MASTER'S THESIS IN LIBRARY AND INFORMATION SCIENCE
FACULTY OF LIBRARIANSHIP, INFORMATION, EDUCATION AND IT
2024

Digital Libraries and Sustainability:
A Bibliometric Analysis Using Web of Science and Scopus

FOTINI LEVENTAKOU

UNIVERSITY
OF BORÅS

English title:	Digital Libraries and Sustainability: A Bibliometric Analysis Using Web of Science and Scopus
Author(s):	Fotini Leventakou
Completed:	2024
Abstract:	This study aims at contributing to the advancement of the field of Library and Information Science, academia, and the wider society by exploring the scientific research production on the topic of digital library sustainability as this is represented in two bibliometric databases: *Web of Science* and *Scopus*. It also attempts to make a methodological point about how scientific domains should be investigated using bibliometrics. Lastly, it investigates the extent to which the Sustainable Development Goals (SDGs), set out by the United Nations in 2015, constitute a study object in the existing literature within the particular field of study. To achieve these aims, a bibliometric analysis of academic literature (i.e., articles, review articles, books, book chapters, and conference papers) published between 1997 and 2023 was conducted. Moreover, existing research in relation to the SDGs was analyzed using the relevant filtering options available at each of the two databases. As a means for triangulation, the data sets exported from the two databases were examined separately and their results compared. The results showed an upward trend in scientific production on the topic. Nevertheless, the discussion around it is still at preliminary stages as proven by the limited number of (a) publications on the topic, (b) authors systematically researching it, and (c) inter-author and inter-country collaborations. The study also revealed significant differences in the way SDGs are mapped in each of the two databases allowing for questioning whether SDG database mappings may be trusted, and highlighting the need for more research on the issue.
Keywords:	sustainability, digital libraries, sustainable development goals, SDG mapping, bibliometric analysis

Contents

List of tables

List of figures

List of abbreviations

Abbreviation	Definition
ACG	American College of Greece
ALA	American Library Association
DCH	Digital Cultural Heritage
DLF	Digital Library Federation
DOI	Digital Object Identifier
ICT	Information and Communication Technology
IFLA	International Federation of Library Associations and Institutions
IS	Information Services
LIS	Library and Information Science
LISA	Library & Information Science Abstracts
LISS	Library & Information Science Source
LISTA	Library, Information Science & Technology Abstracts
MDGs	Millennium Development Goals
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SDGs	Sustainable Development Goals
UN	United Nations
UOB	University of Borås
WoS	Web of Science

1 Introduction

The issue of sustainability is a much-discussed topic in recent years; one that touches all aspects of human life. With the advent of the Internet and the transformation of information from print to digital, and of its dissemination through digital and web-based means, questions started arising regarding how sustainable such services are. The word "services," in this context, reflects what is usually referred to as digital libraries, repositories, archives, or even information systems.

Consequently, soon enough scientists started exploring aspects of sustainability in the context of various fields of human development. On top of that, governments and international organizations also started dealing with sustainability primarily in relation to the environment, but also extending their interest towards sustainability's social, economic, and even cultural implications, and highlighting its significance for the survival of society and the planet at large.

This whole discussion on the topic along with my interest in digital libraries motivated my desire and decision to explore it further. Therefore, the present document constitutes a bibliometric study on the topic of digital library sustainability, conducted in the context of the *Master's in Library and Information Science, Digital Library and Information Services* held at the University of Borås, Sweden.

1.1 Research problem and research questions

The concept of sustainability has always been a core value of librarianship. This is why the American Library Association (ALA) has been active in publishing guidelines, policies, and resolutions towards this direction (Council of the ALA, 2015, 2017, 2019, 2021). Furthermore, the International Federation of Library Associations and Institutions (IFLA, 2017) has been actively involved with the creation and promotion of the United Nation (UN)'s 2030 Sustainable Development Goals (SDGs) that aim at ensuring that people's present needs are met without undermining ours and our planet's future. The UN has highlighted the importance of the scientific community's contribution towards the achievement of these goals (UN, 2023a).

Research on libraries and sustainability has primarily focused on physical rather than digital libraries (see for example Li & Yang, 2023; Moreno et al., 2022). According to Chowdhury (2012) and Nolin (2010), the concept of sustainability in the context of information science and more specifically with regard to digital libraries seems to have preoccupied less scientists and researchers. This highlights a need to understand how the conversation around digital library sustainability is evolving, how this conversation is translated into scientific production on the topic, and if the UN SDGs have been considered and studied within the existing relevant scientific research.

Consequently, the present research project will attempt to answer the following research questions, which based on Wildemuth's (2017) suggestion have

originated from the researcher's personal experience and a noticed gap in the existing literature. The research questions are:

1. How is the topic of digital library sustainability represented in current academic literature? What is the particular field's overall structure?

2. What trends can be identified in relation to digital library sustainability research? What is the field's intellectual, social, and conceptual structure?

3. Do the UN SDGs constitute a study object within digital library sustainability research? How is this represented in the way different bibliometric databases map their content to the UN SDGs?

1.2 Aim and expected outcomes

The above-mentioned research problem guided me towards articulating the purpose of the present thesis. This is to explore an apparently understudied field in order to determine what research has been conducted to date on the topic of digital library sustainability; to get an overall picture of the scientific research trends on the topic; and to examine the extent to which the UN SDGs constitute a study object in the area of digital library sustainability research. There is also a methodological aim to reflect on the bibliometric method as a way of studying a topic.

Doing all of the above will: (a) increase the scientific community's understanding of the amount and quality of research that has already been conducted on the topic of digital library sustainability; (b) clarify the knowledge framework of research in this field; (c) highlight the need, or not, for more research related to digital libraries and the UN SDGs; (d) propose future research directions; and (e) contribute to the understanding and enhancement of bibliometrics as a methodological approach. The ultimate goal is to determine how the Library and Information Science (LIS) world can contribute to what has been called sustainable development, defined by the UN (2023b) as "how we must live today if we want a better tomorrow, by meeting present needs without compromising the chances of future generations to meet their needs" (para. 1).

The outcomes of this study are expected to show that there is some amount of research on the topic of digital library sustainability, and that this research focuses mostly on the environmental aspect of digital libraries. It is also expected that the study will reveal gaps with regard to social, economic, and cultural digital library sustainability research. Lastly, regarding the UN SDGs, it is expected that the results of the study will reveal some amount of scientific production related to them, but also that there is room for more research.

1.3 Relevance and significance

Even though the concept of sustainability in the sense that we understand it today did not appear until the late 20th century when Meadows et al. (1972) highlighted the need for a sustainable world system, it has been a topic of interest historically from as early as the 17th and 18th centuries (Purvis et al.,

2019). Nevertheless, as the limited published research on the topic suggests, the theme of sustainability in relation to digital libraries is quite new.

To the best of my knowledge, this is the first bibliometric study on the topic of digital library sustainability. As will be discussed in Chapter 3, some other bibliometric studies have been conducted, but those explored either the one or the other concept. None of them attempted to explore digital library sustainability as a unified topic. Consequently, the project's significance lies in the fact that it attempts to highlight areas within the topic under discussion that need to be further investigated and researched. Moreover, the results and conclusions reached will be useful to the discipline of LIS. The study will also contribute to the advancement of academia in general, as it may pave the way for more research to improve digital libraries and make them more financially robust. It will also help the broader society by exploring how digital libraries are positioned in the wider discussion on sustainability, and what steps need to be taken so that they become more socially and environmentally sustainable, thus contributing to the international effort towards a more sustainable future. Lastly, the study will contribute to the advancement of bibliometrics as a methodological approach.

1.4 Structure of the study

The study comprises seven chapters. The current chapter (i.e., Chapter 1) aims at introducing the reader to the topic. It also presents the research problem on the basis of which the study was built, along with the study's purpose, research questions, and expected outcomes. It briefly discusses the study's relevance to LIS, academia, and society, and concludes with the structure of the present document.

In Chapter 2, the theoretical framework that underlies this study is presented. This involves a thorough discussion of citation theory, its history, and development.

Chapter 3 provides a description and a historical review of the concepts that constitute the basis of the study, which are *digital libraries* and *sustainability*. Furthermore, the chapter presents a review of the relevant literature on the topic, that is relevant bibliometric studies as well as studies on digital library sustainability that have engaged in other methodological procedures.

Chapter 4 presents and analyzes the study's methodology. It begins with a brief definition of bibliometrics, describes the process of source selection and the reasons for selecting the specific sources; moves on to a detailed description of how data were chosen and cleaned; and concludes with an explanation of the bibliometric techniques that were applied along three methodological axes.

Chapter 5 thoroughly describes the study's results. It begins with a description of the bibliometric tools used for the analysis of the data; continues with a comparison of the two examined data sets; and presents the results along the three methodological axes described in Chapter 4 through visualizations and tables.

Chapter 6 presents a discussion, analysis, and interpretation of the main findings regarding the current scientific production, the intellectual, social, and conceptual trends, and the representation of SDGs within the relevant literature. Moreover, the limitations and problems encountered in the research process are discussed.

In conclusion, Chapter 7 presents my overall concluding remarks as well as suggestions for future research.

2 Theoretical framework

The *Oxford English Dictionary* has defined the word *citation* as "the action or an act of quoting or referring to a passage, text, author, legal precedent, etc., esp. as an authority or in support of an argument; quotation" ("Citation, n., sense 2.a.," n.d.). Cronin (1981) provided a metaphorical and rather poetic description of the term. He referred to citations as "frozen footprints in the landscape of scholarly achievement; footprints which bear witness to the passage of ideas" (p. 16). Mulkay (1974) mentioned that citations constitute relatively objective data and Singleton (1976) referred to them as being, at least at first sight, quantifiable and able to be easily processed by computers. Citations have been characterized as a "quality-sensor machine" (Lindsey, 1978, p. 137) that may determine the quality and the impact of scientific oeuvres through citation analysis. The gradually increasing popularity of citation analysis highlighted a need for the creation of the so-called citation indexes such as the proprietary *Web of Science* (WoS) and *Scopus* databases, the open access *CiteSeer*, the free to use *Google Scholar*, as well as various subject specific ones, such as *PubMed*.

Nevertheless, throughout the years, many scientists have expressed their doubts with regard to the level of trust the scientific community can show to citation analysis. Crane (1972) highlighted the fact that how scientific publications interrelate on the basis of their citations cannot be defined in absolute terms. She also argued that scientists decide to cite other scientists on the basis of numerous different factors, which renders the process rather subjective. Whitley (1969) came to support this too, claiming that high citation rates are not always the direct result of the cited work's quality, but may depend on social factors and subjective evaluations. This is exactly what Chubin and Moitra (1975) referred to as the phenomenologically "*private* process by which authors choose references" (p. 426). Gilbert (1977) explicitly stated that scientists do not really have a clear idea of what citation analysis allows them to measure.

This debate around the strengths and limitations of citations and citation analysis soon came to underline the need for a theory of citing, which up to that point was mostly implied (Mulkay, 1974), and which would assist in more standardized citation analysis procedures. Gradually, two different viewpoints emerged. The first was in favor of a theoretically informed framework that would explain the role of citations (e.g., Luukkonen, 1997), while according to the second, a citation theory model was not needed as citations are by themselves a quality indicator (e.g., van Raan, 1998). For a theory of citation to be developed, though, one key question had to be answered and this question was: Why do scientists and scholars cite other scientists and scholars? As has been proven to date, even though it sounds simple, this is not a quite easy question to answer and has given rise to endless theory-building discussions (Leydesdorff, 1998).

Cronin (1984) was the first to explicitly refer to a citation theory in his book *The Citation Process: The Role and Significance of Citations in Scientific Communication*. Initially, Cronin focused his discussion around two theoretical frameworks: Merton's normative and positivistic view, according to which

citations are a means for recognizing and giving credit to others for their work, and a post-modernist social constructivist view according to which norms do not have a value. Back then, Cronin mostly supported the second viewpoint with regard to bibliometrics. This is something that Small (2016) negated, claiming that a social constructivist view of scientific knowledge would mean knowledge is not based on reality, but on subjective claims and rhetorical persuasion.

In a later paper, Cronin (1998) discussed the issue from three different theoretical perspectives: (a) a functionalist approach, according to which scientists cite other scientists to provide additional evidence, support or refute scientific assumptions; (b) a normative approach, according to which scientists cite each other because this is required by the scholarly communication system; and (c) a phenomenological approach, according to which scientists cite each other because of personal socio-psychological motives. By doing this he attempted to explain that a combination of the three theories would best describe why scientists cite other scientists. This perspective of a more unifying theoretical framework was in accordance with Leydesdorff's (1998) acceptance that citation analysis allows for movement "between the cognitive, the textual, and the social dimensions of science in terms of socio-cognitive interactions" (p. 19).

Building upon the work by Small (1978), who was the first to talk of citations as conceptual symbols, Cronin (2000) moved towards a more normative position attempting to link citation theory to semiotics, posing the rhetorical question "What, after all, are references and citations if not signalling devices?" (p. 440). This claim led him to admit that this was something he had overlooked when he originally talked about citations as the "frozen footprints" mentioned at the beginning of this chapter. It also led him to the conclusion that seeing citations as signs might assist the bibliometrics and scientometrics research community in realizing that these signs are not objectively indicative of things such as quality or impact but may carry multiple symbolic meanings. He concluded that semiotics may not provide a theory for understanding citation, but it could provide a framework within which scientists could understand the pros and cons of the different theories around it.

More recently, Macfarlane (2017) referred to the term performativity which emerged in Academia a few decades ago as an expression of the "scholarly" audit culture. It is the quest on the part of academics for better and higher teaching and research performance data; data that, on the one hand, help academics stand out, and on the other, assist universities in claiming or attracting more funds. Within this newly created reality, citations and citation analysis indicators are treated as a means for evaluating the quality and impact of scientific work and, consequently, as something that researchers are actively seeking to receive. However, as Aksnes et al. (2019) pointed out the interpretation and validity of citations as quality evaluation devices may be questioned simply because the concept of research quality encompasses much more than simple citation counts. It presupposes qualities such as the plausibility of argumentation, the originality of a research topic, and the research's contribution to the advancement of science and society.

Despite the fact that there are still opposing views regarding citation theory, it cannot be doubted that the study of citations may offer valuable scientific insights. Cronin (1984) identified three main functions of citations. These are (a) the dissemination of knowledge, (b) the preservation of standards, and (c) the recognition of other scientists' work and contributions to the development and evolution of science. As such, citations and their study and examination may play a significant role in drawing conclusions about how human knowledge progresses. This, along with the fact that bibliometrics, the methodological approach, used in the present study, rely on citations and citation analysis, is what links the discussion about citation theory to the present study.

3 Definitions and literature review

As has been mentioned in the introduction of this report, the present study relies on two basic concepts: *digital libraries* and *sustainability*. To proceed with a deeper exploration of these two concepts there is a need to first delve into how each one of them emerged and how they have been defined. This will be done in the following two sections of this chapter. In the third section, a review of relevant literature on the topic will be presented.

3.1 Emergence and definitions of the concept "digital libraries"

The concept of a digital library is one that has been much discussed in recent times. Calhoun (2014) placed the first appearance and emergence of digital libraries in year 1991 as this was the year when the US National Science Foundation first launched a number of workshops on the topic aiming at putting the foundations for the creation of digital libraries. Obviously the two and a half decades that preceded this were the ones during which a big number of technological innovations that constituted requirements for the development of digital libraries came into existence. Lesk (2005) has mentioned developments such as digital storage and retrieval, speedy processors, connectivity, natural language processing, and document conversion technologies, such as scanning and optical character recognition, to name but a few.

As the first digital libraries started emerging, various definitions started coming to light along with them. Nonetheless, until today, there is still a great deal of confusion regarding what a digital library actually is (Cleveland, 1998). Fox et al. (1995) attributed this confusion to the fact that "the phrase 'digital library' evokes a different impression in each reader" (p. 24). Nürnberg et al. (1995) provided yet another explanation for the confusion, supporting that it stems partially from within the library community, which has been using multiple different terms, such as electronic library, virtual library, etc. to denote the same thing; and partially from the fact that throughout the years digital libraries became the center of interest for many research fields (e.g., library science, information retrieval, hypertext technology, information delivery) each of which defined the term slightly differently.

Borgman (1999, 2000) attempted to group the various definitions of the term digital library in terms of whether they (a) were research- or practice-oriented; (b) approached the concept of a library narrowly, as a simple collection of information, or broadly, as a social structure; or (c) addressed digital libraries as simple information retrieval systems or as services and institutions.

In general, what is noticed is a transition from definitions that focused on the technologies of digital libraries to those that emphasized the social role of digital libraries. One of the definitions that represent the original viewpoint, focusing on the technological aspects of digital libraries is that by Lesk et al. (1993), who defined a digital library as one that is "based on electronic data," contains "both text and graphics," and is "widely accessible via electronic

networks" (p. 12). Towards the same direction is the definition by Borgman (1993) who supported that a digital library is "a service … an architecture … a set of information resources, databases of text, numbers, graphics, sound, video, etc. … a set of tools and capabilities to locate, retrieve, and utilize the information resources available" (p. 122). A few years later, Lesk (1997) came to provide a more enhanced definition of the term claiming that "digital libraries are organized collections of digital information. They combine the structuring and gathering of information, which libraries and archives have always done, with the digital representation that computers have made possible" (p. xix).

In the meantime, definitions that emphasized the social role of digital libraries started emerging. The first comprehensive, concise, and clear definition expressing the perspective of librarians was the one developed by the Digital Library Federation (DLF). According to it:

> Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. (DLF, as cited in Waters, 1998, para. 3)

Lyman (1996) stated that "the definition of the *digital library* will require an understanding of the role and nature of public institutions in a postindustrial society" (p. 4). He also referred to digital libraries as "a realm of free speech and association as well as an information marketplace" (p. 29) highlighting the fact that, just like physical libraries, they maintain roles traditionally assigned to libraries. Borgman (2000) provided one of the most influential definitions in the field of digital libraries, which combines the idea that they are "a set of electronic resources and associated technical capabilities for creating, searching, and using information" with the idea that they are "constructed—collected and organized—by [and for] a community of users, and their functional capabilities support the information needs and uses of that community" (p. 42). Van House et al. (2003) gave a concise but comprehensive definition describing them as "*sociotechnical systems*—networks of technology, information, documents, people, and practices" (p. 1)

Lastly, a more recent definition by Calhoun (2014) provided a two-fold explanation of what the term digital library means adding to it more nuances. First, he defined it as a multidisciplinary "field of research and practice" (p. 18) and second, as "systems and services … that (a) support the advancement of knowledge and culture; (b) contain managed collections of digital content … intended to serve the needs of defined communities; (c) often use an architecture that first emerged in the computer and information science/library domain …" (p. 18).

3.2 Emergence and definitions of the concept "sustainability"

Just like the term digital library, sustainability is another term that has been defined in multiple ways over the years. Etymologically the word has derived

from the Latin verb *sustinēre*, which means to hold up, in other words to support (Jeronen, 2013). Jeronen saw it as a paradigm, the aim of which is to achieve good quality of life by offsetting environmental, social, and economic issues.

The very concept of sustainability started developing gradually. Purvis et al. (2019) traced its origins back to (a) the forestry experts of the 17th and 18th centuries, who first coined the idea of the sustainable yield; (b) the early political economists, who doubted the limits of economic and demographic growth and talked about the inevitable compromise between wealth and social justice; and (c) the 19th and 20th century ecologists, who stressed the significance of combining natural resource conservation with nature preservation.

The idea of sustainability as we know it today appeared in the late 1970s. Among the first documents that mentioned the notion of a sustainable world system was *Limits to Growth: A report for the Club of Rome's project on the predicament of mankind* (Meadows et al., 1972) and according to Grober (2012) this was the first time the word "sustainable" was used in the sense and with the meaning it has today. From there on, more concepts such as that of a *sustainable society* or the *three pillars of sustainability* started arising. Sustainability started being represented as a three-fold concept, encompassing an economic, an environmental, and a social aspect. According to Purvis et al. (2019) these three aspects have been referred to in multiple ways in academic research. Scientists have talked about the three pillars (Boyer et al., 2016; Zijp et al., 2015), dimensions (Moir & Carter, 2012; Mori & Christodoulou, 2012; Waas et al., 2011), components (Du Pisani, 2006), stool legs (Dawe & Ryan, 2003; Vos, 2007), aspects (Lozano, 2008; Tanguay et al., 2010), or perspectives (Arushanyan et al., 2017; Brown et al., 1987) of sustainability. Chowdhury's (2013a, 2013b, 2013c, 2014a, 2014b) research on digital library sustainability also referred to three sustainability pillars or forms.

Gradually, sustainability started becoming the center of attention of governments and institutions around the world, and soon enough another concept, that of *sustainable development* made its appearance. Sustainable development became intertwined with sustainability and according to Waas et al. (2011) was, and still is, at times, treated as its synonym and, at others, as a distinct term. The term originally appeared in what has become famous as the Brundtland Commission Report, which defined it as development that "meets the needs of the present without compromising the ability of future generations to meet their own needs" (Brundtland, 1987, Chapter I, Section 3, para. 27).

The UN soon became interested in promoting sustainability and sustainable development around the world and started taking actions towards this direction. In 1992, they set Agenda 21, a comprehensive global action plan with the aim to support sustainable development (UN, 1992). The UN also established the Commission on Sustainable Development to ensure, on one hand, the monitoring of the implementation of Agenda 21, and, on the other, that all local, national, regional, and international agreements would be met. In 2001, the UN launched the eight so called Millennium Development Goals (MDGs) aiming at having the global community achieve them by 2015 (UN, 2001). The eight MDGs were to: (a) eradicate extreme poverty and hunger; (b) achieve

universal primary education; (c) promote gender equality and empower women; (d) reduce child mortality; (e) improve maternal health; (f) combat HIV/AIDS, malaria, and other diseases; (g) ensure environmental sustainability; and (h) develop a global partnership for development. These were accompanied by 18 targets and 48 indicators that would assist in their implementation.

