provision. Starting in 1985, a Microsoft Access based system was commissioned by the Ministry of Health and Child Care (MoHCC) to collect and report on routine data, but it soon became slower and less efficient. From 2011 to 2012, the country implemented the District Health Information Software version 1.4 (DHIS1). In 2012 the MoHCC got help from The United States President's Emergency Plan for AIDS Relief (PEPFAR) to pilot a DHIS-2 in Manicaland, and was scaled to other provinces in 2013 with the help of funding from UNDP, PEPFAR, and other partners (UNDP, 2020). Another system of note is the FrontlineSMS system launched by the MoHCC, a mobile-phone-based system used for transmitting early infant diagnosis of HIV from selected health facilities to the Zimbabwe's National Microbiology Reference Laboratory.

In 2012 the government, through the MoHCC and in partnership with various donor agencies such as World Health Organization (WHO), UNDP, Joint United Nations Programme on HIV/AIDS, the National AIDS Council, Centre for Disease Control, and the Research Triangle Institute, started piloting an electronic patient management system for HIV and TB. The first phase began in 2013, with the system installed at 85 sites and covering 65% of all national patients; 161 more sites were added in 2014, taking the total to 246. The plan was to implement Phase 3 of the project between 2015 and 2016, adding the remaining 184 facilities and bringing the number of patients covered to 97% (UNDP, 2014a). At Chitungwiza Hospital, a system using the System Applications and Products in Data Processing (SAP) Health Software application had approximately 100 desktop computers in all departments and wards, and these were used as access points. The system was linked to a central server at MoHCC headquarters with a fast fibre-optic link. Members were given login details to access the system. The system was used for tasks such as clinical information storage, capture of patient demographic data, and handling accounts of patients (Techzim, 2014).

1.3. Problem Statement

Challenges with regard to the old manual and current systems include difficulties in maintaining manual records and in the preparing and sharing of reports, a lack of standard names and codes to be used in service delivery, difficulty in monitoring and evaluation of hospital performance without real-time data, and a lack of adequate data to perform evidence-based program management (Asemahagn, 2017; Tan, 2009). The need to have HMISs in hospitals is driven by the growing need for connectivity among healthcare providers, for patients to participate in their personal care, for the managing of market pressures associated with healthcare provision, and for exploiting the great improvements in IT that have improved systems interoperability (International Center for AIDS Care and Treatment Programs at Columbia University, 2015; Tan, 2009). The challenges being experienced with old technologies such as tapes and folders is now pushing executives and healthcare givers to consider the implementation of innovative HMISs to improve healthcare delivery (Herrick et al., 2010; Ross & Venkatesh, 2016).

Though the benefits of using HMISs are known, few studies have dealt with the critical success factors (CSFs) for their adoption in Zimbabwe. The ICT infrastructure has developed to an extent where it is possible to computerize some or all of the repetitive information processing in hospitals and to meet the required information standards in minimal time with less effort on the part of staff, patients, and healthcare professionals. The opportunity to access health care information, including patient hospital records, which is fundamental to the delivery of healthcare services, has been hampered by a lack of suitable information systems. Studies documenting the issues around the adoption of HMIS in healthcare institutions in Zimbabwe are not available. Proper evaluations and proof regarding the perceived benefits and challenges to adoption in Zimbabwe are generally lacking.

The goal of this research study is to ascertain the CSFs that drive the adoption of an HMIS in Zimbabwe. The factors from the technology, organization, and environment (TOE) theory are used with additional constructs from other organizational and individual theories, such as the diffusion of innovation (DOI) theory, theory of reasoned behavior (TRB), theory of reasoned action (TRA), and the technology acceptance model (TAM), in order to construct a modified theoretical framework for explaining the CSFs influencing the adoption of HMISs at the organizational level.

 https://doi.org/10.1633/JISTaP.2023.11.2.6

1.4. Theoretical Framework

The main theory underlying this study is the TOE model, modified with constructs from other adoption models. The TOE framework is the work of Tornatzky and Fleischer (1990) and depicts the process a firm goes through during the adoption and implementation of technological innovations in the context of technological, organizational, and environmental factors. The internal and external technologies available and relevant to a firm constitute the technological context. The resources available and the internal characteristics of the organization constitute the organizational context. The industry's structure and size, the existing competition to the firm, the macroeconomic environment, and the regulatory framework and environment is what constitutes the environmental context (Tornatzky & Fleischer, 1990).

The TRA posits that an individual's behavior is a function of their intention, which in turn is a function of their attitude and subjective norms (Fishbein & Ajzen, 1975).

The theory of planned behavior (TPB) by Ajzen (1985) is similar to the TRA as it shows the link between an individual's behavior as influenced by intention, which in turn is a function of attitude, subjective norms, and perceived behavioral control (PCB).

The TAM by Davis (1989) has been widely used and accepted in adoption studies and it explains how perceived ease of use (PEOU) and perceived usefulness (PU) are regarded as the primary factors influencing an individual's intention to use a new innovation.

The TAM2 developed by Venkatesh and Davis (2000) is an extension of the TAM model and incorporates additional constructs, namely social influence and cognitive instrumental processes. TAM2 uses subjective norm, voluntariness, and image as the three interrelated social forms in determining whether an individual will accept or reject a new innovation.

The DOI theory from the work of Rogers (1962) states that different people in a society will not adopt an innovation uniformly but will do so at different times and stages. Rogers (1962) identifies the five main factors responsible for affecting the adoption of an innovation as relative advantage, compatibility, complexity, trialability, and observability, which have a different extent of influence in the five adopter categories. These are the five factors used from DOI to assess the technological category in the next section on conceptual framework and constructs.

The relationships between the models used are shown in Fig. 1.

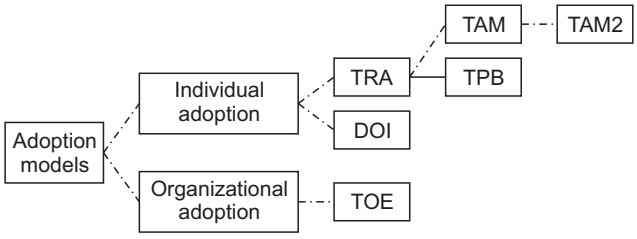


Fig. 1. Relationships between adoption models. TRA, theory of reasoned action; DOI, diffusion of innovation; TOE, technology, organization, and environment; TAM, technology acceptance model; TPB, theory of planned behavior.

1.5. Conceptual Framework and Constructs

The main research aim of this study was to find the technological, organizational, environmental, and individual factors that are critical for the successful adoption of HMISs in Zimbabwean healthcare institutions. Appropriate research constructs validated in prior studies were borrowed from the literature and used for questionnaire development.

The TOE model by Tornatzky and Fleischer (1990) was adapted to determine the factors influencing adoption of IT innovations at the organizational level. While numerous studies have used the model with varying modifications to explain how organizations adopt technology, not all the modifications have been related to the health industry or have included individual behavioral models such as TPB and TRA. To bridge this knowledge gap, the TOE model was integrated with individual behavioral models, and additional constructs, such as attitude (TRA), intention to use, PEOU (TAM), and PU (TAM), to get a better understanding of the dynamics of technological adoption in healthcare institutions. The constructs to be operationalized and measured are provided in Table 1.

The focus of this research is therefore to test the relationships among the constructs of HMIS adoption in the context of technology, organization, environment, and individual characteristics. In the investigation of technological, organizational, environmental, and individual factors affecting the adoption of HMIS by health institutions, 21 constructs were used in order to extend the existing knowledge base of technology adoption.

1.6. Research Questions

The specific research questions that are addressed in this article are as follows:

1. What are the organizational factors that influence the adoption of HMIS in Zimbabwean healthcare

Table 1. Conceptual framework constructs

Construct	Definition
Organizational	
Communication processes	The communication channels and how information flows in an organization (TOE)
Size	The larger the firm, the better the chances of surviving failures of new innovations (TOE)
Management support	Top management responsibility for cultivating the vision of the organization and encouraging success of technological innovation (TOE)
Organizational readiness	The availability of resources to allow for adoption of the said technology (TOE)
ICT expertise	The experience of the organizational members with ICT innovations
Technological	
Availability	How easily the required technology can be acquired (TOE)
Relative advantage	How an innovation is better compared to current practices (DOI)
Compatibility	How the innovation will match with the current systems and culture (DOI)
Complexity	How difficult it is to integrate and learn the new innovation for use if adopted (DOI)
Trialability	How easy it is to experiment with an innovation (DOI)
Observability	How easy it is to prove or show the results of an innovation (DOI)
Environmental	
External support	Other policy influences by external supporters to the organization and availability of support for implementation of the technology (TOE)
Technology infrastructure Support	The availability of the requisite infrastructure to support the use of the intended technology (TOE)
Government regulation	Government policies on use of the intended technology (TOE)
Industry and competitive pressure	Type of industry and competition that influences technological adoption (TOE)
Market scope	The market being served (TOE)
Behavioral/individual	
Perceived ease of use	The degree to which someone believes it is easy to learn to use the application (TAM/TAM2)
Perceived usefulness	The degree to which someone believes how important the use of the applications is and what the perceived benefits are, i.e., how it will enhance one's work (TAM/TAM2)
Attitude	Negative or positive feeling about the decision to use an application (TRA)
Intention to use	A person's readiness to adopt ICT innovations (TRB)

TOE, technology, organization, and environment; ICT, information and communication technology; DOI, diffusion of innovation; TAM, technology acceptance model; TRA, theory of reasoned action; TRB, theory of reasoned behavior.

institutions?

2. What are the environmental factors that influence the adoption of HMIS in Zimbabwean healthcare institutions?
3. What are the technological factors that influence the adoption of HMIS in Zimbabwean healthcare institutions?
4. What are the individual factors influencing the adoption of HMIS in Zimbabwean healthcare institutions?
5. How can the CSFs be integrated into an adoption model for HMIS in public health institutions in Zimbabwe?

1.7. Research Model and Hypotheses

The model was derived from the combination of TOE, DOI, TRA, TPB, TAM, and other adoption models and was tested for relevance and confirmation in public health provision settings. The model and hypotheses are shown in Fig. 2.

