Digital Divide: A Bibliometric Approach to Existing Knowledge

Jorge Cruz-Cárdenas
ESTec Research Center
Universidad Indoamérica
Quito, Ecuador
jorgecruz@uti.edu.ec

Andrés Palacio-Fierro
ESTec Research Center
Universidad Indoamérica
Quito, Ecuador
andrespalacio@uti.edu.ec

Ekaterina Zabelina
Department of Psychology
Chelyabinsk State University
Chelyabinsk, Russia
katya_k@mail.ru

Olga Deyneka
Department of Political Psychology
St. Petersburg State University
St. Petersburg, Russia
osdeyneka@yandex.ru

Carlos Ramos-Galarza
Centro de Investigación Mist
Universidad Indoamérica
Quito, Ecuador
carlosramos@uti.edu.ec

***Abstract*—The topic of digital divide—the gap between those who have and those who do not have access to digital technologies—is of growing interest for both public and private sectors. This interest is reflected in an increasing number of academic and scientific publications focused on the topic. This study presents an overview of the digital divide research using a bibliometric approach. Analyzing 9,523 documents from the Scopus database, this study found that 1) there has been an accelerated growth in the publication rate on digital divide since 2018; 2) digital divide is a multidisciplinary field, led by the United States and other developed countries and some emerging countries; 3) the dominant term used to refer to the field, as per our analysis of the relevant keywords, is "digital divide"; and 4) as per our analysis of the evolution of keywords' importance, while topics related to technologies were of interest several years ago, current interest revolves around the effects of the Covid-19 pandemic, the elderly, education, and healthcare access.**

Keywords— Digital Divide, ICTs, Covid, Bibliometric Analysis

I. INTRODUCTION

Digital divide is the gap that exists between those who have access to technologies, particularly information and communication technologies (ICTs), and those who do not [1]. To measure digital divide, various products and services associated with ICTs are taken as parameters. Some studies measured it based on internet access; some were more specific in their focus, examining, for example, access to a fixed broadband connection; and some others focused on a combination of devices and services, such as mobile phones [2].

The United Nations (UN) alludes to the digital divide in one of its sustainable development goals (SDGs) [3], specifically in SDG number 10, which is called *reduced inequalities*. Recently, the COVID-19 pandemic, which lasted for around three years, augmented the state of digital gaps and the state of digital exclusion of broad sectors of the population, particularly the poorest and the elderly [4].

The concept of digital divide can be applied to different degrees of individual aggregation [1], [4]. At the macro level, the focus is on the digital divide between continents or countries; at the meso level, on the aggregate differences within a country (e.g., between cities, in rural–urban dimensions, between social classes, between genders, between industries, etc.); and at the micro level, on organizations, groups, and individuals (e.g., the existing gaps between educational institutions, between students in a class, etc.).

Digital divide can also be conceived as a series of levels [5], [6]: the first level is the gap in physical or material access to technologies; the second, the gap in the skills and abilities needed to use technologies; and the third, the gaps in the benefits received from using technologies.

Studying digital divide is of great interest to the public and private sectors and the for-profit and non-profit sectors. For the public sector, it is crucial to understand the most disadvantaged groups when it comes to the digital divide and the reasons for it in order to conduct campaigns to bridge the divide [3]. These campaigns can be supported by non-profit organizations. On the other hand, the lucrative sector, represented by companies, must understand the digital gap among the population to better direct their products [7].

Digital divide's immense significance and impact has attracted a growing academic and scientific interest, which is reflected in a rapidly growing number of publications [5], [6]. Since such large volumes of knowledge demand corresponding efforts of organization and systematization, this study focused on the following general objective:

- To conduct an analysis of the activity and generate the mapping of science in the area of digital divide using bibliometric techniques.

This paper is divided into several sections. The *Introduction* section is followed by the *Method* section, where the bibliometric method is explained and implemented. In the *Analysis and Results* section, two types of results are presented: activity measurement and science mapping. Finally, in the *Discussion and Implications* section, the most important results are discussed in comparison with that of other studies, and the implications for practice and for future studies are established.

