Digital transformation in tourism: bibliometric literature review based on machine learning approach

Peter Madzík
Department of Management, Faculty of Management,
Comenius University Bratislava, Bratislava, Slovakia
Lukáš Falát
Department of Macro and Microeconomy,
Faculty of Management Science and Informatics, University of Zilina, Zilina, Slovakia
Lukáš Copuš
Department of Management, Faculty of Management,
Comenius University Bratislava, Bratislava, Slovakia, and
Marco Valeri
Faculty of Economics, Niccolò Cusano University, Rome, Italy and
Magellan Research Center Iaelyon, Jean Moulin University, Lyon, France

Abstract

Purpose – This bibliometric study provides an overview of research related to digital transformation (DT) in the tourism industry from 2013 to 2022. The goals of the research are as follows: (1) to identify the development of academic papers related to DT in the tourism industry, (2) to analyze dominant research topics and the development of research interest and research impact over time and (3) to analyze the change in research topics during the pandemic.

Design/methodology/approach – In this study, the authors processed 3,683 papers retrieved from the Web of Science and Scopus. The authors performed different types of bibliometric analyses to identify the development of papers related to DT in the tourism industry. To reveal latent topics, the authors implemented topic modeling based on latent Dirichlet allocation with Gibbs sampling.

Findings – The authors identified eight topics related to DT in the tourism industry: City and urban planning, Social media, Data analytics, Sustainable and economic development, Technology-based experience and interaction, Cultural heritage, Digital destination marketing and Smart tourism management. The authors also identified seven topics related to DT in the tourism industry during the Covid-19 pandemic; the largest ones are smart analytics, marketing strategies and sustainability.

Originality/value – To identify research topics and their development over time, the authors applied a novel methodological approach – a smart literature review. This machine learning approach is able to analyze a huge amount of documents. At the same time, it can also identify topics that would remain unrevealed by a standard bibliometric analysis.

Keywords Digital, Digitization, Digitalization, Digital transformation, Tourism, Bibliometric, Topic modeling, Machine learning, LDA, Latent Dirichlet allocation, Covid, Pandemic

Paper type Literature review

1. Introduction

A considerable number of companies and organizations around the world were aware of the importance of digital transformation (DT) in the pre-Covid years (Vial, 2019; Frank *et al.*, 2019; Hinings *et al.*, 2018; Matt *et al.*, 2016; Alkhatib and Valeri, 2022; Deb *et al.*, 2022), but they did not always respond to these calls. At that time, the relatively stable market environment can be included among the reasons for partially ignoring this trend. However, the Covid pandemic outbreak at the turn of 2019 and 2020 significantly disrupted this stability (Adhikari *et al.*, 2020; Harel, 2021). Thus, under the influence of the pandemic and significant uncertainty, companies, in many cases, implemented such significant changes that they would probably not dare to implement in the near future (He and Harris, 2020; Verma and Gustafsson, 2020). Tourism was one of the most affected sectors of the economy during the pandemic (Škare *et al.*, 2021; Gössling *et al.*, 2020; Sigala, 2020). While in the pre-pandemic year of 2019, it was 1.5 billion international tourist arrivals; in 2020, there were only 381 million, equal to the level of 1990 (Lončarić *et al.*, 2022; United Nations World Tourism Organization, 2021). Thus, organizations sought ways to overcome current problems and build their long-term resilience into the future (Ntounis *et al.*, 2021; Prayag, 2020). However, since the organization had no experience with the pandemic, there was no generally accepted guide on overcoming a crisis of this extent. Organizations operating in tourism therefore sought and applied their own methods, including the area of information and communication technologies (ICT) (Podzharaya and Sochenkova, 2022; Botti and Monda, 2021).

DT is mentioned more significantly in the literature especially during the last decade (Chatterjee *et al.*, 2022; Jayawardana *et al.*, 2022; Zheng *et al.*, 2022). Verhoef *et al.* (2021) state in their study that digitalization has also undergone development. Its initial phase in some industries already took place in the distant past. The drivers from the information technologies (IT) point of view are digital technology, digital competition and digital customer behavior. In the DT itself, he mentions three phases: digitization, digitalization and DT. Digitization, as the first phase, represents the transformation of analog information into digital information. Digitalization involves using digital technologies to transform existing business processes (Li *et al.*, 2016). Finally, DT represents a company-wide change that leads to the development of new business models (Verhoef *et al.*, 2021; Pagani and Pardo, 2017; Kane *et al.*, 2015; Iansiti and Lakhani, 2014).

The studies published so far focus on different phases of DT, either individually or cross-sectionally. Naturally, it can be assumed that the future trend will cover the last phase of DT. However, several organizations in the tourism industry have gone more significantly through initial phases only in the recent pandemic period. Perry *et al.* (2022) introduce the connection between digitization and sustainability in parks with recreation indicators. Their results show that both practice and research should focus more precisely on the highest level of DT – that is, a company-wide change that will lead to the development of new approaches to business.

We can see several partial examples of DT in tourism under the influence of the Covid-19 pandemic in the literature. For example, Rana *et al.* (2022) present an application of blockchain technology. Nabila *et al.* (2021) mention various collaborative systems that could bring long-term sustainability and resilience to organizations operating in the tourism industry. Akhtar *et al.* (2021) draw attention to digital tourism, stating that virtual tourism was a practical and valuable option for mass tourism during the Covid-19 outbreak and can replace mass tourism after the pandemic. Toubes *et al.* (2021) looked at the impact of the pandemic on tourism from the point of view of marketing. Their study reported that online information sources gained weight over consulting friends and relatives during the pandemic. A significant advance in digitalization is expected, where online platforms will replace physical travel agencies. Ohe (2022) also covered a similar area, stating that Covid-

19 promoted digitalization, such as online travel agencies and expanded the field of e-hospitality based on digital technology in addition to traditional on-site personal hospitality. Fontanari and Traskevich (2022) presented a model for applying smart technologies to prevent over-tourism and develop destinations' resilience in the post-Covid-19 period.

However, elements of DT in tourism can be found in the literature even before the pandemic. For example, Gazi *et al.* (2016) looked at social media, which was becoming a popular tool in tourism at the time, evaluating the role of social media tools for tourism services in the country concerning the perception of people with disabilities. A similar topic, but from a different point of view, was also addressed by Keil *et al.* (2017), namely the optimization of the user interface design and interaction paths for the destination management information system. Gómez *et al.* (2018) dealt with the integration of ICT in European Union countries in terms of booking accommodation in the tourism industry. The results suggest different behavior patterns in managing digital accommodation and how ICT is managed. Finally, Gyódi (2019) looked at sharing economy services such as Airbnb, and his results show that Airbnb and traditional hotels compete for travelers in a wide variety of market segments.

Understanding the development of digitalization in the tourism industry is becoming an essential tool for building long-term resilient organizations. In addition, digitalization has become critical under the influence of the pandemic, and pointing out trends and areas of research is highly topical. Therefore, the aim of this study is to create a scientific map of DT in the tourism industry, which will be closely connected and applicable in research and practice.

2. Literature review
2.1 Bibliometric reviews overview
Similar to other areas, also in tourism, many authors are devoted to the topic of DT, and naturally, different literature reviews are published. Since this topic is relatively new, we present an overview of a significant literature review based on bibliometrics analysis published in the recent past. At the same time, based on our opinion, we try to indicate which phases of DT the particular publication is likely to address.

One of the largest and newest bibliometrics reviews related to tourism in connection with DT is a study by Molina-Collado *et al.* (2022). They identified and analyzed 2,424 scientific journal publications indexed in the Web of Science and Scopus databases from 1988–2021. Their primary focus, based on the search terms, was on ICT in connection with tourism, and thus it was primarily the first or the second phase of DT. Several possible research topics for the future emerge from their results, for example, electronic word-of-mouth, user-generated context, self-service technologies, robotics, smart tourism or virtual reality.

One of the newest literature reviews is also a study by Verma *et al.* (2022). The authors focused on the past, present and future of virtual tourism, and thus it was about all three phases of DT. They linked quantitative (science mapping) and qualitative (intellectual structure mapping) methodologies. They analyzed a total of 1,652 articles published in the years 2000–2021. However, they focused specifically only on virtual tourism and analyzed elements such as augmented reality, virtual reality or big data. Their conceptual model also brings future research directions, including mobile devices and smart tourism, internet-based interactions and destination management, and virtual reality and augmented reality-based tourism.

A similar approach from the point of view of past, present and future, but in the topic of smart tourism destinations, chose in their bibliometric analysis Bastidas-Manzano *et al.* (2020). Again, it was about all three phases of DT. However, the number of analyzed articles was

significantly lower (258 in total), which was also because they only focused on the period 2013–2019. They introduced research topics such as smart city, sustainability or tourist experience measuring through Big Data and IoT.

Smart tourism is also the topic of a bibliometric review by Chen *et al.* (2021). They focused on studies published between 2010–2021, analyzing 441. They consider the mentioned years as the start of smart tourism and cover the subsequent entry of 5G mobile technology and the impact of Covid-19 on tourism. From our point of view, all three phases of DT are probably covered. Based on their results, it is recommended that research should focus more on the practical implications of topics such as IoT, artificial intelligence, cloud computing, big data and biometrics.

As part of the tourism industry, e-tourism was a topic of the bibliometric study by Singh and Bashar (2021). They analyzed 146 publications from 2004–2020 and thus probably included all three phases of DT in this e-tourism topic. In addition to results like authors' institutions, journals or most cited papers, they also identified the main trends and topics of e-tourism that could be addressed in further research. It is, for example, smart technology, virtual reality, augmented reality, or digital architecture.

Bibliometric analysis by Ndou *et al.* (2022) focuses on using technology tools to deal with Covid challenges in organizations in tourism (probably the two first phases of DT). They analyzed a total of 319 publications from 2019–2022. Their study provided evidence that Covid-19 has increased the use of various technologies in the tourism value chain. The analysis also pointed to the main research topics that should be pursued further.

Applying artificial intelligence in tourism, i.e. the highest phase of DT, through bibliometric review was the topic of the study by Knani *et al.* (2022). They analyzed a total of 1,035 publications from the years 1984–2021 from the Web of Science and Scopus databases. Their analysis focused on common areas, such as authors, institutions, and countries. They also created the thematic map, with which they expressed possible research topics (e.g. big data, service robots, forecasting tourism models and others).

Bibliometric analysis of big data in tourism (the highest phase of DT) was addressed in the study by Li and Law (2020). The authors focused on the period of 2008–2017 while analyzing a total of 1999 scientific publications indexed in the Web of Science database. Based on the results, they state that similar topics are essential in tourism as well as in other areas using big data, especially privacy, data quality and appropriate data use.

2.2 Research gap

The studies we listed above generally aimed to analyze some direct or indirect aspects of DT in the tourism industry. However, systematic literature reviews and bibliometric reviews also have certain limitations. Studies of the systematic literature review type are in-depth and usually process several dozens of documents, which they analyze sufficiently. A systematic literature review results are then more narrowly oriented (Moher *et al.*, 2015; Page *et al.*, 2021). On the other hand, bibliometric reviews are oriented toward a broader scope of the researched area, and their goal is more priority trends (Cobo *et al.*, 2011; Eck and Waltman, 2010). Currently, even bibliometric studies can contain more in-depth information, e.g. by analyzing keywords, co-occurrence of authors/keywords, etc. However, one of the top trends in the field of bibliometric reviews is the use of machine learning to identify latent patterns in textual data (Zhang *et al.*, 2017; Han, 2020; Mariani and Baggio, 2022).

Our study focuses precisely on the use of the potential of processing academic papers with automated tools. The aim of this study is to create a science map of DT in the tourism industry. To do this, we use a novel machine-learning-based methodological approach to identify latent topics from a vast number of academic documents. The study analyzes 3,683 papers published in the last ten years (2013–2022) retrieved from Scopus and Web of Science databases.