Hypotheses 1-3 are direct derivatives of the original TOE model with additional constructs from DOI; hypothesis 4 is an addition, adapted from the mixture of TOE and technology acceptance theories (TRA, TPB, TAM,

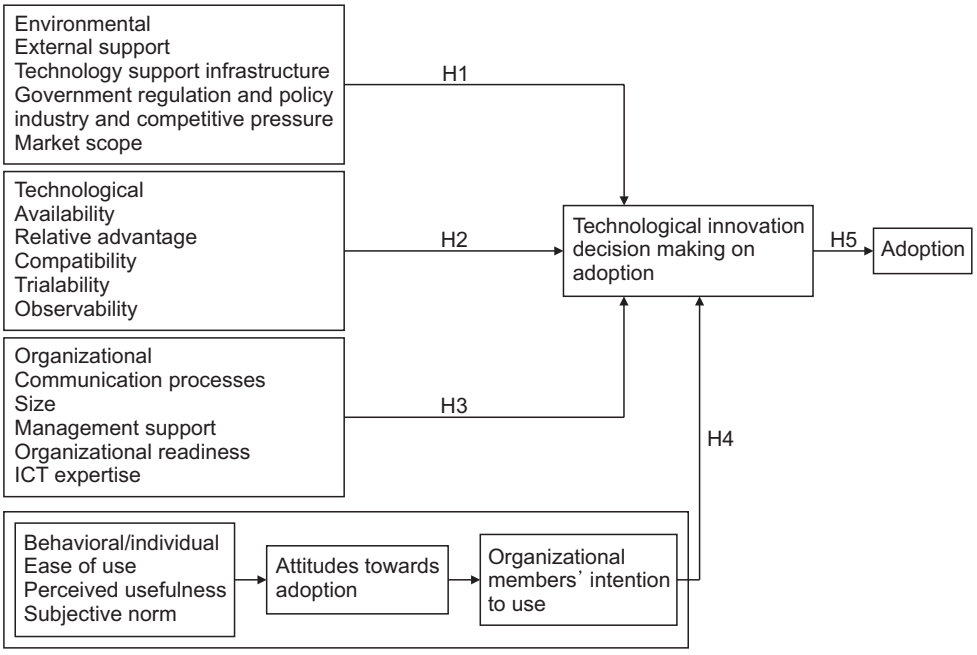


Fig. 2. Research model and hypotheses: H1, Environmental factors influence technological decision making on adoption; H2, Technological factors influence technological decision making on adoption; H3, Organizational factors influence technological decision making on adoption; H4, Individual factors influence technological decision making on adoption; H5, Technological decision making on adoption influences adoption. ICT, information and communication technology.

TAM2), and H5 is from the TOE model.

H1. Environmental factors influence technological decision making on adoption.

H2. Technological factors influence technological decision making on adoption.

H3. Organizational factors influence technological decision making on adoption.

H4. Individual factors influence technological decision making on adoption.

H5. Technological decision making on adoption influences adoption.

2. RELATED MATERIALS

This section presents a literature review of factors that influence the adoption of information systems in the healthcare sector. Understanding the factors that influence the adoption of innovations and new ideas enables efficient dissemination of potential innovations. The decision to adopt an innovation is an engaging, interactive, and dynamic process that can be influenced by various individual, technological, environmental, and organizational factors. Alam et al. (2016) has concluded that organizational, technological, environmental, and human factors are responsible for the adoption pattern of a human resource information system in Bangladesh hospitals.

Technological factors are regarded as the characteristics of the existing technology and the proposed innovations that affect the adoption decision in an organization. Relative advantage refers to whether the technological innovation is better than the one it is replacing. It has been found to be important in the adoption of various health related information systems, e.g., in the study conducted by Hung et al. (2010) on customer relationship management in hospitals and that by Proscoviour (2017) on HMIS adoption in hospitals. Compatibility shows the extent to which the new system will operate with the current technology and fit into the organizational culture and processes. A study by Khobi et al. (2020) of DHIS2 adoption in Uganda, one by Marsan and Paré (2013) of adoption of open system software (OSS), and one by Proscoviour (2017) of HMIS adoption in Kenya all found compatibility to be important for the adoption of health-related systems. Eger et al. (2001) concurred that the amount of change needed to use the technology affects the adoption of ICTs by physicians; this factor was listed as a system interoperability issue by Kralewski et al. (2008). Complexity is the measure of how difficult it is to integrate and learn to use a newly adopted innovation. Complexity was found to be an important technological factor by Khobi et al. (2020) and by Proscoviour (2017).

Trialability, which refers to the opportunity for potential users to test the application before actual installation in the organization, has been found to be important for the adoption of innovations (Rogers, 1962). This technological construct was included in the study by Marsan and Paré (2013) but proved to be insignificant. Included in

the same study was the construct of observability, which relates to how easy it is to show the results of an innovation and can be a measure of the popularity and use of the innovation in the industry. However, observability was reported to be insignificant for innovation adoption in this study. Availability relates to how easily the required technology can be acquired, involving the organizational availability of funds and the availability of the innovation at reasonable and affordable prices. Availability was used as an antecedent in the study by Dolan et al. (2020) and was given as the availability of electricity and the Internet in adoption of e-immunisation registries. Njeri (2018) used affordability in a study of adoption of HMIS, and Dixon and Dixon (1994) used availability of tools in their study of adoption decisions by physicians.

Organizational factors include the attributes of the organization and the resources it has available in its quest to adopt an innovation. Management support refers to management activities that influence the adoption of an innovation. Top management is responsible for cultivating the vision of the organization and encouraging success in the adoption processes. It was used as an antecedent by Khobi et al. (2020) and specified as the innovativeness of senior executives by Hung et al. (2010), and as support from management by Ifinedo (2018). These studies show that management support is important for adoption of innovation decisions. Communication processes represent the channels and mediums used within an organization and how they can potentially affect the adoption decision. Communication channels and information flow in an organization influence the type, rate, and success of innovation adoption in an organization. Dolan et al. (2020) used this factor in their studies and termed it as engagement among partners. Zhou et al. (2019) used this construct as interaction within communities in the organization, and found it to be important. Wang et al. (2005) used two constructs related to communication processes, that is, information processing and communication channels. Organizational size is a construct representing the influence of size on adoption decisions. It has been used in many studies and has been interpreted to mean that the larger the firm, the better the chance of surviving failures in an adoption process. The construct was used by Hung et al. (2010), who called it the size of the organization, and by Zhang et al. (2013), who used the number of beds to determine the size of hospitals. Both these studies concluded that size was important for the adoption of health-related innovative systems.

ICT expertise is a construct that defines the experience of the organizational members with ICT, and many studies have examined the influence of ICT expertise on the decision to adopt an ICT innovation. Marsan and Paré (2013) used the term for technological skills and knowledge, while Hung et al. (2010) listed it as IT capabilities of staff. Dixon and Dixon (1994) referred to it as the capabilities of the physicians in using the technology. These studies highlight the importance of ICT expertise in the adoption of innovations. Organizational readiness refers to the general preparedness of the organization to adopt an innovation and looks at the availability of resources (human, capital, and technology) in the organization and how they affect the adoption decision. Khobi et al. (2020) refer to ICT expertise as technology readiness, Proscoviour (2017) as organizational readiness, and Eger et al. (2001) as computer readiness. Organizational readiness covers all organizational resources that can affect the adoption of an innovation and has been shown to be an important antecedent in the studies mentioned above.

The environmental factor category is made up of all factors external to the organization but which have a bearing on its operations, and it has been indicated as a contributory factor in innovation adoption decisions. External support is one of the factors in this category, which refers to policy influences exerted on the organization by external supporters or the availability of IT support from outside organizations. The availability of external support tends to influence the adoption of ICT innovations positively. Ross et al. (2016) concluded that external support may facilitate adoption of technology and that it can be offered in the form of incentives by government or any other external stakeholder. Incentives could be financial and may be in the form of start-up funds to cover initial costs, financial sponsorship packages, compensations for adoption, and pay-for-performance initiatives. Marsan and Paré (2013) showed that external support is important when they used the construct of availability of external OSS expertise for the adoption of OSS by physicians. Alam et al. (2016) has used the construct of technology vendor support, which partly refers to external support, and this has proved to be a significant predictor of innovation adoption.

Technology infrastructure support refers to the availability of the requisite infrastructure to allow for the implementation and use of the intended technological innovation. Ross et al. (2016) note that many authors agree that availability of resources affects the decision to adopt and implement ICT innovations. Infrastructure includes access to computers, electricity supply, available band-

width, reliable Internet connectivity, and access to phone lines and mobile phones. Park and Lee (2014) reported that small hospitals showed a more positive correlation between adoption of electronic medical records and a fully installed IT infrastructure than any other type of hospital.

Government regulation refers to government policies and influence on the use of the new technology. Ross et al. (2016) noted in their review of the literature that the shortcomings of legislature, policies, and standards can possibly derail the implementation of health information systems at any level, be it professional or organizational. Industry and competitive pressure refer to the type of industry and the pressure it can exert on an organization to adopt ICT. As noted by El Kandri (2018), the healthcare industry is generally under the control of governments, and hence there is no competition and no motivation to innovate and improve patient experience. The role of government is blamed for the lack of resources needed for the adoption of innovations due to constrained budgets, in contrast with countries where health care is controlled by the private sector and where there is more innovation. Market scope refers to the market being served and the influences it can have on the readiness of an organization to adopt an ICT innovation. Miraz and Habib (2016) found that the connection between market scope and adoption of ICTs is significant when they showed that the activities of the market, such as partner or customer adoption trends, can influence an organization to adopt.

The individual factor category describes factors that pertain to the behavior of an individual or a collection of individuals and how they influence the adoption of innovation in an organization. PEOU is an individual's perception and belief regarding how easy it is for someone to learn to use an application; it is a measure of the effort required to be able to use the application. Ease of use is regarded as one of the factors in users' adoption of new technologies. PU is a measure of how the new technology will enhance one's work if adopted. It is a measure of how the innovation is better than the technology it succeeds. Although referring to studies outside of the health sector, Wanyoike et al. (2012) found a positive correlation between PU and an enterprise's intention to adopt new Internet technologies. Attitude is the negative or positive feeling about the decision to use an application. Khan and Woosley (2011) stated that the main themes that emerged in their study of the literature regarding models that fit adoption in the health sector were attitude, efficiency of use, and usability, which were covered in TAM, DOI, and Unified Theory of Acceptance and Use of Technology. Their conclusion is that by understanding the attitudes and satisfaction of the physicians, the adoption of health IT can be increased.

