Aug 10th, 12:00 AM

Bibliometric Analysis of Sustainability Information Systems Research

Nesreen El-Rayes
Kean University, nelrayes@kean.edu

Thomas Abraham
Kean University, tabraham@kean.edu

Viet T. Dao
Shippensburg University, vtdao@ship.edu

Bibliometric Analysis of Sustainability Information Systems Research

Completed Research

Nesreen El-Rayes
Kean University
nelrayes@kean.edu

Thomas Abraham
Kean University
tabraham@kean.edu

Viet Dao
Shippensburg University
vtdao@ship.edu

Abstract

As IS literature on Sustainability has evolved significantly, reviews of the literature tend to focus more on Green IT, and Green IS research and utilize a smaller body of literature. To help enhance understanding of the evolution of IS and Sustainability literature, this research examines an extensive body of knowledge. We used a comprehensive list of keywords to retrieve research related to Green IT, Green IS, and Sustainability IS. We searched for work beyond a limited list of outlets using the SCOPUS database with an extended timeframe. We then conducted a bibliometric analysis using text-mining tools to provide a clearer view of the field and its themes, thus it helps address what, where, and when questions about the nature of research in SIS literature, providing insights on what topics have been studied, where the research originated from, how the body of research has evolved over the years, and suggested future research.

Keywords

Sustainability information systems, Green IS, Green IT, Text-mining, bibliometrics.

Introduction

Global economic progress over the last centuries has delivered broad prosperity and lifted billions of people out of poverty in many countries. However, during recent decades, it has been increasingly acknowledged that our economic activities have contributed to climate change and other pressing social issues such as increasing inequality. As a result, business practitioners and researchers have increasingly embraced sustainability practice and research to help businesses address such environmental and social issues. Business research has found that while challenges abound, companies that genuinely embrace sustainability over the long-term outperform their counterparts in all three dimensions (environmental, social, and financial) of the triple bottom line (TBL) (Busch & Friede, 2018; Eccles et al., 2014; Flammer et al., 2019).

Since the 2000s, Information Systems (IS) practitioners and researchers have increasingly focused their efforts on studying how IS can contribute to addressing sustainability (Elliot & Webster, 2017; Singh & Sahu, 2020). Over time, IS research has broadened from an initial focus on ways to reduce the environmental impacts of IT infrastructure usage (Green IT) to studying the role that information systems play in enabling environmental sustainability initiatives across a firm and its partners (Green IS) (Elliot & Webster, 2017; Sedera et al., 2017). More recent research has started to examine the role of IS in enabling sustainability holistically, addressing all three dimensions of the TBL (Sustainability IS) (Abraham & Dao, 2019; Dao & Abraham, 2021; Dzhengiz & Niesten, 2020).

As IS literature on Sustainability has evolved significantly, researchers have made efforts to review the literature (Esfahani et al., 2015; Idrissi & Corbett, 2016; Singh & Sahu, 2020). Such reviews have tended to

focus more on Green IT and Green IS research and utilized a smaller body of literature (either in data sources searched or time frame studied). This research examines a more extensive body of work in several ways to help enhance our understanding of the evolution of the IS and Sustainability literature. We used a more extensive list of keywords to search for research that includes Green IT, Green IS, and Sustainability IS (SIS). We searched for work beyond a limited list of outlets using the SCOPUS database, and we used an extended timeframe for our search. We then conduct a bibliometric analysis using text-mining tools to provide what we hope will be a clearer view of the field and its themes. By doing so, we aim to help provide further answers to several questions with regards to IS and sustainability literature: 1) what topics have been studied?, 2) where does the research originate from?, and 3) how has the body of research evolved over the years? Such knowledge would help provide deeper understanding of the literature and could help us suggest areas for future research.

Literature Review

As IS literature on Sustainability evolved from Green IT to Green IS to Sustainability IS (SIS), researchers have expanded the scope of studies on the role of IS in enabling sustainability to address sustainability holistically, addressing all three dimensions of the TBL. Recent research has taken different approaches to review the literature to summarize our collective knowledge of IS and Sustainability and identify future research directions. Esfahani et al. (2015) were one of the earlier systemic literature reviews on Green IT and Green IS. Building on the belief-action framework, Henkel & Kranz (2018) reviewed research on pro-environmental behavior, and Green IS. Adopting a modernity perspective, Idrissi & Corbett (2016) identify four patterns of modernity manifested in Green IS research.