Our approach will make it possible to comprehensively capture the areas of DT in the tourism industry and thus offer a science map of this dynamically developing area. We developed three interconnected research questions (RQs) to operationalize the main aim of our paper:

RQ1. What is the development of academic papers related to digital transformation in the tourism industry?

○ Digitalization is growing dramatically in practically all research areas, not excluding the tourism industry. Therefore, a bibliometric overview of the development of the number of research papers, top journals and most cited papers could be a fresh probe into the development of this dynamic area.

RQ2. What are the dominant research topics, and how do their research interest and research impact develop over time?

○ The number of documents published in the last ten years related to digitalization or DT in tourism is enormous. Therefore, identifying the topics that these documents explore is practically impossible in the usual way. However, using machine-learning procedures, we can not only extract (identify) these topics from a considerable number of documents but also record their development over time.

RQ3. How have research topics changed under the influence of the pandemic?

○ The Covid-19 pandemic was a game-changer in virtuality in all industries. In the tourism industry, the pandemic accelerated digitalization and probably profiled separate scientific topics that we can identify using machine-learning topic modeling.

3. Methodology
3.1 Data protocol
In order to perform a relevant analysis of the current state of DT in tourism, we needed relevant resources. To obtain resources we selected the scientific databases Web of Science and Scopus. These are the two largest and most prestigious scientific databases that form a source of world knowledge.

We performed the data acquisition process similarly in both databases. After selecting the information sources, we defined the search query. We decided to define it as follows: *("digital" OR "digitization" OR "digitalization" OR "DT") AND tourism*. The reason for choosing the mentioned search query was the fact that according to Verhoef *et al.* (2021) DT is a process that has multiple phases. The first phase is digitization, the second phase is digitalization and the final phase is the DT itself. In addition, we added the word "*digital*" to the search query. The reason for including this word in the search query was because this word is very closely related to DT and is very often part of it.

We then applied the defined search query into the search tools of both databases. In both cases, we set to search for the defined query in the titles of articles, abstracts and keywords. Since we wanted to capture the development of this topic over the last 10 years, we also defined a time restriction related to the year of publication of articles (2013–2022). We performed the search in both databases on September 23, 2022. The search criteria defined in this way returned 5,233 results on that day (2,160 results in the Web of Science database and 3,073 in the Scopus database).

After obtaining the data, the initial editing phase followed. We excluded articles that did not have the "*Author*" item defined (i.e. proceedings). Since a large number of articles are indexed in both databases, it was necessary to discard those records that were in both databases. When deleting duplicate records, we kept the record that had a higher number of citations. This was because in the next phase we tried to capture research impact and the number of citations best represented this metric. Finally, we removed articles that were editorials and retracted papers. The resulting dataset, with which we continued to work, had a size of 3,683 papers. The entire process of data acquisition and data handling is clearly shown graphically in the form of a flowchart in Figure 1.

3.2 Latent Dirichlet allocation

To answer RQ2 and RQ3, we used a machine learning approach. The reason was that our dataset contained many studies, and manual processing of documents through a systematic literature review approach would be lengthy and inefficient. Using machine learning, we performed the so-called smart literature review, which allowed us to analyze a large number of articles while maintaining the depth of the analysis.
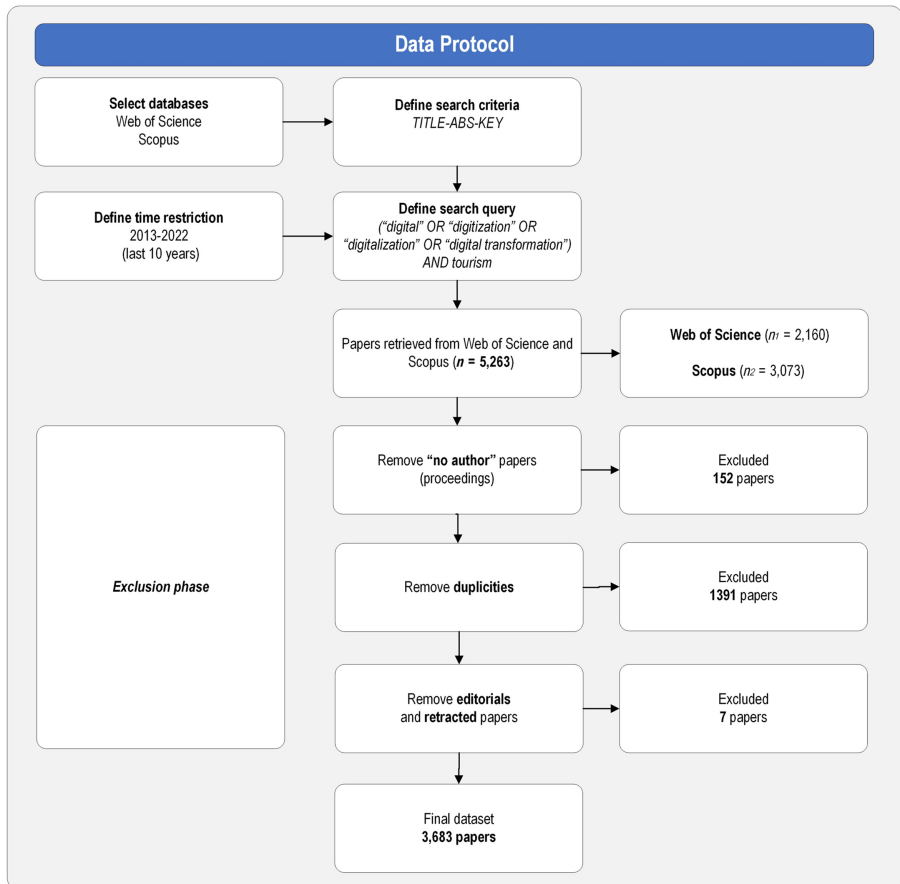

Figure 1.
Data acquisition and data handling protocol

Source(s): Author's own creation/work

For the smart literature review, we used an unsupervised machine learning approach based on topic modeling. Topic modeling is one of the most powerful tools in the area of text mining. Topic models are able to find relationships among huge amounts of text documents (Jelodar *et al.*, 2019). Topic modeling is based on the analysis of a large number of text documents. The purpose of this method is to segment documents into several groups based on their similar structure.

We implemented topic modeling with latent Dirichlet allocation (LDA) (Blei *et al.*, 2003). LDA, which is a generative probabilistic model of a corpus (Blei *et al.*, 2003), is one of the most popular methods for topic modeling (Jelodar *et al.*, 2019) and is widely used in various fields.

The basic concept of LDA is that documents are composed of multiple topics. Moreover, topics are composed of many words. Each document is a mixture of topics, and each topic is a mixture of words. More mathematically, in the LDA model, documents are seen as random mixtures over latent topics. In addition, each topic is represented by a distribution over words. In other words, every topic is represented by word probabilities (Blei *et al.*, 2003).

From the structure of words in documents latent topics can be inferred. The idea is as follows. Each document is generated by a process. This process has a Dirichlet distribution with α vector parameter. From this Dirichlet distribution with α parameter, θ is sampled. θ is then used in multinomial (θ) distribution. This distribution is then used to generate (or to sample) a topic for each of N steps (N = number of words in the document). Finally, based on the sampled topic, each word is sampled from the topic-specific multinomial distribution, i.e. $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

As stated above, θ is a random variable with Dirichlet distribution conditioned on α (parameter) and is defined as

$$p(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{k}\alpha_i\right)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\theta_1^{\alpha_1-1}\ldots\theta_k^{\alpha_k-1} \tag{1}$$

The topic proportions for a given document are then generated using this θ random variable with the Dirichlet probability distribution based on α parameter (Blei and Lafferty, 2009; Ponweiser, 2012).

Using the marginal distribution of a document, we can calculate the probability of a document (Blei *et al.*, 2001) as

$$p(w) = \int_{\theta}\left(\prod_{n=1}^{N}\sum_{z_n}p(z_n|\theta)p(w_n|z_n,\beta)\right)p(\theta|\alpha)d\theta \tag{2}$$

To put it together, if we take the product of all marginal probabilities of single documents, we get the total probability of a corpus which is defined according to (Blei *et al.*, 2001) as

$$p(D|\alpha,\beta) = \prod_{d=1}^{M}\int p(\theta_d|\alpha)\left(\prod_{n=1}^{N}\sum_{z_n}p(z_n|\theta_d)p(w_n|z_n,\beta)\right)d\theta_d \tag{3}$$

3.3 Topic modeling protocol
We performed the whole topic modeling process in the R programming language and the process consisted of the following phases: (1) dataset reduction (2) preprocessing, (3) corpus transformation, (4) topic modeling, (5) visualization of topics.

We performed topic modeling on a sample of abstracts from articles that we obtained following the procedure described in chapter 3.1. However, after initial analysis of the dataset, we found that 38 articles retrieved from the Web of Science and Scopus databases had no text defined in the abstract. Therefore, we decided not to take these articles into account.

Smart literature review was thus performed on a sample of 3,645 abstracts of articles related to DT in the tourism industry.

The second step was the creation of a corpus and data preprocessing. Within this phase, we performed standard procedures for text cleaning such as: retyping to lowercase letters, removing punctuation, removing numbers, removing extra spaces. In addition, in this phase we also deleted the so-called stopwords. They are meaningless words that do not add value to the topic modeling process. The reason for these steps was the fact that by removing such characters in the text, the entire process of text analysis and modeling becomes more efficient and faster, while the quality of the analysis remains preserved.

In addition to the standard preprocessing of our text corpus, we also analyzed the most frequently occurring words in the corpus of papers. Subsequently, we identified those words that have a general meaning in the field of DT in tourism. We also identified words [1] that we thought were not relevant to the field. We then removed the words identified in this way in the entire corpus. The reason was the higher quality of the subsequent topic modeling process.

Although it is a subjective choice of words that we decided to remove, we tried to proceed objectively. Although one of the reasons was that some words were unimportant, another, more important reason why we removed some words before the LDA analysis was that without this step, there would be biased results. LDA analysis takes into account words in individual documents and, based on this, classifies individual documents into groups (topics). Since there are words that appear in many documents, their frequency is often significantly higher than other relevant words. For that reason, these words are highly significant when creating topics. These are, for example, the words: "tourism", "digit", or "tourist". If we had not removed these words, it could have happened that the result of the LDA analysis would have been only 1 or 2 topics, since words like tourism or digit appeared in almost all documents and their relative importance in the creation of topics itself was high. In order to prevent our results from being distorted by words that appeared in almost all documents, we removed these words. We selected these words based on a previously performed frequency analysis of terms, when we analyzed the most frequently occurring words in the corpus of documents (abstracts).

In addition, based on term frequency analysis, we also removed other words before LDA modeling itself. In this case, these were words that were not relevant and important before the topic of DT in tourism itself. Examples of these words were: study, paper, research, result, method, approach, etc. As follows from the very nature of the words, these are generic words that appear in almost all scientific articles or abstracts. In this case too, these were words that appeared in a large number of abstracts. If we did not remove these words before modeling the topics, the relevant results would again be distorted. An example could be a topic where there would be articles in which documents with words such as paper, method, approach or analysis would appear. However, these words are generic words in scientific documents. In the case of creating topics based on these words, the result of dividing the documents into topics would not correspond to the topic of DT in tourism. The results themselves would not be relevant and correct for our research. We also determined the generic words that we removed based on the frequency analysis of the most numerous words in the corpus of abstracts. For greater credibility and objectivity, we have listed these words.

The next step was the transformation of the corpus. The reduced and modified data frame was subsequently transformed into a document-term matrix, which contained the frequencies of individual words in individual documents. The mentioned procedure is necessary for the correct implementation of topic modeling. The matrix that was created after the transformation had a size of 3,645 × 18,796. Since this matrix contained more than 68 million elements, we decided to reduce it. The reduction was made to increase the speed of the analysis. We performed the reduction in two ways. We set the minimum word frequency to three and the maximum frequency to the number of documents. In addition, we limited the minimum and maximum word length (minimum = 4, maximum = 20). The resulting matrix

had a dimension of $3{,}645 \times 5{,}149$ elements. With the matrix modified in this way, we then performed topic modeling process.