Intention is a person's readiness to adopt an innovation and the actions he or she makes to signal readiness to use an application. Macharia and Maroa (2014) in their study of adoption of HMIS in Kenya's private hospitals concluded that behavioral intention has a significant role in predicting the usage of new innovations. Christodoulakis et al. (2017) listed the most notable barrier to adoption of remote patient monitoring technologies as the lack of user acceptance and their behavioral intention to adopt. The same study, however, indicated that the lack of adoption of electronic patient records was due to poor design of the application, inadequate use by clinicians, and socio-organizational factors such as lack of support from colleagues, which was supported by Khan and Woosley (2011). The constructs of PEOU and PU were found to have a significant influence on user intentions and behaviors in the adoption of cloud technologies in public service organizations using TAM (Shin, 2013).

Furusa and Coleman (2018) carried out a study on the elements influencing adoption of e-health and its implementation by public hospital medical doctors in Zimbabwe. The results showed that adoption and implementation of e-health was largely motivated by internal and external elements. The internal factors reported to be significant influencers were technical support, ICT infrastructure, security concerns, ICT skills and knowledge, refusal to accept change, practitioner-patient relationships, basic medical facilities shortages, and demographic data such as age. The external factors found to be significant influencers were funding, bureaucracy, and health policy.

However, most of the studies have been carried out in other regions and not in Zimbabwe, and are limited to enumerating the systems in use, accompanied by benefits, challenges, and implementation strategies of HMISs (Mapesa, 2016; RTI International, 2021; UNDP, 2014a; 2014b). The factors from the various studies, although using different approaches, frameworks, and methodologies, and in some instances being referred to by different names, can be grouped into a set of encompassing categories. This article reports on the synthesis of factors from the various adoption models to come up with a modified theoretical model to explain the adoption of HMISs in Zimbabwe, which could be used as a basis for carrying out similar studies in other regions.

3 METHODOLOGY

3.1. Research Approach

The mixed methods (pragmatic) researchapproach was used for the empirical study. It is a methodology that uses both quantitative and qualitative research techniques in one study (Creswell, 2014). Quantitative and qualitative methods were used concurrently for collecting both forms of data, labelled as the parallel-databases variant. The data from the two methods were kept distinct, that is, the data collection and analysis for the two approaches were done independently, and only combined at the interpretation stage, where the results were given equal priority in drawing conclusions. The collection and analysis of both sets of data occurred at approximately the same time.

The researcher compared the results for convergence during the interpretation stage. Where divergent results were obtained in quantitative and qualitative findings, the researcher considered the potential complementarity and completeness of the findings. Confirmation of findings was addressed in a later phase of the study by use of a third set of data collected to evaluate the model developed after the first two sets of data collection and analysis. The approach is depicted in Fig. 3.

3.2. Phase 1: Quantitative Data Collection and Analysis

This phase discusses the sampling, data collection, and data analysis of the quantitative data collected using the CSF questionnaire.

3.2.1. Target Population

The target population included those who work in various capacities in these health centres, such as the administrators, nurses, and doctors. The province has a population of about 1,486,604, as reported by 2012 Zimstat online (https://www.zimstat.co.zw/). There is a physician density of 0.16, nurse density of 0.72, and an overall density of 1.23 health staff per 1,000 population in Zimbabwe. According to the WHO, density=population/workers, which in this study gives an estimate of 1,829 health workers comprising the target population for the province.

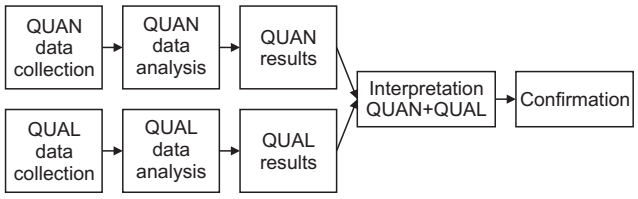


Fig. 3. Convergence model of mixed method data analysis for the study. QUAN, quantitative; QUAL, qualitative. Adapted from Creswell, 2014. Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Sage Publications.

3.2.2. Sampling

A combination of convenience and random sampling was employed. Three hospitals were conveniently selected, and the participants were then randomly selected for the CSF questionnaire. The hospitals include Masvingo general hospital, Chivi District hospital, and Mogenster mission hospital in Masvingo Province. The questionnaires were distributed to the employees of the selected institutions until the target was reached, using systematic random sampling. Each institution was allocated the same number of questionnaires. The sample for the study comprised 360 participants.

3.2.3. Quantitative Data Collection

The CSFs for the adoption of the HMIS questionnaire comprised six sections. Section A comprised general demographic questions such as age, gender, department, qualifications, and experience with computer systems. These generally helped to show the coverage of the employees who participated in the study and helped to demonstrate the randomness of the process of respondent selection. Section B was designed to solicit information about technological factors, while Section C was designed to collect information on organizational factors and Section D to cover environmental factors. Section E was designed to cover individual factors and Section F was designed to cover general hospital operations. A 5-point Likert scale was used with 1 denoting strongly disagree and 5 denoting strongly agree to rate the questions in the main sections of the questionnaire.

The total number of distributed questionnaires was 360 as per the calculated sample size distributed to 3 participating institutions, and each had an allocation of 120 questionnaires. At the end of the data collection period, 122 questionnaires were deemed usable from those received. Usable questionnaires are those that were correctly filled in and with missing information, if any, deemed to be insignificant to distorting the results. The participant profile is given in Table 2.

3.2.4.Quantitative Data Analysis

Inferential statistics techniques were used for the analysis of quantitative data. These techniques help researchers

Table 2. Participants profile for critical success factors questionnaire

Variable	Classification of variables	Frequency	Percentage
Sex	Male	39	32.0
	Female	83	68.0
Age (yr)	18-25	23	18.9
	25-30	35	28.7
	30-39	21	17.2
	39-49	25	20.5
	>49	18	14.8
Education	Doctorate	3	2.5
	Masters	4	3.3
	Undergraduate	33	27.0
	Diploma	68	55.7
	Certificate	12	9.8
	Other	2	1.6
Job title	Nurse	69	56.6
	Physician	13	10.7
	Allied health	7	5.7
	Admin	14	11.5
	Technical services	11	9.0
	Specialist doctor	8	6.6
Site	1	47	38.5
	2	39	32.0
	3	36	29.5

to draw conclusions and find trends in data about a population, helping them to study the connections between variables in a sample and to make generalisations and predictions about the relationships between the variables within the entire population. The inferential statistics used for this study included factor analysis, correlation analysis, and regression analysis.

3.2.4.1. Factor analysis.

Factor analysis allows for the reduction of many variables to a smaller number of factors. The assumption is that the set of variables have not been measured perfectly. A factor is the name given to the unobserved or latent variables that make up common variance; thus, the test is named factor analysis (Garth, 2008). Factor analysis, more properly termed exploratory factor analysis, provides a mechanism for explaining the relationships of observed variables $x_1, x_2, \ldots, x_n$ using a few unobservable factors, commonly referred to as common factors or latent variables $f_1, f_2, \ldots, f_m$ where $m<n$ and preferably is much less (Landau & Everitt, 2004). Factor analysis was performed first in the study to validate which factors were important in the adoption of HMISs.

3.2.4.2. Regression analysis.

Regression analysis technique provides an estimation of relationships between two variables, with one known as the dependent variable (DV) and the other as the independent variable (IV), and would be useful for modelling future relationships as well (De Smith, 2015). Regression analysis provides a reliable way to identify which variables have an impact on a subject under investigation. Carrying out regression analysis allows one to identify the important factors and their influence on the subject under study. This test was performed on the CSF questionnaire using the IBM SPSS Statistics version 20.0 (IBM Co., Armonk,

NY, USA).

3.3. Phase 2: Qualitative Data Collection and Analysis

This phase comprised sampling, data collection, and data analysis for the qualitative data collected, that is, for the CSF interviews.

3.3.1. Sampling

Purposive sampling was used for the CSF interviews. Those interviewed were senior personnel in the medical fraternity and a small number of participants from administration or service departments, namely, the medical superintendent, sisters-in-charge, accountants, human resource administrators, and IT personnel. The total number of participants interviewed was 10. This number was deemed sufficient and representative as there are generally fewer management and administrative staff members in relation to the entire staff complement at the hospital.

3.3.2. Qualitative Data Collection

Ten interviews were held with the senior staff and administrators in the selected institutions. Semi-structured in-depth interviews were used to collect data on the factors influencing the adoption of HMISs from a managerial viewpoint and on the role played by the technological, environmental, organizational, and individual factors. The interviews were held at times convenient to the participants and at their places of work so as not to inconvenience their work schedules and to minimise costs for the researcher. The researcher interviewed two participants from one hospital and four participants from each of the other two hospitals. All interviewees agreed to accommodate follow-up questions by e-mail or telephone if clarity on their responses was needed. The participants' profile is shown in Table 3.

Table 3. Summary of participant profile

Variable	Classification of variables	Frequency
Sex	Male	4
	Female	6
Age (yr)	18-25	0
	25-30	1
	30-39	3
	39-49	4
	>49	2
Education	Doctorate	1
	Masters	2
	Undergraduate	4
	Diploma	2
	Certificate	0
	Other	
Site	1	4
	2	4
	3	2

3.3.3. Qualitative Data Analysis

The thematic analysis technique was used to analyse the data collected during the in-depth interviews. Crowe et al. (2015) define thematic analysis as the procedure of analysing qualitative data to discover patterns of meaning. To analyse qualitative data using thematic analysis, themes are identified in the data, labelled, organised, and analysed, with the findings finally being reported on. The data analysis software R-QDA (https://rqda.r-forge.r-project.org/) was employed in analysing the qualitative data.

The analysis procedure for the qualitative data was a combination of content and thematic analysis. The process went through the familiarisation, identification of a thematic framework, coding, charting, mapping, and interpretation stages, which were performed beyond describing what participants said. The ten interview scripts were converted into ten plain text documents and loaded into R-QDA software. Codes and code categories were created based on the information obtained from the review of literature and the underpinning theory. The codes were generated from the conceptual model and hypotheses and matched with emerging themes from the scripts, as shown in Table 4.