The 2002 World Summit on Sustainable Development highlighted the need to integrate the three pillars of environmental, economic, and social development as three components of sustainable development (UN, 2002). The discussion on sustainable development continued in the years that followed, until 10 years later, in the UN Conference on Sustainable Development, the world's leaders decided to adopt a 10-year framework of actions on a voluntary basis, to ensure sustainable consumption and production (UN, 2012). In the same conference, the foundations for the development of SDGs were laid. The new SDGs would be set based on the MDGs that had been launched a few years earlier in the context of Agenda 21, aiming at improving the deficiencies of their "predecessors."

After this, in 2015, leaders from around the world came together at the UN Summit on Sustainable Development held in New York and set the renowned 2030 Sustainable Development Agenda. The agenda was an action plan that aimed at protecting the five Ps of a sustainable future, namely the *people*, the *planet*, *prosperity*, *peace*, and *partnership* (UN, 2015). To achieve this, they put together the 17 well-known SDGs, which were accompanied by 169 targets that would pave the way as to how the world would proceed after 2015 and as to what should have been achieved by 2030. The 17 SDGs they set were the following: (a) no poverty; (b) zero hunger; (c) good health and well-being; (d) quality education; (e) gender equality; (f) clean water and sanitation; (g) affordable and clean energy; (h) decent work and economic growth; (i) industry, innovation, and infrastructure; (j) reduced inequalities; (k) sustainable cities and communities; (l) responsible consumption and production; (m) climate action; (n) life below water; (o) life on land; (p) peace, justice, and strong institutions; (q) partnerships for the goals.

Returning to the definition of sustainability and sustainable development, one may conclude that the two concepts are intricately linked to each other. One may support that sustainability is the goal the world is aiming at achieving, while sustainable development and the 17 UN SDGs are the means towards achieving this goal. Reaching sustainable development is a multi-factor process, which will only be successful if everyone contributes to the best of their ability. Ever since their appearance, libraries have been the gatekeepers of information and knowledge, and as such they may play a crucial role towards ensuring access to information for all as well as towards educating people on how they may contribute to the achievement of SDGs. IFLA and the ALA have engaged actively in promoting the 2030 agenda and the SDGs by creating the ALA Task Force on UN 2030 SDGs and delegating to it the responsibility of increasing awareness about the SDGs through libraries.

3.3 Literature review

Research has always been the mechanism for the advancement of humanity and the world. It is a mechanism through which people can understand the present, identify its problems, and find solutions to address these problems. The significance of scientific research is undoubted. As a consequence, I will next attempt to outline the scientific research that has been conducted on the topic of digital library sustainability.

A simple search for the query *"digital librar*" AND sustainability* in two of the most well-respected abstracting and indexing databases on LIS, namely *Library, Information Science & Technology Abstracts (LISTA)* and *Library & Information Science Abstracts (LISA)* yielded 130 unique research papers. Searching in just two databases obviously does not cover the full range of relevant research; it is, though, a good indication that allows one to assume that there is not much scientific literature on the topic. This comes in contrast to a wider trend in current scientific literature, which shows that most scientific disciplines produce big amounts of research relating to sustainability every year and that sustainability is seen as a social contract between science and society (Lubchenco, 1998).

A deeper examination of the retrieved results revealed that there are even fewer bibliometric and document studies on the topic. What follows is an attempt to go through the most representative of these studies. I will begin with a review of some document studies, continue with a discussion of bibliometric studies that focus on either digital libraries or sustainability, and finish with studies that combine the two concepts. Some of the reviewed literature was collected from the databases mentioned in the previous paragraph and from *Google Scholar*. Nevertheless, most of it was gathered by applying the technique of *backward citation chaining* (Ellis, 1989), else referred to as *pearl-growing* (Hawkins & Wagers, 1982). This is a technique that allows for the retrieval of more literature by examining the references of a key publication.

3.3.1 Document and bibliometric studies

To my knowledge, one of the few document studies that combined the topics of digital libraries and sustainability was the one by Eschenfelder et al. (2016), who attempted to highlight how the concept of digital archive, repository, and library sustainability is represented in academic literature published between 2000 and 2015, and indexed in *Library & Information Science Source (LISS)*, *LISTA*, and *LISA* databases. The researchers highlighted the significance of the concept for LIS research especially with regard to its economic aspect. Nevertheless, according to the results of their study, current literature has touched upon the issue of digital library sustainability rather superficially, as indicated by the average of four publications per year, and extra effort should be put on developing methodologies that will help promote the sustainable development of information systems.

Apart from this, the closest to the topic document study that has been conducted was one by Meschede and Henkel (2019). The scientists attempted a mapping of publications in the field of LIS that directly dealt with sustainability and sustainable development. They focused on the entire LIS

field and not on digital libraries in particular. Their data set was obtained from Scopus database. In the absence of a subject area in the field of LIS, the authors used the *SCImago Journal Rank*[1] to identify relevant sources. They ended up reviewing 81 documents. Their results revealed that sustainability-related research within the LIS field focused mostly on two broad topics: (a) libraries and (b) information and communication technology (ICT). They also suggested that future research could focus on education and information dissemination with regard to sustainability, and on developing frameworks for the assessment of sustainable development within LIS.

Lastly, Li and Yang (2023) conducted a bibliometric study that touched upon the topic of sustainability in relation to green libraries. The study did not focus specifically on digital libraries. With this study, the researchers aimed at raising awareness about the significance of creating green libraries, motivating action towards their development, and indicating future research pathways. They used the WoS database to collect their data, and their data set included 134 items. Some of their primary findings were that (a) there was a steady increase in research output on green libraries since 2007, but the overall number is still small; (b) research showed interdisciplinarity, with Library Science playing a central role; (c) the USA and China were the most productive countries in terms of research output; (d) cooperation among countries was small; and (e) topic trends showed a shift from pure environmental considerations to topics such as library strategy and services.

In terms of bibliometric studies, no studies that combined the topics of digital libraries and sustainability have been conducted to my knowledge, but a few studies that focused on either the one or the other concept do exist and will be discussed next.

3.3.2 Bibliometric studies on “digital libraries”

Singh et al. (2007) conducted a bibliometric study aiming to explore the existing literature on digital libraries. They examined 1.000 articles published between 1998 and 2004 and retrieved from the LISA Plus database. Among their findings were that most of the examined items focused on digital library initiatives, whereas some other prominent topics related to music digital libraries, digital repositories of theses and dissertations, and digital preservation. They also highlighted the dominance of the USA as the primary producer of literature on the topic.

Ahmad et al. (2018) conducted another bibliometric study on digital libraries. Aiming at quantifying the prominence and impact of digital library literature, they examined 4.206 documents published between 2002 and 2016 and retrieved from the ISI WoS database. They concluded that the number of published documents on the topic in general showed a rapid increase in research productivity and, similarly to Singh et al. (2007), indicated the USA as a primary contributor.

[1] More information about the SCImago Journal Rank may be found at: https://www.scimagojr.com/.

Lastly, Borgohain et al. (2022) examined the global trends in digital library research using data from Scopus database. They focused on the period from 2016 to 2020 and their data set included 5.576 papers. Once again, the USA proved to be the highest-ranking country both in terms of productivity and impact, whereas 2019 was highlighted as the most productive year. Borgohain et al. came to confirm Ahmad et al.'s (2018) finding that there is a rapidly increasing trend with regard to scientific production on the topic.

It is worth mentioning that sustainability did not emerge as a trending topic in any of the above-discussed studies. In spite of this, scientists have been conducting bibliometric studies on sustainability. Some of them are discussed in the following section.

3.3.3 Bibliometric and other studies on "sustainability"

As discussed in Section 3.2, in a world faced with increasingly more and diverse global social, economic, and environmental challenges, the UN has assumed a leading role towards the promotion of sustainability and sustainable development. They have developed a sustainable development agenda for the implementation and monitoring of the 17 SDGs by the year 2030, and consider the involvement of the scientific community and the advancement of research crucial for the achievement of these goals (UN, 2023a). Consequently, the UN SDGs have become a topic of interest for many researchers as well as academic and research institutions.

Recently, Mishra et al. (2024) conducted a bibliometric analysis of the scientific production related to the SDGs. The researchers used the WoS database and analyzed a big data set of 12.176 documents published between 2015 and 2022 in a variety of academic disciplines. The study's goal was to examine the progress, challenges, opportunities, trends, and prospects of the SDGs. The results showed that there is extensive interest in the topic of SDGs as indicated by the large number of authors (i.e., 45.345) and the high levels of co-authorship. The study also revealed the following as the most significant topics addressed in the literature: SDGs, climate change, Agenda 2030, the circular economy, poverty, global health, governance, food security, sub-Saharan Africa, the MDGs, universal health coverage, indicators, gender, and inequality. Digital libraries did not emerge as one of the topics related to SDGs.

Chowdhury and Koya (2017) also dealt with the UN SDGs. Specifically, they explored how the *iSchools* could assist in achieving them. iSchools is an international non-profit organization that originated in 2005 (iSchools, n.d.). It currently numbers 120 member universities from around the world. The organization's vision is to "strengthen all aspects of research in information and information science for members across the globe" ("Vision" section). This vision is linked to the fact that information and data are crucial to the achievement of the SDGs. Chowdhury and Koya (2017) concluded that by linking people to the appropriate information, iSchools could assist the world in achieving the SDGs and help develop a sustainability-centered culture in relation to information practices and services. They specifically identified four teaching and research areas they should focus on, namely sustainable (a) information systems and infrastructure; (b) information practices; (c)

information policies and governance; and (d) user education, training, and literacy.

Grosseck et al. (2019) were among the scientists who have conducted bibliometric studies on sustainability and sustainable development. Specifically, their study focused on education for sustainable development. They used the WoS database and examined 1.813 papers published between 1992 and 2018. The results of their study indicated a growing interest in the field as attested by the growing number of publications, authors, and journals dealing with the topic as well as by the gradually increasing collaborations among authors and countries around the world.

More recently, Koya and Chowdhury (2020) highlighted the significance of culture and cultural heritage information practices in achieving sustainable development. In doing this, they conducted a thematic analysis of 25 UN official documents and came up with 14 themes through which cultural sustainability may be achieved. They classified these 14 themes in five main categories, that is information management skills (including information access, collection & distribution, curation, quality, and standardization), technology skills (including information exchange, platforms, and transformation), application skills (including cross-sectional information), leadership skills (including information accountability and broadcast), and people/user skills (including information seeking, sharing, and usage training). Koya and Chowdhury stressed the importance of incorporating these skills into information science curricula and cultural heritage information education so that knowledge and awareness on the topic increases and cultural sustainability is eventually achieved.

3.3.4 Studies on digital library sustainability

As has already been mentioned, the interest of the scientific community in issues of sustainability has pervaded almost every scientific discipline. Consequently, the topics of digital library sustainability and digital library sustainable development have slowly started becoming a point of interest in contemporary LIS research, even though they still appear to be understudied.

Spink (1995) was one of the first scholars, who highlighted the fact that there was an emerging need for the digital library community to consider how digital libraries should contribute to the international debate around sustainable development. In a conference presentation, she pointed out that the concepts of digital libraries and sustainable development had to be commonly addressed. Spink proposed the development of an interdisciplinary research agenda that would investigate connections between the two.

Hamilton (2004) identified the issue of economic sustainability as a concern of primary importance for many digital libraries. The researcher proposed that becoming an integral part of their respective parent institution is what will help digital libraries achieve economic sustainability as in this way they will stop being regarded and treated as temporary special projects. To achieve this, digital libraries need to invest on the following key ingredients: (a) creating a widely valuable product; (b) finding strong and influential supporters; and (c) developing a solid funding strategy that would aim for integration and could

include options such as charging for access or receiving sponsorships and in-kind support.

Research for literature on the topic revealed G. Chowdhury as a scholar who has dealt a lot with digital library sustainability. This is why what follows is a review of his relevant work. His original work on digital library sustainability focused mainly on environmental sustainability. Chowdhury (2010) supported the idea that a turn towards the digital distribution of information might be a solution against the massive amounts of CO2 emissions resulting from several activities involved in the print knowledge products. He claimed that moving from print to digital solutions would make the system environmentally sustainable. He later moved on to propose an agenda of green information services (IS) that would be based on cloud computing (Chowdhury, 2012). Pointing out that the information systems used in higher education and research constitute sources of huge amounts of green-house gas emissions, he thought cloud computing could become the solution for the development of green IS. In another one of his studies, he attempted to propose ways to estimate the environmental costs of digital libraries given the rapid and continuous increase in digital content (Chowdhury, 2016). He concluded that energy consumption is huge at two levels: servers and clients. Consequently, he came up with a two-fold suggestion. On the one hand, he proposed using better equipment and improving policies to reduce server energy consumption, and on the other, he suggested more research related to user information behavior and usage patterns to determine ways in which user search and access time could be reduced.

Chowdhury's later studies brought into the discussion the two other pillars of sustainability apart from environmental, namely economic and social sustainability. In one of his conference presentations, Chowdhury (2013c) discussed what the three aspects of sustainability mean with regard to digital libraries. He specifically mentioned that in the context of digital libraries economic sustainability would be to "ensure cheaper, easier and better access to digital information;" social sustainability would be to "ensure equitable access to information in order to build a better (well informed) and healthy society;" and environmental sustainability would be to "ensure reductions in the environmental impact of digital information" (p. 2). Based on this, he proposed a conceptual model that considers all three forms of sustainability supporting that the three of them are intertwined and none of them should be disregarded if digital libraries wish to achieve sustainable development (Chowdhury, 2013c, 2014b).

In the same year, Chowdhury also published an article that focused exclusively on the social aspect of digital library sustainability (Chowdhury, 2013a). In this article, he analyzed a number of factors that relate to digital library sustainability and proposed a number of indicators (e.g., rights to information, equal access to information, etc.) for measuring it. He also stated that to develop socially sustainable digital libraries a number of socio-cultural and technical factors need to be considered, and confirmed once again that it is impossible to achieve social sustainability independently of the other two previously mentioned sustainability pillars.

Liew and Chowdhury (2016) extended the discussion on social sustainability to Digital Cultural Heritage (DCH) services. Through semi-structured interviews, they tried to determine the factors that assist in rendering DCH services socially sustainable. To do that they selected 16 users of the following three DCH services: *New Zealand Electronic Text Collection*[2]; *Kete Horowhenua*[3]; and *New Zealand History Online*[4]. What they concluded was that for a DCH service to be socially sustainable four important attributes need to be met: (a) strategy and policy; (b) advocacy and community engagement; (c) equity, cultural sensitivity, and literacy; and (d) assessment and evaluation.

Chowdhury (2013b) also studied research and policy documents in an attempt to identify relations between information research and the sustainability of digital information services. His findings showed that at the time he conducted his study, sustainability was not an established topic within the information science discipline, and he proposed that it had to become one. He also proposed a model on the basis of which digital information systems and services would be able to achieve environmental, economic, and social sustainability.

Lastly, Chowdhury also focused a lot on the open access (OA) movement and on how this could affect digital library sustainability (Chowdhury, 2014a). He specifically examined to what extent the implementation of OA policies might affect the social, environmental, and economic sustainability of digital libraries highlighting the need for more research, and claiming that the way in which OA policies are implemented may have an effect on the sustainability of OA digital libraries.

[2] The *New Zealand Electronic Text Collection* may be accessed at: https://nzetc.victoria.ac.nz/.
[3] The *Kete Horowhenua* may be accessed at: https://horowhenua.kete.net.nz/.
[4] The *New Zealand History Online* may be accessed at: https://nzhistory.govt.nz/.

4 Research design and method

As mentioned in Chapter 1, the methodology applied in the present study is bibliometric analysis. Before moving on to a discussion of the specifics of the study's design, it would be wise to briefly discuss what is meant by the term *bibliometrics*.

4.1 Bibliometrics and its use in the present study

Bibliometrics is a term that dates back to 1969. It was coined by Alan Pritchard in an attempt to replace the term *statistical bibliography* (Hulme, 1923), which had been used sparingly up to then, but which was deemed inadequate and confusing (Pritchard, 1969). Pritchard defined bibliometrics as "the application of mathematics and statistical methods to books and other media of communication" (p. 349). In 1955, Garfield first referred to the concept of *citation indexing* in an effort to facilitate scientific information retrieval by scientists. This revolutionary concept in combination with the idea of *citation analysis*, which he later came to develop, contributed to the development of bibliometrics, as this was the first attempt to create an index for the association of ideas (Garfield, 1955, 1979).

Ever since its invention and despite the fact that it is considered a controversial methodology that has provoked intense arguments for and against it over the years, the bibliometric approach has become a commonly used approach for the evaluation of science and scientific research and production. Okubo (1997) mentioned that it is a methodology used in various fields such as the history of science, the social sciences, documentation, and science policy.

Since the bibliometric methodological approach has emerged as an effective type of methodology for the evaluation of research and scientific production, it was selected as the most suitable quantitative methodological approach for the present study. More specifically, conducting a bibliometric analysis permitted the measuring, monitoring, and evaluation of scientific and scholarly production (Gumpenberger et al., 2012). The methodological framework for the present work was based on the guidelines provided by de Oliveira et al. (2019), Andrés (2009), and Ball (2018) on how to conduct a bibliometric analysis and can be briefly summarized in the following steps: (a) defining the topic to be studied; (b) selecting search platforms; (c) mining bibliometric data; (d) importing data into visualization and analysis tools to analyze them; and (e) identifying literature gaps and trends, and reaching conclusions.

4.2 Source selection and reasons for selection

The data for the current study were collected from two of the biggest multidisciplinary bibliographic databases, the *WoS* by Clarivate and *Scopus* by Elsevier. A bibliographic database contains bibliographic records, which in turn contain bibliographic metadata about scientific documents.

The two databases were selected for a number of reasons. Firstly, both fulfilled the criteria of usefulness outlined by Neuhaus and Daniel (2008), that is coverage, consistency and accuracy of data, sufficient data field availability,

browsing and searching options, analytical tools, and saving and exporting options. Secondly, the two selected databases have proven themselves to be extremely impactful for the scientific community despite the fact that both also have disadvantages (Pranckutė, 2021). For example, both have been criticized for lack of transparency as well as inaccuracies in classifying documents (Wang & Waltman, 2016). Moreover, WoS has been accused of not having as strong a representation of the LIS field as Scopus does (Pranckutė, 2021; Wang & Waltman, 2016). Similarly, Scopus has been charged with being weaker than WoS in indexing some papers, meaning that it, more frequently than WoS, fails to index certain articles (Franceschini et al., 2016), a serious error that often results in losing the citations that these articles have given or obtained.

According to de Oliveira et al. (2019), the use of more than one database for the collection of scientific data in a bibliometric study may lead to a more consistent bibliometric analysis and more robust results. This was the third reason for selecting two databases instead of just one. In the context of the present study, a fourth reason for which it was decided to use both platforms was that comparing the results from both sources of data (Denzin, 1970, as cited in Bryman, 2016) could function as a type of triangulation method. This would enhance the validity of the research output. It would also restrict possible biases that could result due to the objective restrictions each of the two databases have been proven to have, as mentioned in the previous paragraph.

The metadata of both WoS and Scopus databases are copyright protected. As a result, a subscription is required in order for someone to acquire access to the content of the two resources. Access to both platforms was ensured by both the University of Borås (UOB) and my workplace (i.e., The American College of Greece [ACG]).

With regard to the Scopus database both institutions' subscriptions were identical as the content of Scopus is available with a single subscription (Pranckutė, 2021). Nevertheless, the case was not the same with regard to the WoS database. WoS allows institutions to select a customized subscription from among the 10 available WoS Core Collection sub-data sets, which vary in terms of content and coverage timespan (Liu, 2019).

After exploration of the UOB's and the ACG's subscriptions, it was discovered that there were differences in terms of content and coverage between the two institutions' database versions. In terms of content, the ACG subscription included two Book Citation Indexes in the Core Collection, which were not included in the UOB subscription. In terms of coverage, the UOB subscription covered a much broader date range than the ACG one. A comparison of the two versions appears in Table 1.

Executing the same preliminary queries in both WoS versions pointed out that the number of final results was slightly higher in the ACG subscription. Since the total number of items was not extremely high and in order to take advantage of the best of the two versions, it was decided to run the exact same searches in both the ACG and the UOB versions and combine the two data sets. Details about how this was done are explained in Section 4.3.1.

Table 1

Differences in content and coverage between the ACG and the UOB WoS subscriptions

Index	ACG Coverage	UOB Coverage
Arts & Humanities Citation Index	2003-Present	1975-Present
Book Citation Index-Science	2005-Present	-
Book Citation Index-Social Sciences & Humanities	2005-Present	-
Conference Proceedings Citation Index-Science	2003-Present	1990-Present
Conference Proceedings Citation Index-Social Science & Humanities	2003-Present	1990-Present
Emerging Sources Citation Index	2019-Present	2019-Present
Science Citation Index Expanded	2003-Present	1900-Present
Social Sciences Citation Index	2003-Present	1956-Present

4.3 Data selection

The collected data were bibliographic information, keywords, abstracts, and citation information of published research studies on the topic of digital library sustainability. More specifically, the studies included in the data sets were *articles, review articles, books, book chapters,* and *conference papers*. Even though English is a universally used language in scientific production (de Oliveira et al., 2019) and it is frequently used as a limiter in bibliometric studies, it was decided not to limit the studies only to those written in the English language. The reason for this was that all the studies included in WoS and Scopus databases, irrespectively of the language in which their text has been written, are always accompanied by bibliographic information and abstracts written in English. Since only the bibliographic information and abstracts of the studies would be examined, there was no reason to further limit the number of included studies by language.

