



Sustainability communication of tourism cities: A text mining approach

Valentina Marchi^{a,*}, Alessandra Marasco^b, Valentina Apicerni^c

^a National Research Council of Italy - Institute of BioEconomy (CNR-IBE), Via G. Caproni 8, 50145, Florence, Italy

^b National Research Council of Italy - Institute of Heritage Science (CNR - ISPC), Via G. Sanfelice 8, 80134, Naples, Italy

^c National Research Council of Italy - Institute for Research on Innovation and Services for Development (CNR-IRISS), Via G. Sanfelice 8, 80134, Naples, Italy

ARTICLE INFO

Keywords:

Tourism cities
Sustainable tourism
Sustainability communication
Websites
Text mining
Dictionary building

ABSTRACT

Urban destinations have enjoyed increasing sustainable tourism attention and a consequent need to address challenges on promoting sustainable behavior in recent years. Tourism destinations' websites play an important role in the communication of sustainable tourism products and behaviors and have been receiving increasing attention in research on sustainability communication in tourism. However, the development of a holistic approach to online sustainability communication at the destination level remains challenging and requires further research. This study introduces a text-mining approach based on the creation of an ad-hoc dictionary to provide insights into how 10 European tourism cities communicate sustainability on their websites. Based on this approach, an assessment of online sustainability communication in these cities is performed through the definition and measurement of specific indicators relating to presence, depth and dispersion. The results show how this approach can be used to automatically assess the depth and the relative balance of communication across the environmental, socio-cultural, economic, and general dimensions of sustainability. Based on a large-scale, comparable analysis of sustainability-related online content, the proposed approach provides a tool to assist the design and improvement of websites for promoting sustainable tourist behavior and supporting sustainable destination management towards the green transition.

1. Introduction

Sustainability communication is an emerging area in tourism research and practice (Tölkes, 2018) with relevant implications in light of the growing importance of increasing awareness of sustainability issues, the visibility of sustainability tools (such as eco-labels) as well as the demand for sustainable options in line with the Transition Pathway for Tourism (European Commission, 2022). The Covid-19 pandemic contributed to rethinking urban tourism (Pasquinelli, Trunfio, Bellini, & Rossi, 2022) and to raising awareness of tourism sustainability issues, but closing the gap between tourists' attitudes and sustainable behavior remains a challenge (Eichelberger, Heigl, Peters, & Pikkemaat, 2021; European Travel Commission, 2021, 2022).

Urban destinations are at the forefront of sustainability challenges as blended places of life and travel in spatial alike (Chung, Lee, Ham, & Koo, 2021) and technological advances can make tourism cities more livable and sustainable (Ivars-Baidal, Celdrán-Bernabeu, Femenia-Serra, Perles-Ribes, & Vera-Rebollo, 2023; Šegota et al., 2019). Promoting visitors' sustainable behaviors is a key challenge in the development of sustainable urban tourism destinations (Miller, Merrilees, & Coghlan,

2015). However, the academic debate on sustainable tourism has primarily focused on rural and coastal areas rather than on urban settings, overlooking the environmental, socio-economic and cultural sustainability issues associated with urban destinations (Day, 2020). Notwithstanding the greater challenges to sustainability posed by urban tourism, the operationalization of sustainable tourism indicators in these destinations is considerably overlooked (Diéguez-Castrillón, Gueimonde-Canto, & Rodríguez-López, 2022).

Destination websites represent an important channel of sustainability communication (Ghanem & Elgammal, 2017; Spinelli, 2021; Tölkes, 2018). Official websites are often travelers' first point of contact with the destination and play an important role in the promotion of sustainable tourist behavior by raising awareness and encouraging sustainable choices (Ali & Frew, 2014; Garbelli, Adukaite, & Cantoni, 2017). Despite the growing literature in this area, there is still limited understanding of how to design effective online content for sustainability communication through websites (Ghanem & Elgammal, 2017; Spinelli, 2021; Tölkes, 2018). In particular, research remains largely fragmented due to the different criteria and approaches adopted for investigating sustainability communication through destination websites (Ghanem & Elgammal,

* Corresponding author.

E-mail addresses: valentina.marchi@ibe.cnr.it (V. Marchi), alessandra.marasco@cnr.it (A. Marasco), v.apicerni@iriss.cnr.it (V. Apicerni).

2017). Further, website sustainability content analysis primarily relied on non-automated methods, which are subject to personal bias (Law, Qi, & Buhalis, 2010). In the age of digital transformation, online environmental content can be usefully captured by means of a text analytics approach (Mariani & Borghi, 2021).