3.4. Phase 3: Model Confirmation

In this phase, another set of data was collected in order to confirm the results that had been collected in the first two phases and to clarify all discrepancies.

3.4.1. Sampling

Purposive sampling was the choice of the researchers in selecting respondents for the questionnaire designed to evaluate the HMIS adoption model, as the researcher sought participants who had managerial and ICT insights.

3.4.2. Data Collection

The evaluation questionnaire was made up of closed and open-ended questions. The questionnaire was made

Table 4. Summary of themes, subthemes, and codes

Code	No. of interviewees	No. of mentions	Sub themes	Themes
External support	10	22	Government support NGOs and corporate support Financial support Human resource support	Need for financial and resources support from government and corporates (Partners)
Technology support infrastructure	10	21	Need for adequate or requisite infrastructure Support for acquisition Support for installation Support for technology shareability Policy and regulation	Supporting infrastructure development
government regulation and policy	10	22	Policy pushes for technology use Political environment necessary for technology boom Executive support Legislature support	Government policies and regulation at the core of adoption
Industry and competitive pressure	0	0	Horizontal integration Vertical Integration Industry standardisation	Integration of departments and with outside organizations
Market scope	6	6	Technology to suit clients' needs How related organizations are dealing with same technology Reach of clients	Market focus determines choice
Technology availability	10	15	Availability on market Affordability of technology Availability of funds	Availability, affordability, and availability of capital
Technology relative advantage	10	52	Improved processes and user benefits Increased revenue/ Decreased costs Client satisfaction	Technology brings more benefits than costs
Technology compatibility	10	12	Equipment compatibility Process compatibility Values compatibility Human resource compatibility	How technology fits into organization
Technology trialability	10	13	Pre-testability Pre-implementation installation Possibility of scaled implementation	Certification of product before implementation
Technology observability	10	12	Understanding the value of a technology through: Personal experience Third party experience	Identification and visualization of benefits
Technology complexity	10	12	Scale of systems (components, people) System features complexity Processes complexity	Inter-dependence of components, routines, individuals, and resources
Management support and practices	10	17	Management practices and priorities Resource mobilization & provision Participation	Importance of management activities
Organizational size	3	3	Organizational size measured as: Departmental size Staff size	Size's effects on organizational activities
Organizational communication processes	10	23	Communication channels Adequate information & Relevant information	Communication processes

Table 4. Continued

Code	No. of interviewees	No. of mentions	Sub themes	Themes
Organizational readiness	10	27	Patient readiness Members' readiness Resources availability	Organizational components readiness
Organizational ICT expertise	10	35	In-house expertise Training possibility Hiring possibility	Knowledge on organizational ICT systems
Perceived ease of use	10	16	Systems learnability Interface usability System trainability	Effort at using the system
Perceived usefulness	10	23	Benefits or usefulness Extend of benefits derived versus cost Functionality User needs compared to system functions	Technology solves the current problems
Subjective norm	3	1	The need to keep in line with trends can also affect decisions Individual norm Organizational norm	Norms of other organizations
Attitude	10	19	Attitudes of members can affect adoption personal attitude Group attitude	Feelings about the technology
Intention to use	10	21	Possibility of resistance to system Personal willingness Group willingness	Willingness and readiness to use

ICT, information and communication technology.

up of six sections, which were designed for evaluating the fit of the various factors in the model. The first section was designed to test for the general agreement of the participants with the model's adequacy, completeness, and applicability. The second section was used to confirm the organizational factors, while the third section was used to confirm the technological factors. The fourth section was used to confirm the environmental factors, and the fifth to confirm the individual factors. The sixth section was used as an open-ended questionnaire for use in clarification of their choices by the participants. The outcome would confirm the results from the first two data sets. For the structured questions, the researcher chose to use a five-point Likert scale rating.

The participants included 36 hospital staff and four academic experts (employed in the academic field and with an interest in adoption studies as shown by their publications). The researcher contacted the participants and sought their consent. Those who indicated their willingness were sent the questionnaire together with the informed consent form by electronic means. Thirty-one (31) questionnaires were returned in total and were usable, of which 28 were from hospital employees while three were from academic experts.

3.4.3. Data Analysis

3.4.3.1. Correlation Analysis.

Correlation provides a picture of the association or lack of it between two variables of interest, which can be positive or negative. If the two variables' relationship moves in one direction, that is, a decrease in one variable corresponds to a decrease in the other variable or vice-versa, then the relationship is deemed to be positive; if the variables move in opposite directions, that is, an increase in one variable corresponds to a decrease in another or vice-versa, then the relationship is deemed to be negative. There are various correlation tests that can be used in data analysis, and these include the Spearman's rank, Karl Pearson's product-moment, the coefficients of concurrent deviations, or scatter diagrams (Sereno, 2021). Correlation tests were performed on the confirmation questionnaire

as a way of proving the relationships that had been identified through the CSF questionnaire and the interviews.

3.4.3.2. Thematic Analysis.

The open response questions in the model confirmation questionnaire were analysed using the thematic analysis technique using R-QDA software.

3.5. Reliability and Validity

Reliability and validity were ensured by monitoring the parameters of the instruments to maintain the same properties. A pilot test was used to ensure reliability by giving the instrument to different respondents in different health institutions to check the consistency of the results prior to the main data collection exercise. Validity can be ensured in two ways, namely internal and external. Internal validity ensures the obtained results are legitimate, and for this study it was ensured by the way that the respondents were selected (health workers), and the way that the data was handled, i.e., recorded and analysed.

The sampling criteria used provided the ability to externally validate or generalise the results. Although random sampling was used, the information can be deemed to be reliable to localisation of the research to a particular area where the respondents were typical or representative of the whole population. A high level of validity (internal and external) was achieved through strict protocol execution and study design.

4. RESULTS

The results of the analysis are presented in this section.

4.1. CSF Questionnaire Results

The 122 completed questionnaires were captured into IBM SPSS statistics 20 for analysis and the results are presented below:

4.1.1. CSF Questionnaire Instrument Validation

Construct validity was performed using factor analysis. The techniques applied include principal-components factor analysis, extraction technique, and varimax orthogonal rotation method. To achieve a reliable scale from the available set of survey items, the Cronbach's alpha coefficient of internal consistency had to be determined. Reliability can be guaranteed for higher values of the Cronbach's alpha, which has a range of 0-1. The factor loadings for the factors accepted for further tests range from 0.640 to 0.955, which meet the recommended minimum value of 0.5 suggested by Chin et al. (1997) (Table 5).

Table 5. Factor analysis results

Construct	Sub-construct	Items included	Cronbach's alpha
Environmental	External support	ESP 2-5	0.804
	Government regulation and policy	GR 3-4	0.854
	Industry and competitive pressure	PRES 2-3	0.922
Technological	Availability	TA 5-7	0.766
	Relative advantage	RA 1-3 (organization)	0.895
		RA 5-11 (patients)	0.854
	Compatibility	CBT 1-2	0.655
	Observability	OB 1-2	0.737
Organizational	Support and skills		0.929
	Organizational readiness (general)	OR 5-6 & SI 1-2	0.807
	Organizational readiness (aim/vision)	OR 2, 4	0.663
Individual	Usefulness and ease of use	PU 1-3 & PEOU 1-3	0.955
	Subjective norm	SN 1-2	0.685
Adoption	Adoption	F 2.3 & F 2.4	0.640

ESP, external support; GR, government regulation, PRES, industry and competitive pressure; TA, availability; RA, relative advantage 9; CBT, compatibility; OB, observability; OR, organisational readiness; SI, support and skills; PU, perceived usefulness; PEOU, perceived ease of use; SN, subjective norm; F, adoption.

The constructs retained after the Cronbach's test were the ones deemed to be internally consistent. They were subjected to correlational and regression testing in a bid to come up with the existing relationships between the constructs and the DV as well as to ascertain the predictors of HMIS adoption. Those with a Cronbach's value of less than 0.5 were not deemed to be internally consistent and hence were removed from further statistical testing, leaving only the remaining 13 constructs for further statistical analysis.

4.1.2. Hypotheses Test

Regression analysis was used to test the structural links of the research model and the results are presented in Table 6. The estimated coefficients are given along with the associated *p*-value. Calvello (2020) explains the use of correlation and regression as providing a quick and simple summary of the direction and strength of the relationship between two or more numeric values, looking to predict, optimise, or explain a number response between the variables, respectively.

The regression tests show the relationships between DVs and IVs. The closer the *p*-value is to 0, the stronger is the relationship, meaning that it affects adoption. The constructs were taken as the IVs with adoption as the DV. As indicated in Table 6, the tests showed the following 'significant influences' from the tested hypotheses:

1. The more agreement there is that industry and competitive pressure (external pressure) is good for the organization, the higher the adoption of HMIS technologies, p=0.014.
2. The more agreement that current technology gives the organization relative advantage, the higher the adoption of HMIS technologies, p=0.031.
3. The more agreement that there is management support and available skills, the higher the adoption of HMIS technologies, p=0.003.

Table 6. Hypotheses results

Dependent variable	Independent variable	Coefficient		Model			
		Beta	*p*-value	R^2	F	df	*p*-value
Adoption	External support	-0.052	0.570	0.003	0.324	1; 120	0.570
Adoption	Government regulation and policy	-0.060	0.510	0.004	0.437	1; 120	0.510
Adoption	Industry and competitive pressure	-0.223	0.014*	0.050	6.282	1; 120	0.014*
Adoption	Availability	-0.047	0.606	0.002	0.268	1; 120	0.606
Adoption	Relative advantage-organization	-0.196	0.031*	0.038	4.779	1; 120	0.031*
Adoption	Relative advantage-patients	0.044	0.631	0.002	0.232	1; 120	0.631
Adoption	Compatibility	-0.078	0.392	0.006	0.738	1; 120	0.392
Adoption	Observability	0.052	0.571	0.003	0.324	1; 120	0.571
Adoption	Management support and skills	-0.263	0.003*	0.069	8.952	1; 120	0.003*
Adoption	Organizational readiness (general)	0.036	0.692	0.001	0.158	1; 120	0.692
Adoption	Organizational readiness (aim/vision)	-0.111	0.224	0.012	1.497	1; 120	0.224
Adoption	Usefulness and ease of use	-0.003	0.978	<0.0005	0.001	1; 120	0.978
Adoption	Subjective norm	-0.010	0.916	<0.0005	0.011	1; 120	0.916

df, degree of freedom.
*p<0.005.

 https://doi.org/10.1633/JISTaP.2023.11.2.6

4.1.3. Explaining the CSF Questionnaire Results

From the quantitative results, only three factors within the technological, organizational, and environmental context proved to be significant and to have an influence in the adoption of HMIS by public hospitals in Zimbabwe.