We performed LDA Topic modeling using the Gibbs sampling method (Griffiths and Steyvers, 2004; Grun and Hornik, 2011). We set the number of LDA modeling iterations to 1,000. In order not to distort the probability distribution, we abstracted from the first 100 iterations. Due to possible autocorrelation, we used every 40th observation for further calculation.

We used experimental testing to choose the optimal number of topics k. The entire LDA process was run five times for each k, and only the best solution was saved as the final result. In order to avoid a possible distortion of the results for all k, we always chose the same seeds (79; 68; 48; 11; 222). Subsequently, we used an expert approach to evaluate the created solutions. This procedure, which was based on human judgment, consisted in the manual assessment of the structure of the created topics for an individual number of topics. We gradually expertly evaluated the created solutions for the number of topics from 7 to 14. The solution with the number of eight topics was chosen as the final solution. The result of LDA modeling was the probabilistic division of all papers into individual topics.

We used alpha and beta default values in the topic models package to set our hyperparameters before LDA modeling. According to Griffiths and Steyvers (2004), the LDA implementations in the topic models package in R use a default value for alpha equal to 50/k, i.e. it is inversely proportional to the number of topics. As for the beta hyperparameter, denoted as delta by Griffiths and Steyvers (2004), the default value of 0.1 is used. The procedure for optimizing LDA parameters, i.e. for interring the posterior probabilities, was performed using Gibbs sampling.

Parameter optimization is implemented indirectly in the case of Gibbs sampling. The number of iterations for which Gibbs sampling of the LDA model parameter estimation is to be implemented is determined. Based on the selected criterion, it is subsequently checked whether the result is satisfactory. The criterion can be a selected statistical metric, e.g. perplexities. In our case, we focused on the criterion of interpretability of individual topics.

For Gibbs sampling, in our case, we chose 1,000 iterations since, according to several sources, this is a sufficiently high number of iterations. Obviously, after training for 5,000 iterations, we could get a slightly different solution with different topics. And that could be slightly different from the solution with 8,000 iterations. This is understandable since it is an unsupervised machine learning method that is used to detect unknown patterns. For the mentioned reason, it is not possible to compare the selected solution with the expected one, since there is no model for comparison. Based on the above, it can be concluded that individual solutions will probably differ, since there is no such thing as a correct answer in unsupervised learning. The recommended way to solve an unsupervised problem (such as topic modeling) is to experimentally test different values of k and according to the selected evaluation criterion to find the optimal grouping of topics in this way (Liu, 2015).

For a better interpretation and evaluation of individual solutions, we used the LDAvis library (Sievert and Shirley, 2014). The library enables the visualization of topics within the 2D space using the principal component map.

3.4 Topic modeling in papers related to pandemic
The input dataset for the analysis in section 4.3 was filtered as follows. We defined the search query as "Covid19" OR "Covid-19". Subsequently, we searched for the mentioned search query in the Title, Abstract and Author keywords. If a match was found in any of the listed variables, the given article received the Covid flag. This was done for all 3,645 articles.

Using this filtering method, we identified 321 articles related to the pandemic (29 articles were from 2020, 155 articles were from 2021 and 124 articles were from 2022.13 articles did not have a year specified). Covid-19 is related to the topic of DT in tourism. The abstracts of

these articles were subsequently entered into the LDA analysis, where we modeled individual new pandemic topics from these abstracts.

Before the modeling of the topics itself, we performed the transformation into a corpus and preprocessing of the text in a similar way as in the previous topic modeling. In addition to the words we defined above, we also removed the following words from the corpus of words: "covid19", "covid-19", "covid", "corona", "coronavirus", "sars-cov-2", "sarscov", "pandemic". The mentioned words were removed in order not to distort the results and the relevance of pandemic topics.

The LDA analysis itself was performed as follows: To quantify the parameters of the LDA model, we used the Gibbs sampling method. The LDA parameters were set the same as in the previous LDA modeling (n. of iterations: 1,000, burning period: 100, thin parameter: 40). As in the previous case, we performed LDA analysis with a different number of topics (k). For each k, we performed five runs of the algorithm, and the best solution was always saved as a result. We subsequently assessed these solutions from an expert point of view, while we preferred the interpretability of the topics to a statistical approach. When modeling abstracts related to Covid-19, we finally selected the number of topics k = 7 based on the mentioned procedure.

4. Results

A bibliometric approach provides detailed insight regarding the development of particular scientific fields. In our case, we processed 3,645 papers published between 2013 and 2022 related to DT in the tourism industry and retrieved them from Scopus and Web of Science databases. We processed the data with three RQs in mind. The following chapters contain the results of the basic bibliometric analysis, the results of topic modeling for the entire period and the results of topic modeling in the pandemic years 2020–2022.
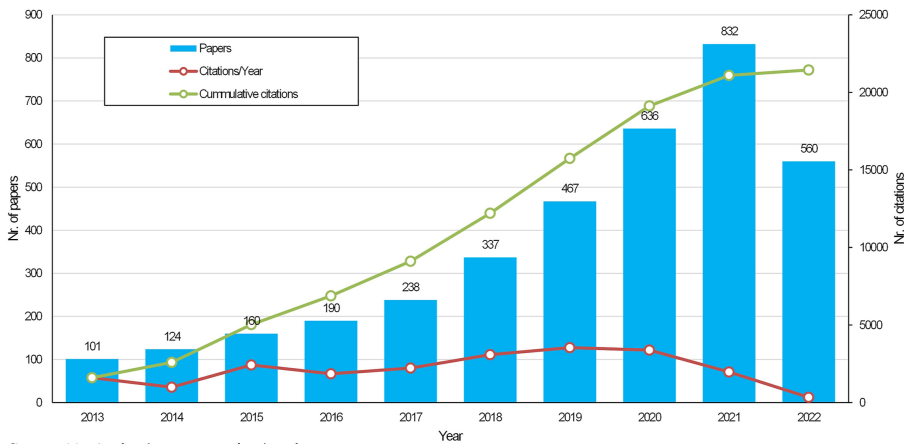
4.1 Development of academic papers related to DT in the tourism industry

The development of published papers related to digitization, digitalization, or DT in the tourism industry shows that the increase in the interest of scientists in this area is significant. Figure 2 provides an overview of the number of published articles over the last ten years, including citations and the cumulative number of annual citations.

The data was collected as of September 20, 2022, so it can be assumed that the number of papers in 2022 will still grow. Nevertheless, it is possible to see a clear trend of increasing scientific interest. The research impact of digitalization in tourism (measured through the number of citations) appears stable in the monitored period. However, it should be noted that, according to some similar bibliometric studies, the increase in citations becomes apparent only two to three years after a significant increase in the number of publications (Min et al., 2016). For example, in 2022, 560 papers were published, for which "only" 345 citations were recorded. Therefore, it is highly likely that the number of citations (and also research impact) will increase in later years.

Bibliographic analysis of journals provides a picture of outlets that have contributed most to the digitalization of tourism. We performed this analysis from three perspectives – research interest (number of papers), research impact (number of citations) and top cited papers. Table 1 shows the results of the analysis.

The journal with the highest research impact is Tourism management, in which 28 papers were published with more than 1,500 citations. The most significant study was by Hudson et al. (2015), with 272 citations. The study focuses on two research areas – the management of music festivals and the influence of social media on customer relationships. The authors developed a conceptual model by structural equation modeling. The authors concluded that

Source(s): Author's own creation/work

Journal	Published papers	Citations of papers	Top paper
Tourism Management	28	1,527	Hudson *et al.* (2015)
International Journal of Hospitality Management	11	712	Wang and Nicolau (2017)
Sustainability (Switzerland)	70	546	Encalada *et al.* (2017)
Journal of Destination Marketing and Management	8	534	Marine-Roig and Anton Clavé (2015)
Current Issues in Tourism	35	473	Jovicic (2017)
Annals of Tourism Research	18	442	Dolnicar (2019)
Journal of Sustainable Tourism	26	377	Gössling and Hall (2019)
Information Technology and Tourism	20	343	Pencarelli (2020)
Journal of Travel Research	14	331	Önder *et al.* (2016)
Tourism Management Perspectives	9	329	Xiang (2018)

Source(s): Author's own creation/work

interacting with the brand using social media had a direct effect on emotional attachment to the festival, and emotional attachment has a direct effect on word of mouth.

Eleven studies were published in the International Journal of Hospitality Management, while 712 citations to these studies were reported. The study with the highest impact was published by Wang and Nicolau (2017), which had 288 citations. In this study, the authors aimed to identify the price determinants of sharing economy based on accommodation offers in the digital marketplace. They find five categories of price predictors: host attributes, site and property attributes, amenities and services, rental rules and the number of online reviews and ratings.

The topic of digitization in the tourism industry was also relatively widely elaborated in the journal Sustainability, where up to 70 such focused papers with 546 citations were published. The most impactful study was conducted by Encalada *et al.* (2017) with 58 citations. In this study, the authors focused on digital imprints for identifying tourist places of interest. The authors demonstrated the potential of digital tools for sustainable smart city concepts and provided a practical perspective on this vital topic.

Another journal with a high research impact is the Journal of Destination Marketing and Management, in which only 8 studies were published, but they reached 534 citations. Marine-Roig and Anton-Clavé (2015) published a practically oriented study highlighting the usefulness of big data analytics to support smart destinations by studying the online image of a particular city (Barcelona was analyzed in this study). They concluded that user-generated content analytics should be essential for destination smartness. 201 citations were reported for this study.

Current Issues in Tourism is a reasonably popular journal for publishing digitalization topics in the tourism industry. In the last ten years, 35 papers were published in this journal, for which 473 citations were reported. 125 citations were reported on the paper by Jovicic (2017). The paper reviews the evolution of key tourism destination concepts. The author devotes special emphasis to the concept of smart tourism destinations since this is a recent concept that strongly relies on the systemic concept and represents an entirely different understanding of a destination than the traditional one.

Another journal with a relatively good research impact is the Annals of Tourism Research, which published 18 papers related to digital issues in tourism in the last decade. With 442 citations, this journal is sixth in research impact. The study with the most impact (106 citations) was published by Dolnicar (2019). The author developed a review study to create a knowledge map reflecting key areas of academic insight into the phenomenon of paid online peer-to-peer accommodation. She also identifies under-researched areas and under-utilized research designs in current literature.

Journal of Sustainable Tourism is another leading journal in the field of digitalization of the tourism industry. Twenty-six papers were published in this journal, and 377 citations were reported on these papers, whereas the study by Gössling and Hall (2019) has had the most impact – 96 citations. The paper investigates the ICT developments and conceptualizes the sharing economy in comparison to the wider collaborative economy.

Journal Information Technology and Tourism also contributed to developing digital issues in the tourism industry. 20 papers with 343 citations were published in this journal. The paper by Pencarelli (2020) was the one with the most impact – 76 citations. In this interesting paper, the author focused on providing a point of view and some preliminary answers to current questions related to the digital revolution in tourism. The author concluded that soon, it would not be possible for tourism ecosystems and territories to only take into account digital innovations. However, they will have to include smart tourism perspectives like sustainability, circular economy, quality of life, and social value. They should also aim to enhance tourism experiences and increase smart tourism destinations' competitive advantage.

The Journal of Travel Research published 14 studies with 331 citations. The study by Önder et al. (2016) reported 77 citations. In this study, the authors show the potential of the tourists' digital footprint to provide a valuable indicator for tourism demand. In addition, the authors showed practical analysis to provide vital marketing insight on data collection and data analysis.

The last of the top 10 journals is Tourism Management Perspectives. Nine studies were published in this journal with 329 citations. The study by Xiang (2018) focused on digitization, and the age of acceleration was the one with the highest number of citations – 80 citations. The author emphasizes the shift in our view of information technology in tourism research from a primarily marketing-driven tool to a knowledge creation tool due to new technological conditions such as the smartphone, drones, wearables, new connectivity and big data.