II. METHOD

When aiming to describe and organize the existing knowledge in a certain field, the methodology should be based on the volume of literature generated in that field. Bibliometric analysis is apt if the volume is large [8], and systematic literature review if the volume is small. Since the published documents on digital divide reaches a figure in the range of thousands [6], bibliometric analysis was chosen.

979-8-3503-3823-2/23/$31.00 ©2023 IEEE

The current study was deployed in several stages, very frequent in bibliometric studies and literature reviews [8], [9]: 1) formulating the study objective; 2) crafting strategies for access to information; and 3) analyzing and discussing the results.

After establishing the study objective in the *Introduction* section, the next step was to decide the strategies for searching information, which involves several decisions [8]: selecting the database, structuring the search chain, selecting the inclusion/exclusion criteria, and recovering the document body.

The Scopus database was selected, as it is reliable, widely used, has great coverage, and includes almost twice as many journals with impact metrics as the other large database, Web of Science, [10].

The search string—adopted from Lythreatis et al. [6], who developed it to conduct a systematic review of the literature on digital divide and achieved excellent results—was ("digital divide" OR "digital inequality" OR "digital inequalities" OR "digital gap" OR "digital division" OR "digital disparities"). The search, conducted in May 2023, yielded 10,542 documents, and two filters or delimitations were applied to them. The first involved selecting articles, conference papers, books, and book chapters and excluding editorials and reviews (they are, generally, less subject to major editorial reviews and shorter in length). The second delimitation involved selecting English-only documents (which were, in fact, the vast majority), as documents must have a common language to map science [11]. No temporal delimitation was made. After applying the filters and delimitations, 9,523 documents were selected.

III. Analysis and Results

The first part of data analysis focused on the academic/scientific activity in the field of digital divide. The second part focused on mapping science.

First, the evolution of scientific production over time was analyzed.

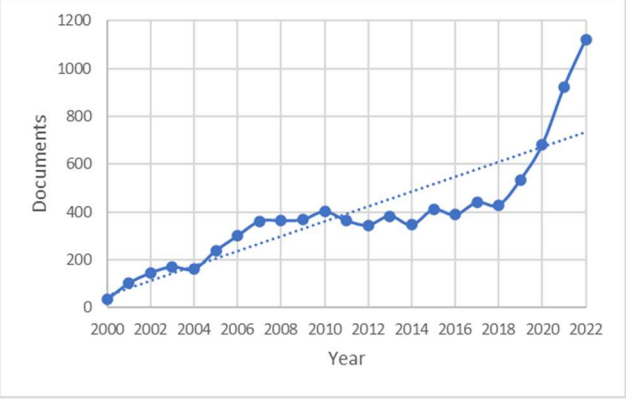

Fig. 1. Annual evolution of the number of published documents (Source: Scopus)

As can be seen in Fig.1, the long-term trend line (dotted line) is increasing. Already in the curve of the number of documents over time, three stages can be clearly observed. The first stage—2000 to 2006—is characterized by moderate growth. In the second stage—2007 to 2017—the annual number of documents remains relatively constant. Finally, in the third stage—2018 to 2022—there is an accelerated growth in the number of published documents, with 1,122 articles published in 2022 alone. However, these figures—it is necessary to clarify—do not necessarily mean that digital divide has deepened but that interest in studying it has increased.

Second, the type of publications were analyzed, and it was found that articles were the main publication format, with 6,003 documents (63.0%) in the analysis period (until May 2023), followed by conference papers, with 2,307 documents (24.2%), then book chapters, with 1,056 documents. (11.1%), and, finally, books, with only 157 documents (1.6%).

Analyzing the fields of publication, it was found that the social sciences dominated, with 5,055 documents (53.1%), followed by computer science, with 3,805 documents (40.0%). None of the other fields exceeded 15%. It should be noted that the same document can be linked to several fields of knowledge simultaneously. Thus, evidently, digital divide research is multidisciplinary, fed by knowledge from the social sciences and computer sciences (a trait common to other topics related to society and technology as well) [7].