First, industry and competitive pressure (p=0.014) produced a significant result, showing that it influences the decision to adopt. This has been demonstrated in various studies, such as that by Zhang et al. (2013), who studied the elements responsible for the adoption of Health Information Technology in US acute care hospitals and found competitive pressure given as market competition as one of the factors. Secondly, management support (p=0.03) was shown to be a critical factor for adoption of HMISs and this is consistent with the studies of Oliveira et al. (2014) and Borgman et al. (2013) in cloud adoption studies. Oliveira stated that by providing financial and other resources, top management would show support for the adoption of cloud computing. The third significant factor for the adoption of HMISs was relative advantage (p=0.031), and this is in line with studies by Premkumar and Roberts (1999) in the adoption of innovative information technologies by enterprises, by Sin Tan et al. (2009) in the adoption of online data access, and by Zhu et al. (2006) in the study of digital transformation for European companies.

These three significant factors have been proven in previous studies while the other hypotheses failed to hold.

Table 7. Themes, codes, and RQ

RQ	Supporting themes	Matching codes
RQ1: What are the environmental factors that influence the adoption of HMIS in Zimbabwean healthcare institutions?	Theme 1: Need for financial and resource support from government and corporates (partners)	External support
	Theme 2: Supporting infrastructure development	Support infrastructure
	Theme 3: Government policies and regulation at the core of adoption	Government support
	Theme 4: Integration of departments and partnerships with outside organizations	Market scope
	Theme 5: Market focus determines choice	Industry and competitive pressure
RQ2: What are the technological factors that influence the adoption of HMIS in Zimbabwean healthcare institutions?	Theme 6: Availability, affordability, and availability of capital	Relative advantage
	Theme 7: Technology brings more benefits than costs	Trialability
	Theme 8: How technology fits into the organization	Technology availability
	Theme 9: Certification of product before implementation	Observability
	Theme 10: Identification and visualisation of benefits	Compatibility
	Theme 11: Inter-dependence of components, routines, individuals, and resources	Complexity
RQ3: What are the organizational factors that influence the adoption of HMIS in Zimbabwean healthcare institutions?	Theme 12: Importance of management activities	Management support
	Theme 13: Size's effects on organizational activities	Size
	Theme 14: Communication processes	Communication processes
	Theme 15: Organizational components readiness	Organizational readiness
	Theme 16: Knowledge on organizational ICT systems	ICT expertise
RQ4: What are the individual factors influencing the adoption of HMIS in Zimbabwean healthcare institutions?	Theme 17: Effort in using the system	PEOU
	Theme 18: Technology to solve the current problems	PU
	Theme 19: Feelings about the technology	Attitude
	Theme 20: Willingness and readiness to use	Intention to use

RQ, research question; HMIS, health management information system; ICT, information and communication technology; PEOU, perceived ease of use; PU, perceived usefulness.

However, this is in line with the previous conclusions of other researchers, who stated that each study setting can have its own set of constructs that are unique to the technology and organization under study. This view is upheld by Oliveira and Martins (2011, p. 112) who stated: "The TOE framework has a solid theoretical basis, consistent empirical support and the potential of application to IS innovation domains, though specific factors identified within the three contexts may vary across different studies." An example of variations within the three contexts of the TOE model was reported by Ruivo et al. (2016), who proposed a model with five constructs: compatibility and training as technological factors; size and best practices as organizational factors; and competitive pressure as an environmental factor. Another example was reported by Chiu et al. (2017) in their study of adoption of broadband by enterprises, which revealed that the significant factors were as follows: relative advantage, trialability, observability, and compatibility within the technological context; information intensity, top management support, employees' knowledge, and absorptive capability within the organizational context; and business partners, competitive pressure, and external support within the environmental context.

4.2. CSF Interview Results

This section presents the results of the interviews for answering the research questions and hypotheses.

After the coding of the responses from the interviewees, the themes and codes were identified and used to answer the research questions, as shown in Table 7. All the identified themes were mapped to one or more identified codes.

The results showed that environmental (external support, support infrastructure, government support, market scope, industry and competitive pressure), technological (relative advantage, trialability, technology availability, observability, compatibility, complexity), organizational (management support, size, communication processes, organizational readiness, ICT expertise) and individual factors (PEOU, PU, attitude, intention to use) do have an influence on the adoption of HMIS in Zimbabwe, as shown in Table 7.

More information on themes, sub-themes, and codes is provided in Table 4. All the codes and themes answer all four research questions as well as the fifth research question ("How can the CSFs be integrated into an adoption model for HMIS in public health institutions in Zimbabwe?"), helping us to come up with a model that explains the factors at play in the organization's adoption of technology, as shown in the Fig. 4.

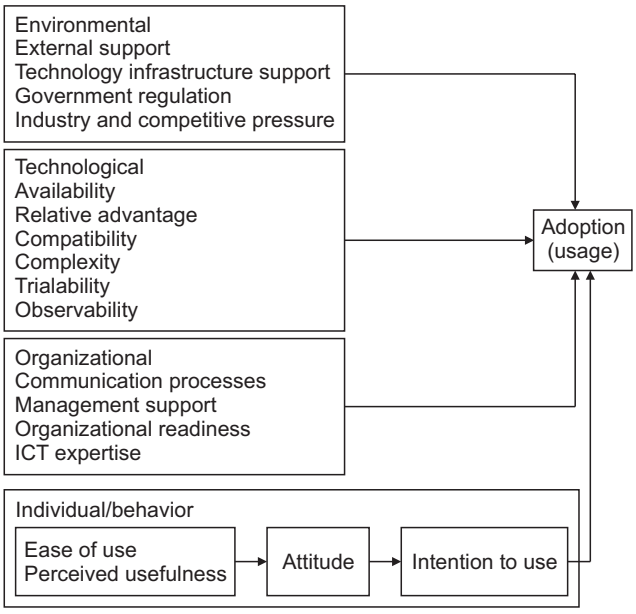


Fig. 4. Final health management information system adoption model. ICT, information and communication technology.

4.3 Model Confirmation Results

Because of the discrepancy between the CSF questionnaire and CSF interview results and in order to achieve completeness, complementarity, and triangulation, another instrument was designed and administered. The results of the two methods of correlation and thematic analyses are presented in the two sections below.

4.3.1. Confirmation Correlation Analysis

The quantitative part of the questionnaire was analysed using the Wilcoxon signed ranks test (non-parametric equivalent to the one-sample t-test), in which the median is tested against the central score of '3' to determine significance of agreement or disagreement.

The tests conducted proved that there is significant agreement that organizational factors are important for adoption of HMISs (communication processes M=4.52, $p<0.0005$; management support M=4.35, $p<0.005$ & M=4.29, $p<0.005$; organizational readiness M=4.48, $p<0.005$ & M=4.39; and ICT expertise M=4.65, $p<0.005$).

The results also show that there is significant agreement that technological factors are important for adoption, as shown by the mean scores and the Wilcoxon signed ranks test (relative advantage M=4.55, $p<0.005$; trialability M=4.29, $p<0.005$; observability M=4.16,

Table 8. Main factor level test results

Factor	Items included	Cronbach's alpha	Mean agreement score	*p*-value (testing for sig agreement/disagreement)
Organizational	2.1-2.6	0.929	4.45	<0.0005
Technological	3.1-3.3, 3.5-3.7, 3.9-3.10	0.753	4.09	<0.0005
Environmental	4.1-4.3, 4.5-4.9	0.844	4.21	<0.0005
Individual	5.1-5.4	0.843	4.12	<0.0005

$p<0.005$ & M=3.84, $p<0.005$; availability M=3.77, $p<005$ & M=4.06, $p<005$; compatibility M=4.29, $p<0.005$ & M=4.26, $p<005$; and technology complexity M=4.13, $p<0.005$ & M=3.94, $p<0.005$).

The environmental factors important for adoption of an HMIS are as follows: external support policies M=4.13, $p<0.005$; government regulation M=4.16, $p<0.005$ & M=4.13, $p<0.005$; industry and competitive pressure M=3.81, $p<0.005$ & M=4.35, $p<0.005$ & M=4.06, $p<0.005$; and technology infrastructure support M=4.26, $p<0.005$ & M=4.35, $p<0.005$ & M=4.19, $p<0.005$.

The individual factors important for adoption of an HMIS are as follows: PU M=4.35, $p<0.005$; PEOU M=3.81, $p<0.005$; attitude M=4.13, $p<0.005$; and intention to use M=4.19, $p<0.005$. Thus, they are important and are integrated in the final model based on these results.

4.3.2. Confirmation of Main Factors Significance

Reliable composite measures were computed for the main factor levels, namely organization, technology, environment, and individual, to measure the significance of the main factors for adoption of HMIS. The results are given in Table 8.

The results show that the participants agreed that the organizational (M=4.45, $p<0.005$), environmental (M=4.21, $p<0.005$), technological (M=4.09, $p<0.005$), and individual (M=4.12, $p<0.005$) factors are important for adoption of HMIS.

The confirmation results support the results from the quantitative and qualitative results, giving the model output shown in Fig. 4.

4.4. Summary of the Results

The results of the CSF questionnaire proved only three factors as significant for the adoption of HMIS, namely, the environmental factor construct of industry and competitive pressure, the technological factor construct of relative advantage, and the organizational factor construct of management support and available skills. The interview data showed that the following factors influenced HMIS adoption in public hospitals in Zimbabwe: the organizational factor constructs of communication processes, management support, organizational readiness, and ICT expertise; the technological factor constructs of technology availability, relative advantage, compatibility, complexity, trialability, and observability; the environmental factor constructs of external support, technology infrastructure support and government regulation; and the individual factor constructs PEOU, PU, attitude, and intention. Because of the discrepancy between the results of the interviews and the questionnaires, a third instrument was designed, including constructs that had been validated in either the interviews or the questionnaire. The set of the constructs were validated as shown in the HMIS model in Fig. 4.