Furthermore, the examined timespan was decided upon. De Oliveira et al. (2019) have suggested that "for a complete map of the state of the art, it is suggested that all studies regardless of their year of publication be analyzed" ("Mining and analysis" section, para. 15). They have also advised that a date range may be selected if the topic under investigation has been studied for a long period of time, hence, too much research exists about it. Since the topic examined in the present study was relatively new, it was decided not to set any limitations with regard to the starting date of the published research. A limit was though placed to the ending date, which was set to 2023, since the content of the WoS Core Collection is being updated on a daily basis (Clarivate, 2024) and this would affect the content of the data set in case it needed to be downloaded anew.

The original data sets were extracted on February 11, 2024. Data were selected by executing the same queries in both selected databases. Moreover, the same filtering options (to the extent possible) were used on both platforms. Keyword selection was made very carefully to ensure that all relevant documents would be retrieved. Taking into account the two main concepts, namely *digital*

libraries and *sustainability*, and in accordance with de Oliveira et al.'s (2019) recommendation that when selecting keywords "the specific terms of the field under study should express the subject, its synonyms, and different spellings" ("Mining and analysis" section, para. 14), I decided to include in the search query any combination of the words *digital*, *electronic*, *online*, and *virtual* (term set A) with any of the words *library*, *archive*, *repository*, *collection* (term set B). As mentioned in Chapter 3, sustainability has been frequently discussed along three aspects: environmental, economic, and social. Aiming at grasping all these aspects of the concept, I decided not to restrict the term to any of the three. Thus, the keywords selected for the second concept were broad and limited to *sustainability* and *sustainable development* (term set C). The search terms were searched in the *article titles*, *keywords*, and *abstracts*. The exact procedures followed during the data selection phase for each database are outlined in Sections 4.3.1 and 4.3.2.

4.3.1 Web of Science query

In WoS, the following query was run in the WoS Core Collection of both the ACG and the UOB subscriptions.

TS=((digital OR electronic OR virtual OR online) NEAR/0 (library OR archive OR repository OR collection)) AND TS=(sustainability OR "sustainable development")

In the above query *TS* represents the "Topic" search field, which when selected allows the researcher to search for a term, in the *Title*, *Abstract*, *Author Keywords*, and *Keywords Plus* fields (Clarivate, n.d.-c). *Keywords Plus* is a technology that is unique to WoS database. It "supplies additional search terms extracted from the titles of articles cited by authors in their bibliographies and footnotes" (Garfield, 1990, p. 295). In other words, these key terms are generated by an algorithm without the authors necessarily being aware of them. This has made many scholars wonder whether such keywords are indeed representative of the content of an article (Zhang et al., 2016) and whether articles that include keywords only in the specific field should be excluded from a bibliometric analysis. Such articles were retained in the present study, first, due to the small sample size, and second, because according to Zhang et al., even though they are less descriptive of the content of a document, they have been found to be as effective as author keywords in terms of the knowledge structure of scientific fields.

Moving back to the query, the *NEAR/0* proximity operator was used in order to determine the relationship between term set A (i.e., digital OR electronic OR virtual OR online) and term set B (i.e., library OR archive OR repository OR collection). NEAR/0 signified that any term from term set A had to be adjacent to any term from term set B (Clarivate, n.d.-b). The Boolean operator *AND* was used to determine that any combination of terms from term sets A and B had to be present along with any of the terms/phrases from term set C (i.e., sustainability OR "sustainable development").

After running the search query, the results were refined using some of the available filtering options. Specifically, the *Publication Years* and the *Document Types* filters were used. For the reasons explained in Section 4.3, the

Publication Years filter was set to exclude the year 2024. Anything published prior to that date was included in the data set. Furthermore, the following publication types were retained: *Article, Proceeding Paper*, *Review Article*, *Book Chapter*, whereas *Early Access*, *Editorial Material*, and *Book Review* were excluded.

The above-described procedure was followed both in the ACG and in the UOB versions of WoS. Two data sets, one from each institution's subscription were exported. The data sets were downloaded in *plain text* (recommended requirement by Biblioshiny, see Section 5.1) and *tab-delimited* (recommended requirement by VOSviewer, see Section 5.1) format including the Full Record and Cited References as per the guidelines provided by Aria (2021) and van Eck and Waltman (2023). The instances included in both were compared based on the WoS ID and all unique instances were retained. Since the items that were unique in each of the two data sets were few, the two data sets were manually combined. This resulted in the final WoS data set.

4.3.2 Scopus query

Similar procedures were followed in Scopus. The following query was executed:

TITLE-ABS-KEY((digital OR electronic OR virtual OR online) W/0 (library OR archive OR repository OR collection)) AND TITLE-ABS-KEY (sustainability OR "sustainable development")

In the above query *TITLE-ABS-KEY* was used, which allows the researcher to search for a term in the *Title, Abstract,* and *Keywords* search fields (Elsevier, 2024). The *W/0* proximity operator was applied to determine the relationship between term set A (i.e., digital OR electronic OR virtual OR online) and term set B (i.e., library OR archive OR repository OR collection). W/0 signified that any term from term set A had to be adjacent to any term from term set B (Elsevier, 2024). The Boolean operator *AND* was used to determine that any combination of terms from term sets A and B had to be present along with any of the terms/phrases from term set C (i.e., sustainability OR "sustainable development").

After running the search query, the results were refined using some of the available filtering options. Specifically, the *Year*, the *Document type*, and the *Publication stage* filters were used. Similarly to WoS, it was decided to include publications that were published by the end of 2023. Therefore, the Year filter was set to exclude the year 2024. Anything published prior to that date was included in the data set. Furthermore, the following publication types were retained: *Article, Conference paper, Review, Book chapter, Book,* and *Short survey*, whereas *Editorial, Note, Erratum,* and *Conference review* were excluded. The publication stage filter was used to exclude articles in press. The data set was downloaded in *comma separated value* (recommended requirement by both Biblioshiny and VOSviewer, see Section 5.1) as per the guidelines provided by Aria (2021) and van Eck and Waltman (2023).

4.4 Data cleaning

The two data sets had to be free from duplications and irrelevant items in order for them to be as representative as possible. Therefore, they underwent quantitative (in terms of duplications), qualitative (in terms of content), and metadata (in terms of normalization) checks to ensure accuracy and validity of the results. The data cleaning process comprised the following four phases:

1. Removal of duplicate items.
2. Source-based screening.
3. Keyword-based screening.
4. Author, source title, and affiliation normalization.

The *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA; Page et al., 2021) methodological framework was used during the first three phases of the cleaning process. PRISMA is an internationally recognized standard used to enhance transparency when conducting mostly systematic reviews and meta-analyses. Nevertheless, it has been used successfully in bibliometric studies in the past (see for instance Khin Khin Oo & Rakthin, 2022; Maier et al., 2020), and was, therefore, selected as appropriate to be used in this study. More specifically, the PRISMA flow diagram was used to map the number of originally identified records, the number of those eventually included in the analysis, and the number of those excluded along with the reasons for exclusion. The PRISMA flow diagrams for each one of the two data sets may be seen in Figures 1 and 2.

As far as the qualitative check is concerned, it was deemed to be mandatory, because the term *sustainability* may be encountered in multiple disciplines and the term *digital libraries* is frequently used in a different sense rather than the one with which it is used in the present study. During this check, items relevant to the topic under examination were retained and the rest were removed. Also, items were retained no matter whether digital library sustainability was their main or secondary focus.

Since both data sets were fairly small, I manually did the data cleaning. No automation tools were used. Each one of the four phases of the data cleaning stage is thoroughly described in Sections 4.4.1 to 4.4.4.

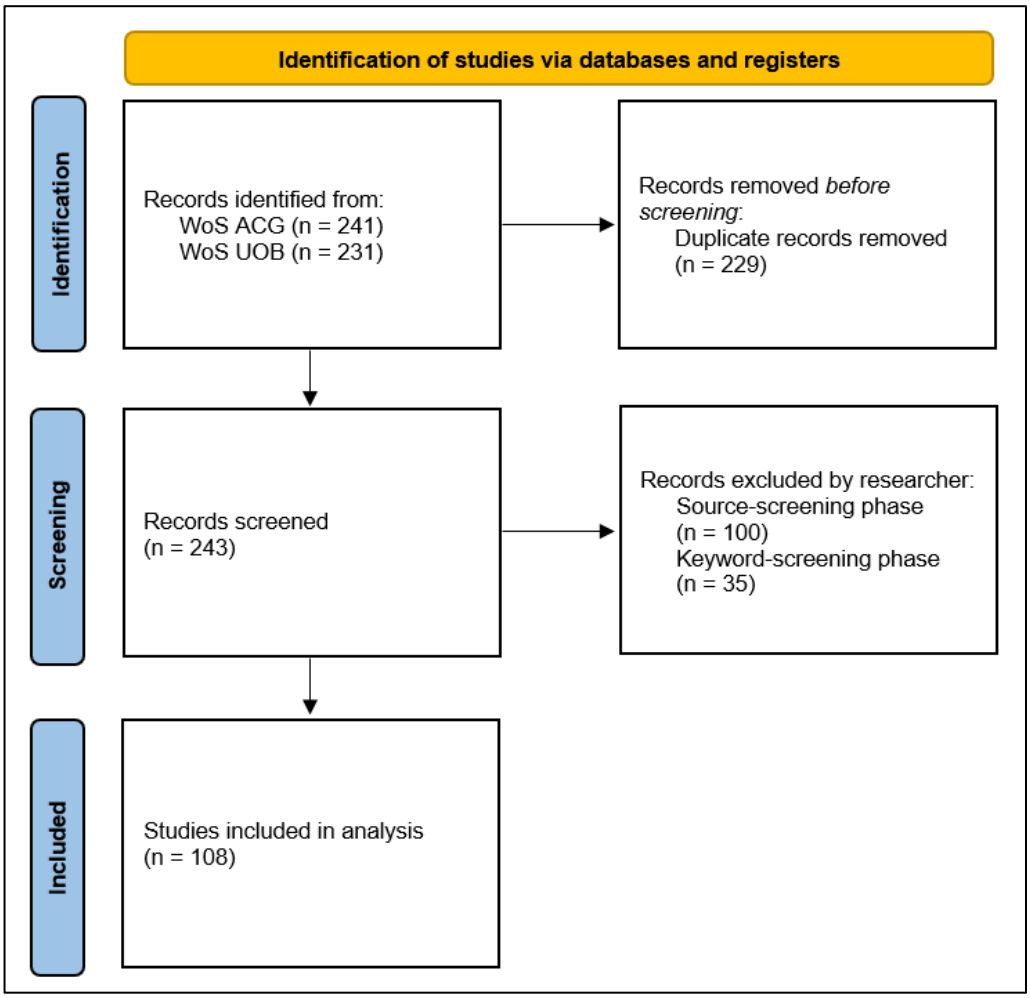

Figure 1: WoS PRISMA flow diagram. The figure shows how the data screening process evolved until the final WoS data set was reached. The flow diagram is the result of an adjustment of the "PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only" (https://www.prisma-statement.org/prisma-2020-flow-diagram).

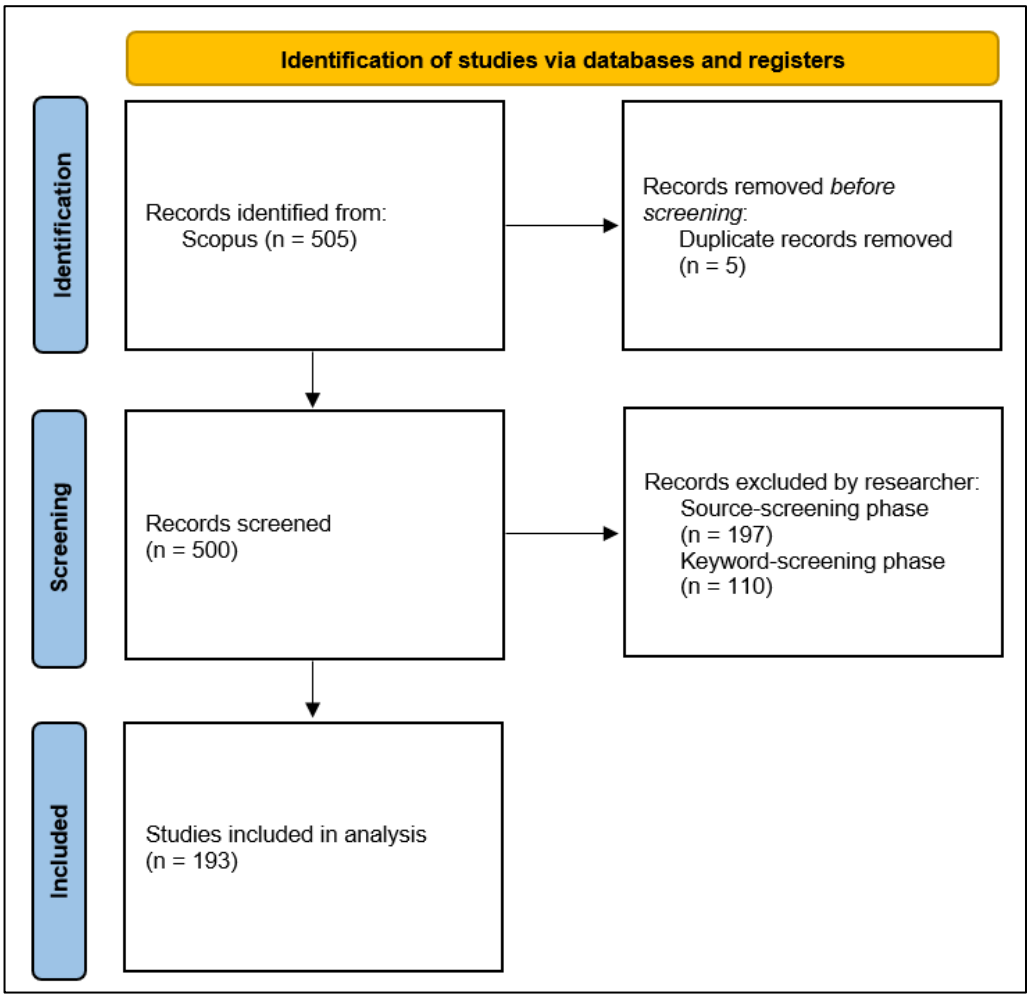

Figure 2: Scopus PRISMA flow diagram. The figure shows how the data screening process evolved until the final Scopus data set was reached. The flow diagram is the result of an adjustment of the "PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only" (https://www.prisma-statement.org/prisma-2020-flow-diagram).

4.4.1 Removal of duplicate items

During the first phase of the cleaning process, both the WoS and Scopus data sets were checked for duplicate items. In WoS, this was done by checking the *Accession Number*, the Digital Object Identifier (*DOI*), the item *Title*, the *Author Keywords*, and the item *Abstract* fields for duplicate values. The duplicate records found at this stage were the result of the content overlap between the ACG and the UOB WoS versions. No other duplicate records were found.

In Scopus, a similar procedure was followed. The Scopus *Document Identifier*, the *DOI*, the *Document Title*, the *Author Keywords*, and the item *Abstract* were checked. Checking the Digital Object Identifier resulted in two duplicate records; checking the Document Title, Author Keywords, and Abstract resulted in one duplicate record per field. Therefore, a total of five duplicate records were found and removed. In the case in which one of the two duplicate records had been cited in other publications, the cited record was retained. In all other cases, the record with the most recent publication date was retained. In one

case, one of the duplicates carried a title in a non-English language and because of this it was discarded.

4.4.2 Source-based screening

During the source-based screening phase, the title, keywords, and abstract of items published in sources, or presented in conferences, which did not relate to library and information science, or computing were examined. If the title, the keywords, or the abstract included any combination of term sets A, B, and C, as these have been described in Section 4.3, then the item was retained. If an item included any of the query terms, but in a different context, it was excluded. An item was also excluded if the query terms were indeed mentioned in any of the three examined fields as part of a title (e.g., IEEE Xplore Digital Library or ACM Digital Library), but the content of the study was totally irrelevant to digital library sustainability. A document was also excluded if it referred to sustainability or SDGs, but not in the context of digital libraries. Titles that included any of the terms that appeared in the original search query were retained and further examined if their exact content was not explicit just by the title. The items whose titles included the query terms, but it was not explicit if they were relevant were further examined in terms of keywords, source, and abstract. The source-based screening resulted in removing 100 records from the WoS data set and 197 records from the Scopus data set.

4.4.3 Keyword-based screening

The keyword-based screening was performed on the remaining records of each data set after removing the items that were excluded during the source-screening phase. This second screening was done starting with the author keywords, since they constitute a concise and indicative description of the item's content. Items whose author keywords included both the term sustainability (or any of its derivatives) and any combination of term sets A and B, as these were described in Section 4.3, were retained. The rest were further examined on the basis of their title and abstract. In case an item did not have keywords, it was judged based on the rest of the criteria, that is the title and the abstract. The keyword-based screening resulted in removing 35 additional records from the WoS data set and 110 additional records from the Scopus data set.

4.4.4 Author, source title, and affiliation normalization

In the last phase of the data cleaning process author names, source titles, and affiliations were normalized. Due to the small size of the data sets the procedure was done manually. In cases when author names, source titles, and affiliations appeared in multiple variations, one of these variations was retained and all the rest were replaced by the selected one. The alterations made in the WoS data set appear in Appendix A and those made in the Scopus data set appear in Appendix B.

4.5 Bibliometric and other techniques used

Since the content of the WoS and Scopus data sets was significantly different (for a comparison of the two data sets, see Section 5.2), it was decided to conduct a separate bibliometric study on each data set. This was reinforced by the fact that (a) the differences in their metadata did not permit the combination of the two and (b) this way, the results extracted from the two databases could be compared. Consequently, all the selected bibliometric indicators were applied on both data sets and the results from both are presented in Chapter 5.

In an attempt to describe the existing literature, determine its impact, trace collaboration patterns, identify major research areas and trends, and explore to what extent the UN SDGs have been a topic that has been addressed in the examined literature, a variety of bibliometric and other techniques and metrics were put to practice. More specifically, I moved methodologically along three basic axes, which I developed based on the set research questions. These focused on the following:

1. Examination of the overall structure of the current literature on the topic with the purpose of highlighting and recording the current situation (referred to as the *descriptive axis*).

2. Analysis of the intellectual, social, and conceptual structure of knowledge with the intent to pinpoint current trends (referred to as the *analytical axis*).

3. Exploration of SDG mappings to the examined literature with the intent to determine (a) the extent (if any) to which researchers dealing with the topic under discussion have been preoccupied with the UN SDGs, and (b) how different bibliometric databases deal with SDG mapping (referred to as the *mapping axis*).

More details about the examination of each one are provided in Sections 4.5.1 to 4.5.3, which follow.

4.5.1 Descriptive axis

The exploration of the overall structure of the current literature on the topic of digital library sustainability involved the examination of the annual scientific production along with an analysis of the documents, authors, publications, institutions, and countries. The specific items were selected for the analysis because they would provide a good overview of the topic under examination and determine the current state of published literature.

4.5.2 Analytical axis

Analysis of the knowledge structures to identify current trends in the field was conducted at three levels: intellectual, social, and conceptual. Citation and co-citation analysis were conducted in order to determine the impact of scientific production on generating new knowledge, thus new intellectual wealth. Co-authorship analysis as well as examination of the collaborations among countries were conducted in order to establish the social knowledge structure.

Co-word and thematic map analysis were applied to establish the conceptual knowledge structure.

4.5.3 Mapping axis

Both WoS and Scopus databases recognizing the significance of the SDGs have mapped their content, whenever applicable, to them.

WoS uses a three-level hierarchical classification system to organize items by topic. With the help of a clustering algorithm topics are divided into macro-topics, meso-topics, and micro-topics moving from broader to more specific subject areas. The SDG schema used by WoS follows a category-to-category mapping system to map sixteen out of the seventeen UN SDGs (SDG17, "Partnerships for the goals," is excluded) to various micro-topics and the publications associated with them (Clarivate, n.d.-a).

Scopus maps SDGs to documents by matching terms from a record's title, keywords, key descriptors, journal subject areas, and abstracts against predetermined queries designed by Elsevier's data scientists, supplemented by a predictive machine learning element (Elsevier, 2023). *Times Higher Education*, who originally requested from Elsevier the above, use Elsevier's SDG mapping in their Impact Rankings. Since 2023, the 17th SDG has been included in the Scopus classification scheme, even though, just like in WoS, it was originally excluded as it had been deemed difficult to quantify (McCullough, 2020, 2023).

To examine whether the UN SDGs constitute a study object within the currently existing literature on digital library sustainability research, I measured how many documents in the used data sets had been mapped to one or more UN SDGs. I also measured which SDGs were the most prominent within the examined data sets. In WoS, this was done by using the "Sustainable Development Goals" filtering option and documenting which SDGs were listed and how many articles from the data set were mapped to each one of them. Given that in the Scopus database no such filtering option existed, a slightly different procedure was followed. Documents were examined one by one. Each item's record was examined to determine if the "Sustainable Development Goals 2023" field was available in it. For the items in which the specific field was present, the SDGs mapped to the item were documented. In the end, the total number of items that included each SDG in their record was calculated.

To further examine the way in which the two databases map SDGs, and to determine the trustworthiness of bibliographic database mapping procedures, I moved on to comparing the mappings of the documents that were common in the two data sets. This process involved: (a) determining the common items between the two data sets based on item titles; (b) finding the respective SDGs for each title; (c) identifying mapping similarities and differences between the two data sets.

5 Data analysis and results

5.1 Tools used for data analysis

For the analysis of data along the descriptive and analytical axes, I used two open-source programs: *Biblioshiny* by Bibliometrix and *VOSviewer* by Leiden University. Both tools have been recommended by Moral-Muñoz et al. (2020) as especially useful and effective tools to be used in a bibliometric study.