As seen from the above-mentioned overview of leading journals and papers, digitalization issues in the tourism industry cover several more or less related topics. To identify the main

research topics, we used a machine-learning text-mining approach. The results of this analysis are presented in the next chapter.

4.2 Dominant research topics and the development of research interest and research impact over time

Using LDA algorithms, we analyzed abstracts of 3,645 research papers. This approach compared the mutual occurrence of words in each paper, which made it possible to create groups of co-occurring words – i.e. topics. Each topic contains particular terms with different frequencies. One term could appear in multiple topics, however, with different frequencies. Therefore, we carried out several experiments to determine the correct number of topics. Using the algorithm, we analyzed the results for the number of topics 7, 8, 9, 10, 11, 12, 13 and 14. For each experiment, we subsequently assessed the statistical reliability of the results (measured by perplexity) and the interpretive reliability of the results (by expert evaluation of the relatedness of the words in the given topic). In this combined way, we determined the final number of topics. The best interpretable results with a sufficient level of perplexity were identified for eight topics, which are shown in Figure 3.

In the picture, you can see research interest and research impact of particular topics. The highest research interest was recorded for Topic-07. At the same time, however, there are two topics of similar size – Topic-02 and Topic-08, in which research interest and research impact are extremely large. The intertopic distance map represents a multidimensional scaling view of topics (the axis represents principal components). We can see that there is no obvious overlap between the topics, which indicates that the clustering procedure was able to extract unique topics. When naming topics, we considered the frequency of terms in particular topics. It should be noted that each topic contained terms such as "digital", "digitization", "digitalization", or "DT" and at the same time, the term "tourism" (these terms were excluded from LDA for a clearer extraction of topics). Based on the composition of terms and their occurrence, we named the topics as follows:
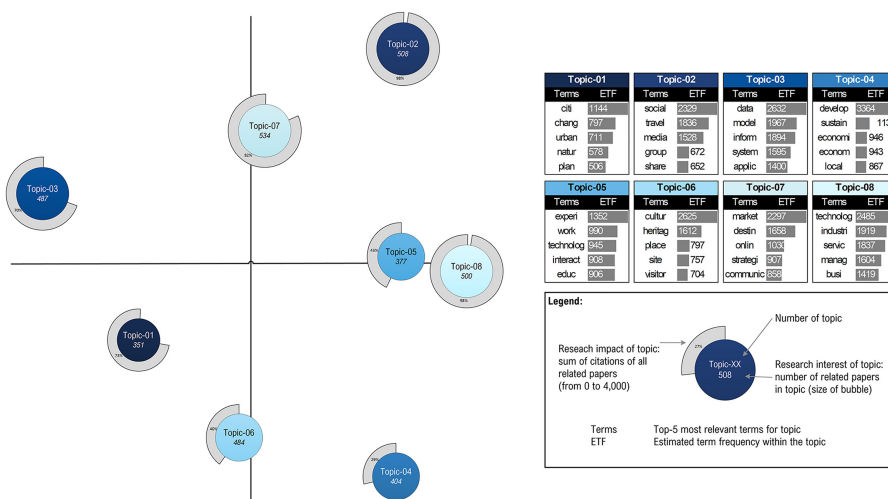
Source(s): Author's own creation/work

Figure 3.
Intertopic distance
map (left) and
frequency of terms in
particular topics (right)

(1) *City and urban planning (Topic-01).* This topic is represented by terms that are closely related to planning, namely "citi", "chang", "urban", "natur" and "plan". It is a relatively minor topic, but a relatively good research impact characterizes it. Within this topic, research focuses, for example, on cultural ecosystems (Martínez Pastur *et al.*, 2016), urban planning and formation of biocomfort (Cetin, 2019) or socioeconomic influences on biodiversity (Hou *et al.*, 2014).

(2) *Social media (Topic-02).* The content of this topic was relatively consistently made up of terms related to sharing experiences via social media. The most frequent terms were "social", "travel", "media", "group" and "share". We can see that this topic's research impact and interest are huge. Studies that belong to this topic focus, among other things, on social media research (Hudson *et al.*, 2015), sharing economy analysis (Cheng and Foley, 2018) or several aspects of social media marketing (Minazzi, 2015).

(3) *Data analytics (Topic-03).* The articles that belong to this topic are closely related to the use of various hardware and software technologies to support tourism. The most represented terms in this topic are "data", "model", "inform", "system" and "applic". It is a larger topic with a reasonable research impact. In this topic, we can come across studies devoted to technological solutions such as microblogging platforms (Mocanu *et al.*, 2013), blockchain (Bodkhe *et al.*, 2020) or tracking technologies (Shoval and Ahas, 2016).

(4) *Sustainable and economic development (Topic-04).* In this topic, the term "develop" dominates. After looking at other less frequently occurring terms, "sustain", "economi", "econom", and "local", we can see that this topic is closely related to sustainability. It is a topic with a medium research interest, but its research impact is relatively small compared to other topics. Research in this topic is closely related to ICT and their role in sustainable development goals in tourism (Gössling and Hall, 2019), smart cities (Kunzmann, 2020) or digital tourism (Watkins *et al.*, 2018).

(5) *Technology-based experience and interaction (Topic-05).* The composition of the articles in this topic suggests research in the field of consumer experience and interaction. This is also confirmed by the most frequent terms: "experi", "work", "technolog", "interact" and "educ". It is a small topic with less research impact. Research in this is often related to peer-to-peer concept (Dolnicar, 2019), new technologies such as virtual reality (Hudson *et al.*, 2019) or gamification in tourism (Skinner *et al.*, 2018).

(6) *Cultural heritage (Topic-06).* It is a reasonably consistent topic in terms of meaning, dominated by the terms "culture" and "heritag". The following three terms only specify the semantic definition of the topic – "place", "site", and "visitor". It is a medium-sized topic with a medium-sized research impact. In this topic, we can find articles focused on, for example, selfie-taking as touristic looking (Dinhopl and Gretzel, 2016), developing models for immersive heritage tourism experiences (Bec *et al.*, 2019) or the role of geoinformatics in the promotion of cultural heritage (Xiao *et al.*, 2018).

(7) *Digital destination marketing (Topic-07).* Even in this case, it is a content-unique topic, the most frequent terms of which are closely related – "market", "destin", "onlin", "strategi" and "communic". This is the biggest topic with high research interest. Research in this topic can cover multiple digital marketing issues. For example, electronic word-of-mouth (Sotiriadis and van Zyl, 2013), tourism analytics

(Marine-Roig and Anton-Clavé, 2015) or smart tourist destination research (Jovicic, 2019).

(8) *Smart tourism management (Topic-08).* This is a topic within which tourism frameworks related to digitalization issues are usually examined. The most frequently occurring terms are "technolog", "industry", "servic", "manag" and "busi". At the same time, it must be said that practically only in this topic was the term "smart", which, however, was in the seventh place in terms of occurrence, so it is not visible in the list of words. It is a big topic with practically the highest research impact. Research in this topic, for example, focuses on smart tourism ecosystems (Gretzel *et al.*, 2015), knowledge transfer in smart tourism destinations (Del Chiappa and Baggio, 2015) or smart aspects of urban mobility (Lyons, 2018).

Research interest and research impact of these eight topics are not static over time. Since we included each research paper in one of these topics, it was also possible to assess the development of individual topics over time. Figure 4 contains an overview of the development of eight topics over the monitored period (2013–2022).

It can be seen from the picture that the development of individual topics is really not static over time. In general, there are three trends – a stable topic, a growing topic and a declining topic. Social media (Topic-02) and Technology-based experience and interaction (Topic-05) can be included among stable topics that change relatively little over time. Social media have been a frequent research topic in the tourism industry for the last decade, and there has not been a leap shift in their principles that would lead to an increase or decrease in interest in the topic. However, it must be said that concerning the research impact (bottom of Figure 4), this topic is extremely interesting for the scientific community. The role of social media in tourism research will, therefore, likely remain significant. Technology-based experience and interaction is a topic that receives steady attention over time. ICTs, like social media, are connected practically to each research area; therefore, the stable interest in this topic is not surprising, even in tourism research. The research impact of this topic is still relatively small, although a slight increase can be observed since 2016. We can therefore assume that in the future, ICT in tourism industry research will still be a relevant topic for researchers.

Among the declining topics belong City and urban planning (Topic-01), Data analytics (Topic-03) and Cultural heritage (Topic-06). City and urban planning experienced a relatively significant decline over 10 years. While in 2013, more than 17% of all papers in tourism research were devoted to this topic, in 2022, it was only around 7%. We assume that this can be caused by the "shift" of attention to the topic of Smart tourism management, which has almost exactly the opposite trend. In City and urban planning, not only research interest is decreasing, but also research impact. So we can quite objectively believe that this topic will become a minority in the future and be replaced by other more current topics – e.g. smart tourism issues. Another declining topic is Data analytics, which has the most significant decline of all – from almost 30% in 2013 to less than 10% in 2022. Not only the number of published articles is decreasing, but also the number of citations to articles in this topic. This decline can probably be explained by the independent development of data analytics research. While in the past, data analytics tools were developed and researchers pointed out their application possibilities, nowadays, it can be seen that the topic has reached a certain limit of development, beyond which the potential for further research is limited. If an essential aspect of data analytics does not appear in the future, which has not yet been investigated in the tourism industry, we can expect that this topic will continue to decline gradually. The third decreasing topic is Cultural heritage. The decline of interest in this topic is visible,
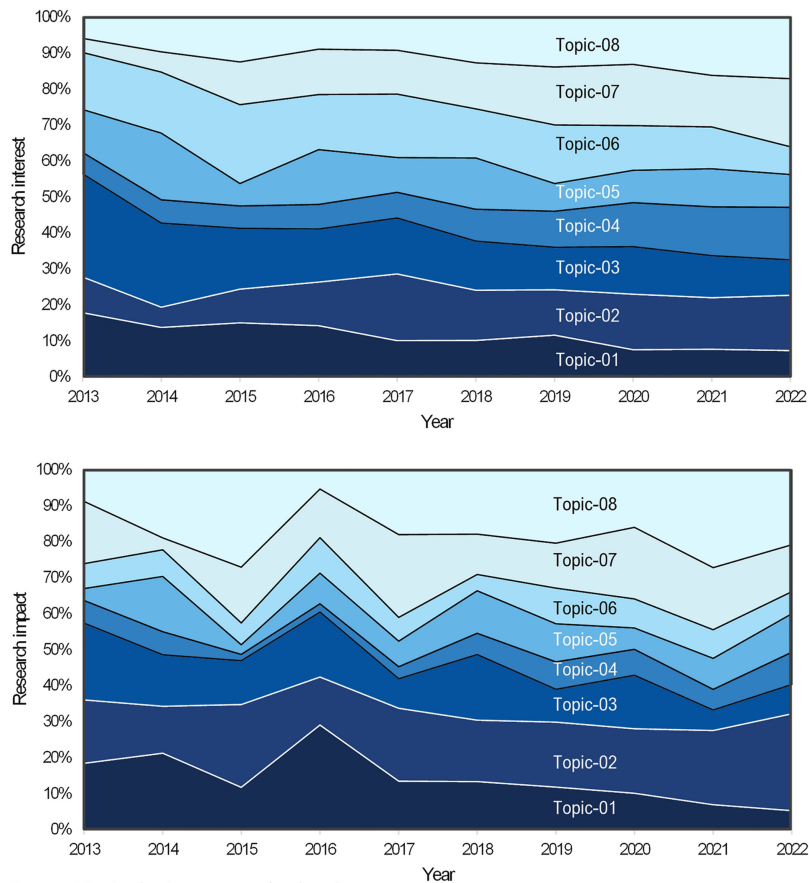
Figure 4.
Topics evolution –
research interest (top)
and research impact
(bottom)

Source(s): Author's own creation/work

especially in the last two years. Although the research impact remains stable, it is relatively small. Even in this case, we can expect a gradual decrease in interest in the given topic.