One of the challenges of mapping the field is the lack of a common name. Variously called Green IS, Green IT, Environmental Sustainability and IS, and so on, there is even a lack of common keywords authors use to describe their work. Sometimes research is overlooked because of these inconsistencies. Most reviews have focused mainly on Green IT, and Green IS. For example, Singh and Sahu (2020) use a classification approach to review the green IS literature. However, they used such descriptors as "Environmental Sustainability" and "Green Computing," which are broad and can return results from many disparate fields only tangentially related to SIS. Meanwhile, Vidmar et al. (2021) identified a small set of 61 papers. They used Leximancer to conduct an automated content analysis on a more narrowly defined field of sustainable business models and information systems. Therefore, our understanding of the literature on IS and Sustainability is limited to a smaller body of work and tends to focus more on the environmental aspects of sustainability. We conducted a more thorough literature review in this research, using a broader set of keywords, a larger set of database entries, and a longer time frame to extract a larger body of work relevant to Information Systems and Sustainability.

Methodology

Drawing from the bibliometric studies in (Donthu et al., 2020; Zhang et al., 2020), we present a bibliometric overview of sustainability information systems (SIS) literature in the forms of both descriptive and network analyses. We accessed bibliographic data used in this study from the SCOPUS database. We began by conducting a SCOPUS database search using "Green IS" keywords. We pulled data on publications' titles, abstract, and keywords for our analysis. Afterward we conducted a validation process of the results using a list of references from a few recently published papers. During this process, it became clear that we would need to expand the list of keywords and search on the abstract and title in addition to the keywords to capture the complete list of papers in the reference lists.

We performed a four-cycle process to explore and evaluate sustainable information systems research status. First, we brainstormed to identify frequently cited papers in the field of Green IS, and a sample of 25 publications was selected to combine and extract the keywords related to our study. We identified forty terms from this first step. Second, we retrieved all the abstracts from the SCOPUS database across the 40 terms for all available years (1900-2022). The total number of publications across the 40 terms was approximately 650,000 (including duplicates across the themes). Third, we excluded unrelated areas such as Medicine, Mathematics, Earth and Planetary Science, Agriculture, Physics and Astronomy, Health Professions, Biochemistry, Energy, Nursing, Arts and Humanities, Psychology, etc. A paper could be classified under more than one area in SCOPUS. In such situations, we include documents that are classified

under relevant fields. For example, if a paper is about IS in the medical area, it will be included in our dataset if SCOPUS classifies the paper under Computer Science and Medicine. Fourth, 16 terms were selected as the final set to study in-depth (see Table 1.). The entire data set comprised 4,488 publications, and after removing the duplicates, the final data set is 3,885 publications. Table 1 shows the search terms, the range of years the research was done, and the number of publications found.

	Terms in ascending order	Year	Publications
1	"Energy informatics"	2002-2021	105
2	"Environmentally sustainable IS"	2000-2022	9
3	"Firms' green practices"	2011-2021	3
4	"Green ICT"	2008-2021	183
5	"Green Information System"	2008-2022	169
6	"Green Information Technology"	2008-2021	141
7	"Green IS"	1949-2022	706
8	"Green IT"	1971-2022	911
9	"Green practices-IS"	2011-2021	23
10	"Integrated sustainable-value"	2011-2022	4
11	"Sustainability Systems"	1991-2022	83
12	"Sustainable ICT"	2003-2021	75
13	"Sustainable Information Systems"	1999-2021	40
14	"Sustainable IS"	1988-2022	254
15	"Sustainable IT"	1992-2022	309
16	"Sustainability" AND "Information Technology"	1990-2022	1473
		Total	4488

Table 1. List of searches and results (the year 2022 covers February)

We analyzed this corpus of research using both descriptive and network methods. Our descriptive analyses include the number of publications by country, keyword, and time zone. This allows us to answer where, what and when questions about the nature of research in this field.