This study is one of the first to propose a text mining approach for systematic, large-scale and comparable assessment of online sustainability communication at the destination level. The sustainability of 10 leading European tourism cities has been monitored through the application of the dictionary.

Specifically, the proposed approach is based on the creation of an ad-hoc dictionary through a semi-automatic dictionary-building process, combining a manual analysis of previous studies and sustainable tourism communication indicators systems including the Global Sustainable Tourism Criteria (GSTC) and the European Tourism Indicators System (ETIS) with an automated text analysis of destination official websites. The dictionary has been tested on the content published on the websites of European cities using a set of indicators to understand how and if the sustainability is communicated. These indicators, as integral elements of tourism planning and management (Rasoolimanesh, Ramakrishna, Hall, Esfandiari, & Seyfi, 2020), have been defined to assess various aspects related to communication methods, such as presence, depth, and dispersion across the environmental, economic, and socio-cultural dimensions of tourism sustainability. The paper shows the support provided by these tools in assessing adopted strategies and information, empowering city managers to make informed decisions that promote and encourage sustainable and responsible behaviors in urban cities. This aligns with the growing importance of using indicators in decision-making, reflecting the trend towards informed decision-making using unstructured and big data (Kourtiti & Nijkamp, 2018; Nijkamp & Kourtiti, 2023).

The rest of the paper is structured as follows: the next section describes the theoretical background of the study through a review of previous research on sustainable tourism dimensions and online sustainability communication in tourism with a focus on destination websites. Section 3 describes the specific methodological process for text mining. In section 4 findings are presented and discussed, followed by the implications theoretical and managerial of the study, the limitations and recommendations for future research (Section 5).

2. Research background

2.1. Sustainable tourism dimensions and challenges in urban destinations

Since its first definition by the WCED (1987), the concept of sustainability has been extensively investigated in many scientific areas, particularly in tourism research, given its intrinsic link with the exploitation of natural resources (Ghanem & Elgammal, 2017; Santos, Veiga, Santos, & Águas, 2022; Zolfani, Sedaghat, Maknoon, & Zavadskas, 2015). Indeed, the specific term “sustainable tourism” has been used since the beginning of the 90s, opening two decades of growing interest from different academic backgrounds (e.g., geography, sociology, economics, and environmental management) and fueling definitions, critiques, and reconceptualizations.

In the last decade, sustainable tourism research has focused more on the outcomes of research concerning the practicalities of sustainability by moving towards suggesting solutions to concrete challenges (Buckley, 2012; Lane, 2018). Although there is still a lack of agreement on a unique definition of sustainable tourism (Asmelash & Kumar, 2019; Buckley, 2012; Santos et al., 2022; Solís-Radilla, Hernández-Lobato, Callarisa-Fiol, & Pastor-Durán, 2019), the one provided by the World Tourism Organization (1998, p.21) is commonly accepted, remaining an essential reference in tourism research: “Sustainable Tourism Development meets the needs of present tourists and host regions while protecting and enhancing opportunities for the future. It is envisaged as leading to management of all resources in such a way that economic,

social, and aesthetic needs can be fulfilled while maintaining cultural integrity, essential ecological processes, biological diversity, and life support systems”.

Despite the wide range of definitions, it is well-established that sustainable tourism is a multidimensional concept. The early dimensions identified by the United Nations Conference on Environment and Development (1992) were environmental preservation, social equality, and economic growth, which can be considered the three original pillars of sustainable development. However, the main challenge of delivering sustainable development involves integrating and finding a balance within them, which are mutually interdependent while recognizing their specificities (UNEP and UNWTO, 2005).

Economic sustainability means generating socio-economic benefits that are fairly distributed at different levels of society. It includes ensuring the viability and long-term stability of enterprises and stakeholders while providing income and social services to the host local communities. Social sustainability means respecting human rights and different cultures and avoiding any form of inequality and exploitation. It implies maintaining and strengthening the life support system for all in society, focusing on the welfare of local communities and living cultural heritage. Environmental sustainability means conserving and managing local resources to preserve natural heritage, landscapes, and biodiversity. It requires policies and actions to minimize the exploitation of resources and the emissions generated by the tourism sector. It should be highlighted that while the environmental dimension has remained a fundamental pillar in sustainable tourism research, given the awareness of climate change and the impact of tourism activities (Peng & Chen, 2019; Santos, Veiga, Águas, & Santos, 2019), the rise of the cultural dimension occurred only later. It is rooted in the recognition that culture shapes what we mean by development and, given its compatibility with the other three dimensions, can create solid bridges between them (United Cities and Local Governments, 2010). This is also confirmed by UNESCO (2013), which recognized culture as a driver for sustainable social, economic, and environmental development by reaffirming it as the “fourth pillar” of sustainability. Nowadays, the importance of culture in sustainable tourism is witnessed by the growing attention towards protecting and upgrading artistic, archaeological, worship sites, and cultural heritage to pass them on to future generations. Indeed, sustainable tourism also implies preserving and enhancing both tangible and intangible assets of past and living cultures representing their lifestyles, value systems, spiritual features, beliefs, and traditions (UNWTO, 2020).