However, through time analysis, we also identified topics with a growing tendency. These are Sustainable and economic development (Topic-04), Digital destination marketing (Topic-07) and Smart tourism management (Topic-08). Sustainable and economic development is one of the long-term growing topics in general. Therefore, it is unsurprising that this topic was also identified with digitalization in the tourism industry. According to the trends in Figure 4, research interest in this topic is growing, while research impact has been growing to a lesser extent yet. Under the influence of various environmental and economic trends, we can expect that sustainable development issues will increase in intensity in tourism research in the future. Another topic with a growing tendency is Digital destination marketing. From the point of view of trends, this topic has grown dramatically – while in 2013, only 4% of all articles on digitalization in tourism were published in this area, in 2022, it was almost 20%. The research impact of this topic is relatively high and stable over time; therefore, we can expect that this area will also be attractive for a range of researchers in the future. The last growing topic is Smart tourism management. This topic has also seen a significant increase in

scientific interest among those with the greatest research impact (measured by the number of citations). Trends indicate that smart issues will play a key role in tourism research in the future.

4.3 Change of research topics under the influence of the pandemic
The Covid pandemic was a game-changer in practically all economic sectors. This was naturally reflected in the research on the tourism industry. Covid has built severe implications, as countries shut their borders to restrain the coronavirus spread (Albattat *et al.*, 2020; Chemli *et al.*, 2020; Connor, 2020). Many papers assert how the 2nd wave of Covid could look like, whereas destinations desire to relaunch tourism for the sake of economic survival (Kleczkowski, 2020; Valeri and Baggio, 2020a, 2020b). Afterward a crisis, tourism recovery is narrated to administrating public insight over intercommunication and details channels (Beirman, 2003). Essentially to have public sensibility about crises offering realistic assessments of potential risks without building emphasize and anguish (Boin and McConnell, 2007). Studies have additionally proved that media play a major part in destination image and tourists' intentions to visit (Gartner, 1993; Gartner and Shen, 1992; Govers *et al.*, 2007). As wander restrictions were introduced, details over media became a critical software for motivating individuals to take preventive measures (UNWTO, 2021). Correct details by the media can play down misinformation and reducing public anguish and fears. The media give details about public health problems as a catalyst to community insight of risk (Lin and Lagoe, 2013). Television headlines (Chang, 2012) as general media are intends to influence perceived risk. Many wander risk dimensions are studied (Lenggogeni *et al.*, 2019), containing disease risk and wander price tag and pains (Rittichainuwat and Chakraborty, 2009). Two major dimensions are severity and susceptibility (Pask and Rawlins, 2016). Severity mirrors the insight of how risky the disease might be (El-Toukhy, 2015), while susceptibility consults to one's insight of the probability of learning the disease.
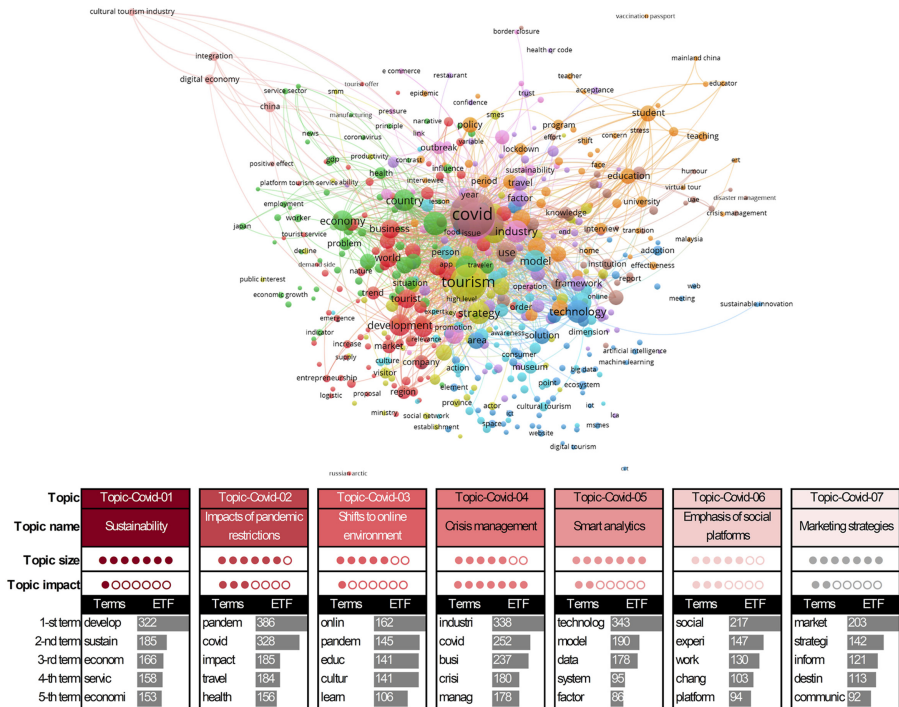
One of the ways to track the change that was caused by the pandemic is the possibility of modeling three discrete periods – before, during and after the pandemic period. However, since it is not entirely obvious how to divide the individual articles in terms of time, we decided on a different approach. In our approach, we modeled the development of individual topics continuously. We looked at how individual topics gradually developed over time. Perhaps the biggest advantage of this approach is the consistency of topics and the possibility of comparing the size and proportions of individual topics before, during and after the pandemic. If we were to model three separate discrete periods, this would not be possible, as individual groups of topics would differ from each other.

In addition, as part of RQ3, we also analyzed articles related to Covid-19 in the field of DT in tourism and modeled these into individual pandemic topics. The aim of this was to analyze in detail individual topics related only to the Covid-19 topic in this area.

The disadvantage of this second analysis was that the topics related to Covid-19 were different from the main topics. Therefore, we decided to perform a third analysis – an analysis of the connection between large (main) topics and topics from articles related to Covid-19. In this way, it is possible to see the relationships and interconnectedness of individual topics. Figure 5 shows the results in the form of co-occurrence maps and characteristics of individual topics.

Research on digitalization in the tourism industry was focused on seven topics at the time of the pandemic. These are the following topics, which were named following the frequency of occurrence of individual terms and the analysis of papers that fall into the given topic:

(1) *Sustainability (Covid Topic-01).* The topic focuses on issues related to sustainability and economic aspects of digitalization in tourism. This is very similar to the topic identified in the entire corpus of documents.

(2) *Impacts of pandemic restrictions (Covid Topic-02).* In this topic, the research focuses primarily on the negative impacts of the pandemic on the tourism industry.

(3) *Shifts to online environment (Covid Topic-03).* The articles in this topic mostly point to the possibilities or the need to move from a physical environment to an online environment. As a parallel to tourism, online teaching is often used in schools, which several countries had to apply.

(4) *Crisis management (Covid Topic-04).* This new topic focuses on ways to manage crisis situations, and the pandemic was undoubtedly one of them. The topic appeared naturally as a consequence of the strong need to solve the negative impacts that Covid-19 brought with it. It is a topic with the highest research impact.

Topic	Topic-Covid-01		Topic-Covid-02		Topic-Covid-03		Topic-Covid-04		Topic-Covid-05		Topic-Covid-06		Topic-Covid-07	
Topic name	Sustainability		Impacts of pandemic restrictions		Shifts to online environment		Crisis management		Smart analytics		Emphasis of social platforms		Marketing strategies	
Topic size	●●●●●●●		●●●●●●○		●●●●●○○		●●●●●○○		●●●●●●●		●●●●○○○		●●●●●●●	
Topic impact	●○○○○○○		●●○○○○○		●○○○○○○		●●●●●●●		●●○○○○○		●●○○○○○		●●○○○○○	
	Terms	ETF	Terms	ETF	Terms	ETF	Terms	ETF	Terms	ETF	Terms	ETF	Terms	ETF
1-st term	develop	322	pandem	386	onlin	162	industri	338	technolog	343	social	217	market	203
2-nd term	sustain	185	covid	328	pandem	145	covid	252	model	190	experi	147	strategi	142
3-rd term	econom	166	impact	185	educ	141	busi	237	data	178	work	130	inform	121
4-th term	servic	158	travel	184	cultur	141	crisi	180	system	95	chang	103	destin	113
5-th term	economi	153	health	156	learn	106	manag	178	factor	86	platform	94	communic	92

Note(s): Due to irrelevance, the following words were removed from the co-occurrence analysis: study, paper, research, analysis, article, approach, context, need, relationship, implication, case, methodology, number, literature, theory, concept, aspect, insight, form, field, author, addition, contribution, example, type, part, basis, show, chapter, focus, outcome, mean, relation, topic, line, series, highlight, view, document, fact, reason, hypothesis, kind, thing, word, lot
Source(s): Author's own creation/work

Figure 5.
Co-occurrence map of most frequent terms (top) and identified LDA topics (bottom)

(5) *Smart analytics (Covid Topic-05)*. This topic is closely related to smart solutions and their use in tourism during the pandemic. Among the most common terms are "technolog", "model", and "data", which are related to analytics, but the term "smart" itself appears only in this topic.

(6) *Emphasis of social platforms (Covid Topic-06)*. Articles in this topic generally cover the increasing importance of social media and related platforms and their role in the DT of the tourism industry.

(7) *Marketing strategies (Covid Topic-07)*. This is a relatively large topic focused on changes in the marketing aspects of tourism caused by the pandemic.

As can be seen from the description of research topics during the pandemic, some previously identified topics were preserved, but some new ones were also created. Figure 6 provides an overview of the interconnection between the original and new topics in Circos view (Krzywinski *et al.*, 2009).

The figure shows several relatively strong relationships between old and new topics. For example, if we look at the old Topic-04 (Sustainable and economic development), we see that it has split into two areas. The first is Topic-Covid-01 (Sustainability). It means that the topic of sustainability is dominant even in times of pandemic. However, the second is Topic-Covid-02 (Impacts of pandemic), which may indicate that the very concept of sustainability could be the basis for the direction of research focused on the impacts of the pandemic. Such a finding is interesting since the tourism sector was one of the most affected areas, and the pandemic revealed its vulnerability and opened the question of its sustainability. The relationship between old and new topics can also be seen in the case of Topic-05 (Technology-based experience and interaction) and Topic-Covid-03 (Shifts to online environment). This connection well illustrates the transit that occurred between a relatively isolated topic from the pre-pandemic period and a new topic covering new possibilities for using technologies in the pandemic period. From the figure, we can also see how Topic-07 (Digital destination marketing) has moved to practically all new topics, but above all to Topic-Covid-07 (Marketing strategies). This shift points to the potential of digital marketing in creating longer-term strategies. An interesting finding is that Topic-08 (Smart tourism management) moved the most to Topic-Covid-04 (Crisis management). Since the new topics also include Smart analytics (Topic-Covid-05), one could expect a shift mainly to this area. However, that did not happen, and it can be concluded that smart solutions developed in the past helped with crisis management during the pandemic.

5. Discussion
5.1 Summary
Our study focused on creating a science map of DT in the tourism industry. To this end, we formulated three RQs, the answers to which can be found in chapters 4.1 to 4.3. The most significant findings include the following:

(1) Research interest in digitization in the tourism industry is overgrowing. The number of published articles on this topic increases by approximately 25% annually.

(2) Top journals in tourism research fall in tourism areas e.g. Tourism management or International Journal of Hospitality Management.