Biblioshiny is a web-based application included in the wider Bibliometrix package. It is a flexible bibliometrics tool programmed in R (Aria & Cuccurullo, 2017). The Biblioshiny app allows non-programmers to use R through a user-friendly and easy to use interface.

VOSviewer is a piece of software programmed in the Java programming language (van Eck & Waltman, 2023). It was created for the analysis of bibliometric networks, but also supports the creation, visualization, and exploration of maps from imported data. For more information on the software, one may refer to van Eck and Waltman (2010).

As per the recommendations of the creators (Aria, 2021), data from WoS were imported in *plain text* format and data from Scopus were imported in *comma separated value* format in the Biblioshiny web-based application. Similarly, data from WoS were imported in *tab delimited* format and data from Scopus were imported in *comma separated value* format in VOSviewer as per the recommendation of van Eck and Waltman (2023).

For the analysis of data along the SDGs mapping axis, I used Microsoft Excel and its treemap and column chart creation feature.

5.2 WoS and Scopus data set comparison

The two final data sets were compared in terms of content in order to determine whether there was an overlap between them. The comparison was made based on the DOI, which is a unique and persistent digital identifier assigned to any kind of object, numerous published items included (doi Foundation, n.d.). For items that were not assigned a DOI, the comparison was made based on the title and other bibliographic information. After the comparison of the two data sets, it was discovered that 77 items were common in both, constituting 34,38% of the total number of titles examined. The combined percentage of unique titles in both data sets exceeded 65% of the total number of examined titles. For a more detailed description see Table 2.

Table 2

Number of unique and common items in the WoS and Scopus data sets

	No. of items	%
WoS unique titles	31	13,84%
Scopus unique titles	116	51,79%
Common titles	77	34,38%
Sum	224	100,00%

Note. Over half of the items included in the Scopus data set were unique, whereas the percentage of unique items in the WoS data set was significantly lower.

As expected, since the Scopus data set was bigger than the WoS one, it outnumbered the second in terms of total number of documents, sources, authors, authors of single-authored documents, author keywords, and references (see Figures 3 and 4). Nevertheless, it can be noticed that the respective numbers in the two data sets were proportional.

In terms of the rest of the indicators appearing in Figures 3 and 4, the results were more or less similar, with Scopus being about one unit higher than WoS in terms of annual growth rate, and about one and a half unit lower than WoS in terms of international co-authorship rate. Document average age was estimated to be 8,33 years in WoS versus 9,66 years in Scopus, while co-authorship per document was almost the same in both data sets. Lastly, the average number of citations per document was slightly lower in WoS as compared to that of Scopus.

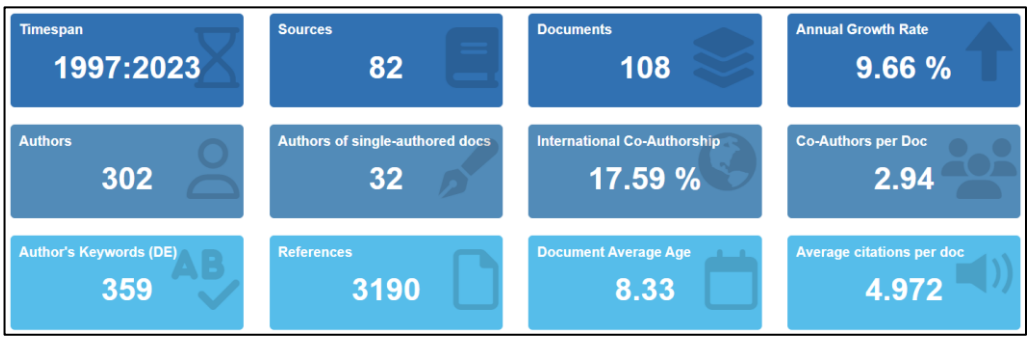

Figure 3: Overview plot of the WoS data set. Figure extracted from Biblioshiny.

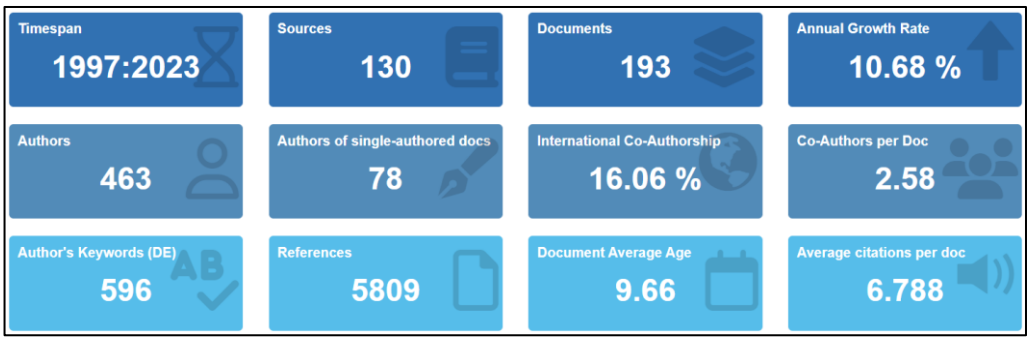

Figure 4: Overview plot of the Scopus data set. Figure extracted from Biblioshiny.

What follows is a detailed presentation of the study's results. The results are presented along the three methodological axes (i.e., descriptive, analytical, mapping) that have been described in Sections 4.5.1 to 4.5.3. Each of them attempts to answer the present study's research questions, starting with the topic representation in current published literature, continuing with the current trends, and concluding with the representation of the SDGs in each one of the two data sets.

5.3 Descriptive axis results

To achieve an overview of the two data sets, I first examined and analyzed the annual scientific production on the topic. I continued with an analysis at the level of documents, authors, publications, institutions, and countries. It should be noted that the two databases are using different classification schemes for their indexed content (Pranckutė, 2021), which may have an effect on the results.

5.3.1 Annual scientific production

The scientific production trends were first analyzed on the basis of annual scientific production. The timeframe of published research on the topic of digital library sustainability may be roughly divided into three periods, which may be discerned in the results of both data sets. These may be characterized as the *initial*, the *steady rise*, and the *high growth* phase. As shown in Figure 5, in the WoS data set, the first phase extends from 1997 to 2005, the second from 2006 to 2012, and the third from 2013 to 2023. In the Scopus data set (see Figure 6), the initial phase is a little shorter extending from 1997 to 2002, the steady rise phase (2003-2009) is the same in terms of length but starts and ends earlier than what is the case in WoS, and the high growth phase (2010-2023) starts a little earlier and continues to show increase. Both in WoS and in Scopus, the last phase presents fluctuations. Also, in both data sets, the growth of scientific production is exponential, with an annual growth rate of 9,66% in WoS and 10,68% in Scopus (see Figures 3 and 4, p. 37).

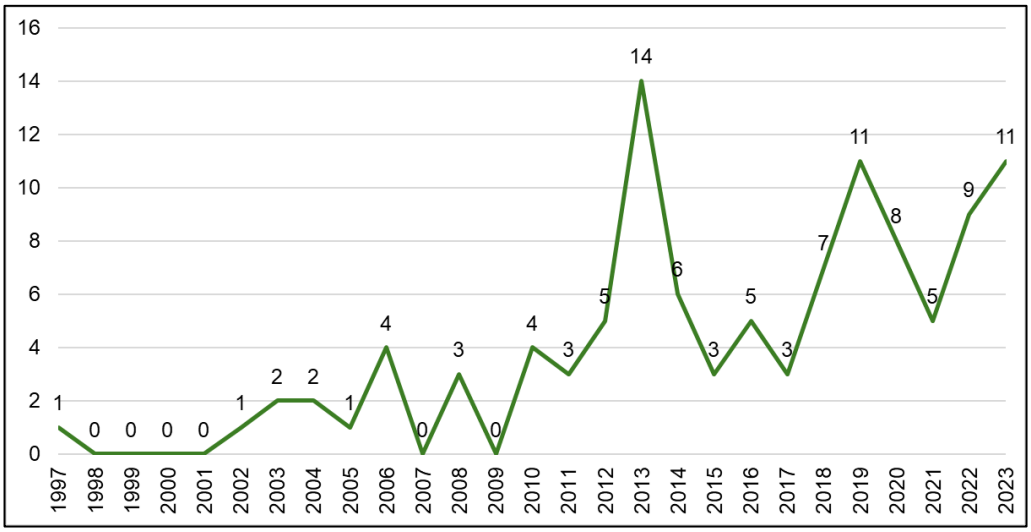

Figure 5: Temporal evolution of scientific productivity in the WoS data set. Three phases (i.e., 1997-2005, 2006-2012, and 2013-2023) of scientific productivity on the topic of digital library sustainability may be noticed.

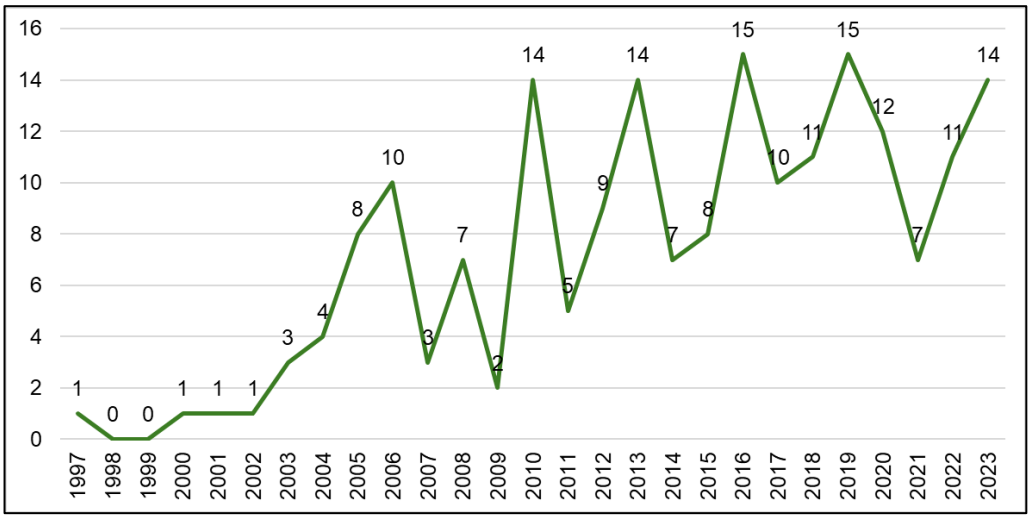

Figure 6: Temporal evolution of scientific productivity in the Scopus data set. Three phases (i.e., 1997-2002, 2003-2009, and 2010-2023) of scientific productivity on the topic of digital library sustainability may be noticed.

5.3.2 Document types

The 108 documents included in the WoS data set were divided into four document types, whereas the 193 documents included in the Scopus data set were divided into five document types. In both data sets, the most frequent document type was articles, and the vast majority of documents were written in English. More details about the document types and the languages in which documents were written appear in Tables 3 and 4.

Table 3

Number of items per document type in the WoS and Scopus data sets

Document type	No. of items (WoS)	No. of items (Scopus)
Articles	59	108
Proceeding/Conference papers	35	49
Book chapters	8	20
Reviews	6	14
Books	0	2

Note. There is an analogy in the number of items per document type in the two data sets.

Table 4

Number of items per document language in the WoS and Scopus data sets

Document language	No. of items (WoS)	No. of items (Scopus)
English	103	186
Spanish	2	2
Italian	1	2
Chinese	0	2
Catalan	1	1
Russian	1	0

5.3.3 Authorship

For the period under examination, 302 authors were identified in the WoS data set. Out of them, 32 were authors of single-authored documents. The top authors were G. Chowdhury with four papers and P. Simpson with three papers, whereas 3,64% of authors wrote two papers and the vast majority of 95,7% wrote only one paper each.

Comparably, 463 authors were identified in the Scopus data set. Out of them 78 were authors of single-authored documents. The top authors were G. Chowdhury with six papers, P. Simpson, A. Adams, K. R. Eschenfelder, P. Ngimwa, and K. Shankar with three papers each, whereas 4,10% of authors wrote two papers and the vast majority of 94,6% wrote only one paper each.

5.3.4 Publication source production

The papers included in the WoS data set were published in 82 sources. One third (36) of these papers were published in the top 10 publication sources. On the other hand, the papers included in the Scopus data set were published in 130 sources and 56 of these papers were published in the top 10 publication sources.

Among the 10 most prominent sources that were common in both data sets were one book (i.e., the *Handbook of Digital Library Economics: Operations, Collections and Services*) and four journals (i.e., *Electronic Library*, *Program-Electronic Library and Information Systems*, *Sustainability*, and *Library Hi Tech*). More details regarding the top 10 publication sources in each one of the examined data sets are presented in Table 5.

Table 5

Top 10 publication sources in the WoS and Scopus data sets

Document source	No. of items (WoS)	No. of items (Scopus)
Lecture Notes in Computer Science	-	8
Electronic Library	***6***	***7***
Handbook of Digital Library Economics: Operations, Collections and Services	***6***	***6***
Proceedings of the Association for Information Science and Technology	-	6
Program-Electronic Library and Information Systems	***5***	***5***
Proceedings of the ACM/IEEE Joint Conference on Digital Libraries	(1)	5
Sustainability	***4***	***4***
Communications in Computer and Information Science	-	4
Library Hi Tech	***3***	***5***
International Journal of Heritage Studies	3	(1)
Journal of Librarianship and Information Science	3	(2)
Developing Sustainable Digital Libraries: Socio-Technical Perspectives	-	3
International Journal on Digital Libraries	-	3
Archiving 2013: Final Program Proceedings	2	(1)
Digital Library Perspectives	2	(2)
Journal of Documentation	2	(1)

Note. The sources presented in bold italics were common in both data sets. The numbers in parentheses indicate that the relevant source was not among the top 10 in the respective data set. In the Scopus data set, two sources shared the 10th position having produced the same number of items, thus, both were included in the table.

5.3.5 Institutional production

Institutional production was examined in relation to the most relevant author affiliations. The WoS data set included 171 institutional affiliations, and the top five institutions within it published 12,84% of the total number of documents on the topic. Similarly, the Scopus data set included 195 institutional affiliations, and the top five institutions within it published 14,72% of the total number of publications. The only common institutional affiliation among the most productive ones in the two data sets was *University College London*. More details regarding the institutions with the highest number of document production on the topic appear in Table 6.

Table 6

Five most productive institutions in the WoS and Scopus data sets

Document source	No. of items (WoS)	No. of items (Scopus)
Natural Resources Canada	9	(4)
University of Innsbruck	(1)	8
Universitat Politecnica De Valencia	(1)	7
University of California	-	7
University of Aberdeen	4	-
University College London	***3***	***7***
Kent State University	3	-
University of Southampton	3	(1)
University of Malaya	3	-
University of London	3	(2)
Washington State University	(1)	6
Universitat Rostock	(2)	6
Universitat De València	(1)	6
Open University	(1)	6

Note. The institution presented in bold italics was common in both data sets. The numbers in parentheses indicate that the relevant institution was not among the top five in the respective data set. In both data sets, more than one institution shared the fifth position having produced the same number of items, thus, all were included in the table.

5.3.6 Country production

Considering the country where authors worked when they published their research, Table 7 presents the most highly producing countries on the topic. The dominant among the 37 (for WoS) and 41 (for Scopus) identified countries was the USA, while nine among the most highly producing countries were common in the two data sets.

The maps presented in Figures 7 and 8 are geographic visual representations of the scientific production on the topic under discussion based on the WoS and Scopus data sets, respectively. The highest percentage of countries dealing with the topic were European in both data sets (WoS: 55,26%; Scopus: 51,22%). Asian countries came next (WoS: 18,52%; Scopus: 26,83%), followed by African (WoS: 10,53%; Scopus: 9,76%) and North American countries (WoS: 10,53%; Scopus: 7,32%).

Table 7

Ten most productive countries in the WoS and Scopus data sets

Document country	No. of items (WoS)	No. of items (Scopus)
USA	***41***	***151***
UK	***38***	***62***
Germany	***14***	***28***
China	***14***	***21***
Malaysia	***10***	***11***
Canada	***9***	***22***
Spain	***7***	***22***
India	***7***	***15***
Australia	***6***	***13***
Romania	6	(4)
Austria	(2)	12
Greece	(2)	11
Italy	(1)	11

Note. The countries presented in bold italics were common in both data sets. The numbers in parentheses indicate that the relevant countries were not among the top 10 in the respective data set. In the Scopus data set, three countries shared the 10th position having produced the same number of items, thus, all three were included in the table.

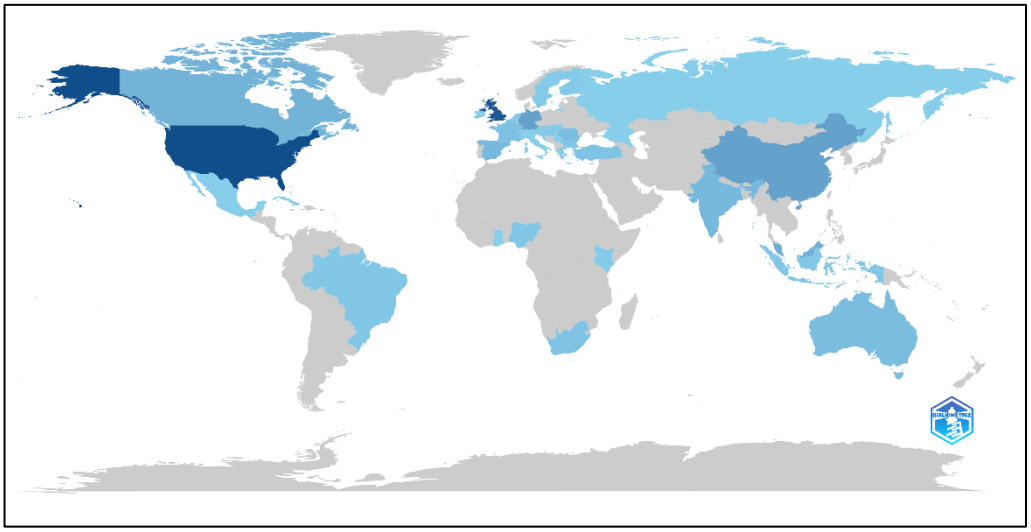

Figure 7: Geographic visual representation of scientific production in the WoS data set. The intensity (darkness) of the color indicates a higher number of publications by the specific country.

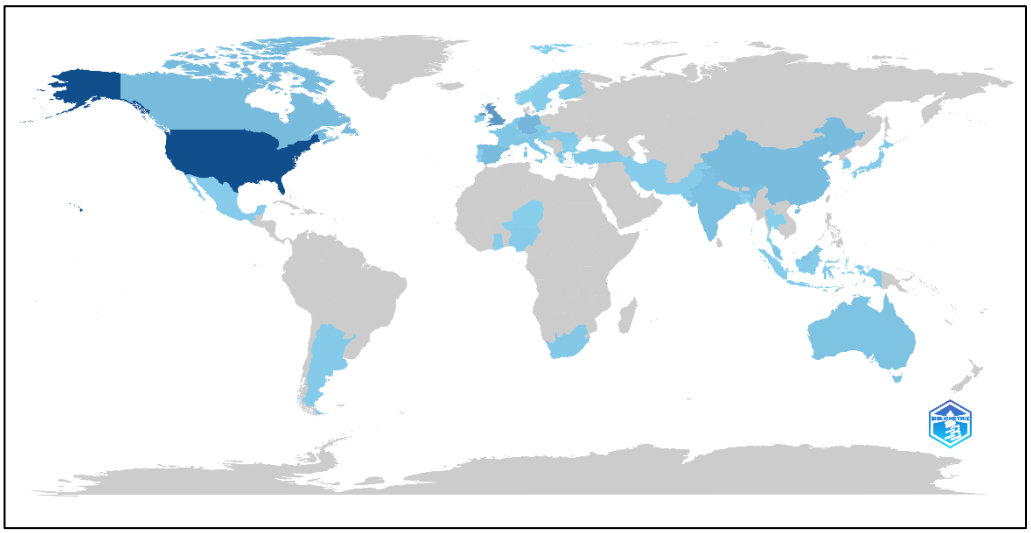

Figure 8: Geographic visual representation of scientific production in the Scopus data set. The intensity (darkness) of the color indicates a higher number of publications by the specific country.

5.4 Analytical axis results

Analyzing the knowledge structure of academic literature at the intellectual, social, and conceptual level contributed to the understanding of research trends on the topic. An analysis at the intellectual level revealed the documents and authors that have contributed most to the advancement of the field under discussion; an analysis at the social level provided insights regarding the degree of cooperation in the field; lastly, an analysis at the conceptual level highlighted trending topics of discussion within the field.

It should be noted that I tried to use the default thresholds (i.e., criteria for excluding the least important authors, sources, or documents) for most metrics. Nevertheless, due to the small size of the data sets this was not always possible and there were cases when the thresholds were altered to achieve better and more meaningful visualizations. Whenever this is the case, it is mentioned in the analysis of the results. I also attempted to maintain the same thresholds for both data sets to achieve comparability between the WoS and Scopus results. However, there were cases when this was not possible, and thresholds were adjusted to achieve similar levels of detail in the respective visualizations.

5.4.1 Intellectual structure

The intellectual structure of the scientific literature on the topic of digital library sustainability was examined based on citation and co-citation analysis.

5.4.1.1 Citation analysis

Biblioshiny counts two types of citations, namely *global citations* and *local citations* (Bibliometrix, n.d.). Global citations represent the total number of citations that an article, included in the examined data set, has received from documents indexed in the entire database. Local citations are citations an article has received from other articles within the examined data set. Even though local citations show inter-domain citation relationships and would,

therefore, be more relevant to investigate for the purposes of this study, I examined global citations. The reason for this was that the Scopus data set returned zero local citations and could, therefore, not be evaluated, neither compared to the respective one emerging from WoS.

The most globally cited papers on the topic within each database are presented in Table 8. Six among the 10 most globally cited papers of each database were common in both WoS and Scopus.