Finally, the participation of countries was estimated based on the authors' institutions of affiliation. The United States stood out, being present in 2,793 documents (29.3%). The other countries on the list were the United Kingdom, Spain, Australia, India, and China, their percentages of presence between 4% and 10%. It is worth clarifying that, since co-authorship and multiple affiliations are quite frequent, one document could be linked to more than one country. The above result reflects the leading role of developed countries and the so-called emerging countries in the digital divide research and the much smaller role of developing countries. Such a finding has also been observed in other topics concerning society and technology [12].

In the second part of data analysis, to map the science of the digital divide, two types of analyses were conducted—the first to identify the main keywords using the bibliometric package for R language [13], and the second to establish the historical evolution of the keywords' importance using VOSviewer 1.6.19 [11].

In total, 31,184 keywords were identified. The five most prominent keywords and their number of occurrences are shown in Table I.

Table I. Most frequent keywords

Keyword	No. of occurrences	%
Digital divide	5697	18.3%
Internet	1812	5.8%
ICTs	873	2.8%
Older adult/aged	570	1.8%
Covid	406	1.3%

Evident from Table I, "digital divide" is the most frequently used keyword by authors and indexers, indicating that this is the dominant name of the field of study. The next most-used keywords are "internet" and "ICTs", which are the two most-used technologies to measure first-level digital divide [2]. These keywords are followed by "Older adult/aged", which is currently one of the populations most affected by digital divide [14]. Finally, the keyword "Covid"

refers to the most important disruptive event of the last decade [7].

Though such an analysis offers a global picture of keywords, this is also a static picture, one based on cumulative keyword usage over a period of nearly 25 years of research. Therefore, from a dynamic perspective, an overlay visualization of keyword analysis, one focused on the evolution of keywords' importance, was conducted. This analysis presents a mapping of science.

For the overlay visualization analysis, VOSviewer 1.6.19 [11] was used. Considering the large volume of keywords (31,184), it was specified that a keyword should appear at least 30 times to be considered [12]. Additionally, the terms that could interfere with the interpretation of results were eliminated [11]. The keywords referring to geographical locations (e.g., Africa, Australia, etc.) and associated with the research methodology or with the characteristics of the document (e.g., survey, in-depth interviews, articles, etc.) were eliminated. Moreover, the keywords with similar meanings were grouped together as in a thesaurus (e.g., "ICT," "ICTS," and "information and communication technologies" were grouped as ICTS). The overlay visualization map thus obtained is presented in Fig. 2.