(3) More than 3,500 research papers were devoted to digitization in tourism, while the research can be divided into eight separate topics: City and urban planning; Social media; Data analytics; Sustainable and economic development; Technology-based
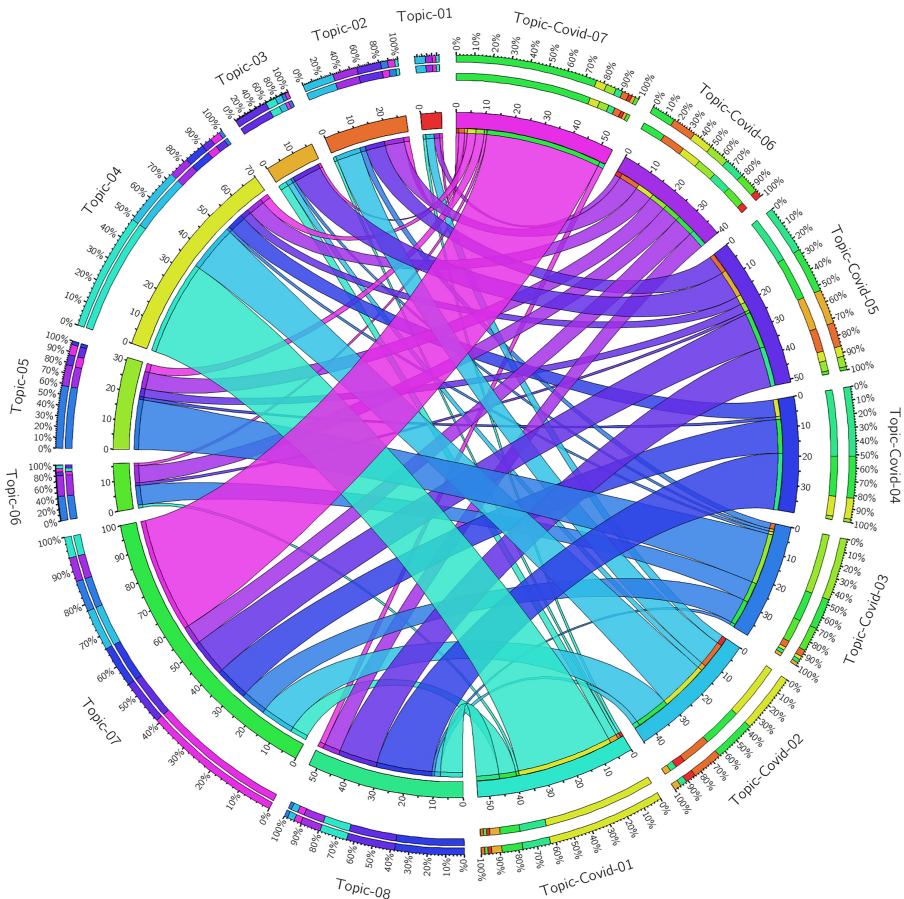
Figure 6.
Relationships between
topics

Source(s): Author's own creation/work

experience and interaction; Cultural heritage; Digital destination marketing; Smart
tourism management.

(4) Over the past ten years, the meaning of particular topics has changed. Research
interest decreases, especially in City and urban planning, Data analytics, and Cultural
heritage. On the other hand, the importance is growing in Sustainable and economic
development, Digital destination marketing, and Smart tourism management. Topics
Social media and Technology-based experience and interaction are among the stable
ones.

(5) In the pandemic years 2020–2022, several hundred studies were published, the
content of which can be divided into the following topics: Sustainability; Impacts of
pandemic; Shifts to online environment; Crisis management; Smart analytics;
Emphasis of social platforms and Marketing strategies.

5.2 Implications
The increase in research interest in digitization in the tourism industry is evident. The outcome of this study will influence the decision makers to understand the potentiality of DT in tourism industry. Though there are some shortcomings and risks associated with DT adoption, it will benefit the users in the long term. The competitive advantage of DT is much higher than traditional activities (Zheng *et al.*, 2022). In crisis times, DT represents an evolutionary process that leverages digital capabilities and technologies to create unique value for business models, operational procedures, and customer experiences (Jayawardana *et al.*, 2022).

Our results confirm the opinions of several authors who state that DT will strongly affect the tourism industry in the future (Pencarelli, 2020; Cuomo *et al.*, 2021). Our bibliometric study also showed that, as a rule, research is exclusively published in journals dedicated to the tourism industry. In a way, this finding is interesting, as one would expect that relevant scientific journals would also be those whose focus is on information or digital technologies. However, this is not the case, and digitalization in the tourism industry is still the domain of thematic tourism magazines Molina-Collado *et al.* (2022). However, we can assume that this state will not be permanent. The gradual adoption of digital technologies in practically all aspects of society will probably also cause more studies in scientific journals, the domain of which is not tourism but digital technologies.

Through the analysis of textual data, we have identified eight relatively unique topics that characterize research on digitization in the tourism industry. These topics are very diverse, which indicates that DT has several relevant but different aspects in tourism that are worth investigating (Pencarelli, 2020). Over the past ten years, the importance of some topics has changed significantly – it has grown or risen. Considering that the number of published articles related to digitization in the tourism industry grows every year, we can assume that the number of such unique topics will grow in the future. At the same time, as the number of topics increases, we assume that some currently less important topics will disappear. However, our analysis of trends also indicates that research interest may shift to more specific topics for some topics, where the research potential is more current. For example, we can report on the topic City and urban planning, from which the interest was gradually transferred to Smart tourism management. According to the development of trends simultaneously, it can be assumed that the most important topics will be Smart tourism management and Social media. In recent years, these topics showed a significant increase in research interest and, above all, in research impact.

The pandemic was a game-changer in the area of digitization in the tourism industry. Even though several topics researched so far have been preserved even during the pandemic, several new and unique research directions have emerged. In a way, it can be said that the pandemic has changed the structure of aspects of tourism's DT. For example, a completely new topic Crisis management was created, and it has an extremely high research interest among researchers. The pandemic also caused topics to split into more specific areas – e.g. Sustainable and economic development split into two topics: Sustainability and Impact of pandemic.

At the same time, our approach shows the potential of machine learning tools for processing unstructured text data. An indirect implication of our research is that identifying topics with such analytical tools can be effective and valuable in terms of content. Implementing a review study on a certain scientific topic does not have to be very laborious. Our results show that automated LDA procedures can reveal latent topics that would remain unrevealed by standard literature review procedures. For example, very few review studies identified the topic Cultural heritage (Bec *et al.*), while the LDA algorithms identified it very unambiguously. This indicates that text mining tools are becoming relevant tools for literature review processing.

5.3 Limitations and future research

Several research limitations can also be identified in our research. The first limitation concerns the fact that our results are not at the same qualitative level as the results of a standard systematic literature review (SLR). It is obvious that the systematic literature review is at a higher quality level as each paper is analyzed by human judgment and manually. However, this procedure is very time-consuming and does not allow the analysis of a large set of documents. Our approach based on machine learning makes it possible to analyze a huge number of articles and at the same time to identify valuable and insightful information that would not be possible to detect through bibliometric analysis.

Another limitation concerns the possible irrelevance of some papers that were selected in our sample. Because we analyzed a large number of articles, it was not possible to manually assess each article for its relevance to the topic. However, we believe that the defined search query provides a strong enough reason to believe that the set of articles selected for analysis is sufficiently and highly relevant.

Our goal was to analyze all articles in the Scopus and Web of Science databases that are related to DT in the tourism industry over the last 10 years. We set the relevant search query defined according to Verhoef *et al.* (2021). In addition, we also defined the word "*digital*" in the search query. However, despite this fact, there is still a small probability that some articles from this area were not included in our analysis. However, we assume that this is such a small and negligible percentage of documents that it does not fundamentally affect our results.

An important aspect of topic modeling is setting the appropriate number of topics. From a mathematical point of view, it is an optimization task that can be solved by various criterion functions (e.g. minimization of perplexity, maximization of topic coherence). In our case, however, we did not apply a statistical approach to solve this problem. We used an expert approach based on manual human judgment of topics. Although it is possible that the given number of topics will not be optimal from a statistical point of view, a much more important criterion for us was the explainability of the topics and their interpretability.

Certain limitations also result from the topic naming process. Topic is a mixture of a large number of words, and therefore, it is not possible to name a topic with two or three words in such a way to describe its essence completely accurately. However, our effort was to take into account the most numerous words in the topics and choose the name in such a way that the topic name would reflect the essence of the topic from a holistic point of view.

The answers to RQ2 and RQ3 are the result of topic modeling analyzes performed on the text corpus, which consisted of abstracts of scientific articles extracted from the Web of Science and Scopus databases. It is obvious that by analyzing the entire articles, we would achieve a higher credibility of the results. However, this would significantly increase the processing time. Moreover, the added value is not guaranteed since the abstract is a representative sample of the article. Therefore, we assume that the results would not change in a fundamental way. On the other hand, it was possible to approach topic modeling in several ways. For RQ3, we divided the dataset according to the keyword Covid-19. We also considered dividing the dataset according to years into two or three periods – pre-pandemic, pandemic (and post-pandemic). However, there is still no established scientific consensus in the literature regarding the end of the pandemic. It would not be possible to reliably determine which studies already belong to the post-pandemic period (also taking into account the time between the study preparation and its publication may be longer). By dividing the dataset into time periods, we could only compare the given topics subjectively, not statistically, since the articles would be assessed separately. Each of the mentioned paths to topic modeling has its advantages and disadvantages. After careful consideration, we chose an approach based on analyzing papers containing the word Covid-19.

Our approach also carries with it a logical limitation. Our study is bibliometric and aims to cover the widest possible scope of research on DT in tourism. Since we processed a large

number of articles, it was not possible to analyze all of them in detail, as is the case with studies of the systematic literature review type. Although this is a limitation, the topics that have been identified can be analyzed later through a systematic literature review.

There are several possibilities for further research that our study implies. First of all, it is possible to expand the given analysis with other scientific databases (except Scopus and Web of Science), or with gray literature – reports, student theses, projects outputs, etc. The dataset can also be expanded by adding additional words to the search query that are directly or indirectly related to digitization in the tourism industry. After a certain period, when it will be possible to say with certainty that we are living in a post-pandemic era, it would be possible to analyze the development of topics that appeared before, during and after the pandemic. To realize this, however, it is necessary to have enough studies in the dataset that would fall into the post-Covid period. Our research identified several rapidly developing topics such as Smart tourism management or Social media. Examining directions and areas of research can represent a vital opportunity for researchers in these topics. The bibliometric analysis that we have carried out can also be extended by other types of analysis, such as co-authorship, countries analysis, keywords analysis, etc.

Crisis management is a relatively interesting area for future research on digitization in the tourism industry. This topic appeared only in recent years and its research impact is huge. This opens up a whole range of research possibilities for a better understanding of the role of digital technologies in tourism. Last but not least, it would be interesting to observe longer-term trends in individual topics. The analysis of topics showed that the pandemic changed the direction of topics. Whether the focus on a new emerging topic is only temporary or whether it is an unchanging state can be one of the interesting and challenging future research opportunities.

Note

1. "tourism", "digit", "studi", "research", "tourist", "paper", "result", "provid", "analysi", "also", "base", "area", "increas", "present", "sector", "method", "approach", "author", "need", "will", "find", "main", "show", "well", "articl", "focus", "potenti", "identifi", "content", "propos", "signific", "creat", "contribut", "analyz", "purpos", "role", "becom", "improv", "futur", "implement", "context", "field", "current", "understand", "make", "explor", "year", "case", "howev", "build", "examin", "right", "within", "high", "limit", "concept", "adopt", "possibl", "allow", "various", "investig", "methodolog", "first", "part", "indic", "particular", "conduct", "analys", "three", "type", "springer", "number", "exist", "help", "mean", "enhanc", "point", "suggest", "recent", "term", "among", "implic", "determin", "sourc", "issu", "aspect", "addit", "follow", "featur", "take", "reveal", "toward", "involv", "therefor", "repres", "avail", "Switzerland", "relev", "obtain", "associ", "establish", "achiev", "highlight", "characterist", "state", "general", "especi", "play", "function", "view", "recommend", "thus", "better", "describ", "found", "enabl", "exampl", "theoret", "accord", "larg", "initi", "intent", "made", "scientif", "address", "theori", "live", "final", "sinc", "demonstr", "element", "like", "rapid", "period", "real", "conceptu", "chapter", "topic", "advantag", "question", "reflect", "introduc", "document", "refer", "around", "small", "necessari", "insight", "special", "empir", "defin", "great", "situat", "ieee", "along", "variabl", "four", "statist", "still", "subject", "attent", "scale", "common", "concern", "even", "european", "interpret", "china", "second", "last", "face", "conclus", "given", "consist", "comprehens", "report", "past", "account", "higher", "good", "previous", "languag", "name", "across", "total", "originalityvalu", "represent", "start", "index", "bring", "often", "give", "taylor", "although", "essenti", "must", "furthermor", "conclud", "theme", "unit", "fact", "best", "incorpor", "idea", "narrat", "gain", "moreov", "express", "elsevi", "analyt", "seek", "encourag", "mine", "similar", "organis

References

Adhikari, J., Timsina, J., Khadka, S.R., Ghale, Y. and Ojha, H.R. (2020), "COVID-19 impacts on agriculture and food systems in Nepal: implications for SDGs", *Agricultural Systems*, Vol. 186, p. 102990.