Table 8

Ten most globally cited papers in the WoS and the Scopus data sets

	Citations	
Paper	**WoS**	**Scopus**
Smith M, 2003, D-Lib Mag 10.1045/january2003-smith	-	181
Chowdhury G, 2012, J Am Soc Inf Sci Tec 10.1002/asi.21703	47	-
Muehlberger G, 2019, J Doc 10.1108/JD-07-2018-0114	***33***	***58***
Causer T, 2012, Lit Linguist Comput 10.1093/llc/fqs004	***33***	***49***
Xie I, 2016, Discov Digit Libr: Theory and Pract	-	38
Chowdhury G, 2014, Int J Digital Libr 10.1007/s00799-014-0116-0	-	34
Baker S, 2015, Int J Herit Stud 10.1080/13527258.2015.1041414	***31***	***33***
Singh M, 2020, Int J Inform Manage 10.1016/j.ijinfomgt.2020.102147	***27***	***35***
Rizor Sl, 2014, J Scholarly Publ 10.3138/jsp.45.4.01	***26***	***28***
Hamilton V, 2004, Libr Rev 10.1108/00242530410556210	-	28
Chowdhury G, 2012, Inform Process Manag 10.1016/j.ipm.2012.02.003	***26***	***36***
Tait E, 2013, Aslib Proc 10.1108/AP-05-2013-0039	21	(27)
Meschede C, 2019, J Doc 10.1108/JD-02-2019-0021	20	-
Fanea-Ivanovici M, 2020, Ieee Access 10.1109/ACCESS.2020.2975542	17	-
Spoor S, 2019, Database-Oxford 10.1093/database/baz077	17	(19)

Note. The papers presented in bold italics were common in both data sets. The numbers in parentheses indicate that the relevant paper was not among the 10 most highly cited in the respective data set. In the WoS data set, two papers shared the 10th position with the same number of citations, thus, both were included in the table.

Next, I examined the most cited authors as these may be considered to be the ones who have contributed the most to the advancement of knowledge on the topic under discussion. Out of the 247 authors in the WoS data set, nine met the thresholds, with the minimum number of documents of an author set to two and the minimum number of citations of an author set to one. The same thresholds were retained in the Scopus data set. In this data set, out of the 408 authors, 23 met those thresholds. In both data sets, the most cited author was found to be G. Chowdhury, while the only other name that was common between the two was that of P. Simpson. The rest of the most highly cited authors appear in Table 9.

Table 9

Ten most cited authors in the WoS and Scopus data sets

	Citations	
Author	**WoS**	**Scopus**
Chowdhury, G.	***89***	***98***
Xie, I.	-	47
Eschenfelder, K.	-	33
Shankar, K.	-	33
Fanea-Ivanovic, M.	27	-
Salo, D.	-	27
Williams, R.	-	27
Zhang, M.	-	27
Churchill, E.	-	24
Simpson, P.	***14***	***21***
Castelli, D.	-	19
Ocon, D.	14	-
Buttigieg, P. L.	5	-
Pearlman, J. S.	5	(12)
Chiang, J. K.	4	-
Bruder, I.	1	(1)
Heuer, A.	1	(1)

Note. The authors presented in bold italics were common in both data sets. In WoS, only nine authors met the thresholds. In Scopus, the numbers in parentheses indicate authors who met the thresholds but were not among the 10 most highly cited.

5.4.1.2 Co-citation analysis

Small (1973) has defined co-citation as "the frequency with which two items of earlier literature are cited together by the later literature" (p. 265). It is a measure that allows researchers to monitor how scientific fields develop and how they interrelate. It is assumed that the more two documents or authors are cited together, the closer they relate to each other, which may be an indication of emerging research fronts. According to Persson (1994) research fronts are formulated by the citing articles, whereas the cited articles are the ones that form an intellectual base.

Figure 9 presents how authors extracted from the WoS data set interrelate. Authors are presented in five clusters indicated by different colors. The nodes represent the researchers, while the size of the nodes represents the number of publications meeting the analysis criteria and authored by the specific researcher. The bigger the node the more the publications. Lastly, the distance between researchers indicates the level of interrelation between them. It should be noted that due to size restrictions, some of the labels in such visual representations may not be displayed on the exported maps.

In the case of the WoS data set out of the 2.537 authors, 15 met the threshold, with a minimum number of citations of an author set to six. The default threshold (i.e., 16) was not retained because due to the small data set it resulted in an extremely low number of authors (i.e., three). Even the usually applied

threshold of 10 was not applied in this case because it resulted in just four authors. Among the five most influential co-cited authors were G. Chowdhury (46 citations), UNESCO (18 citations), the European Commission (16 citations), B. Stiegler (nine citations), and D. Baker (eight citations).

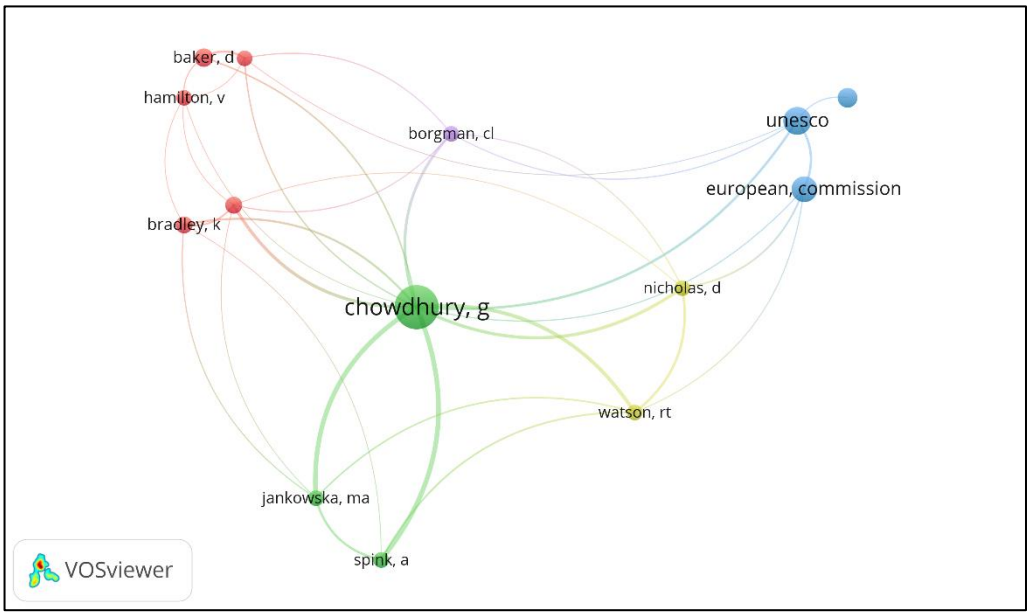

Figure 9: Author co-citation network in the WoS data set.

Figure 10 presents how authors extracted from the Scopus data set interrelated. Authors were presented in three clusters indicated by different colors. Out of the 6.267 authors, 36 met the threshold, with a minimum number of citations of an author set to 10. Among the five most influential co-cited authors were G. Chowdhury (55 citations), C. L. Borgman (28 citations), I. Xie (23 citations), C. Lagoze (22 citations), E. A. Fox, and M. Lesk (17 citations each).

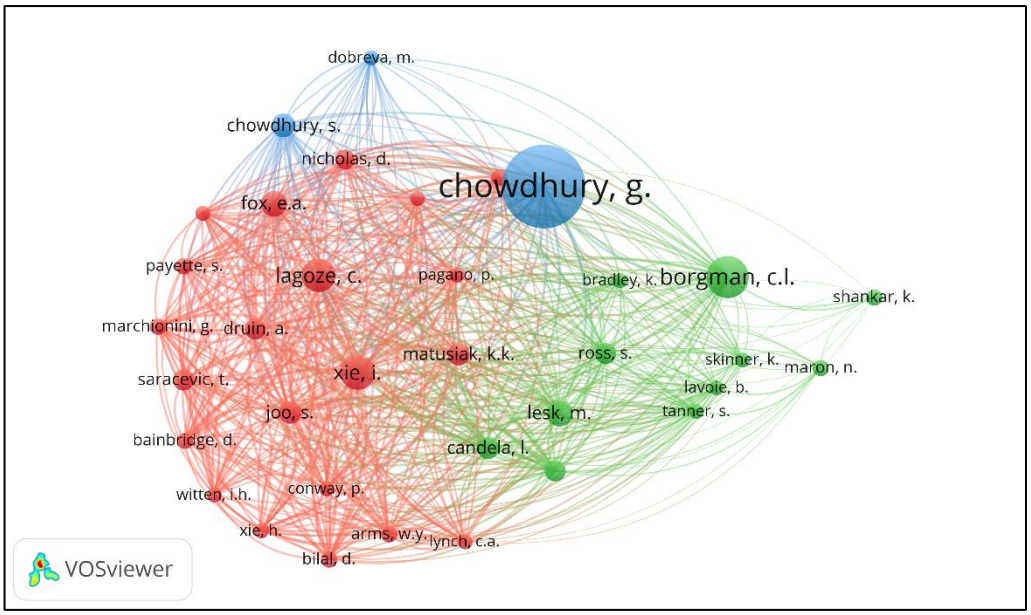

Figure 10: Author co-citation network in the Scopus data set.

Apart from authors, a co-citation analysis of the most highly cited papers was also conducted. In the WoS data set, out of 3.185 cited references, seven met the threshold with a minimum number of five citations of a cited reference. In

the Scopus data set, out of 5.751 cited references, five met the threshold with a minimum number of three citations of a cited reference. The results from both data sets are presented in Table 10.

Table 10

Most co-cited references in the WoS and Scopus data sets

	Citations	
Cited reference	**WoS**	**Scopus**
chowdhury g, 2013, j doc, v69, p602, doi 10.1108/jd-08-2012-0104	***7***	***3***
chowdhury g, 2010, j doc, v66, p934, doi 10.1108/00220411011087878	***5***	***3***
chowdhury g, 2014, int j digit librarie, v14, p181, doi 10.1007/s00799-014-0116-0	5	-
chowdhury g, 2016, j assoc inf sci tech, v67, p2379, doi 10.1002/asi.23599	5	-
hamilton v., 2004, lib rev, v53, p392	***5***	***3***
jenkin ta, 2011, inform organ-uk, v21, p17, doi 10.1016/j.infoandorg.2010.09.003	5	-
nolin j, 2010, inform res, v15	5	-
bollen j, 2002, d-lib mag, v8, doi 10.1045/june2002-bollen	-	3
conway p, 2010, lib q, v80, p61, doi 10.1086/648463	-	3

Note. The references presented in bold italics were common in both data sets.

5.4.2 Social structure

An attempt to capture the social structure and patterns appearing in the literature on the topic of digital library sustainability was made. This was done by tracing patterns of collaboration between authors and countries.

5.4.2.1 Co-authorship

The co-authorship network analysis in both data sets showed that not much collaboration exists among researchers. Figures 11 and 12 present co-authorship networks resulting from the WoS and Scopus data sets, respectively. In WoS, authors were divided into seven clusters, whereas in Scopus they were divided into 14 clusters indicated by different colors. Authors with similar colors are connected with stronger co-authorship links. Out of the 247 authors in the WoS data set, 12 met the thresholds with the minimum number of documents of an author set to two and the minimum number of citations of an author set to zero, whereas out of the 408 authors in the Scopus data set, 24 met the same thresholds. The results in both data sets show a rather dispersed picture of authors, which is an indication of the fact that not much cooperation exists between them.

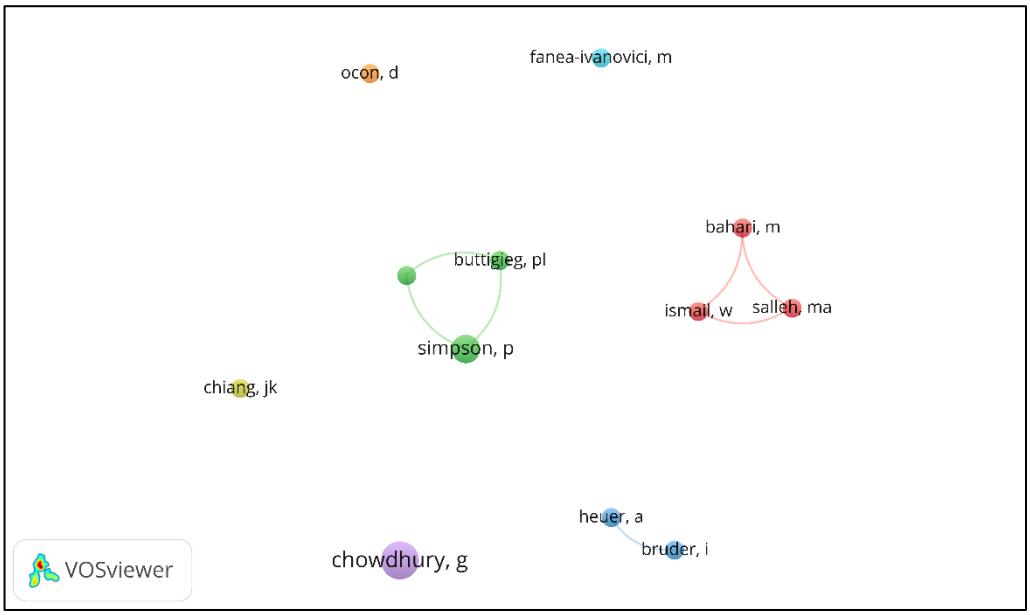

Figure 11: Network of co-authorship links in the WoS data set. Links represent co-authorship between authors or those that have co-authored with them.

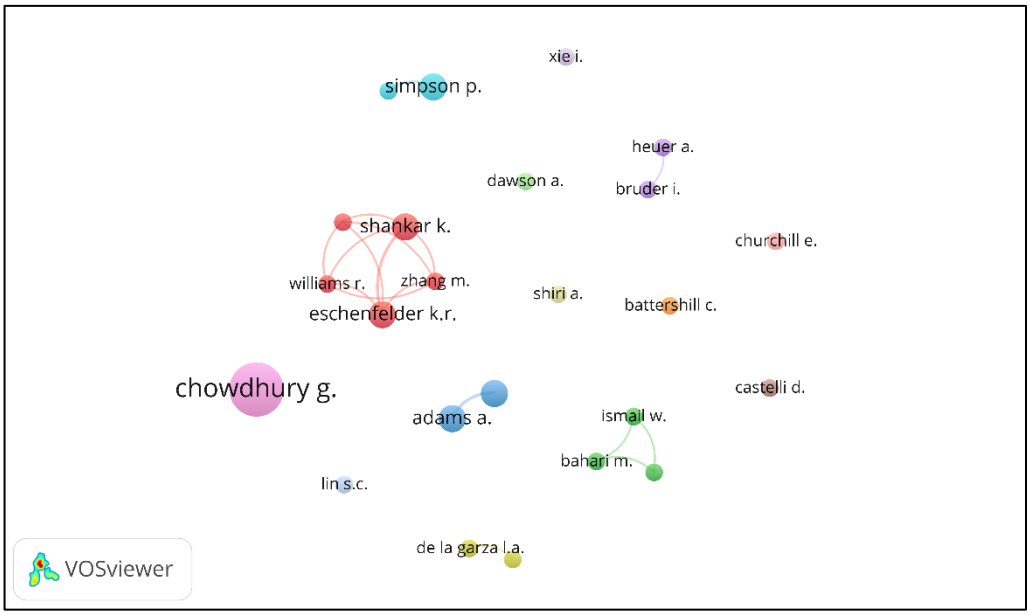

Figure 12: Network of co-authorship links in the Scopus data set. Links represent co-authorship between authors or those that have co-authored with them.

5.4.2.2 Collaboration between countries

Collaborative patterns were also analyzed at the level of country collaboration. Similarly to the co-authorship networks, here each node represents a country. Countries that are linked with a line cooperate with each other and the thickness of the line indicates the strength of collaboration between them. Figures 13 and 14 present country collaboration networks resulting from the WoS and Scopus data sets, respectively. Out of the 42 countries within the WoS data set, eight met the thresholds with the minimum number of documents of a country set to five and the minimum number of citations of a country set to zero. Similarly, out of the 50 countries within the Scopus data set, 10 met the same thresholds. The USA had the highest performance in

international collaborations with 24 collaborative documents in WoS and 69 in Scopus, and collaborations with the UK, Australia, Germany, and France. The Scopus data set also revealed additional collaborations between the USA and Spain, Canada, and China, which were not revealed by the WoS results.

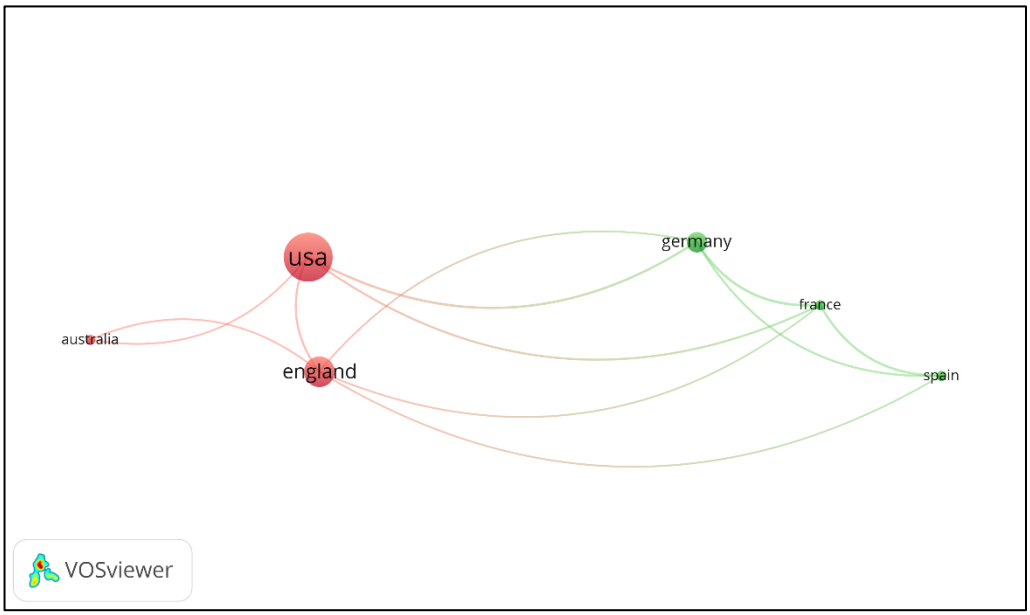

Figure 13: Country collaboration network in the WoS data set.

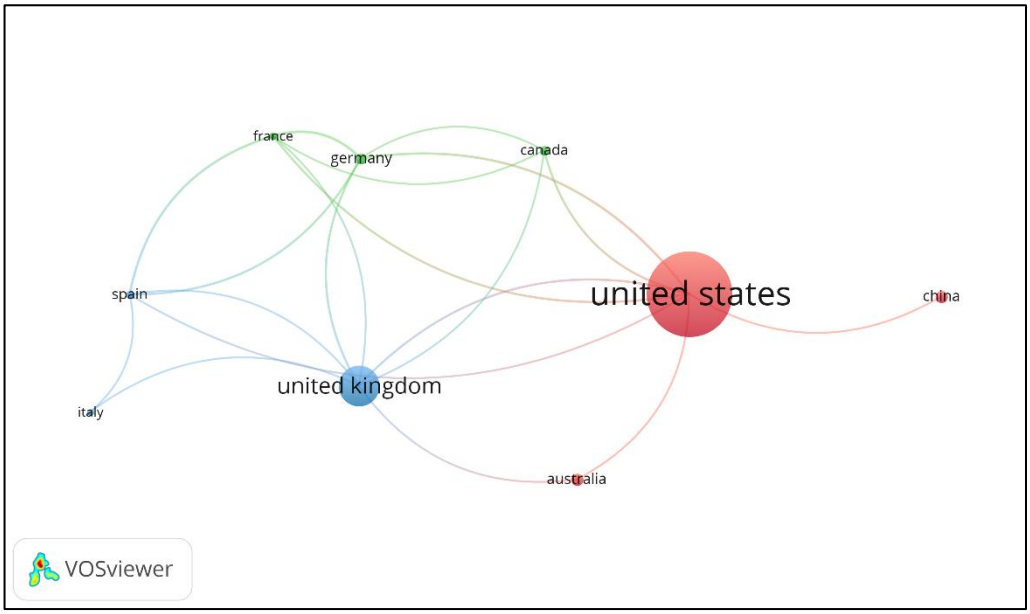

Figure 14: Country collaboration network in the Scopus data set.

5.4.3 Conceptual structure

Co-occurrence analysis was used in order to identify key research areas and emerging thematic trends in the field. More specifically, the techniques of co-word analysis and thematic mapping were used.

5.4.3.1 Co-word analysis

Running a co-word analysis allowed me to detect the most prominent subject areas within the examined data sets as well as the temporal evolution of these

subjects. Since author keywords function as a concise description of a piece of research, their analysis allows for the identification of trending topics in a field of study (Mishra et al., 2024; Zhang et al., 2016). Therefore, the analysis was done on the basis of author keywords.

Out of the 359 terms that were identified in the WoS data set, 14 met the threshold with the minimum number of occurrences of a keyword being four (not the default threshold). On the other hand, out of the 596 terms that were identified in the Scopus data set, 24 met the same threshold. The top four most frequently occurring author keywords (i.e., *sustainability*, *digital libraries/digital library*, *cultural heritage*, and *sustainable development*) were common in the two data sets. The complete results appear in Table 11.

Table 11

Most frequently occurring author keywords in the WoS and Scopus data sets

Author keywords	WoS occurrences	Scopus occurrences
sustainability	***24***	***42***
digital libraries/digital library	***22***	***44***
cultural heritage	***8***	***7***
sustainable development	***7***	***12***
institutional repositories	5	-
digital archives	4	9
digital repositories/digital repository	4	9
digital humanities	-	9
open access	4	8
academic libraries	4	7
collaboration	-	7
digital preservation	4	7
economic sustainability	4	7
metadata	-	6
digitization	4	5
libraries	4	5
archives	-	4
collections management	-	4
digital storage	-	4
environmental sustainability	-	4
higher education	-	4
public libraries	-	4
social sustainability	-	4

Note. The keywords presented in bold italics were common in both data sets. The terms that appeared in both plural and singular form have been combined.