Our network analyses included text mining to identify important terms and co-occurrences to identify research clusters. We used VOSviewer software to carry out the mapping analyses in this study (Van Eck & Waltman, 2020). VOSviewer, a visualization tool, was used for text mining and cluster analysis. VOSviewer algorithms are based on the Apache OpenNLP library. First, copyright statements and stop words are removed from the abstract, titles, and keywords, Then, a sentence detection algorithm is used to split the abstract into sentences, and 'part-of-speech' tagging breaks down the sentences by noun, verb, adjective, preposition, etc. For the cluster analysis, VOSviewer uses two standardized weights, the number and total strength of the links, to visualize the nodal network graphically. The size of the nodes denotes the frequency of occurrence of the identified terms.

Results

This section reports the results of our descriptive, text mining, and network analysis.

Descriptive Statistics

The map in Figure 2 shows the distribution of publications over countries. The top ten countries in terms of frequency of publications are the United States (666), Germany (351), the United Kingdom (323),

Australia (256), China (247), India (238), Malaysia (155), Canada (145), Italy (137), and Spain (116). Regarding the publications by continent, Europe has the highest number of publications (1904), followed by Asia (1450), North America (836), Oceania and Australia (259), Africa (189), and South America (108).[1]

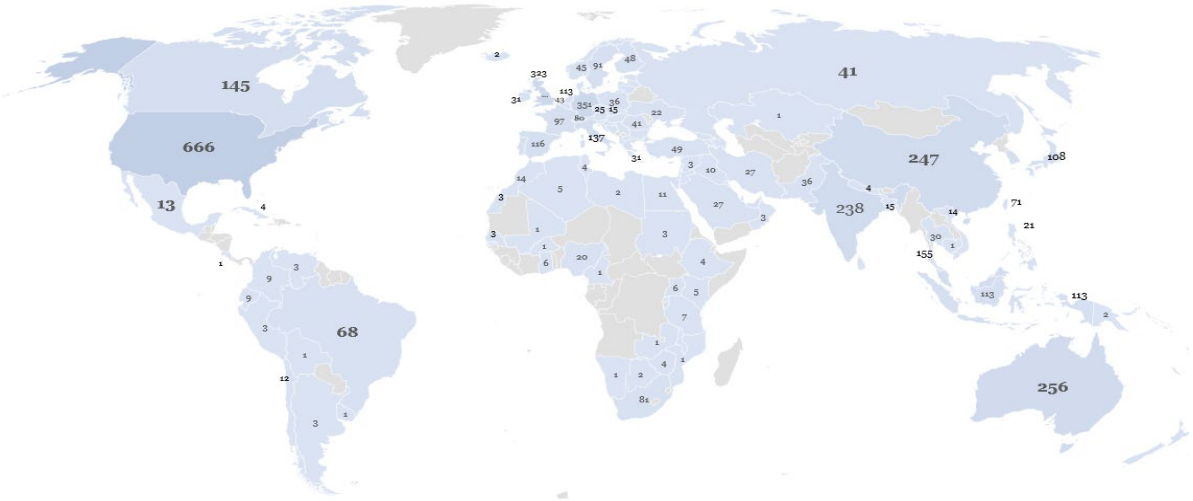
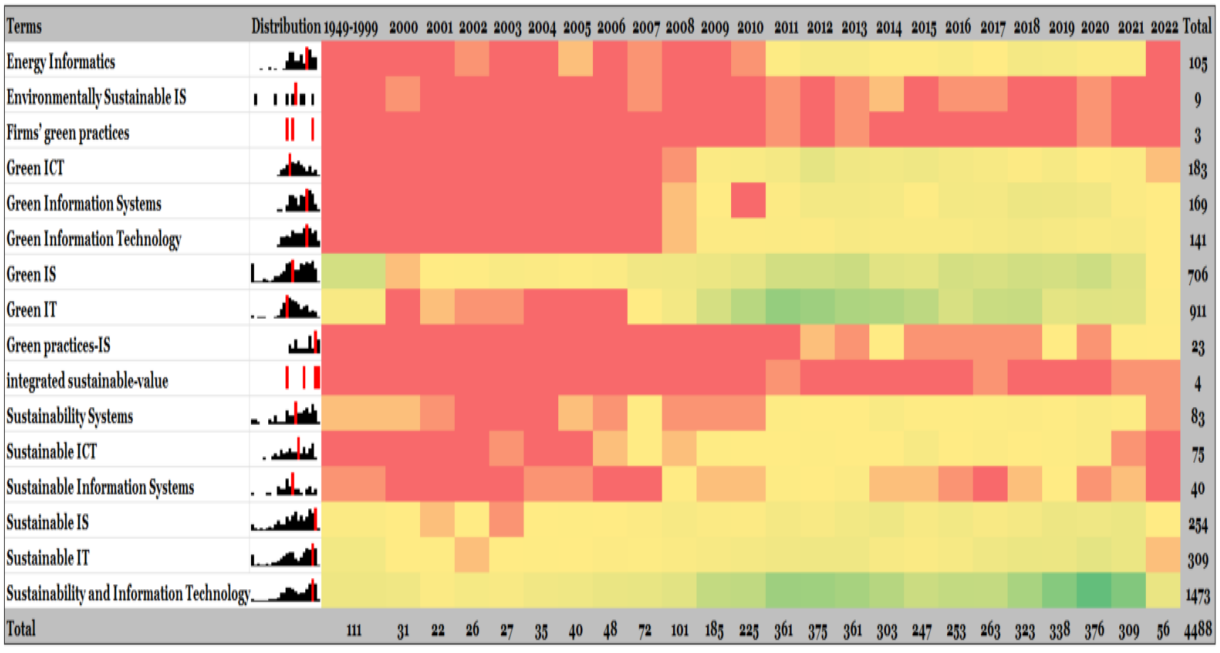