Akhtar, N., Khan, N., Mahroof Khan, M., Ashraf, S., Hashmi, M.S., Khan, M.M. and Hishan, S.S. (2021), "Post-COVID 19 tourism: will digital tourism replace mass tourism?", *Sustainability*, Vol. 13 No. 10, p. 5352.

Albattat, A., Jamaludin, A., Zuraimi, N.S.M. and Valeri, M. (2020), "Visit intention and destination image in post- covid- 19 crisis recovery", *Current Issues in Tourism*, Vol. 24 No. 17, pp. 2392-2397, doi: 10.1080/13683500.2020.1842342.

Alkhatib, A. and Valeri, M. (2022), "Can intellectual capital promote the competitive advantage? Service innovation and big data analytics capabilities in a moderated mediation model", *European Journal of Innovation Management*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/EJIM-04-2022-0186.

Bastidas-Manzano, A., Sánchez-Fernández, J. and Casado-Aranda, L. (2020), "The past, present, and future of smart tourism destinations: a bibliometric analysis", *Journal of Hospitality and Tourism Research*, Vol. 45, pp. 529-552.

Bec, A., Moyle, B., Timms, K., Schaffer, V., Skavronskaya, L. and Little, C. (2019), "Management of immersive heritage tourism experiences: a conceptual model", *Tourism Management*, Vol. 71, pp. 117-120.

Beirman, D. (2003), *Restoring Tourism in Crisis: A Strategic Marketing Approach*, Cambridge, CABI Publishing.

Blei, D.M. and Lafferty, J.D. (2009), "Topic models", in Sahami, A. and Srivastava, M. (Eds), *Text Mining: Theory and Applications*", pp. 71-93, Taylor & Francis.

Blei, D., Ng, A. and Jordan, M. (2001), "Latent Dirichlet allocation", *Advances in Neural Information Processing Systems*, Vol. 14, pp. 601-608.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.

Bodkhe, U., Tanwar, S., Parekh, K., Khanpara, P., Tyagi, S., Kumar, N. and Alazab, M. (2020), "Blockchain for Industry 4.0: a comprehensive review", *IEEE Access*, Vol. 8, pp. 79764-79800.

Boin, A. and McConnell, A. (2007), "Preparing for critical infrastructure breakdowns: the limits of crisis management and the need for resilience", *Journal of Contingencies and Crisis Management*, Vol. 15 No. 1, pp. 50-59.

Botti, A. and Monda, A. (2021), *ICT Platforms and the Restaurant Business in Times of Covid-19: The Case of Italy*, RIIFORUM, Athens, pp. 393-401.

Cetin, M. (2019), "The effect of urban planning on urban formations determining bioclimatic comfort area's effect using satellitia imagines on air quality: a case study of Bursa city", *Air Quality, Atmosphere and Health*, Vol. 12, pp. 1237-1249.

Chang, C. (2012), "News coverage of health-related issues and its impacts on perceptions: taiwan as an example", *Health Communication*, Vol. 27 No. 2, pp. 111-123, doi: 10.1080/10410236.2011.569004.

Chatterjee, S., Chaudhuri, R., Vrontis, D. and Jabeen, F. (2022), "Digital transformation of organization using AI-CRM: from microfoundational perspective with leadership support", *Journal of Business Research*, Vol. 153, pp. 46-58.

Chemli, S., Toanoglou, M. and Valeri, M. (2020), "The impact of Covid-19 media coverage on tourist's awareness for future traveling", *Current Issues in Tourism*, Vol. 25 No. 2, pp. 179-186, doi: 10.1080/13683500.2020.1846502.

Chen, S., Tian, D., Law, R. and Zhang, M. (2021), "Bibliometric and visualized review of smart tourism research", *International Journal of Tourism Research*, Vol. 24 No. 2, pp. 298-307.

Cheng, M. and Foley, C. (2018), "The sharing economy and digital discrimination: the case of Airbnb", *International Journal of Hospitality Management*, Vol. 70, pp. 95-98.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2011), "Science mapping software tools: review, analysis, and cooperative study among tools", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 7, pp. 1382-1402.

Connor, P. (2020), More than nine-in-ten people worldwide live in countries with travel restrictions amid COVID-19, available at: https://www.pewresearch.org/?p=361112

Cuomo, M.T., Tortora, D., Foroudi, P., Giordano, A., Festa, G. and Metallo, G. (2021), "Digital transformation and tourist experience co-design: big social data for planning cultural tourism", *Technological Forecasting and Social Change*, Vol. 162, 120345.

Deb, S.K., Nafi, S. and Valeri, M. (2022), "Promoting tourism business through digital marketing in the new normal era: a sustainable approach", *European Journal of Innovation Management*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/EJIM-04-2022-0218.

Del Chiappa, G. and Baggio, R. (2015), "Knowledge transfer in smart tourism destinations: analyzing the effects of a network structure", *Journal of Destination Marketing and Management*, Vol. 4 No. 3, pp. 145-150.

Dinhopl, A. and Gretzel, U. (2016), "Selfie-taking as touristic looking", *Annals of Tourism Research*, Vol. 47, pp. 126-139.

Dolnicar, S. (2019), "A review of research into paid online peer-to-peer accommodation Launching the Annals of Tourism Research curated collection on peer-to-peer accommodation", *Annals of Tourism Research*, Vol. 75, pp. 248-264.

Eck, V.N.J. and Waltman, L. (2010), "Software survey: VOSviewer, a computer program for bibliometric mapping", *Scientometrics*, Vol. 84 No. 2, pp. 523-538.

El-Toukhy, S. (2015), "Parsing susceptibility and severity dimensions of health risk perceptions", *Journal of Health Communication*, Vol. 20 No. 5, pp. 499-511, doi: 10.1080/10810730.2014.989342.

Encalada, L., Boavida-Portugal, I, Ferreira, C. and Rocha, J. (2017), "Identifying tourist places of interest based on digital imprints: towards a sustainable smart City", *Sustainability (Switzerland)*, Vol. 62, pp. 120-131.

Fontanari, M. and Traskevich, A. (2022 In press), "Smart-solutions for handling overtourism and developing destination resilience for the post-covid-19 era", *Tourism Planning and Development*, doi: 10.1080/21568316.2022.2056234.

Frank, A.G., Dalenogare, L.S. and Ayala, N.F. (2019), "Industry 4.0 technologies: implementation patterns in manufacturing companies", *International Journal of Production Economics*, Vol. 210, pp. 15-26.

Gartner, W.C. (1993), "Image formation process", in Uysal, M. and Fesenmaier, D. (Eds), *Communication and Channel Systems in Tourism Marketing*, pp. 191-215, Haworth Press, New York.

Gartner, W. and Shen, J. (1992), "The impact of tiananmen square on China's tourism image", *Journal of Travel Research*, Vol. 30 No. 4, pp. 47-52.

Gazi, Z.A., Saner, T., Bahcelerli, N.M. and Aksal, F.A. (2016), "The role of social media tools: accessible tourism for disabled citizens", *Educational Technology and Society*, Vol. 19 No. 1, pp. 89-99.

Gómez, L.M., Fernández, L.R. and Navío-Marco, J. (2018), "Application of communication technologies (ICT) within the tourism industry in the European Union", *Tourism: An International Interdisciplinary Journal*, Vol. 66 No. 2, pp. 239-245.

Gössling, S. and Hall, M.C. (2019), "Sharing versus collaborative economy: how to align ICT developments and the SDGs in tourism?", *Journal of Sustainable Tourism*, Vol. 30 No. 5, pp. 1080-1101.

Gössling, S., Scott, D. and Hall, C.M. (2020), "Pandemics, tourism and global change: a rapid assessment of COVID-19", *Journal of Sustainable Tourism*, Vol. 29, pp. 1-20.

Govers, R., Go, F.M. and Kumar, K. (2007), "Promoting tourism destination image", *Journal of Travel Research*, Vol. 46 No. 1, pp. 15-23.

Gretzel, U., Werthner, H., Koo, C. and Lamsfus, C. (2015), "Conceptual foundations for understanding smart tourism ecosystems", *Computers in Human Behavior*, Vol. 50, pp. 558-563.

Griffiths, T.L. and Steyvers, M. (2004), Finding scientific topics, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, pp. 5228-5235.

Grun, B. and Hornik, K. (2011), "Topicmodels: an R package for fitting topic models", *Journal of Statistical Software*, Vol. 40 No. 13, pp. 1-30.

Gyódi, K. (2019), "Airbnb in European cities: business as usual or true sharing economy?", *Journal of Cleaner Production*, Vol. 221, pp. 536-551.

Han, X. (2020), "Evolution of research topics in LIS between 1996 and 2019: an analysis based on latent Dirichlet allocation topic model", *Scientometrics*, Vol. 125 No. 3, pp. 2561-2595.

Harel, R. (2021), "The impact of COVID-19 on small businesses' performance and innovation", *Global Business Review*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1177/09721509211039145.

He, H. and Harris, L.C. (2020), "The impact of Covid-19 pandemic on corporate social responsibility and marketing philosophy", *Journal of Business Research*, Vol. 116, pp. 176-182.

Hinings, B., Gegenhuber, T. and Greenwood, R. (2018), "Digital innovation and transformation: an institutional perspective", *Information and Organization*, Vol. 28 No. 1, pp. 52-61.

Hou, Y., Zhou, S., Burkhard, B. and Muller, F. (2014), "Socioeconomic influences on biodiversity, ecosystem services and human well-being: a quantitative application of the DPSIR model in Jiangsu, China", *The Science of the Total Environment*, Vol. 490, pp. 1012-1028.

Hudson, S., Roth, M.S., Madden, T.J. and Hudson, R. (2015), "The effects of social media on emotions, brand relationship quality, and word of mouth: an empirical study of music festival attendees", *Tourism Management*, Vol. 47, pp. 68-76.

Hudson, S., Matson-Barkat, S., Pallamin, N. and Jegou, G. (2019), "With or without you? Interaction and immersion in a virtual reality experience", *Journal of Business Research*, Vol. 100, pp. 459-468.

Iansiti, M. and Lakhani, K.R. (2014), "Digital ubiquity: how connections, sensors, and data are revolutionizing business", *Harvard Business Review*, Vol. 92 No. 11, pp. 90-99.

Jayawardana, C., Albattat, A., Valeri, M. and Jaharadak, A.A. (2022), "Technology acceptance antecedents in digital transformation on hospitality industry", *International Journal of Hospitality Management*, Vol. 108 January 2023, doi: 10.1016/j.ijhm.2022.103350.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L. (2019), "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey", *Multimedia Tools and Applications*, Vol. 78, pp. 15169-15211.

Jovicic, D.Z. (2017), "From the traditional understanding of tourism destination to the smart tourism destination", *Current Issues in Tourism*, Vol. 22 No. 3, pp. 276-282.

Jovicic, D.Z. (2019), "From the traditional understanding of tourism destination to the smart tourism destination", *Current Issues in Tourism*, Vol. 22 No. 3, pp. 276-282.

Kane, G.C., Palmer, D., Philips, A.N., Kiron, D. and Buckley, N. (2015), "Strategy, not technology, drives digital transformation", *MIT Sloan Management Review*, Vol. 14, pp. 1-25.