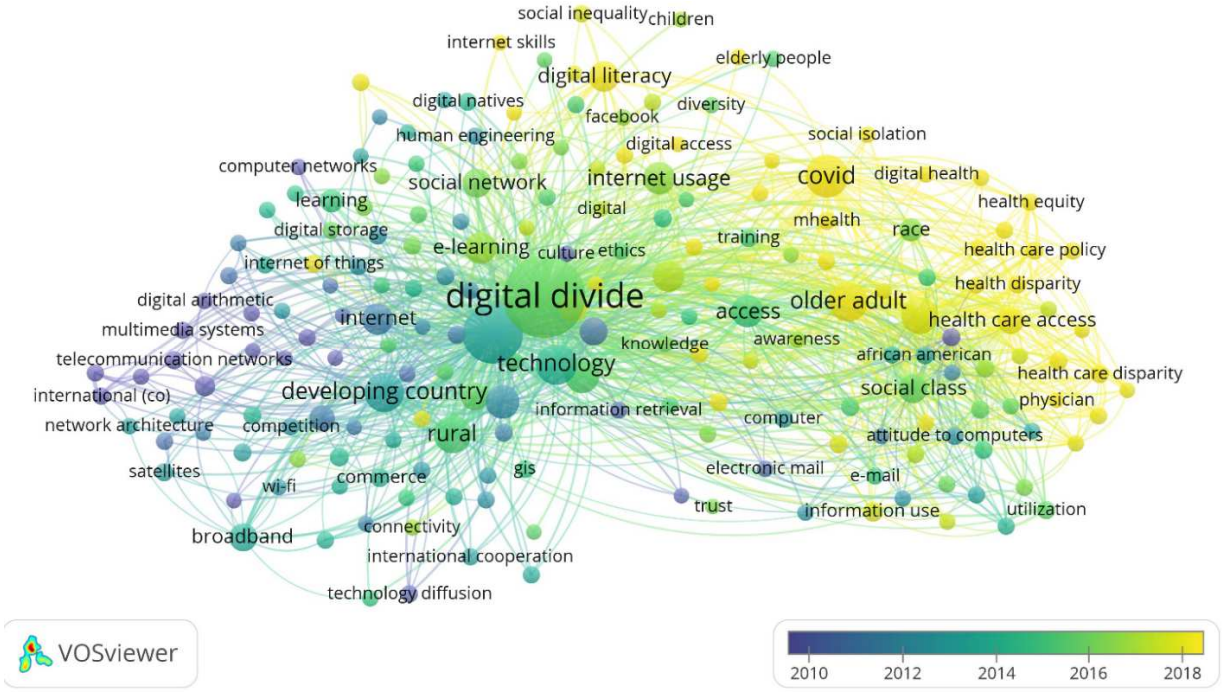

Fig. 2. Overlay visualization map.

In Fig. 2, the nodes associated with each keyword reflect their importance (estimated by their frequency). On the other hand, the color of the nodes shows the keyword's approximate years of validity. The terms or keywords that were most used several years ago (approximately before and around 2010) are shown in purple. Those that are most used currently (2018 to 2023) are in yellow. Finally, the terms located in the middle of these extremes are in different shades of green [11].

As can be seen in Fig. 2, before and around 2010, the most popular keywords in digital divide research are *telecommunication networks*, *multimedia systems*, *digital arithmetic*, and *computer networks*. All these terms are associated with technologies. These terms are followed by *digital divide*, *social network*, *access,* and *social class*. Finally, the current most-used terms are *covid*, *older adult*, *digital literacy*, and *health care access*. Thus, a shift in the importance offered by digital divide studies to a topic—from technical and technological aspects to social aspects and to COVID-19 and excluded groups—can be observed.

IV. Discussion and Implications

Digital divide is a rapidly growing academic and scientific field. This study located the start of this rapid growth in 2018. Its rapid growth can be attributed to the formulation of the SDGs by the UN in 2015 [3], the incessant technological development, and the COVID-19 pandemic that began in 2019 and lasted till 2023. In this sense, there is evidence that the pandemic further complicated the digital divide landscape [15].

Among the other bibliometric analysis studies on digital divide, it is necessary to mention those by Basit et al. [16] and Hoyos-Muñoz and Cardona-Valencia [17]. Though the present study differs from them due to its inclusion of a much larger number of documents (9,523), its three results coincide with their results: 1) digital divide is a growing academic field; 2) articles are the main means of disseminating research on digital divide; and 3) the United States is the leading producer of digital divide research. Among the several unique results of

the current study, its analysis of the evolution of the importance of topics stands out: the first decade of digital divide research (approximately from 2000 to 2010) was dominated by studies focused on the understanding of technologies. Subsequently, topics related to digital divide's social aspects gained importance. Of late, issues related to the pandemic, the groups that are the most excluded, and access to information and services have gained utmost importance. This historical evolution of the importance of different topics is also a good orientation for future studies on digital divide.

As mentioned before, research on the digital divide is led by the United States and other developed countries, such as the United Kingdom, Spain, and Australia, and by emerging countries, such as India and China. Although research in these settings should continue, this study recommends further research in other settings, specifically in developing countries. Research on digital divide in broader scenarios will allow public and private decisionmakers to develop better solutions to bridge the divide. Additionally, research in various scenarios, especially through comparative studies, will enable a greater and better generalization of the theories that are being developed [12].

Examining the frequency of keyword use is a good approach to understand the dominant name of a field of study and its most important topics. Since *digital divide* turned out to be the dominant name—well above other terms such as *digital inequalities*, *digital gap*, *digital division*, or *digital disparities*—we recommend future studies to use it in search strings. The other popular keywords are linked to technologies, such as the *internet* and *ICTs*. Another popular keyword is linked to a population segment, *older adults*, which, in many countries, is the group most affected by the digital divide. Finally, there is *covid*, which has had significant effects on digital divide since its emergence.