Figure 2. Distribution of publications by country

The heat map in Figure 3 reflects the distribution of publications over the 16 terms selected for the study. The number of publications is listed in ascending order: Red-Yellow-Green. The red bars under the distribution reflect the mode of each term over the years. We generated the largest collection of publications from the search term "sustainability" AND "information technology," as you can see from the green area in the heat map. While the terms "green IS" and "green IT" also produce large numbers of publications, the modal values for each appear earlier in the time frame. "sustainable IS" and "sustainable IT" generate a more right-skewed distribution which indicates that these terms are becoming more popular.



[1] *For joint studies across countries or continents, each participant gets a point for each publication.*

Figure 3. Heatmap and Distribution 1949-2022

In the period 1949 through 1989, we found an average of two publications per year. In the Time Zone graph in Figure 4, we label this Zone 1. In each decade since, the average number of publications has increased exponentially, following the polynomial trendline (equation in Figure 4). Extending the trendline, the expected number of publications will be around 400 by the end of 2021.

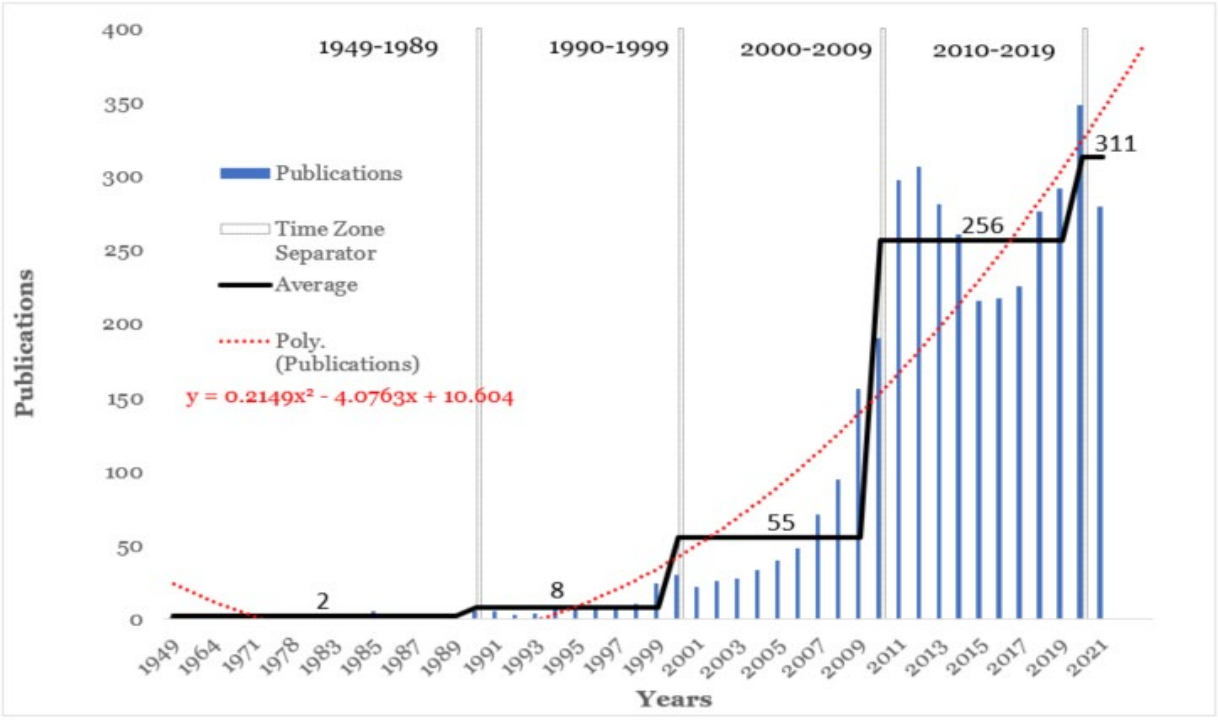


Figure 4. Publications over Time Zones

Text Mining and Network Visualization

The 3,885 publications included in this study comprise 20,226 unique terms in the title, abstract, and keywords (excluding stop words). Of these, 642 terms meet our threshold of ten or more occurrences. After a manual check to exclude any unrelated terms, we are left with a set of 604 terms. We applied a clustering algorithm to our corpus using the software VOSviewer (Van Eck & Waltman, 2020). Figure 5 shows the network visualization of terms that appear more than the threshold set (10 occurrences). The size of the 'node' depends on the total occurrences, while the 'links' reflect the number of terms it is tied to, and the total link strength is a score provided to a term based on its appearance in various publications. The colors reflect the cluster of keywords that co-occur together. There are three significant clusters in that dataset (Blue, Green, and Red). VOSviewer's clustering method is based on an optimization algorithm to determine the number of clusters in each run. The terms were clustered into two, four, five, six, and seven clusters, but three was the optimal one.

The clusters may guide researchers to tackle the areas that are not fully explored yet. In particular, in Cluster 2 (the green cluster), although the overall frequency of terms is not much less than the largest cluster (Red cluster), the size of the majority of the nodes in the green cluster is too small, and there are fewer links between them compared to the other two clusters, which reflects the occurrence of the majority of the terms very few times. To illustrate, more investigation is needed to explore and evaluate the life cycle assessment of sustainable IS, managerial practices, and setting a strategic approach for sustainable development.

Cluster 1: Business (red - 233 Terms)

The red cluster is the largest cluster in figure 5. The top terms in this cluster are 'sustainable development,' which occurred 1158 times, 589 links, with a total link strength of 8135, followed by 'information technology' that occurred 865 times, with a total link strength of 5537. Then 'information systems' occurred 487 times, 462 links, with a total link strength of 3404. Other terms include 'green information systems', 'supply chain management', 'governance', 'corporate social responsibility', and other business-related terms.