With the increase of urban tourism, recent research has started to address these dimensions for sustainable tourism development in cities, in line with a broad approach to sustainability including the environmental, social and economic aspects of sustainable development (Aall & Koens, 2019). A number of international initiatives, such as the Global Platform for Sustainable Cities and the European Capitals of Smart Tourism Competition, have fueled interest on the approaches and best practices to make tourism cities more sustainable.

Despite the relative lack of research on the link between urban tourism and sustainability, the following main challenges can be identified in relation to the sustainable development of tourism in urban contexts (Aall & Koens, 2019; Koens, Melissen, Mayer, & Aall, 2021): the minimization of the social negative impacts of (over)tourism on the quality of life of residents and the preservation of environmental standards and resources, which is heightened by problems associated to climate change.

Studies in this area have also pointed to the need for a convergence between the sustainability discourse and smart city approach, in particular in relation to the need to develop a set of appropriate indicators for supporting progress towards sustainable tourism development in this specific context and the role of big data in this direction (Chung et al., 2021; Ivars-Baidal, Vera-Rebollo, Perles-Ribes, Femenia-Serra, & Celdrán-Bernabeu, 2021). The increasing availability of urban data offers a promising basis for effective city management

(Kourtit, Nijkamp, & Steenbruggen, 2017). However, there is a requirement for the development of easily comprehensible indicators that simplify complex phenomena and assist city managers in setting objectives, performance assessment, monitoring, managing and decision-making purposes (Huovila, Bosch, & Airaksinen, 2019).

The current approach also emphasizes the need to contemplate both supply and demand-side considerations for promoting sustainability within urban tourism destinations, including therefore also an explicit consideration of the role of tourists' sustainable behaviors (recycling, reductions in water use, low-impact transport, heritage preservation) and how to promote them through sustainability innovations and technologies (Marchi, Apicerni, & Marasco, 2021; Müller et al., 2015).

2.2. Online sustainability communication in tourism destinations

Environmental, social, economic and cultural sustainability issues have become increasingly important for organizations in the tourism sectors, including how sustainable tourism products and efforts are communicated (Bogren & Sörensson, 2021). In the last years, there has been a growing interest in sustainability communication in tourism and its potential positive effects on tourists' behavior in the pre-purchase and travel stages (Font, Elgammal, & Lamond, 2017; Tölkes, 2018; Visi-England, 2014). Sustainability communication is part of a wider sustainability marketing strategy aimed to make consumers aware of the availability of sustainable products, stimulate pro-sustainable choices and increase transparency about organizations' sustainability engagement (Tölkes, 2018). Research in this area addressed the challenge of designing effective sustainability messages to motivate sustainable purchases and inform stakeholders about commitment to sustainability goals in relation to various communication channels (Smith & Font, 2015; Tölkes, 2018, 2020; Villarino & Font, 2015; Wehrli et al., 2014).

Tourism organizations' websites represent the most researched channel of sustainability communication (Tölkes, 2018). In the context of digital transformation, websites represent important decision-making tools used by tourists for information search and vacation planning (Pan & Fesenmaier, 2006). Websites can play a key role not only as marketing-driven tools, but more so for improving quality of life, economic prosperity, social well-being, and sustainability (Xiang, 2018), and persuading their target audiences to change their behavior (Font et al., 2017).

Despite the importance of online sustainability communication in promoting desirable behavior from the target audience (Dodds, Novotny, & Harper, 2020; Font et al., 2017; Penz, Hofmann, & Hartl, 2017), previous research showed a great variation in the extent of web usage by tourism organizations for sustainability communication and scarce and/or superficial information on sustainability-related issues (Garbelli et al., 2017; Mura & Sharif, 2015; Pato & Duque, 2021; Pennington-Gray & Thapa, 2004; Santos et al., 2019). Importantly, it highlighted that the websites potential is not being fully utilized to create awareness of sustainability issues (Joseph, Lin, Nichol, & Jussem, 2014) and that sustainability messages have limited persuasiveness (Font et al., 2017; Villarino & Font, 2015).