Furthermore, the most frequently used author keywords were depicted in co-word analysis maps, one per data set. In a co-word analysis map each node represents a term and the size of the node grows proportionally to the number of publications that contain the term in the author keywords field. Also, terms that co-occur tend to cluster near each other. In the WoS data set author keywords were divided into three main clusters, each represented by a different

color (see Figure 15). The most prominent green cluster contained documents focusing on *sustainability* in relation to *institutional repositories* and *digital preservation* and *digitization projects*. The red cluster contained documents that dealt mostly with *economic sustainability* in relation to *digital* and *academic libraries* as well as to the *open access* movement. Lastly, the less prominent blue cluster included documents that dealt with the *sustainable development* of *cultural heritage* projects or institutions, and with *digital archives*.

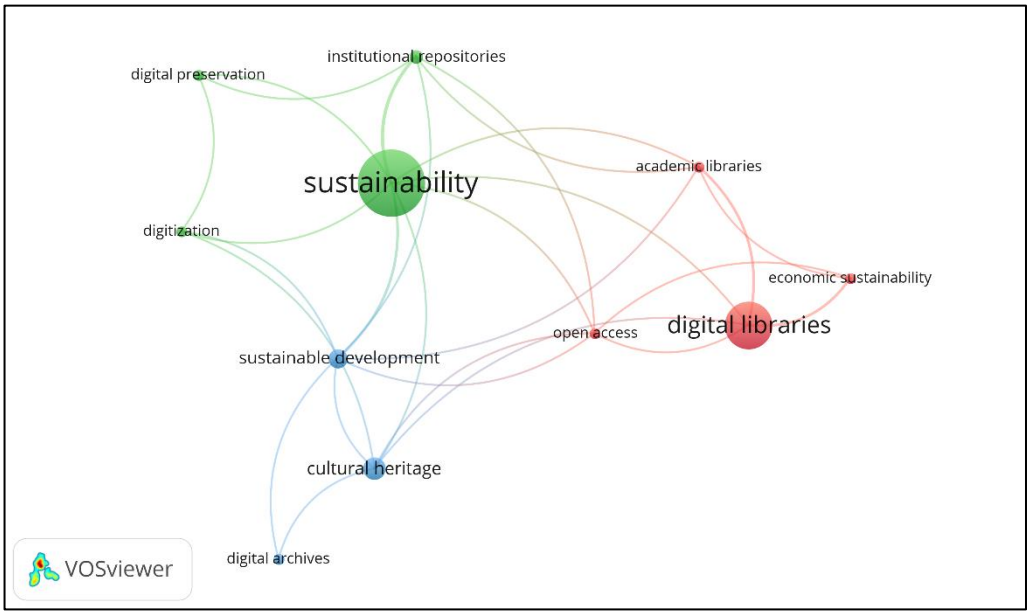

Figure 15: Network visualization of the most frequently used author keywords in the WoS data set.

In the Scopus data set, the most frequently used author keywords were divided into four main clusters, once again represented by different colors (see Figure 16). The most prominent red cluster contained documents focusing on *sustainability* in relation to *digital repositories* and *archives* as well as *digital preservation* and *digitization* projects. It also included documents dealing with the *digital humanities*, *cultural heritage*, and *metadata*. The green cluster contained documents that dealt mostly with *digital libraries*, *economic*, *environmental*, and *social sustainability*, as well as with the *open access* movement and *digital storage*. The blue cluster included documents that dealt with *academic libraries*, *higher education*, and *collaboration*, something that was not present in the WoS data set results. Finally, the least prominent yellow cluster represented documents that focused on *sustainable development* and *collection management*.

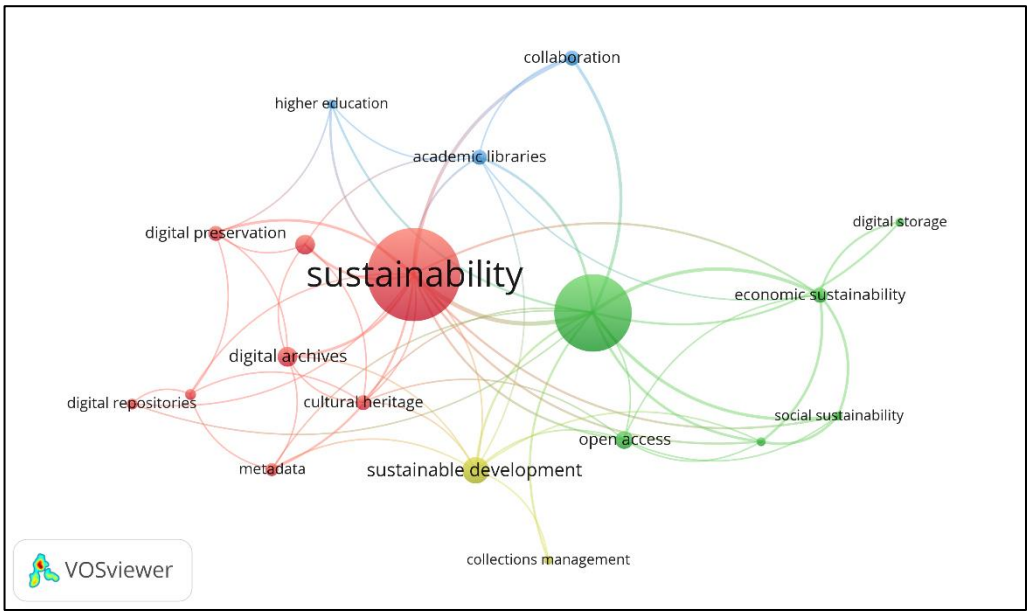

Figure 16: Network visualization of the most frequently used author keywords in the Scopus data set.

Figures 17 and 18 present a temporal view of what has been studied over the years based on the WoS and Scopus data sets, respectively. The lighter the color of the node is, the newer the theme it represents. Consequently, what we can discern is that the topic of *digital library economic sustainability* appeared in WoS as having been discussed in the literature for quite some time. In Scopus, topics such as *digital library economic* and *social sustainability*, *digital storage*, and *collaboration* were among the oldest that have preoccupied scientists. On the other hand, WoS showed that topics such as *digitization*, *institutional repositories*, and *cultural heritage*, represented by the yellowish nodes, have started being addressed more recently, whereas Scopus indicated apart from *cultural heritage*, and *digitization*, also *digital humanities* as some of the most recently researched topics.

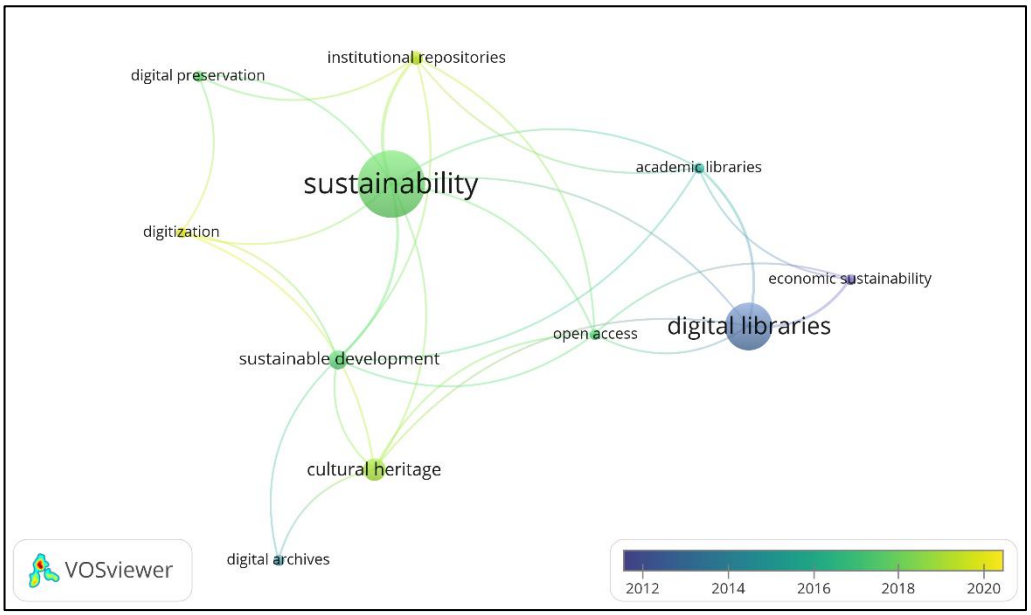

Figure 17: Overlay visualization of the scientific landscape in the WoS data set. The darker the color of the node, the older the topic it represents.

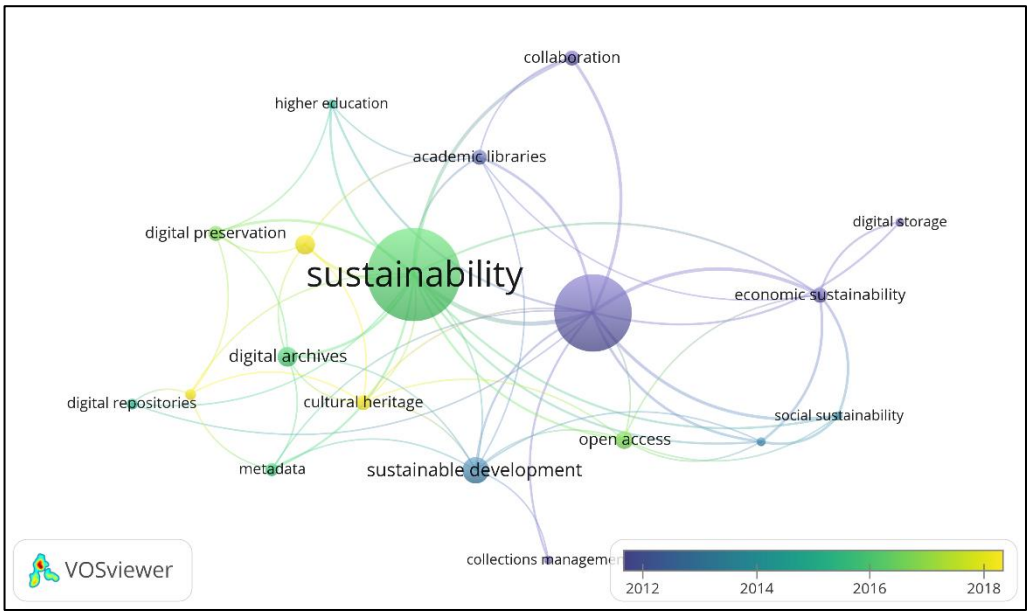

Figure 18: Overlay visualization of the scientific landscape in the Scopus data set. The darker the color of the node, the older the topic it represents.

5.4.3.2 Thematic map

In an attempt to dig deeper into the conceptual structure of the current literature, this was also depicted in thematic maps generated by Biblioshiny. A thematic map is the result of a clustering algorithm applied to the author keywords. According to Callon et al. (1991), the algorithm considers the density (i.e., the degree of the theme's development) and the centrality (i.e., the level of a theme's importance within the research field) of the themes. Even though thematic maps are usually hard to interpret, they may provide useful insights.

A thematic map allows for the visualization of four types of themes, namely the *motor themes*, the *basic themes*, the *emerging or declining themes*, and the *niche themes* (Callon et al., 1991; Cobo et al., 2011). The motor themes constitute the "driving force" themes and are depicted in the upper right quadrant of the thematic map. The basic themes constitute significant crosscutting, but general themes, and appear at the lower right quadrant. The emerging or declining themes constitute themes that are quite marginal and not so much discussed. They appear in the lower left quadrant. Lastly, the niche themes are also marginal basically because they are too specialized, and they appear at the upper left quadrant.

The thematic maps presented below were built based on author keywords, and generally, present both similarities and differences between the two data sets. The thematic map that resulted from WoS is presented in Figure 19. According to it the keywords *collaboration*, *economic models*, *funding* as well as *digital libraries*, *sustainability*, *information management*, and *digital repositories* appeared to be motor themes, whereas the terms *digital cultural heritage*, *digital archives*, *open access*, and *environmental sustainability* were basic themes that seem to be moving towards becoming motor ones. Lastly, *cultural sustainability* constituted a niche theme, while *digitization* and *inclusion* were

two emerging themes and *digital preservation* appeared to be a declining topic of discussion.

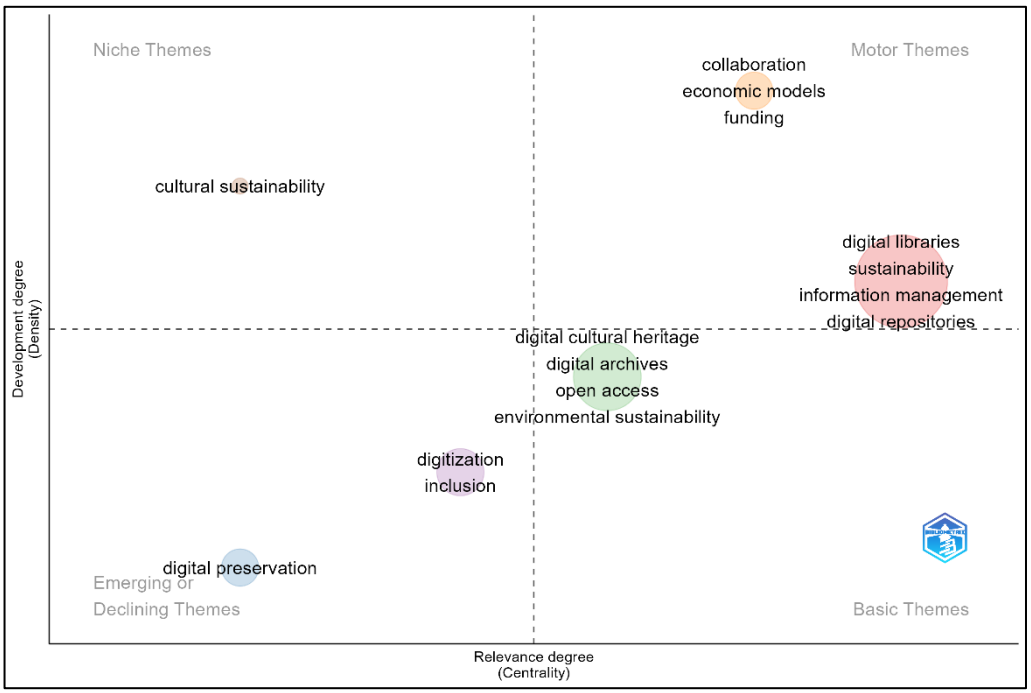

Figure 19: Thematic map of author keywords in the WoS data set. The volume of each node is proportional to the number of documents containing each keyword.

The thematic map that resulted from the Scopus data set is presented in Figure 20. According to it, just like in the WoS thematic map, the keyword *digital repositories* fell within the motor themes. Along with it were the keywords *open access* and *best practices*. The terms *digital libraries*, *sustainability*, and *sustainable development* along with *digital archives*, *digital cultural heritage*, and *digital preservation* were among the basic themes. *Costs* and *publishing*, two topics that did not come up in the WoS data set, were among the emerging/declining themes, while *collaboration* and *funding*, which appeared as motor themes in WoS fell within the niche themes category in Scopus along with *aggregation*, *library management*, *digital humanities*, and *archiving*.

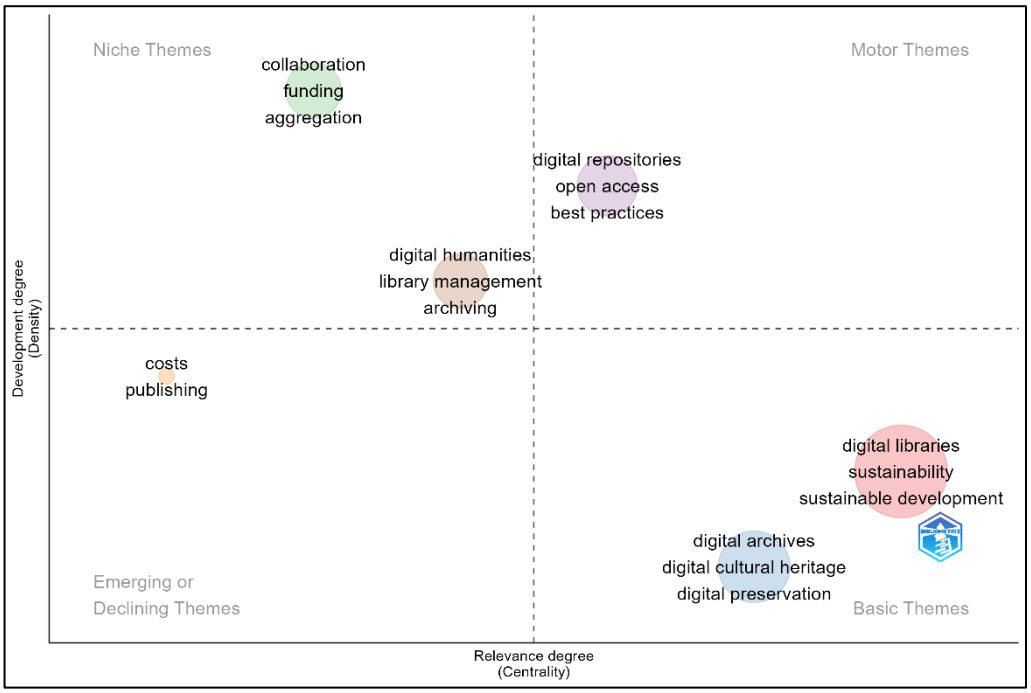

Figure 20: Thematic map of author keywords in the Scopus data set. The volume of each node is proportional to the number of documents containing each keyword.

5.5 Mapping axis results

With regard to the last methodological axis, which was used to determine whether UN SDGs constituted a topic of discussion in the examined data sets, the results showed that about half of the documents in each of the two data sets were mapped to at least one SDG. More specifically, the results showed that in the WoS data set 45,37% of publications were mapped to an SDG. The respective percentage for the Scopus data set was 51.81% including SDG17, and 45,60% excluding it (see Table 12).

Table 12

Percentage of items mapped to SDGs per data set

	WoS	**Scopus including SDG17**	**Scopus excluding SDG17**
Total no. of items	108	193	193
Items mapped to SDGs	49	100	88
Percentage	45,37%	51,81%	45,60%

Note. The table presents the total number of items mapped to one or more SDGs in each one of the two data sets. WoS does not map items to SDG17, whereas Scopus does. For this reason, in the Scopus data set the relevant percentage has been calculated both including and excluding items mapped to SDG17. This way, the results from the two data sets may be compared on the same basis.

Out of the items that were mapped to an SDG in the WoS data set, SDG4 “Quality education” received the highest percentage, followed by SDG3 “Good health and well-being”, and SDG9 “Industry, innovation, and infrastructure.”

In the Scopus data set the results were a little different with SDG17 "Partnership for the goals" receiving the highest percentage, followed by SDG4 "Quality education," SDG9 "Industry, innovation, and infrastructure," and SDG11 "Sustainable cities and communities." The complete and detailed distribution of items per SDG in each data set appears in Table 13. Moreover, detailed visualizations of the most frequently mapped SDGs in WoS and Scopus appear in the treemaps presented in Figures 21 and 22, respectively.

Table 13

Distribution of items per SDG and respective percentages in the WoS and Scopus data sets

SDG	Mappings in WoS	%	Mappings in Scopus	%
1 No poverty	-		1	0,55%
2 Zero hunger	-		1	0,55%
3 Good health and well-being	7	14,29%	1	0,55%
4 Quality education	30	61,22%	38	20,77%
5 Gender equality	1	2,04%	1	0,55%
6 Clean water and sanitation	-		-	
7 Affordable and clean energy	-		7	3,83%
8 Decent work and economic growth	-		10	5,46%
9 Industry, innovation, and infrastructure	4	8,16%	23	12,57%
10 Reduced inequalities	-		6	3,28%
11 Sustainable cities and communities	2	4,08%	20	10,93%
12 Responsible consumption and production	2	4,08%	6	3,28%
13 Climate action	-		7	3,83%
14 Life below water	-		1	0,55%
15 Life on land	2	4,08%	2	1,09%
16 Peace, justice, and strong institutions	1	2,04%	10	5,46%
17 Partnership for the goals	-		49	26,78%
Sum	49		183	

Note. In the case of WoS, each item has been mapped to one SDG, whereas in the case of Scopus, each item has been mapped to more than one SDG.

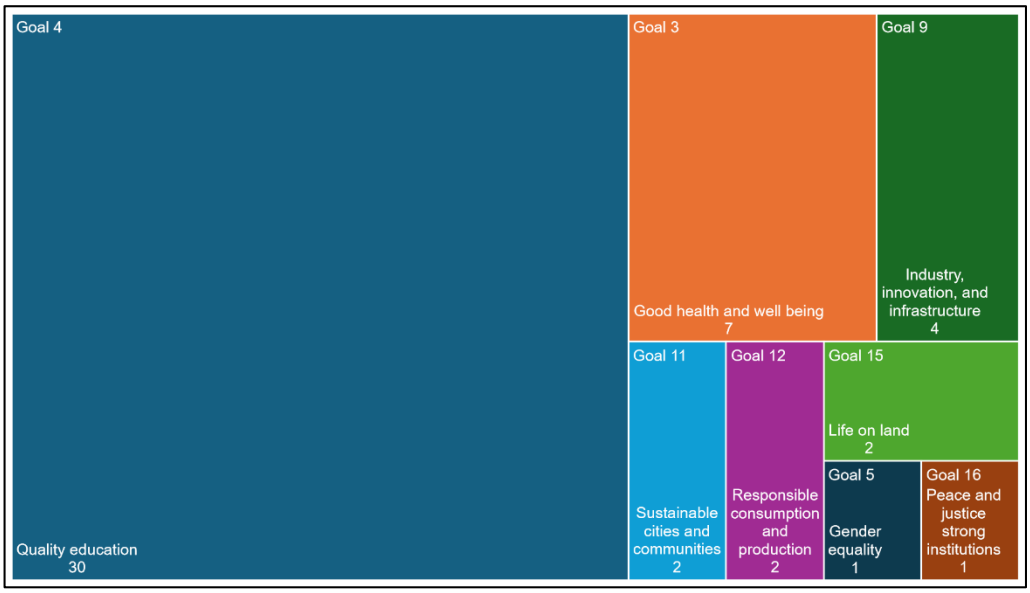

Figure 21: Treemap representing the top SDGs mapped to WoS items. The number below the title of each SDG shows the exact number of items mapped to that SDG.