5. DISCUSSION AND RECOMMENDATIONS

Several conclusions can be drawn from the results presented, and these pertain to the research objectives. Although the sample from which data was gathered is small, especially for qualitative data, the researcher believes that the results provide meaningful insights that could be generalised to adoption experiences of HMIS in public health institutions in Zimbabwe.

The idea of adoption, while seemingly straightforward, is a complex process with many factors at play, as has been shown by the unification of different models or theories to try to explain the phenomena. The results of this study have demonstrated that organizational adoption of HMISs can be influenced by technological, organizational, environmental, and individual factors. The technological, organizational, and environmental factors have been proved by the research done by Tornatzky and Fleischer (1990), and it has been shown that the actual antecedents of these broad factors vary from one setting to another as shown by the removal of the sub-constructs of the market scope under (H1) and size under H3. This may be due to the

nature of the industry, where they are driven by government policies rather than industry standards and client demands. This has also been proved by many studies that have used the TOE model, which come up with different antecedents for the three factors. There is a direct path for technological adoption (H2), organizational factors (H3), and environmental factors (H1). This agrees with studies conducted by Kuan and Chau (2001), Ramdani and Kawalek (2007), Zhu and Kraemer (2005), Zhu et al. (2002), and Zhu et al. (2006), all using the TOE model for the adoption of various technologies, although with varying sub-constructs depending on the technology and other business settings.

Baker (2012) suggests that there is a need to explore ways to evolve the TOE model and find out how individual behaviors and adoption can affect organizational adoption. Accordingly, the researcher included individual factors and they have been proved to have an indirect effect on technological decision making through attitude (H4), which has a direct effect on intention to use (H7), which directly influences decisions on technology adoption. The individual factors from TRA and TRB were shown to affect the adoption of innovation at the organizational level. This is shown by the acceptance of the PEOU and PU in the research results. This shows that there is an interplay of more factors that affect how organizations adopt technology than those stated in the TOE model. Thus, although the TOE model can be useful in interpreting how organizations adopt IT innovations, it does not fully show how the individual behaviors of members can contribute to the adoption matrix. There have been many studies that try to complement the TOE model, but none have used individual/behavioral factors as in this research study.

Most adoption/acceptance studies have separated individual/behavioral studies from organizational studies, with the assumption that adoption in an organization is either on an organizational level or an individual level. At the organizational level, the studies have assumed that when an organization adopts an innovation, the employees have no choice but to use what the organization has adopted; with behaviorbehavioral studies, the assumption is that the use of the technology is not mandatory, and users have the choice to use or reject. In the studies carried out previously, the two adoption phases are recognized as two distinct processes, which do not interact. However, it appears that the two are related and should be treated as a continuum and that, when studies are carried out, they should be undertaken at the same time, especially for mandatory environments. The success of an adoption of an innovation should be measured by implementation and continued use. In this way, the behavioral aspects of the use of the technology would be known and could reduce the chances of the use of the technology being discontinued after adoption (Weng et al., 2018).

Researchers should study the adoption of organizational innovations with individual acceptance in mind, as individual employees can cause disruptive tendencies after organizations adopt a technological innovation. Managers of adoption should also assess all possible factors that can affect the success of the adoption process and should not concentrate on technological, organizational, and environmental factors only but should, in addition, assess the behaviorbehavioral factors of the individuals within the organization. The traditional approach of separating organizational from individual adoption and treating them as primary and secondary should be discontinued to allow the success rate of technological adoption to increase and, more importantly, to enable the impact of adopted technologies for the health sector to result in improved organizational processes and ultimately high-quality patient care.

Hospitals are recommended to implement the HMIS technology as it allows for improved service delivery. The use of HMIS will allow for streamlined resource allocation, effective management and monitoring, and consequently its use will allow for improved services to the patients. Resources are a challenge and hospitals can come up with fundraising initiatives such as billing the patients towards the implementation of the HMIS, so as to improve service delivery through technology. There is a need to integrate the use of ICTs in the curriculum of health professionals' training so as to improve their chances of accepting technology when it is integrated into their environments.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl, & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11-39). Springer.

Alam, M. G., Masum, A. K., Beh, L. S., & Hong, C. S. (2016). Critical factors influencing decision to adopt human resource information system (HRIS) in Hospitals. *PLoS*

One, 11(8), e0160366. https://doi.org/10.1371/journal.pone.0160366

Alwan, K., Awoke, T., & Tilahun, B. (2015). Knowledge and utilization of computers among health professionals in a developing country: A cross-sectional study. *JMIR Human Factors,* 2(1), e4. https://doi.org/10.2196/humanfactors.4184

Ariani, A., Koesoema, A. P., & Soegijoko, S. (2017). Innovative healthcare applications of ICT for developing countries. In H. Qudrat-Ullah, & P. Tsasis (Eds.), *Innovative healthcare systems for the 21st century* (pp. 15-70). Springer.

Asemahagn, M. A. (2017). Determinants of routine health information utilization at primary healthcare facilities in Western Amhara, Ethiopia. *Cogent Medicine,* 4, 1387971. https://doi.org/10.1080/2331205X.2017.1387971

Bacigalupe, G. (2011). Is there a role for social technologies in collaborative healthcare? *Families, Systems & Health,* 29(1), 1-14. https://doi.org/10.1037/a0022093

Baker, J. (2012). The technology–organization–environment framework. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), *Information systems theory: Explaining and predicting our digital society* (pp. 231-245). Springer.

Borgman, H. P., Bahli, B., Heier, H., & Schewski, F. (2013, January 7-10). Cloudrise: Exploring cloud computing adoption and governance with the TOE framework. In R. H. Sprague, Jr. (Eds.), *Proceedings of the 46th Hawaii International Conference on System Sciences* (pp. 4425-4435). IEEE.

Calvello, M. (2020). *Correlation vs. regression made easy: Which to use + why. https://online210.psych.wisc.edu/wp-content/uploads/PSY-210_Unit_Materials/PSY-210_Unit12_Materials/Calvello_Regression_2020.pdf*

Chatterjee, M. (2015). *Towards more efficient service for clients, using HMIS system: Managerial audit of community technology alliance.* (Master's thesis). San Jose State University, San Jose, CA, USA.

Chin, W. W., Gopal, A., & Salisbury, W. D. (1997). Advancing the theory of adaptive structuration: The development of a scale to measure faithfulness of appropriation. *Information Systems Research,* 8(4), 342-367. http://www.jstor.org/stable/23010917

Chiu, C. Y., Chen, S., & Chen, C. L. (2017). An integrated perspective of TOE framework and innovation diffusion in broadband mobile applications adoption by enterprises. *International Journal of Management, Economics and Social Sciences,* 6(1), 14-39. http://hdl.handle.net/10419/157921

Christodoulakis, C., Asgarian, A., & Easterbrook, S. (2017, November 6-8). Barriers to adoption of information technology in healthcare. In J. Featherstone, K. Lyons, J. Wigglesworth & D. Saraf (Eds.), *Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering (CASCON '17)* (pp. 66-75). IBM.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Crowe, M., Inder, M., & Porter, R. (2015). Conducting qualitative research in mental health: Thematic and content analyses. *Australian and New Zealand Journal of Psychiatry,* 49(7), 616-623. https://doi.org/10.1177/0004867415582053

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. https://doi.org/10.2307/249008

De Smith, M. J. (2015). *Statistical analysis handbook: A comprehensive handbook of statistical concepts, techniques and software tools* (2018 ed.). Drumlin Publications.

Dixon, D. R., & Dixon, B. J. (1994, November 5-9). Adoption of information technology enabled innovations by primary care physicians: Model and questionnaire development. In J. G. Ozbolt (Ed.), *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care* (pp. 631-635). American Medical Informatics Association.

Dolan, S. B., Alao, M. E., Mwansa, F. D., Lymo, D. C., Bulula, N., Carnahan, E., Beylerian, E., Werner, L., & Shearer, J. C. (2020). Perceptions of factors influencing the introduction and adoption of electronic immunization registries in Tanzania and Zambia: A mixed methods study. *Implementation Science Communications,* 1, 38. https://doi.org/10.1186/s43058-020-00022-8

Eger, M. S., Godkin, R. L., & Valentine, S. R. (2001). Physicians' adoption of information technology: A consumer behavior approach. *Health Marketing Quarterly,* 19(2), 3-21. https://doi.org/10.1300/J026v19n02_02

El Kandri, M. (2018). *The challenges of technology adoption in healthcare.* https://pharmaboardroom.com/articles/the-challenges-of-technology-adoption-in-healthcare/

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Addison-Wesley Publishing Company.

Furusa, S. S., & Coleman, A. (2018). Factors influencing e-health implementation by medical doctors in public hospitals in Zimbabwe. *South African Journal of Information Management,* 20(1), a928. https://doi.org/10.4102/sajim.v20i1.928

Garth, A. (2008). *Analysing data using SPSS.* Sheffield Hallam University.

Herrick, D. M., Gorman, L., & Goodman, J. C. (2010). *Health information technology: Benefits and problems. Policy report No. 327*. National Center for Policy Analysis.

Hung, S. Y., Hung, W. H., Tsai, C. A., & Jiang, S. C. (2010). Critical factors of hospital adoption on CRM system: Organizational and information system perspectives. *Decision*

Support Systems, 48(4), 592-603. https://doi.org/10.1016/j.dss.2009.11.009

Ifinedo, P. (2018). Empirical study of Nova Scotia nurses' adoption of healthcare information systems: Implications for management and policy-making. *International Journal of Health Policy and Management,* 7(4), 317-327. https://doi.org/10.15171/ijhpm.2017.96

International Center for AIDS Care and Treatment Programs at Columbia University. (2015). *In Lesotho, new health information system provides streamlined, integrated data across health programs.* https://icap.columbia.edu/news-events/in-lesotho-new-health-information-system-provides-streamlined-integrated-data-across-health-programs/

Khan, A., & Woosley, J. M. (2011). Comparison of contemporary technology acceptance models and evaluation of the best fit for health industry organizations. *International Journal of Computer Science Engineering and Technology,* 1(11), 709-717. https://ijcset.net/docs/Volumes/volume1issue11/ijcset2011011106.pdf

Khobi, J. A. M., Mtebe, J. S., & Mbelwa, J. T. (2020). Factors influencing district health information system usage in Sierra Leone: A study using the technology-organization-environment framework. *Electronic Journal of Information Systems in Developing Countries,* 86(6), e12140. https://doi.org/10.1002/isd2.12140

Kralewski, J. E., Dowd, B. E., Cole-Adeniyi, T., Gans, D., Malakar, L., & Elson, B. (2008). Factors influencing physician use of clinical electronic information technologies after adoption by their medical group practices. *Health Care Management Review,* 33(4), 361-367. https://doi.org/10.1097/01.HCM.0000318773.67395.ce

Kuan, K. K. Y., & Chau, P. Y. K. (2001). A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. *Information & Management,* 38(8), 507-521. https://doi.org/10.1016/S0378-7206(01)00073-8

Landau, S., & Everitt, B. S. (2004). *A handbook of statistical analyses using SPSS.* Chapman & Hall.