Keil, D., Höpken, W., Fuchs, M. and Lexhagen, M. (2017), "Optimizing user interface design and interaction paths for a destination management information system", *Design, User Experience, and Usability: Understanding Users and Contexts*.

Kleczkowski, A. (2020), "These Scenarios what a Second Wave of Covid-19 Could Look like", *ScienceAlert*, available at: https://www.sciencealert.com/here-s-what-a-second-wave-of-covid-19-could-look-like (accessed 20 June 2020).

Knani, M., Echchakoui, S. and Ladhari, R. (2022), "Artificial intelligence in tourism and hospitality: bibliometric analysis and research agenda", *International Journal of Hospitality Management*, Vol. 107, 103317.

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009), "Circos: an information aesthetic for comparative genomics", *Genome Research*, Vol. 19 No. 9, pp. 1639-1645.

Kunzmann, K.R. (2020), "Smart cities after covid-19: ten narratives", *disP – The Planning Review*, Vol. 56 No. 2, pp. 20-31.

Lenggogeni, S., Ritchie, B.W. and Slaughter, L. (2019), "Understanding travel risks in a developing country: a bottom up approach", *Journal of Travel and Tourism Marketing*, Vol. 36 No. 8, pp. 941-955, doi: 10.1080/10548408.2019.1661329.

Li, X. and Law, R. (2020), "Network analysis of big data research in tourism", *Tourism Management Perspectives*, Vol. 33, 100608.

Li, F., Nucciarelli, A., Roden, S. and Graham, G. (2016), "How smart cities transform operations models: a new research agenda for operations management in the digital economy", *Production Planning and Control*, Vol. 27 No. 6, pp. 514-528.

Lin, C.A. and Lagoe, C. (2013), "Effects of news media and interpersonal interactions on H1N1 risk perception and vaccination intent", *Communication Research Reports*, Vol. 30 No. 2, pp. 127-136, doi: 10.1080/08824096.2012.762907.

Liu, E. (2015), "Latent dirichlet allocation using Gibbs sampling", available at: https://ethen8181.github.io/machine-learning/clustering_old/topic_model/LDA.html

Lončarić, D., Popović, P. and Kapeš, J. (2022), "Impact of the COVID-19 pandemic on tourism: a systematic literature review", *Tourism: An International Interdisciplinary Journal*, Vol. 70 No. 3, pp. 512-526.

Lyons, G. (2018), "Getting smart about urban mobility – aligning the paradigms of smart and sustainable", *Transportation Research Part A: Policy and Practice*, Vol. 115, pp. 4-14.

Mariani, M. and Baggio, R. (2022), "Big data and analytics in hospitality and tourism: a systematic literature review", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 1, pp. 231-278.

Marine-Roig, E. and Anton Clavé, S. (2015), "Tourism analytics with massive user-generated content: a case study of Barcelona", *Journal of Destination Marketing and Management*, Vol. 4 No. 3, pp. 162-172.

Martínez Pastur, G., Peri, P.L., Lencinas, M.V., García-Llorente, M. and Martín-López, B. (2016), "Spatial patterns of cultural ecosystem services provision in Southern Patagonia", *Landscape Ecology*, Vol. 31 No. 2, pp. 383-399.

Matt, C., Hess, T., Benlian, A. and Wiesbock, F. (2016), "Options for formulating a digital transformation strategy", *MIS Quarterly Executive*, Vol. 15 No. 2, pp. 123-139.

Min, C., Sun, J., Pei, L. and Ding, Y. (2016), "Measuring delayed recognition for papers: uneven weighted summation and total citations", *Journal of Informetrics*, Vol. 10 No. 4, pp. 1153-1165.

Minazzi, R. (2015), "Social media marketing in tourism and hospitality", *Social Media Marketing in Tourism and Hospitality*, Springer, Berlin.

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. and Vespignani, A. (2013), "The twitter of babel: mapping world languages through microblogging platforms", *PLoS ONE*, Vol. 8 No. 4, p. e61981.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M, Shekelle, P. and Stewart, L.A., PRISMA-P Group (2015), "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement", *Systematic Reviews*, Vol. 4 No. 1, p. 1, doi: 10.1186/2046-4053-4-1.

Molina-Collado, A., Gómez-Rico, M.F., Sigala, M., Molina, M.V., Aranda, E. and Salinero, Y. (2022), "Mapping tourism and hospitality research on information and communication technology: a bibliometric and scientific approach", *Information Technology and Tourism*, Vol. 24, pp. 299-340.

Nabila, A.N., Nahdliyah, T., Sensuse, D.I., Suryono, R.R. and Kautsarina (2021), "Collaborative system implementation for tourism: a systematic literature review", *International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 255-262.

Ndou, V., Mele, G., Hysa, E. and Manta, O. (2022), "Exploiting technology to deal with the COVID-19 challenges in travel and tourism: a bibliometric analysis", *Sustainability*, Vol. 14 No. 10, p. 1597.

Ntounis, N., Parker, C., Skinner, H., Steadman, C. and Warnaby, G. (2021), "Tourism and hospitality industry resilience during the covid-19 pandemic: evidence from england", *Current Issues in Tourism*, Vol. 25, pp. 46-59.

Ohe, Y. (2022), "Rural tourism under the new normal: new potentials from a Japanese perspective", *WIT Transactions on Ecology and the Environment*, Vol. 256, pp. 51-62.

Önder, I., Koerbitz, W. and Hubmann-Haidvogel, A. (2016), "Tracing tourists by their digital footprints: the case of Austria", *Journal of Travel Research*, Vol. 55 No. 5, pp. 566-573.

Pagani, M. and Pardo, C. (2017), "The impact of digital technology on relationships in a business network", *Industrial Marketing Management*, Vol. 67, pp. 185-192.

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, E., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and McKenzie, J.E. (2021), "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", *BMJ 2021*, Vol. 372 No. 71, doi: 10.1136/bmj.n71.

Pask, E.B. and Rawlins, S.T. (2016), "Men's intentions to engage in behaviors to protect against human papillomavirus (HPV): testing the risk perception attitude framework", *Health Communication*, Vol. 31 No. 2, pp. 139-149.

Pencarelli, T. (2020), "The digital revolution in the travel and tourism industry", *Information Technology and Tourism*, Vol. 22, pp. 455-476.

Perry, E.E., Xiao, X., Iretskaia, T.A., Li, P., Manning, R.E., Valliere, W.A. and Reigner, N. (2022), "A review of digitalization and sustainability in parks and recreation indicators and thresholds research", *Journal of Outdoor Recreation and Tourism*, Vol. 39, 100550.

Podzharaya, N. and Sochenkova, A. (2022), "Growth of modern technologies usage in tourism due to COVID-19: examples and forecasts", *11th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1-6.

Ponweiser, M. (2012), "Latent dirichlet allocation in R".

Prayag, G. (2020), "Time for reset? Covid-19 and tourism resilience", *Tourism Review International*, Vol. 24, pp. 179-184.

Rana, R.L., Adamashvili, N. and Tricase, C. (2022), "The impact of blockchain technology adoption on tourism industry: a systematic literature review", *Sustainability*, Vol. 14, p. 7383.

Rittichainuwat, B.N. and Chakraborty, G. (2009), "Perceived travel risks regarding terrorism and disease: the case of Thailand", *Tourism Management*, Vol. 30 No. 3, pp. 410-418.

Shoval, N. and Ahas, R. (2016), "The use of tracking technologies in tourism research: the first decade", *Tourism Geographies*, Vol. 25 No. 2, pp. 264-286.

Sievert, C. and Shirley, K. (2014), "LDAvis: a method for visualizing and interpreting topic" *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Association for Computational Linguistics Note, pp. 63-70.

Sigala, M. (2020), "Tourism and COVID-19: impacts and implications for advancing and resetting industry and research", *Journal of Business Research*, Vol. 117, pp. 312-321.

Singh, S. and Bashar, A. (2021), "A bibliometric review on the development in e-tourism research", *International Hospitality Review*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/IHR-03-2021-0015.

Škare, M., Soriano, D.R. and Porada-Rochoń, M. (2021), "Impact of COVID-19 on the travel and tourism industry", *Technological Forecasting and Social Change*, Vol. 163, p. 120469.

Skinner, H., Sarpong, D. and White, G.R.T. (2018), "Meeting the needs of the Millennials and Generation Z: gamification in tourism through geocaching", *Journal of Tourism Futures*, Vol. 4 No. 1, pp. 93-104.

Sotiriadis, M.D. and van Zyl, C. (2013), "Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists", *Electronic Commerce Research*, Vol. 13, pp. 103-124.

Toubes, D.R., Araújo Vila, N. and Fraiz Brea, J.A. (2021), "Changes in consumption patterns and tourist promotion after the COVID-19 pandemic", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 16 No. 5, pp. 1332-1352.

United Nations World Tourism Organization (2021), "COVID-19 and tourism 2020: a year in review".

UNWTO (United Nations World Tourism Organization) (2021), "COVID-19 and tourism 2020: a year in review".

Valeri, M. and Baggio, R. (2020a), "Italian tourism intermediaries: a social network analysis exploration", *Current Issues in Tourism*, Vol. 24 No. 4, pp. 1-14, doi: 10.1080/13683500.2020.1777950.

Valeri, M. and Baggio, R. (2020b), "Social network analysis: organizational implications in tourism management", *International Journal of Organizational Analysis*, Vol. 29 No. 2, pp. 1-12, doi: 10.1108/IJOA-12-2019-1971.

Verhoef, P.C., Broekhuizen, T.L., Bart, Y., Bhattacharya, A., Qi Dong, J., Fabian, N.E. and Haenlein, M. (2021), "Digital transformation: a multidisciplinary reflection and research agenda", *Journal of Business Research*, Vol. 122, pp. 889-901.

Verma, S. and Gustafsson, A. (2020), "Investigating the emerging COVID-19 research trends in the field of business and management: a bibliometric analysis approach", *Journal of Business Research*, Vol. 118, pp. 253-261.

Verma, S., Warrier, L., Bolia, B. and Mehta, S. (2022), "Past, present, and future of virtual tourism-a literature review", *International Journal of Information Management Data Insights*, Vol. 2 No. 2, 100085.

Vial, G. (2019), "Understanding digital transformation: a review and a research agenda", *The Journal of Strategic Information Systsytems*, Vol. 28 No. 2, pp. 118-144.

Wang, D. and Nicolau, J.L. (2017), "Price determinants of sharing economy based accommodation rental: a study of listings from 33 cities on arbnb.com", *International Journal of Hospitality Management*, Vol. 62, pp. 120-131.

Watkins, M., Ziyadin, S., Imatayeva, A., Kurmangalieva, A. and Blembayeva, A. (2018), "Digital tourism as a key factor in the development of the economy", *Economic Annals-XXI*, Vol. 169 Nos 1-2, pp. 40-45.

Xiang, Z. (2018), "From digitization to the age of acceleration: on information technology and tourism", *Tourism Management Perspectives*, Vol. 25, pp. 147-150.

Xiao, W., Mills, J., Guidi, G., Rodríguez-Gonzálvez, P., Gonizzi Barsanti, S. and González-Aguilera, D. (2018), "Geoinformatics for the conservation and promotion of cultural heritage in support of the UN Sustainable Development Goals", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 142, pp. 389-406.

Zhang, Y., Chen, H., Lu, J. and Zhang, G. (2017), "Detecting and predicting the topic change of Knowledge-based Systems: a topic-based bibliometric analysis from 1991 to 2016", *Knowledge-Based Systems*, Vol. 133, pp. 255-268.

Zheng, K., Kumar, J., Kunasekaran, P. and Valeri, M. (2022), "Role of smart technology use behaviour in enhancing tourist revisit intention: the theory planned behaviour perspective", *European Journal of Innovation Management*, Vol. ahead-of-print No. ahead-of-print, doi: 10.1108/EJIM-03-2022-0122.

Corresponding author
Peter Madzík can be contacted at: peter.madzik@gmail.com