Although digital divide studies remain multidisciplinary, the field's center of gravity has turned from computer science and technology toward the social sciences. These conclusions allow us to guide future research and recommend the nature of the multidisciplinary teams of researchers that must be formed.

The present study has some limitations. Due to issues of available space, we used only bibliometric techniques that allow a better general approximation of the topic of digital divide. Future studies can thus consider using a wide range of bibliometric techniques [8] to complement our general vision and delve into certain aspects.

REFERENCES

[1] J.A.G.M. van Dijk, "Digital divide research, achievements and shortcomings." *Poetics*, vol. 34, pp. 221-235, 2006.

[2] Y. S. Chang, S. Jeon and K. Shamba, "Speed of catchup and digital divide: Convergence analysis of mobile cellular, internet, and fixed broadband for 44 African countries." *J. Global Inf. Tech. Man*. vol. 23, pp. 217-234, 2020.

[3] United Nations, "Sustainable development goals", 2015. Available at: https://sdgs.un.org/goals.

[4] A.K. Hollman, T.R. Obermier and P. Burger, P. "Rural measures: a quantitative study of the rural digital divide." *J. Inform. Policy*., vol. 11, pp. 176–201, 2021.

[5] N. Aissaoui, "The digital divide: a literature review and some directions for future research in light of COVID-19." *Glob. Knowl. Mem. Commun.*, vol. 71, pp. 686-708, 2022.

[6] S. Lythreatis, S.K. Singh and A-N. El-Kassar, "The digital divide: a review and future research agenda." *Technol. Forecast. Soc.* vol. 175, pp. 121359, 2022.

[7] J. Cruz-Cárdenas, E. Zabelina, J. Guadalupe-Lanas, A. Palacio-Fierro and C. Ramos-Galarza, "COVID-19, consumer behavior, technology, and society: A literature review and a bibliometric analysis." *Technol. Forecast Soc*., vol. 173. pp. 121179, 2021.

[8] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey and W.M. Lim, "How to conduct a bibliometric analysis: an overview and guidelines." *J. Bus. Res.*, vol. 133, pp. 285-296, 2021.

[9] J. Cruz-Cárdenas, J. Parra-Domínguez, E. Zabelina, O. Deyneka and C. Ramos-Galarza, "Organizational culture and digital transformation: A bibliometric approach," *2022 IEEE Sixth Ecuador Technical Chapters Meeting (ETCM)*, Quito, Ecuador, pp. 1-5, 2022.

[10] R. Pranckuté, "Web of Science (WOS) and Scopus: The titans of bibliographic information in today's academic world." *Publications*, vol. 9, pp. 1–59, 2021.

[11] N.J. van Eck., and L. Waltman, "Manual for VOSviewer version 1.6.19", 2023. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.19.pdf.

[12] J. Cruz-Cárdenas, J. Parra-Domínguez, E. Zabelina, O. Deyneka and C. Ramos-Galarza, "Blended learning and higher education: A bibliometric analysis." *Lecture Notes in Computer Science*, vol. 14041, pp. 456-465, 2023.

[13] M. Aria and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis." *J. Informetr.* vol. 11, pp. 959-975, 2017.

[14] W. Jun, "A study on cause analysis of digital divide among older people in Korea." *Int. J. Environ. Res. Public Health*, vol. 18, pp. 8586, 2021.

[15] Y. Song, C. Qian and S. Pickard, "Age-related digital divide during the COVID-19 pandemic in China." *Int. J. Environ. Res. Public Health.* vol. 18, pp. 11285, 2021.

[16] S.A. Basit, A. Shamshad, M.T. Latif, M.A. Khan and A.G. Miss, "Bibliometric analysis of digital divide using Web of Science." *Libr. Philos. Pract*., 6256, 2021.

[17] J.A. Hoyos-Muñoz and D. Cardona-Valencia, "Trends and challenges of digital divide and digital inclusion: A bibliometric analysis." *J. Inf. Sci.* 2023 (in press).