Macharia, J., & Maroa, C. (2014, May 7-9). Health management information systems (HMIS) implementation characteristics that influence the quality of healthcare in private hospitals in Kenya. *Proceedings of the 2014 IST-Africa Conference* (pp. 1-12). IEEE.

Mapesa, N. M. (2016). *Health information technology implementation strategies in Zimbabwe.* (Doctoral dissertation). Walden University, Minneapolis, MN, USA.

Marsan, J., & Paré, G. (2013). Antecedents of open source software adoption in health care organizations: A qualitative survey of experts in Canada. *International Journal of Medical Informatics,* 82(8), 731-741. https://doi.org/10.1016/j.ijmedinf.2013.04.001

Miraz, M. H., & Habib, M. M. (2016). ICT adoption in small and medium enterprises: An empirical evidence of service sectors in Bangladesh. *Journal of Economics, Business and Management,* 4(8), 482-485. https://doi.org/10.18178/joebm.2016.4.8.439

Ministry of Health and Child Welfare (MOHCW). (2009). Zimbabwe health information system strategy: 2009-2014. MOHCW.

Njeri, M. N. (2018). *Factors influencing adoption of health management information system: Case of Kiambu level 5 hospital, Kiambu County Kenya.* (Master's thesis). University of Nairobi, Nairobi, Kenya.

Oliveira, T., & Martins, M. F. (2011). Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation,* 14(1), 110-121. https://academic-publishing.org/index.php/ejise/article/view/389

Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management,* 51(5), 497-510. https://doi.org/10.1016/j.im.2014.03.006

Park, Y. T., & Lee, J. (2014). Factors affecting electronic medical record system adoption in small Korean hospitals. *Healthcare Informatics Research,* 20(3), 183-190. https://doi.org/10.4258/hir.2014.20.3.183

Premkumar, G., & Roberts, M. (1999). Adoption of new information technologies in rural small businesses. *Omega,* 27(4), 467-484. https://doi.org/10.1016/S0305-0483(98)00071-1

Proscoviour, V. (2017). *Adoption factors of hospital management information systems in selected hospitals in Meru County, Kenya.* (Master's thesis). Kenyatta University, Nairobi, Kenya.

Ramdani, B., & Kawalek, P. (2007, June 14-16). SME Adoption of Enterprise Systems in the Northwest of England. In T. McMaster, D. Wastell, E. Ferneley, & J. I. DeGross (Eds.), *Proceedings of the IFIP TC8 WG 8.6 International Working Conference - Organizational Dynamics of Technology-Based Innovation: Diversifying the Research Agenda* (pp. 409-429). Springer.

Rogers, E. M. (1962). *Diffusion of innovations.* Free Press of Glencoe.

Ross, D. S., & Venkatesh, R. (2016). Role of hospital information systems in improving healthcare quality in hospitals. *Indian Journal of Science and Technology,* 9(26), 1-5. https://doi.org/10.17485/ijst/2016/v9i26/92686

Ross, J., Stevenson, F., Lau, R., & Murray, E. (2016). Factors

that influence the implementation of e-health: A systematic review of systematic reviews (an update). *Implementation Science,* 11(1), 146. https://doi.org/10.1186/s13012-016-0510-7

RTI International. (2014). *Zimbabwe launches new health information system (HIS).* http://static.ictedge.org/node/975.html

RTI International. (2021). *Improving health service delivery through a national health information system and mobile-based disease reporting.* https://www.rti.org/impact/zimbabwe-health-information-and-support-project-zimhisp

Ruivo, P., Rodrigues, J., Johansson, B., Oliveira, T., & Rebelo, J. (2016). Using TOE and RBV theories to define a theoretical model to assess ERP value across Iberian manufacturing and services SMEs. *Procedia Computer Science,* 100, 474-479. https://doi.org/10.1016/j.procs.2016.09.184

Sereno. (2021). *Comparison of Pearson and Spearman correlation coefficients.* https://www.analyticsvidhya.com/blog/2021/03/comparison-of-pearson-and-spearman-correlation-coefficients/

Shin, D. H. (2013). User centric cloud service model in public sectors: Policy implications of cloud services. *Government Information Quarterly,* 30(2), 194-203. https://doi.org/10.1016/j.giq.2012.06.012

Sin Tan, K., Choy Chong, S., Lin, B., & Cyril Eze, U. (2009). Internet-based ICT adoption: Evidence from Malaysian SMEs. *Industrial Management & Data Systems,* 109(2), 224-244. https://doi.org/10.1108/02635570910930118

Tan, J. (2009). Foundation concepts of health management information systems. In J. Tan, & F. C. Payton (Eds.), *Adaptive health management information systems: Concepts, cases, & practical applications* (pp. 3-22). Jones and Bartlett Publishers.

Techzim. (2014). *Zim hospitals introduce a new e-health system: How does it work?* https://www.techzim.co.zw/2014/10/zim-hospitals-introduce-new-e-health-system-work

Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation.* Lexington Books.

United Nations Development Programme (UNDP). (2014a). *Case study - Electronic patient management system ePMS - Zimbabwe*. UNDP.

United Nations Development Programme (UNDP). (2014b). *Electronic patient management system ePMS - Zimbabwe.* https://www.undp.org/sites/g/files/zskgke326/files/migration/zw/UNDP_ZW_HIVAIDS_epms_zimbabwe.pdf

United Nations Development Programme (UNDP). (2020). *Healthy systems for healthy people: UNDP and Global Fund support to strengthen the national health management information system*. https://www.undp.org/sites/g/files/zskgke326/files/migration/zw/UNDP_ZW_GF_HMIS.pdf

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Wang, B. B., Wan, T. T., Burke, D. E., Bazzoli, G. J., & Lin, B. Y. (2005). Factors influencing health information system adoption in American hospitals. *Health Care Management Review,* 30(1), 44-51. https://doi.org/10.1097/00004010-200501000-00007

Wanyoike, D. M., Mukulu, E., & Waititu, A. G. (2012). ICT attributes as determinants of e-commerce adoption by formal small enterprises in urban Kenya. *International Journal of Business and Social Science,* 3(23), 65-74. https://www.researchgate.net/publication/262794622_ICT_Attributes_as_Determinants_of_E-commerce_Adoption_by_Formal_Small_Enterprises_in_Urban_Kenya

Weng, F., Yang, R. J., Ho, H. J., & Su, H. M. (2018). A TAM-based study of the attitude towards use intention of multimedia among school teachers. *Applied System Innovation,* 1(3), 36. https://doi.org/10.3390/asi1030036

Zhang, N. J., Seblega, B., Wan, T., Unruh, L., Agiro, A., & Miao, L. (2013). Health information technology adoption in U.S. acute care hospitals. *Journal of Medical Systems,* 37(2), 9907. https://doi.org/10.1007/s10916-012-9907-2

Zhou, L. L., Owusu-Marfo, J., Asante Antwi, H., Antwi, M. O., Kachie, A. D. T., & Ampon-Wireko, S. (2019). Assessment of the social influence and facilitating conditions that support nurses' adoption of hospital electronic information management systems (HEIMS) in Ghana using the unified theory of acceptance and use of technology (UTAUT) model. *BMC Medical Informatics and Decision Making,* 19(1), 230. https://doi.org/10.1186/s12911-019-0956-z

Zhu, K., & Kraemer, K. L. (2005). Post-adoption variations in usage and value of e-business by organizations: Cross-country evidence from the retail industry. *Information Systems Research,* 16(1), 61-84. http://www.jstor.org/stable/23015765

Zhu, K., Dong, S., Xu, S. X., & Kraemer, K. L. (2006). Innovation diffusion in global contexts: Determinants of post-adoption digital transformation of European companies. *European Journal of Information Systems,* 15(6), 601-616. https://doi.org/10.1057/palgrave.ejis.3000650

Zhu, K., Kraemer, K. L., & Xu, S. (2002, December 15-18). A cross-country study of electronic business adoption using the technology-organization-environment framework. *Proceedings of the 23rd nternational Conference on Information Systems* (pp. 337-348). Center for Research on Information Technology and Organizations.

Call for Paper

Journal of Information Science Theory and Practice (JISTaP)

We would like to invite you to submit or recommend papers to Journal of Information Science Theory and Practice (JISTaP, eISSN: 2287-4577, pISSN: 2287-9099), a fast track peer-reviewed and no-fee open access academic journal published by Korea Institute of Science and Technology Information (KISTI), which is a government-funded research institute providing STI services to support high-tech R&D for researchers in Korea. JISTaP marks a transition from Journal of Information Management to an English-language international journal in the area of library and information science.

JISTaP aims at publishing original studies, review papers and brief communications on information science theory and practice. The journal provides an international forum for practical as well as theoretical research in the interdisciplinary areas of information science, such as information processing and management, knowledge organization, scholarly communication and bibliometrics.

We welcome materials that reflect a wide range of perspectives and approaches on diverse areas of information science theory, application and practice. Topics covered by the journal include: information processing and management; information policy; library management; knowledge organization; metadata and classification; information seeking; information retrieval; information systems; scientific and technical information service; human-computer interaction; social media design; analytics; scholarly communication and bibliometrics. Above all, we encourage submissions of catalytic nature that explore the question of how theory can be applied to solve real world problems in the broad discipline of information science.

Co-Editors in Chief : Gary Marchionini & Dong-Geun Oh

Please click the "Online Submission" link in the JISTaP website (http://www.jistap.org), which will take you to a login/account creation page. Please consult the "Author's Guide" page to prepare your manuscript according to the JISTaP manuscript guidelines.