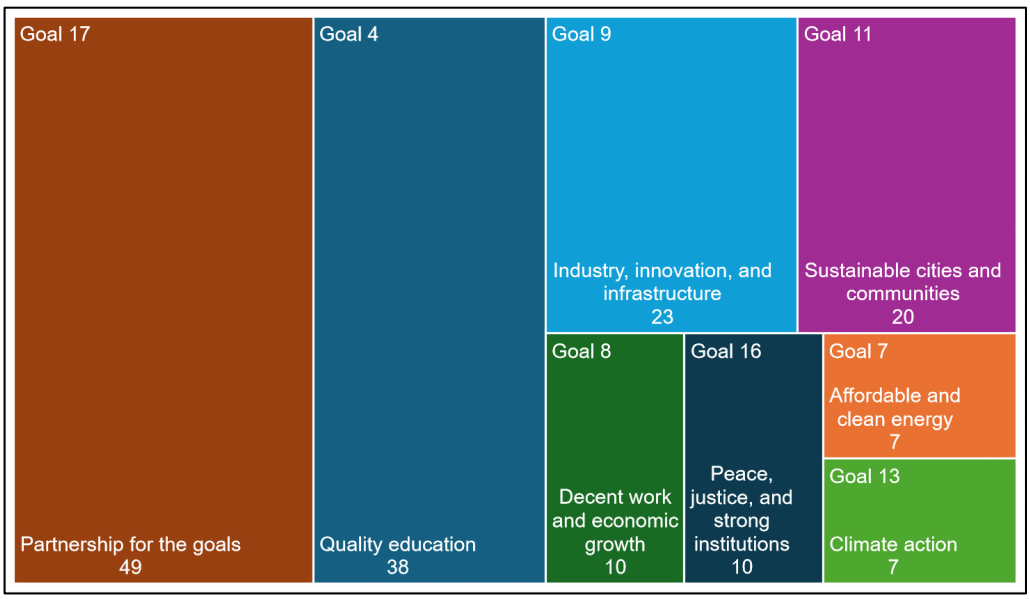

Figure 22: Treemap representing the top SDGs mapped to Scopus items. The number below the title of each SDG shows the exact number of items mapped to that SDG.

As mentioned in Section 5.2, among the documents included in the two data sets, 77 were common in both. A simple comparative analysis of the mappings within the documents shared by both data sets was conducted to determine more specific similarities and differences between the two data sets. Apparently, out of the 77 common documents, WoS had mapped 49% and Scopus 42% to an SDG (see Figure 23). Moreover, 21 documents were not mapped by any of the two databases and only four, a percentage of 5%, were mapped to the same SDG by both databases. It should be noted that even in this case, the mappings were not completely identical as Scopus mapped the four documents with more than just the common SDG. The four documents along with the relevant SDG mappings appear in Table 14.

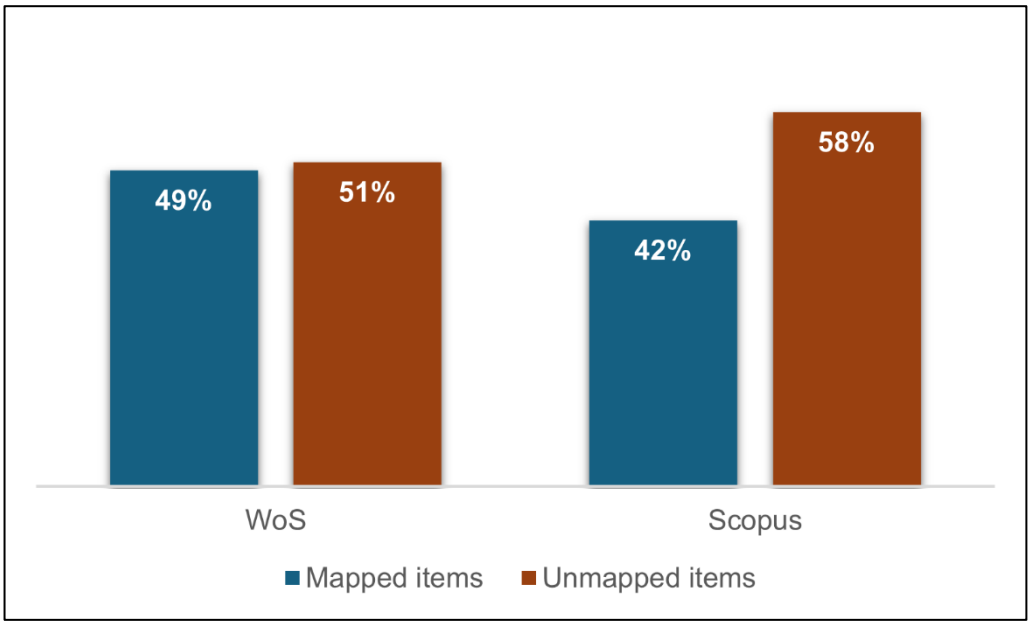

Figure 23: Percentage of SDG mapped and unmapped items in each data set. The percentages have been calculated over the total number of WoS and Scopus data set common documents.

Table 14

Items that received the same SDG mapping in both WoS and Scopus

Item title & DOI	Mapped SDG (WoS & Scopus)	Additional mapped SDGs (Scopus)
A proposed sustainable and digital collection and classification center model to manage e-waste in emerging economies (10.1108/JEIM-02-2020-0043)	SDG12	SDG11
Collaborative working for large digitisation projects (10.1108/00330330610669262)	SDG4	SDG11, SDG17
Digital librarianship practice and open access technology use for sustainable development in Nigeria (10.1108/DLP-01-2021-0007)	SDG4	SDG9, SDG17
Three-dimensional extension of a digital library service system (10.1108/00330331011083202)	SDG4	SDG9

6 Discussion

An analysis of the results presented in Chapter 5 shows that, generally, similar conclusions may be reached from both the WoS and Scopus databases in terms of the descriptive and analytical axes, with just a few noticed differences. Nevertheless, more significant differences may be noticed with regard to the SDG mapping axis. The various conclusions reached in relation to each one of the three methodological axes of analysis are further discussed in the sections that follow.

6.1 Analysis of current scientific production

The first point that can be made is that the topic of digital library sustainability is a rather new topic within the scientific community as proven by the document average age, which ranges from 8,33 years in WoS to 9,66 years in Scopus. The results from both databases show that there seems to be an upward trend in scientific productivity, which has passed from three noticed phases: an initial slow phase, a steady rise phase, and a high growth productivity phase. Even though this is noticed in both databases, the Scopus results show that the high growth phase started around 2010, whereas the WoS results present it to have started around 2013, an admittedly insignificant difference.

Despite this ascending trend, though, scientific production is still at the initial stages as proven by the small number of documents in each one of the two data sets. This confirms my initial suspicion that the topic of digital library sustainability has not yet attracted much interest on the part of the scientific community. The low scientific interest is also proven by the fact that the vast majority of authors (95,7% in WoS and 94,6% in Scopus) have written only one paper on the topic, thing that indicates that there is only occasional and not continuous and intensive involvement with it.

The majority of published documents are journal articles and conference proceedings written in English. Both data sets indicate G. Chowdhury as the author who has contributed the most in the field, whereas in terms of disciplinary interest, it is noticed that the majority of sources dealing with the topic belong to the discipline of LIS.

When it comes to geographic production, the results from both data sets show a similar number of countries dealing with the topic. The dominant country among them is the USA, while the respective dominant continent is Europe. This is an indication that European research is dispersed and divided among many different countries, whereas in America scientific production is more centralized.

At the institutional level, both data sets reveal that the institutions that have mostly dealt with the topic are educational ones. The fact that the top five institutions that have produced research on the topic have published only 13-15% of the total scientific production is also indicative of the fact that digital library sustainability research is at preliminary stages. Moreover, it should be noted that in the case of institutional production there is a noticed difference between the results of the two data sets as only one common institution (i.e.,

University College London) appears among the five institutions with the highest production. This constitutes a clear argument in favor of the methodological proposition that multiple databases should be used for triangulation purposes in bibliometric studies.

6.2 Analysis of intellectual, social, and conceptual trends

Findings related to the knowledge structure of the topic under discussion also indicate that digital library sustainability is still at an initial stage of investigation. One of the things that prove this is the fact that the most cited articles within the examined data sets have received quite a low number of citations (i.e., from 17 to 47) with the exception of one article by Smith et al. (2003), which received a higher number of 181 citations in Scopus. The low citation rates are indicative of the fact that researchers are not in need of consulting and citing such papers, which in turn suggests novelty of the topic.

The author that appears to be the most highly cited with 89 WoS citations and 98 Scopus citations is G. Chowdhury, which confirms the conclusion reached in Section 6.1 that he has contributed the most to the advancement of knowledge on digital library sustainability. Nevertheless, the low number of received citations by other authors, which ranges from 1 to 47, in both databases indicates that researchers have not yet dealt much with the topic.

In terms of author co-citation, Chowdhury is also found to be the most influential co-cited author in both data sets, whereas in terms of article co-citation, two of his articles along with one by Hamilton are in the top five most highly cited papers in both data sets, which is an indication of their influence. Nevertheless, the extremely low number of citations (i.e., three to seven) does not allow for safe conclusions and the fact that there are also very different documents among the most highly cited ones, as these resulted from each one of the two databases, enhances this.

In terms of interrelations between authors, the WoS results show less interaction between authors, whereas the Scopus data set shows slightly more. This may have been influenced by the fact that the two data sets differed in terms of size. As far as collaborations are concerned, co-authorship links indicate limited collaboration among authors, while in terms of country collaboration, both data set results show that there seems to exist some. In both data sets, the USA and the UK are the two countries that present the highest number of collaborations with other countries.

Co-word analysis in both data sets shows that the most prominent topics discussed in the literature are sustainability in relation to institutional repositories and digitization projects, economic sustainability, and sustainable development of cultural heritage institutions and archives. Apart from the above, some additional topics emerge from the Scopus data set. These are academic libraries, education, and collaboration. Among the topics that emerge, themes such as cultural heritage, digital humanities, digitization, and institutional repositories appear to be quite recent. On the other hand, the topics

of economic and social sustainability, digital storage, and collaboration have been addressed for quite some time.

Lastly, the thematic maps present dissimilarities between the two data sets, which confirms the fact that the technique of thematic mapping remains a difficult one to interpret and the results stemming from it are difficult to draw conclusions from. Even though common themes have emerged from the WoS and Scopus thematic maps (Figures 19 and 20), in most cases, these are classified as different types of emerging themes. The only themes that are identified as belonging to the same thematic type in both data sets are "digital repositories," which has been identified as a motor theme; and "digital cultural heritage" and "digital archives," which have been identified as basic themes.

6.3 Analysis of SDG mapping

The analysis of the WoS SDG mappings designates SDG4, SDG3, and SDG9 as the most frequently mapped ones to the articles that were common in the two data sets. In Scopus, the most frequently mapped are SDG17, SDG4, and SDG9. The fact that SDG4 and SDG9 are among the top three most highly mapped SDGs in both data sets is the only noticed similarity between them in terms of SDG mappings.

Excluding the above similarity between databases, the results emerging from the SDG mapping axis present significant differences. Primarily, only eight out of the 17 SDGs are mapped to documents within the WoS data set as opposed to 16 in the Scopus data set. Secondly, it is noticed that WoS does not map SDG17 at all, as opposed to Scopus, which has recently started mapping it. Furthermore, out of the 77 documents that were common in the two data sets only four were found to have been mapped to the same SDG by both databases. Each one of these four documents is also mapped to additional SDGs in Scopus. One of the differences that emerged is that WoS generally maps slightly more (49%) documents than Scopus (42%), but assigns only one SDG per document, whereas the opposite happens with Scopus.

Consequently, it is assumed that the quite different algorithms used by the two databases result in significant differences in terms of the final mapping outcome. The noticed differences in how the two databases map SDGs highlight the difficulty hidden in this procedure. The fact that different databases classify articles differently shows that SDG classifications by commercial services may be unreliable and creates reasons for wondering whether relying on such tools for the evaluation of scientific productivity on SDGs is indeed valid and trustworthy.

These results coincide with the conclusions reached by Armitage et al. (2020), who questioned the validity of commercial database SDG mappings after comparing the Bergen and Elsevier mapping approaches and finding limited similarities in their mappings. According to Armitage et al., interpreting the SDGs, establishing their relevance to documents, and building effective mapping queries constitute big challenges that need to be overcome before database SDG mapping strategies may be relied on.

6.4 Limitations of the study

No matter how well thought-out a research study may be, it is always accompanied by numerous limitations. The present study does not constitute an exception.

One of the limitations that accompanies all bibliometric studies is bibliometric database barriers. Even though bibliometric databases attempt to include the most significant research that is being produced, there are obvious restrictions such as license-related, geographic, and financial ones, which prevent them from including the full range of published literature. A good example that illustrates this is the case of G. Chowdhury, who, as previously discussed, is one of the researchers who have been a lot preoccupied with the topic of digital library sustainability. Chowdhury has produced at least 10 publications on the topic, but not all of them are included in the two databases that were used in this study. Specifically, he is represented by four of his papers in WoS and by six papers in Scopus. Taking this into consideration, one may conclude that even though the data sets extracted from such databases may give us an idea of the research environment on a topic, it is impossible for them to give the complete picture. This should be taken even more seriously into account when examining topics which are fairly new and not much research exists on them, as in the case of the present study.

Another limitation that should be considered is the issue of item metadata. WoS and Scopus, like any database, often contain mistakes and discrepancies in their metadata. This is something researchers conducting bibliometric studies should always consider. The reason is that bibliometric analyses are made based on the extracted metadata and any mistakes and inconsistencies in them may affect a study's final results. In the present study, detailed data cleaning was conducted prior to importing data in the bibliometric analysis tools. Even though the procedures followed were thorough and meticulous, the fact that they were put into practice by one person means that there is always the risk of the researcher having missed something or having neglected to take something else into account. Consequently, the researcher's personal bias should also be listed among the study's limitations both with regard to data processing and cleaning, but also with regard to data collection and analysis.

Lastly, the small number of articles included in the data sets, thus the small sample size, constitutes an additional limitation of the study and its implications should be addressed. Even though it allows one to draw indicative conclusions regarding the topic under discussion, it does not allow for safe conclusions such as the ones that may be reached when bigger data sets with more documents are examined.

7 Conclusion

The present study aims at contributing to the advancement of the field of LIS, academia, and society in general by investigating the scientific production on the topic of digital library sustainability. This is done by recording the current state of relevant academic literature, identifying research trends, and understanding how the conversation around digital library sustainability has evolved, and how it has been translated into scientific production on the topic. On top of this, the study aims at investigating if the UN SDGs have been considered and studied within the existing relevant scientific research and at making a methodological point about the use of bibliometrics as a method for studying a topic.

To achieve the goals of the study two well-respected bibliometrics databases were used (i.e., WoS and Scopus) and numerous bibliometric and other techniques were put into practice. Methodologically, a comparison of the results received from each data set was attempted, something which is proposed as a triangulation method to increase validity of the results especially in cases of small data sets like the ones on which this study was based. In general, the two data sets produced similar results, which confirmed my conclusions. Nevertheless, there were cases when the results from each data set did not coincide, thing that enhances the present study's argument in favor of using multiple sources when conducting bibliometric studies.

Indisputably, the discussion around digital library sustainability is a huge one. Even though it seems to have preoccupied some researchers for quite some time, the low levels of scientific production relating to it prove that there is room for more research and collaborations on the topic. Current trends show that production on the topic presents a slow but steady increase. Nevertheless, it is still at low levels with few authors having contributed to the expansion of knowledge on the topic, and limited inter-author and inter-country collaborations. One of the study's most significant contributions is related to the SDG mappings. Interestingly, their investigation shows big discrepancies between the two examined data sets. This raises questions regarding the level of trust researchers may show in relevant classification methods used by bibliometric databases.

One of the primary strengths of the present study is that it constitutes one of the few bibliometric studies on the topic of digital library sustainability. As a consequence, it may function as a steppingstone that will urge more researchers in LIS and other relevant fields to conduct research on the topic. Primarily, some more qualitative research, such as an extensive literature review on the topic is proposed especially when it comes to the thematic trends appearing in current literature. Another possible future area of investigation by researchers could be exploring the reasons behind the limited scientific production on digital library sustainability, as well as finding ways to overcome possible barriers. Moreover, delving into the reasons behind the limited collaboration among authors and among countries and proposing strategies to foster such collaborations are also proposed as directions for future research. Lastly, the significant finding regarding how databases map SDGs to documents opens the way for further research; research that would explore and determine how SDG

mapping is done and whether it would be worth finding ways to standardize the procedures followed by different databases.

Reference list

Ahmad, K., Jian Ming, Z., & Rafi, M. (2018). Assessing the digital library research output: Bibliometric analysis from 2002 to 2016. *The Electronic Library*, *36*(4), 696–704. https://doi.org/10.1108/el-02-2017-0036

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open*, *9*(1). https://doi.org/10.1177/2158244019829575

Andrés, A. (2009). *Measuring academic research: How to undertake a bibliometric study*. Chandos Publishing.

Aria, M. (2021, March 18). *Bibliometrix: Data importing and converting*. Bibliometrix. https://www.bibliometrix.org/vignettes/Data-Importing-and-Converting.html

Aria, M., & Cuccurullo, C. (2017). *Bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Armitage, C. S., Lorenz, M., & Mikki, S. (2020). Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results? *Quantitative Science Studies*, *1*(3), 1092–1108. https://doi.org/10.1162/qss_a_00071

Arushanyan, Y., Ekener, E., & Moberg, Å. (2017). Sustainability assessment framework for scenarios – SAFS. *Environmental Impact Assessment Review*, *63*, 23–34. https://doi.org/10.1016/j.eiar.2016.11.001

Ball, R. (2018). *An introduction to bibliometrics: New developments and trends*. Chandos Publishing. https://doi.org/10.1016/C2016-0-03695-1

Bibliometrix. (n.d.). *Frequently asked questions*. https://www.bibliometrix.org/home/index.php/faq

Borgman, C. (1993). NEL workshop report. In E. A. Fox (Ed.), *Source book on digital libraries: Prepared for and sponsored by the National Science Foundation* (Version 1.0, pp. 121–142). Virginia Polytechnic Institute and State University. https://fox.cs.vt.edu/DigitalLibrary/DLSB.pdf

Borgman, C. L. (1999). What are digital libraries? Competing visions. *Information Processing and Management*, *35*(3), 227–243. https://doi.org/10.1016/s0306-4573(98)00059-4

Borgman, C. L. (2000). *From Gutenberg to the global information infrastructure: Access to information in the networked world*. The MIT Press. https://doi.org/10.7551/mitpress/3131.001.0001

Borgohain, D. J., Zakaria, S., & Verma, M. K. (2022). Cluster analysis and network visualization of global research on digital libraries during

2016–2020: A bibliometric mapping. *Science & Technology Libraries*, *41*(3), 266–287. https://doi.org/10.1080/0194262x.2021.1993422

Boyer, R. H. W., Peterson, N. D., Arora, P., & Caldwell, K. (2016). Five approaches to social sustainability and an integrated way forward. *Sustainability*, *8*(9), Article 878. https://doi.org/10.3390/su8090878

Brown, B. J., Hanson, M. E., Liverman, D. M., & Merideth, R. W., Jr. (1987). Global sustainability: Toward definition. *Environmental Management*, *11*(6), 713–719. https://doi.org/10.1007/bf01867238

Brundtland, G. H. (1987). *Report of the World Commission on Environment and Development: Our common future*. United Nations. https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Calhoun, K. (2014). *Exploring digital libraries: Foundations, practice, prospects*. Facet Publishing.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, *22*(1), 155–205. https://doi.org/10.1007/bf02019280

Chowdhury, G. (2010). Carbon footprint of the knowledge sector: What's the future? *Journal of Documentation*, *66*(6), 934–946. https://doi.org/10.1108/00220411011087878

Chowdhury, G. (2012). Building environmentally sustainable information services: A green IS research agenda. *Journal of the American Society for Information Science and Technology*, *63*(4), 633–647. https://doi.org/10.1002/asi.21703

Chowdhury, G. G. (2013a). Social sustainability of digital libraries: A research framework. In S. R. Urs, J. C. Na, & G. Buchanan (Eds.), *Digital libraries: Social media and community networks* (pp. 25–34). Springer. https://doi.org/10.1007/978-3-319-03599-4_4

Chowdhury, G. (2013b). Sustainability of digital information services. *Journal of Documentation*, *69*(5), 602–622. https://doi.org/10.1108/jd-08-2012-0104

Chowdhury, G. G. (2013c). Sustainability of digital libraries: A conceptual model. In T. Aalberg, C. Papatheodorou, M. Dobreva, G. Tsakonas, & C. J. Farrugia (Eds.), *Research and advanced technology for digital libraries* (pp. 1–12). Springer. https://doi.org/10.1007/978-3-642-40501-3_1

Chowdhury, G. G. (2014a). OA policies and the sustainability of digital libraries of scholarly information. In K. Tuamsuk, A. Jatowt, & E.