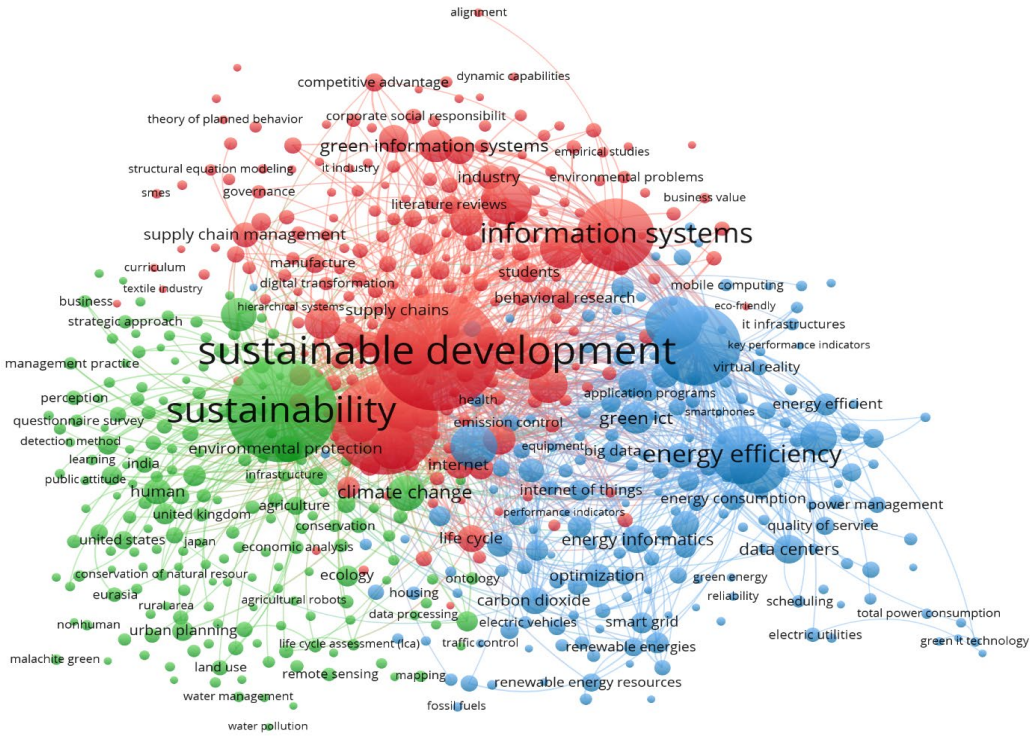


Figure 5. Network Diagram and the top three clusters in the dataset

Cluster 2: Environment (green - 188 Terms)

The top term in this cluster is 'sustainability' which occurred 932 times, has 574 links, with a total link strength of 5511. The next term is 'climate' which occurred 140 times, and 'innovation' which occurred 127 times. Based on the observation of the results, this cluster has publications related to the environment (monitoring, protection, and policies), having a green economy, renewable resources, and life cycle analysis. Some publications related to the food supply and agriculture.

Cluster 3: Technology (blue - 183 Terms)

This cluster is the smallest of the three groups. The top term in this cluster is 'green IT', which occurred 587 times, had 473 links, and had a total link strength of 3700. This was followed by 'energy efficiency', which occurred 333 times, had 414 links, and a total link strength of 2645. Then 'green computing' occurred 288 times, had 387 links, and a total link strength of 2108. Then 'energy utilization', which occurred 276 times, had 383 links and a total link strength of 2291. This cluster had publications related to Industry 4.0 technologies (Artificial Intelligence, Internet of Things, Blockchain, Informatics, Cloud Computing, Virtual Reality), in addition to data mining and machine learning techniques. Still, all these terms show up fewer than 100 times.

	Cluster 1	**Frequency**	**Cluster 2**	**Frequency**	**Cluster 3**	**Frequency**
1	Sustainable development	1158	Sustainability	932	Green IT	587
2	Information technology	865	Climate change	140	Energy efficiency	333
3	Information systems	487	Innovation	127	Green computing	288
4	Green IS	234	Environmental protection	86	Energy utilization	276
5	Environmental sustainability	205	Human	79	Environmental impact	200
6	Information management	190	Information and communication technology	54	Cloud computing	134
7	Decision making	179	Ecology	53	Green ICT	114
8	Information use	167	Humans	53	Data centers	93
9	Information and communication technologies	143	Eco systems	51	Energy informatics	87
10	Environmental management	123	Urban planning	50	Carbon dioxide	85
11	Green information systems	117	China	48	Carbon footprint	84
12	Economics	107	Environmental economics	48	Greenhouse gases	81
13	Research	106	Public policy	43	Energy conservation	77
14	Economic and social effects	105	Agriculture	41	Global warming	76
15	Planning	102	United states	41	Decision support systems	71
16	Environmental technology	97	India	39	Optimization	68
17	Societies and institutions	92	Software	38	Artificial intelligence	67
18	Design	91	Technology adoption	37	Communication	60

19	education	91	Europe	36	Energy efficient	58
20	competition	87	Economic growth	35	Smart grid	57
21	life cycle	86	Economic development	34	Energy consumption	56
22	internet	82	Efficiency	33	Carbon	55
23	developing countries	81	GIS	32	Big data	54
24	commerce	72	Land use	32	Smart power grids	53
25	environment	72	Developing world	31	Software engineering	53