Any question? Seonhee Lee (managing editor) : jistap@kisti.re.kr

JISTaP
Journal of Information Science Theory and Practice

Information for Authors

The Journal of Information Science Theory and Practice (JISTaP), which is published quarterly by the Korea Institute of Science and Technology (KISTI), welcomes materials that reflect a wide range of perspectives and approaches on diverse areas of information science theory, application and practice. JISTaP is an open access journal run under the Open Access Policy. See the section on Open Access for detailed information on the Open Access Policy.

A. Originality and Copyright

All submissions must be original, unpublished, and not under consideration for publication elsewhere. Once an article is accepted for publication, all papers are accessible to all users at no cost. If used for other researches, its source should be indicated in an appropriate manner and the content can only be used for uncommercial purpose under Creative Commons license.

B. Peer Review

All submitted manuscripts undergo a single-blind peer review process in which the identities of the reviewers are withheld from the authors.

C. Manuscript Submission

Authors should submit their manuscripts online via Article Contribution Management System (ACOMS). Online submission facilitates processing and reviewing of submitted articles, thereby substantially shortening the paper lifecycle from submission to publication. After checking the manuscript's compliance to the Manuscript Guidelines, please follow the "Online Submission" hyperlink in the top navigation menu to begin the online manuscript submission process.

D. Open Access and Creative Commons License Statement

All JISTaP content is open access, meaning it is accessible online to everyone, without fee and authors' permission. Open access articles are automatically archived in the Korea Institute of Science and Technology Information (KISTI)'s Open Access repository (KPubS).

All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as long as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

E. Manuscript Guidelines

Manuscripts that do not adhere to the guidelines outlined below will be returned for correction. Please read the guidelines carefully and make sure the manuscript follows the guidelines as specified. We strongly recommend that authors download and use the manuscript template in preparing their submissions.

1. Page Layout :

All articles should be submitted in single column text on standard Letter Size paper (21.59 × 27.94 cm) with normal margins.

2. Length :

Manuscripts should normally be between 4,500 and 9,000 words (10 to 20 pages).

3. File Type :

Articles should be submitted in Microsoft Word format. To facilitate the manuscript preparation process and speed up the publication process, please use the manuscript template.

4. Text Style :

- Use a standard font (e.g., Times New Roman) no smaller than size 10.
- Use single line spacing for paragraphs.
- Use footnotes to provide additional information peripheral to the text. Footnotes to tables should be marked by superscript lowercase letters or asterisks.

5. Title Page :

The title page should start with a concise but descriptive title and the full names of authors along with their affiliations and contact information (i.e., postal and email addresses). An abstract of 150 to 250 words should appear below the title and authors, followed by keywords (4 to 6).

Author1
Affilliation, Postal Address. E-mail
Author2
Affilliation, Postal Address. E-mail

ABSTRACT
A brief summary (150-250 words) of the paper goes here.

Keywords: 4 to 6 Keywords, separated by commas.

6. Numbered Type :

1. INTRODUCTION

All articles should be submitted in single column text on standard letter size paper (21.59 × 27.94 cm) with normal margins[1. Text should be in 11-point standard font (e.g., Times New Roman) with single line spacing. [1 Normal margin dimensions are 3 cm from the top and 2.54 cm from the bottom and sides.

2. SECTIONS

The top-level section heading should be in 14-point bold all uppercase letters.

2.1. Subsection Heading 1

The first-level subsection heading should be in 12-point bold with the first letter of each word capitalized.

2.1.1. Subsection Heading 2

The second-level subsection heading should be in 11-point italic with the first letter of each word capitalized.

7. Figures and Tables :

All figures and tables should be placed at the end of the manuscript after the reference list. To note the placement of figures and tables in text, "Insert Table (or Figure) # here" should be inserted in appropriate places. Please use high resolution graphics whenever possible and make sure figures and tables can be easily resized and moved.

Figure

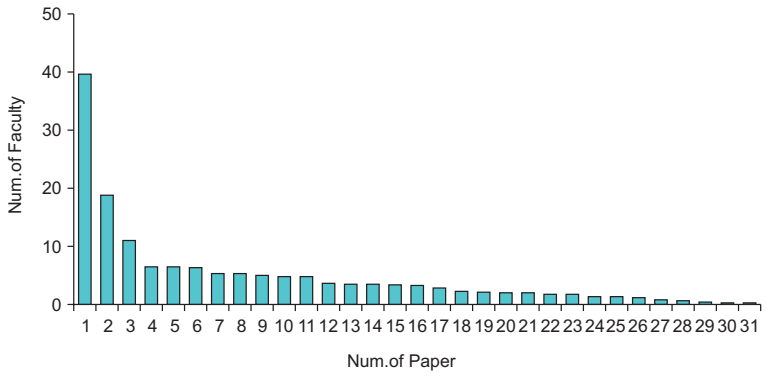


Fig. 1. Distribution of authors over publication count.

Table

Table 1. The title of table goes here

Study	Time period study	Data
Smith Wesson (1996)	1970 - 1995	684 papers in 4 SSCI journals
Reeves [a (2002)	1997 - 2001	597 papers in 3 SSCI journals
Jones Wilson [b (2011)	2000 - 2009	2,166 papers in 4 SSCI journals

[a Table footnote a goes here
[b Table footnote b goes here

8. Acknowledgements :
Acknowledgements should appear in a separate section before the reference list.

9. Conflicts of Interest :
Authors of manuscripts must disclose any potential conflicts of interest at the time of submission. Statements on conflict of interest have no influence on the editorial decision to publish.

10. Citations :
Citations in text should follow the author-date method (authors' surname followed by publication year).
- Several studies found... (Barakat et al., 1995; Garfield, 1955; Meho & Yang, 2007).
- In a recent study (Smith & Jones, 2011)...
- Smith and Jones (2011) investigated...

11. Reference List :
Reference list, formatted in accordance with the American Psychological Association (APA) style, should be alphabetized by the first authors last name.

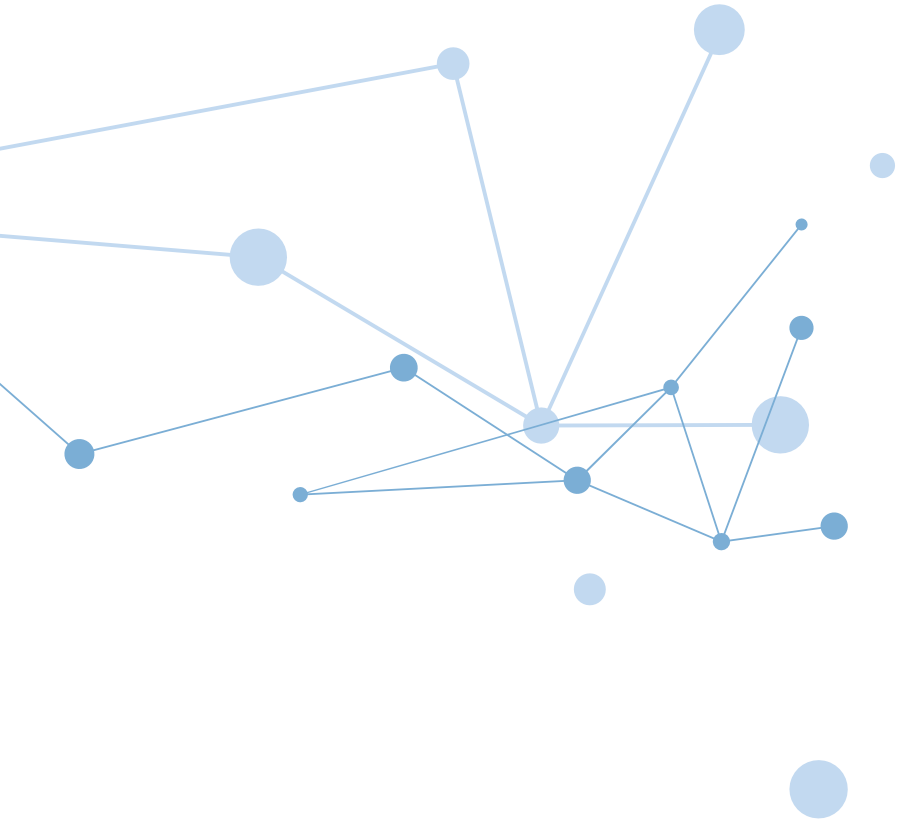
Journal article
- Author, A., Author, B. & Author, C. (Year). Article title. *Journal Title*, volume(issue), start page-end page.
- Smith, K., Jones, L. J., & Brown, M. (2012). Effect of Asian citation databases on the impact factor. *Journal of Information Science Practice and Theory*, 1(2), 21-34.

Book
- Author, A., & Author, B. (Year). *Book title*. Publisher Name.
- Smith, K., Jones, L. J., & Brown, M. (2012). *Citation patterns of Asian scholars*. Sage.

Book chapter
- Author, A., & Author, B. (Year). Chapter title. In A. Editor, B. Editor, & C. Editor (Eds.), *Book title* (pp. xx-xx). Publisher Name.
- Smith, K. & Brown, M. (2012). Author impact factor by weighted citation counts. In G. Martin (Ed.), *Bibliometric approach to quality assessment* (pp. 101-121). Springer.

Conference paper
- Author, A., & Author, B. (Year, Month Day). Article title. In A. Editor & B. Editor (Eds.), *Conference title* (pp. xx-xx). Publisher Name.
- Lin, H., Sun, B., Wu, J. & Xiong, H. (2016, June 24-26). Topic detection from short text: A term-based consensus clustering method. In B. Yang (Ed.), *2016 13th International Conference on Service Systems and Service Management (ICSSSM 2016)* (pp. 1-6). IEEE.

Online document
- Author, A., & Author, B. (Year). *Article title*. URL.
- Smith, K. & Brown, M. (2010). The future of digital library in Asia. *Digital Libraries*, 7,111-119. http://www.diglib.org/publist.htm.



JISTaP

Journal of Information
Science Theory and Practice

KISTI www.kisti.re.kr
66, Hoegi-ro, Dongdaemun-gu, Seoul, Republic of Korea
(T) +82-2-3299-6105 (F) +82-2-3299-6067
http://www.jistap.org