Rasmussen (Eds.), *The emergence of digital libraries: Research and practices* (pp. 65–75). Springer. https://doi.org/10.1007/978-3-319-12823-8_8

Chowdhury, G. (2014b). Sustainability of digital libraries: A conceptual model and a research framework. *International Journal on Digital Libraries*, *14*(3-4), 181–195. https://doi.org/10.1007/s00799-014-0116-0

Chowdhury, G. G. (2016). How to improve the sustainability of digital libraries and information services? *Journal of the Association for Information Science and Technology*, *67*(10), 2379–2391. https://doi.org/10.1002/asi.23599

Chowdhury, G., & Koya, K. (2017). Information practices for sustainability: Role of iSchools in achieving the UN sustainable development goals (SDGs). *Journal of the Association for Information Science and Technology*, *68*(9), 2128–2138. https://doi.org/10.1002/asi.23825

Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, *5*(4), 423–441. https://doi.org/10.1177/030631277500500403

Citation, n., sense 2.a. (n.d.). In M. Proffitt (Ed.), *Oxford English Dictionary*. Oxford University Press. https://doi.org/10.1093/OED/4182619763

Clarivate. (n.d.-a). *Core collection full record details*. Clarivate Web of Science Help. https://webofscience.help.clarivate.com/en-us/Content/wos-core-collection/wos-full-record.htm

Clarivate. (n.d.-b). *Search operators*. Clarivate Web of Science Help. https://webofscience.help.clarivate.com/en-us/Content/search-operators.html?Highlight=proximity%20operator

Clarivate. (n.d.-c). *Web of Science Core Collection search fields*. Clarivate Web of Science Help. https://webofscience.help.clarivate.com/en-us/Content/wos-core-collection/woscc-search-fields.htm?Highlight=fields

Clarivate. (2024, February 5). *Resources for librarians: Web of Science coverage details*. Clarivate Libguides. https://clarivate.libguides.com/librarianresources/coverage

Cleveland, G. (1998). *Digital Libraries: Definitions, issues and challenges [Occasional paper #8]*. International Federation of Library Associations and Institutions.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, *5*(1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002

Council of the American Library Association. (2015, June 28). *Resolution on*

the importance of sustainable libraries. American Library Association. https://www.ala.org/sites/default/files/aboutala/content/governance/council/council_documents/2015_annual_council_documents/cd_36_substainable_libraries_resol_final.pdf

Council of the American Library Association. (2017, June 25). *An American Library Association statement on global climate change and a call for support for libraries and librarians*. American Library Association. https://www.ala.org/sites/default/files/aboutala/content/governance/council/council_documents/2017_annual/cd_41_global_climate%20change%20statement_62517_FINAL.pdf

Council of the American Library Association. (2019, January 27). *Resolution for the adoption of sustainability as a core value of librarianship*. American Library Association. https://www.ala.org/sites/default/files/aboutala/content/Resolution%20for%20the%20Adoption%20of%20Sustainability%20as%20a%20Core%20Value%20of%20Librarianship-FINAL.pdf

Council of the American Library Association. (2021, June 10). *Resolution to achieve carbon neutrality for ALA conferences*. American Library Association. https://www.ala.org/sites/default/files/aboutala/content/ALA%20CD%2053%20RESOLUTION%20to%20Achieve%20Carbon%20Neutrality%20for%20ALA%20Conferences%20Final%20Revised.pdf

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities.* The University of Chicago Press. https://archive.org/details/invisiblecollege0000dian/mode/2up

Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation, 37*(1), 16–24. https://doi.org/10.1108/eb026703

Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. Taylor Graham.

Cronin, B. (1998). Metatheorizing citation. *Scientometrics*, *43*(1), 45–55. https://doi.org/10.1007/bf02458393

Cronin, B. (2000). Semiotics and evaluative bibliometrics. *Journal of Documentation*, *56*(4), 440–453. https://doi.org/10.1108/eum0000000007123

Dawe, N. K., & Ryan, K. L. (2003). The faulty three-legged-stool model of sustainable development. *Conservation Biology*, *17*(5), 1458–1460. https://doi.org/10.1046/j.1523-1739.2003.02471.x

de Oliveira, O. J., da Silva, F. F., Juliani, F., Barbosa, L. C. F. M., & Nunhes, T. V. (2019). Bibliometric method for mapping the state-of-the-art and identifying research gaps and trends in literature: An essential instrument to support the development of scientific projects. In S. Kunosic & E. Zerem (Eds.), *Scientometrics recent advances*. IntechOpen. https://doi.org/10.5772/intechopen.85856

doi Foundation. (n.d.). *The identifier: What is a DOI?* https://www.doi.org/the-identifier/what-is-a-doi/

Du Pisani, J. A. (2006). Sustainable development: Historical roots of the concept. *Environmental Sciences*, *3*(2), 83–96. https://doi.org/10.1080/15693430600688831

Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, *45*(3), 171–212. https://doi.org/10.1108/eb026843

Elsevier. (2023, August 21). *What are sustainable development goals (SDGs)?* Scopus: Access and Use Support Center. https://service.elsevier.com/app/answers/detail/a_id/31662/supporthub/scopus

Elsevier. (2024, February 16). *How can I best use the Advanced search?* Scopus: Access and Use Support Center. https://service.elsevier.com/app/answers/detail/a_id/11365/c/10545/supporthub/scopus/

Eschenfelder, K. R., Shankar, K., Williams, R., Lanham, A., Salo, D., & Zhang, M. (2016). What are we talking about when we talk about sustainability of digital archives, repositories and libraries? *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–6. https://doi.org/10.1002/pra2.2016.14505301148

Fox, E. A., Akscyn, R. M., Furuta, R. K., & Leggett, J. J. (1995). Digital libraries. *Communications of the ACM*, *38*(4), 22–28. https://doi.org/10.1145/205323.205325

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). Empirical analysis and classification of database errors in Scopus and Web of Science. *Journal of Informetrics*, *10*(4), 933–953. https://doi.org/10.1016/j.joi.2016.07.003

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, *122*(3159), 108–111. https://doi.org/10.1126/science.122.3159.108

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359–375. https://doi.org/10.1007/bf02019306

Garfield, E. (1990). KeyWords Plus: ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current Contents on Diskette. *Essays of an Information Scientist*, *13*(32), 295–299. https://garfield.library.upenn.edu/essays/v13p295y1990.pdf

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, *7*(1), 113–122. https://doi.org/10.1177/030631277700700112

Grober, U. (2012). *Sustainability: A cultural history* (R. Cunningham, Trans.). Green Books.

Grosseck, G., Țîru, L. G., & Bran, R. A. (2019). Education for sustainable development: Evolution and perspectives: A bibliometric review of research, 1992–2018. *Sustainability, 11*(21), Article 6136. https://doi.org/10.3390/su11216136

Gumpenberger, C., Wieland, M., & Gorraiz, J. (2012). Bibliometric practices and activities at the University of Vienna. *Library Management, 33*(3), 174–183. https://doi.org/10.1108/01435121211217199

Hamilton, V. (2004). Sustainability for digital libraries. *Library Review, 53*(8), 392–395. https://doi.org/10.1108/00242530410556210

Hawkins, D. T., & Wagers, R. (1982). Online bibliographic search strategy development. *Online, 6*(3), 12–19.

Hulme, E. W. (1923). *Statistical bibliography in relation to the growth of modern civilization: Two lectures delivered in the University of Cambridge in May, 1922*. Butler & Tanner. https://ia801308.us.archive.org/21/items/statisticalbibli00hulmuoft/statisticalbibli00hulmuoft.pdf

International Federation of Library Associations and Institutions. (2017, August). *Toolkit: Libraries, development and the United Nations 2030 agenda*. https://www.ifla.org/wp-content/uploads/files/assets/hq/topics/libraries-development/documents/libraries-un-2030-agenda-toolkit-2017.pdf

iSchools. (n.d.). *About us*. https://www.ischools.org/about

Jeronen, E. (2013). Sustainability and sustainable development. In S. O. Idowu, N. Capaldi, L. Zu, & A. D. Gupta (Eds.), *Encyclopedia of corporate social responsibility* (pp. 2370–2378). Springer. https://doi.org/10.1007/978-3-642-28036-8_662

Khin Khin Oo, N. C., & Rakthin, S. (2022). Integrative review of absorptive capacity's role in fostering organizational resilience and research agenda. *Sustainability, 14*(19), Article 12570. https://doi.org/10.3390/su141912570

Koya, K., & Chowdhury, G. (2020). Cultural heritage information practices and iSchools education for achieving sustainable development. *Journal of the Association for Information Science and Technology, 71*(6), 696–710. https://doi.org/10.1002/asi.24283

Lesk, M. (1997). *Practical digital libraries: Books, bytes, and bucks*. Morgan Kaufmann Publishers. https://archive.org/details/practicaldigital00lesk

Lesk, M. (2005). *Understanding digital libraries* (2nd ed.). Morgan Kaufmann Publishers. https://doi.org/10.1016/b978-1-55860-924-2.x5000-2

Lesk, M., Fox, E., & McGill, M. (1993). A national electronic science, engineering, and technology library. In E. A. Fox (Ed.), *Source book on digital libraries: Prepared for and sponsored by the National Science*

Foundation (Version 1.0, pp. 4–24). Virginia Polytechnic Institute and State University. https://fox.cs.vt.edu/DigitalLibrary/DLSB.pdf

Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, *43*(1), 5–25. https://doi.org/10.1007/bf02458391

Li, S., & Yang, F. (2023). Green library research: A bibliometric analysis. *Public Library Quarterly*, *42*(4), 424–443. https://doi.org/10.1080/01616846.2022.2116886

Liew, C. L., & Chowdhury, G. (2016). Digital cultural heritage and social sustainability. *The Journal of Community Informatics*, *12*(3), 173–196. https://doi.org/10.15353/joci.v12i3.3284

Lindsey, D. (1978). *The scientific publication system in social science: A study of the operation of leading professional journals in psychology, sociology, and social work*. Jossey-Bass Publishers. https://archive.org/details/scientificpublic0000lind

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

Lozano, R. (2008). Envisioning sustainability three-dimensionally. *Journal of Cleaner Production*, *16*(17), 1838–1846. https://doi.org/10.1016/j.jclepro.2008.02.008

Lubchenco, J. (1998). Entering the century of the environment: A new social contract for science. *Science*, *279*(5350), 491–497. https://doi.org/10.1126/science.279.5350.491

Luukkonen, T. (1997). Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics*, *38*(1), 27–37. https://doi.org/10.1007/bf02461121

Lyman, P. (1996). What is a digital library? Technology, intellectual property, and the public interest. *Daedalus*, *125*(4), 1–33. https://www.jstor.org/stable/20027384

Macfarlane, B. (2017). Publication and performativity. In G. A. Postiglione & J. Jung (Eds.), *The changing academic profession in Hong Kong* (pp. 97–108). Springer. https://doi.org/10.1007/978-3-319-56791-4_5

Maier, D., Maier, A., Aşchilean, I., Anastasiu, L., & Gavriș, O. (2020). The relationship between innovation and sustainability: A bibliometric review of the literature. *Sustainability*, *12*(10), Article 4083. https://doi.org/10.3390/su12104083

McCullough, R. (2020, December 9). *Sustainable development goals (SDGs) on Scopus*. Elsevier Scopus Blog. https://blog.scopus.com/posts/sustainable-development-goals-sdgs-on-scopus

McCullough, R. (2023, August 22). *2023 Sustainable development goals (SDGs) now available on Scopus*. Elsevier Scopus Blog. https://blog.scopus.com/posts/2023-sustainable-development-goals-sdgs-now-available-on-scopus

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W., III. (1972). *The limits to growth: A report for the Club of Rome's project on the predicament of mankind*. Universe Books. https://archive.org/details/TheLimitsToGrowth

Meschede, C., & Henkel, M. (2019). Library and information science and sustainable development: a structured literature review. *Journal of Documentation, 75*(6), 1356–1369. https://doi.org/10.1108/jd-02-2019-0021

Mishra, M., Desul, S., Santos, C. A. G., Mishra, S. K., Kamal, A. H. M., Goswami, S., Kalumba, A. M., Biswal, R., da Silva, R. M., dos Santos, C. A. C., & Baral, K. (2024). A bibliometric analysis of sustainable development goals (SDGs): A review of progress, challenges, and opportunities. *Environment, Development and Sustainability*, *26*(5), 11101–11143. https://doi.org/10.1007/s10668-023-03225-w

Moir, S., & Carter, K. (2012). Diagrammatic representations of sustainability: A review and synthesis. In S. D. Smith (Ed.), *Proceedings 28th Annual ARCOM Conference, 3-5 September 2012, Edinburgh, UK* (pp. 1479–1489). Association of Researchers in Construction Management. http://www.arcom.ac.uk/-docs/proceedings/ar2012-1479-1489_Moir_Carter.pdf

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional De La Información*, *29*(1). https://doi.org/10.3145/epi.2020.ene.03

Moreno, E. A., Dutra, A., Junges, I., & Mussi, C. C. (2022). Addressing sustainability in the context of library management: International literature review. *RDBCI: Revista Digital de Biblioteconomia E Ciência Da Informação*, *20*, Article e022025. https://doi.org/10.20396/rdbci.v20i00.8670507

Mori, K., & Christodoulou, A. (2012). Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environmental Impact Assessment Review, 32*(1), 94–106. https://doi.org/10.1016/j.eiar.2011.06.001

Mulkay, M. J. (1974). Methodology in the sociology of science: Some reflections on the study of radio astronomy. *Social Science Information, 13*(2), 107–119. https://doi.org/10.1177/053901847401300206

Neuhaus, C., & Daniel, H.-D. (2008). Data sources for performing citation analysis: An overview. *Journal of Documentation*, *64*(2), 193–210. https://doi.org/10.1108/00220410810858010

Nolin, J. (2010). Sustainable information and information science. *Information Research*, *15*(2), Article 431. https://informationr.net/ir/15-2/paper431.html

Nürnberg, P. J., Furuta, R., Leggett, J. J., Marshall, C. C., & Shipman, F. M., III (1995). Digital libraries: Issues and architectures. In *Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries, Hypermedia Research Lab, Computer Science Department, Texas A&M University, Austin, Texas, USA, June 11-13, 1995*. https://dblp.org/rec/conf/dl/NurnbergFLMS95

Okubo, Y. (1997). *Bibliometric indicators and analysis of research systems* (OECD Science, Technology and Industry Working Papers, No. 1997/01). OECD Publishing. https://doi.org/10.1787/208277770603

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), Article 89. https://doi.org/10.1186/s13643-021-01626-4

Persson, O. (1994). The intellectual base and research fronts of JASIS 1986-1990. *Journal of the American Society for Information Science*, *45*(1), 31–38. https://doi.org/10.1002/(sici)1097-4571(199401)45:1%3C31::aid-asi4%3E3.0.co;2-g

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

Pritchard, A. (1969). Documentation notes: Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*(4), 348–349. https://doi.org/10.1108/eb026482

Purvis, B., Mao, Y., & Robinson, D. (2019). Three pillars of sustainability: In search of conceptual origins. *Sustainability Science*, *14*(3), 681–695. https://doi.org/10.1007/s11625-018-0627-5

Singh, G., Mittal, R., & Ahmad, M. (2007). A bibliometric study of literature on digital libraries. *The Electronic Library*, *25*(3), 342–348. https://doi.org/10.1108/02640470710754841

Singleton, A. (1976). Journal ranking and selection: A review in physics. *Journal of Documentation*, *32*(4), 258–289. https://doi.org/10.1108/eb026628

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, *8*(3), 327–340. https://doi.org/10.1177/030631277800800305

Small, H. (2016). Referencing as cooperation or competition. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication: A Festschrift in honor of Blaise Cronin* (pp. 49–71). De Gruyter. https://doi.org/10.1515/9783110308464-006

Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R., & Walker, J. H. (2003). DSpace: An open source dynamic digital repository. D-Lib Magazine, 9(1). https://doi.org/10.1045/january2003-smith

Spink, A. (1995, June 11–13). *Digital libraries and sustainable development?* [Paper presentation]. Digital Libraries ’95: The Second Annual Conference on the Theory and Practice of Digital Libraries, Austin, Texas, USA. https://www.jcdl.info/archived-conf-sites/dl95/papers/spink/spink.html

Tanguay, G. A., Rajaonson, J., Lefebvre, J.-F., & Lanoie, P. (2010). Measuring the sustainability of cities: An analysis of the use of local indicators. *Ecological Indicators*, *10*(2), 407–418. https://doi.org/10.1016/j.ecolind.2009.07.013

United Nations. (1992). *United Nations Conference on Environment & Development, Rio de Janeiro, Brazil, 3 to 14 June 1992: Agenda 21*. https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf

United Nations. (2001). *Road map towards the implementation of the United Nations Millennium Declaration: Report of the Secretary-General* (A/56/326). https://digitallibrary.un.org/record/448375?ln=en&v=pdf#files

United Nations. (2002). *Report of the World Summit on Sustainable Development, Johannesburg, South Africa, 26 August-4 September 2002* (A/CONF.199/20). https://digitallibrary.un.org/record/478154?v=pdf

United Nations. (2012). *Report of the United Nations Conference on Sustainable Development Rio de Janeiro, Brazil 20-22 June 2012* (A/CONF.216/16). https://digitallibrary.un.org/record/737074?ln=en&v=pdf

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development: Resolution adopted by the General Assembly on 25 September 2015* (A/RES/70/1). https://digitallibrary.un.org/record/3923923?ln=en&v=pdf

United Nations. (2023a). *Times of crisis, times of change: Science for accelerating transformations to sustainable development*. https://sdgs.un.org/sites/default/files/2023-09/FINAL%20GSDR%202023-Digital%20-110923_1.pdf

United Nations. (2023b, August 8). *Fast facts: What is sustainable development?* Sustainable Development Goals. https://www.un.org/sustainabledevelopment/blog/2023/08/what-is-sustainable-development/

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N. J., & Waltman, L. (2023, October 31). *VOSviewer manual*. VOSviewer. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf

Van House, N. A., Peterson Bishop, A., & Buttenfield, B. P. (2003). Introduction: Digital libraries as sociotechnical systems. In A. Peterson Bishop, N. A. Van House, & B. P. Buttenfield (Eds.), *Digital library use: Social practice in design and evaluation* (pp. 1–21). The MIT Press. https://doi.org/10.7551/mitpress/2424.003.0002

van Raan, A. F. J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, *43*(1), 129–139. https://doi.org/10.1007/bf02458401

Vos, R. O. (2007). Defining sustainability: A conceptual orientation. *Journal of Chemical Technology & Biotechnology*, *82*(4), 334–339. https://doi.org/10.1002/jctb.1675

Waas, T., Hugé, J., Verbruggen, A., & Wright, T. (2011). Sustainable development: A bird's eye view. *Sustainability*, *3*(10), 1637–1661. https://doi.org/10.3390/su3101637

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, *10*(2), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Waters, D. J. (1998, July/August). What are digital libraries? *CLIR Issues*, *4*. https://www.clir.org/1998/07/clir-issues-number-4/#dlf

Whitley, R. D. (1969). Communication nets in science: Status and citation patterns in animal physiology. *The Sociological Review*, *17*(2), 219–233. https://doi.org/10.1111/j.1467-954x.1969.tb01183.x

Wildemuth, B. M. (2017). Developing a research question. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (2nd ed., pp. 11–20). Libraries Unlimited. https://www.bloomsbury.com/us/applications-of-social-research-methods-to-questions-in-information-and-library-science-9781440839047/

Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient

adherence research. *Journal of the Association for Information Science and Technology*, *67*(4), 967–972. https://doi.org/10.1002/asi.23437

Zijp, M. C., Heijungs, R., Van der Voet, E., Van de Meent, D., Huijbregts, M. A. J., Hollander, A., & Posthuma, L. (2015). An identification key for selecting methods for sustainability assessments. *Sustainability*, *7*(3), 2490–2512. https://doi.org/10.3390/su7032490

Appendix A

Normalization of author names and affiliations in the WoS data set

Variation 1	Variation 2	Variation 3	Action
Assoc Comp Machinery	ACM		Abbreviation retained
Alemneh, DG	Alemneh, D		Second initial removed
Chowdhury, GG	Chowdhury, G		Second initial removed
Arizona State University-Tempe	Arizona State University		General affiliation retained
Indian Statistical Institute Bangalore	Indian Statistical Institute		General affiliation retained
Kent State University Kent	Kent State University Salem		General affiliation retained
University of North Texas Denton	University of North Texas System		General affiliation retained
University of Tennessee Knoxville	University of Tennessee System		General affiliation retained
University of Wisconsin Madison	University of Wisconsin Milwaukee	University of Wisconsin System	General affiliation retained
University West Indies Mona Jamaica	University West Indies Saint Augustine		General affiliation retained
Geological Survey of Canada - Natural Resources Canada	Natural Resources Canada		Departmental name retained
Lands & Minerals Sector - Natural Resources Canada	Natural Resources Canada		Departmental name retained

Note. In the case of author names, the actions described in the "Action" column were applied no matter whether the initials preceded or followed the last name. In the case of affiliations, campus names were removed. The replacements were made in all fields within the data set.

Appendix B

Normalization of author names and source titles in the Scopus data set

Variation 1	Variation 2	Variation 3	Action
Chowdhury G.G.	Chowdhury G.		Second initial removed
Dawson A.D.	Dawson A.		Second initial removed
Alemneh D.G.	Alemneh D.	Gelaw Alemneh D.	Second initial and full middle name removed
The Electronic Library	Electronic Library		Article removed
Sustainability (Switzerland)	Sustainability		Parenthetical info removed

Note. In the case of author names, the actions described in the "Action" column were applied no matter whether the initials preceded or followed the last name and regardless of the existence of any punctuation. The replacements were made in all fields within the data set.