Table 2. Top 25 keywords by Cluster

Co-authorship Analysis and Network Visualization

We also used VOSviewer to conduct a co-authorship analysis to identify some of the key researchers and their associations. By limiting the authors to ones with at least 100 citations and five documents in the corpus, we found 20 clusters of 66 researchers shown in Figure 6. Of these, only five clusters had four or more linked authors. Cluster 1 (shown in red) has 13 authors in the cluster and appears to have a Chinese/Asian axis and a focus at the intersection of SIS and supply chain management. Cluster 2 (shown in green) has 10 authors and appears to have a European axis and a focus on the intersection of strategy and SIS. Cluster 3 (shown in blue) has nine authors and appears to have an international axis and a focus on energy informatics. Cluster 4 (shown in yellow) has four authors and appears to have an Australian axis and a focus on the technology of green IT. Cluster 5 (shown in purple) has four authors and appears to have a European axis and a focus on the intersection of sustainable development and ICT.

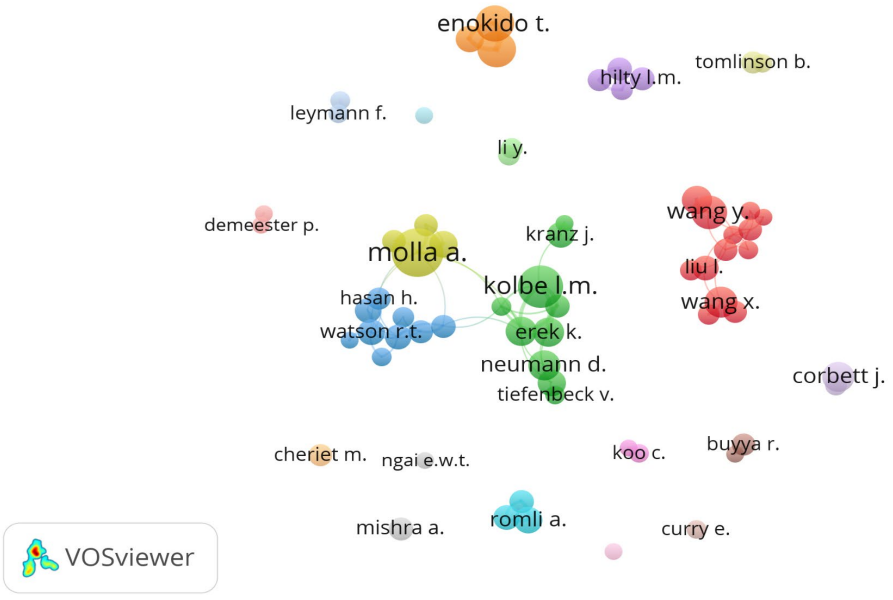


Figure 6. Network Diagram of the twenty author clusters

Findings and Discussion

One important finding is that there is no consistency among the community of researchers in selecting keywords for this area of research. To find our sample reference list, we needed to search 16 different terms. Green IS is popular but does not capture the social dimension of sustainability. It also suffers from the fact that IS is a very common English verb. We recommend that this community of researchers include "sustainability information systems" as a common keyword. This keyword captures the social dimension and avoids the pitfall of returning unrelated publications based on a common verb.

More research is needed with respect to Industry 4.0 technologies. We found fewer than 100 occurrences of terms like IoT and blockchain. 'Social Media' only appeared 16 times in all the 3885 abstracts studied. Furthermore, the keyword was tied with other words that occurred less than ten times. That may be an interesting area to investigate from an information systems or supply chain perspective. Another area that appears under-researched is 'sustainability assessment' and 'sustainability indicators', where each one appeared fewer than 20 times. Without solid measurement and assessment methods, evaluating sustainability will be a challenge.

By identifying the terminology of the field of sustainability information systems (SIS), especially the 604 most-frequently used terms, we hope to point the way to creating useful data dictionaries for future text mining research in this field.

We have found some key clusters of researchers and areas of research. Further analysis of these clusters should help better map the research questions being asked and ones not receiving enough attention. For example, are there any social dimensions being explored here? The topics of poverty and social justice and the United Nations sustainability development goals (SDGs), for instance.

Europe is taking the lead and has the highest number of publications related to sustainable information systems, followed by Asia and North America.

Conclusion

We expanded the corpus of publications in 'Green IS' to conduct a holistic review of sustainability information systems. This expanded review resulted in 3885 publications over one hundred years. Descriptive statistics show that the volume of publications has been increasing at an exponential rate since 1989. Most publications are still found in the US, but Europe, Australia, China, and India produce large numbers of papers. While there are no standard keywords used to describe the literature, it appears that 'sustainability' is starting to replace 'green' as a qualifier. Therefore, we recommend that researchers use the keyword 'sustainability information systems' as the standard. We extracted over 20,226 terms for these publications' titles, keywords, and abstracts using text mining tools. Six hundred and four words appeared over ten times and formed three distinct clusters representing Business, the Environment, and Technology. Industry 4.0 technologies, social media, and assessment are not frequently used, which may point to areas of future research.

REFERENCES

Abraham, T., and Dao, V. 2019. "A longitudinal exploratory investigation of innovation systems and sustainability maturity using case studies in three industries," *Journal of Enterprise Information Management* (32:4), pp. 668-687.

Adamczyk, A. 2021. "Sustainable investments hit record highs in 2020—and they're earning good returns". *CNBC*. Retrieved from https://www.cnbc.com/2021/02/11/sustainable-investments-hit-record-highs-in-2020.html

Busch, T., and Friede, G. 2018. "The robustness of the corporate social and financial performance relation: A second-order meta-analysis," *Corporate Social Responsibility and Environmental Management* (*25*:4), pp. 583-608.

Dao, V. , and Abraham, T. 2021 "An Empirical Examination of the Use of IS-enabled Sustainability Initiatives Across the Integrated Sustainability Framework," *Pacific Asia Journal of the Association for Information Systems* (13:3), pp. 57-85.

Dzhengiz, T., and Niesten, E. 2020. "Competences for environmental sustainability: A systematic review on the impact of absorptive capacity and capabilities," *Journal of Business Ethics* (162:4), pp.881-906.

Donthu, N., Kumar, S., and Pattnaik, D. 2020. "Forty-five years of Journal of Business Research: A bibliometric analysis," *Journal of Business Research* (109), pp. 1-14.

Eccles, R. G., Ioannou, I., and Serafeim, G. 2014. "The impact of corporate sustainability on organizational processes and performance," *Management Science* (*60*:11), pp. 2835-2857.

Elliot, S., and Webster, J. (2017). "Special issue on empirical research on information systems addressing the challenges of environmental sustainability: an imperative for urgent action." *Information Systems Journal* (27:4), pp. 367-378.

Esfahani, M. D., Rahman, A. A., and Zakaria, N. H. 2015. "The status quo and the prospect of green IT and green IS: a systematic literature review," *Journal of Soft Computing and Decision Support Systems* (2:1), pp.18-34.

Flammer, C., Hong, B., and Minor, D. 2019. "Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: Effectiveness and implications for firm outcomes," *Strategic Management Journal* (40:7), pp.1097-1122.

Henkel, C., and Kranz, J. 2018. "Pro-Environmental Behavior and Green Information Systems Research-Review, Synthesis and Directions for Future Research," *International Conference on Information Systems,* San Francisco.

Idrissi, S. and Corbett, J. 2016. "Green IS research: A modernity perspective." *Communications of the Association for Information Systems*, (38:30), pp. 596-623.

Meredith, S. 2020. "Sustainable investment funds just surpassed $1 trillion for the first time on record," *CNBC*. Retrieved from https://www.cnbc.com/2020/08/11/coronavirus-esg-and-sustainable-funds-surpass-1-trillion-for-the-first-time.html

Sedera, D., Lokuge, S., Tushi, B. and Tan, F., 2017. "Multi-disciplinary green IT archival analysis: a pathway for future studies," *Communications of the Association for Information Systems* (41:1), pp. 674-733.

Singh M., and Sahu, G.P. 2020. "Towards adoption of Green IS: A literature review using classification methodology," *International Journal of Information Management (*54), pp. 102-147.

Van Eck, N. J., and Waltman, L. 2020. VOSviewer manual version 1.6.16. Leiden, The Netherlands: Universiteit Leiden.

Vidmar, D., Marolt, M., and Pucihar, A. 2021. "Information Technology for Business Sustainability: A Literature Review with Automated Content Analysis," *Sustainability* (13:1192), pp. 1-24.

Zhang, F., Wu, X., Tang, C. S., Feng, T., and Dai, Y. 2020. "Evolution of operations management research: from managing flows to building capabilities," *Production and Operations Management* (29:10), pp. 2219-2229.