Research explored the use and quality of websites for sustainability communication by destination management organizations, given their role as a first encounter for influencing tourist behavior and also in relation to sustainability (Mura & Sharif, 2015; Pennington-Gray & Thapa, 2004). Previous studies identified a large variety of variables/criteria for the content analysis of online sustainability communication in destination websites based on different sources, including the WTO Code of Ethics, Global Sustainable Tourism Criteria, and UNWTO Sustainable Tourism for Development Guidebook. Content analysis of destination websites for assessing sustainability communication has been generally based on the definition of a checklist and counting the presence of variables/criteria included in the checklist. Ghanem and Elgammal (2017) developed an online sustainability communication checklist including over 30 factors relating to four dimensions of

sustainability: environment, socio-economic, cultural, and general. Through a web content analysis of the top 50 competitive national destinations based on this checklist, they identified a lack of appropriate online approach to informing, motivating and engaging stakeholders in sustainability practices along with an unbalanced communication concerning the environmental, socio-cultural, and economic sustainability dimensions. Another approach for measuring sustainability orientation in destination online communications was developed by d'Angella and De Carlo (2016) through the variable Green D-web score based on a checklist of 35 binary indicators in seven areas (general, infrastructure, transportation, experience, food, shopping, and hospitality). They used this variable to assess the orientation to sustainability of official website communications of tourism cities, revealing different performance levels. Their study highlighted the importance of sustainability communication for differentiating tourism products and attracting new segments of environmentally sensitive tourists. Spinelli (2021) adopted the methodology developed by d'Angella and De Carlo (2016) to explore the sustainability-related content of the Italian regional tourism websites and investigated the potential association between the sustainability orientation and the profile of destinations. His study found that no website explicitly mentioned green, sustainable or responsible travel on its homepage, and just 5 regions devote a specific section to these forms of tourism. In summary, previous studies showed that destination websites tend to provide limited information and lack of a holistic approach to communicating sustainability encompassing all the environmental, economic and socio-cultural aspects. Adequately balancing all sustainability dimensions is highlighted as a key issue in the design and implementation of an effective approach to inform, motivate, and engage stakeholders in sustainable tourism (Ghanem & Elgammal, 2017). In one of the first studies using a web content mining approach, Marchi et al. (2021) assessed the sustainability communication in Italian cultural cities, highlighting the potential of this approach for improving the analysis of online sustainability-related contents.

3. Methodology

As noted above, the aim of this study is to introduce a text-mining approach to provide insights into how cities communicate sustainability on their websites. Previous studies developed dictionaries for different purposes and on various topics, for example, to monitor job descriptions in the digital environment (Park, Lu, & Marion, 2009), to measure corporate social responsibility (Pencle & Mălăescu, 2016), to know the impact of frontline employees' problem solving on customer satisfaction (Marinova, Singh, & Singh, 2018), and to understand how consumer word use varies across contexts (Hovy, Melumad, & Inman, 2021). Dictionaries were also developed to quantify holistic features of a text, such as its sentiment or emotionality (e.g. Pennebaker, Boyd, Jordan, & Blackburn, 2015; Rocklage, Rucker, & Nordgren, 2018).

In previous literature there are few dictionaries developed specifically for the tourism field, in some cases, we can find integration on dictionaries previously created and mainly applied to the hotel industry (e.g. Nie, Tian, Wang, & Chin, 2020; Park, Kang, Choi, & Han, 2020). For this reason, the study develops an ad-hoc sustainable tourism communication dictionary.

3.1. Sustainable tourism dictionary development

To accomplish the goal of developing an automated tool to analyze digital sustainability communication of tourism cities a dictionary was built followed the S-DBP proposed by Deng, Hine, Ji, and Sur (2019). The sustainable tourism communication dictionary is based on an iterative process that includes the existing theoretical bases as well as the adequate contents of the corpus identified. The study carefully followed the steps proposed by Deng et al. (2019) for designing dictionaries and combined a further step to extract additional words based on the Latent Dirichlet Allocation (LDA), a computer-assisted topic modeling

algorithm (Fig. 1).

The first step (Objective clarification) allowed us to clearly define the aim of the dictionary which is to assess online sustainability communication of tourism destinations providing a tool for fast and systematic processing of large amounts of text.

The second phase focused on the identification of categories and categorizing entries. In line with previous research on sustainability and sustainability communication (Ghanem & Elgammal, 2017; Villarino & Font, 2015), the researchers included four entries: environmental, socio-economic, cultural and general. The environmental dimension refers to the promotion of tours/excursions/activities that respect natural sites, green and protected areas, green practices, and the protection of the natural environment and landscape. The socio-economic dimension considers the support to the local economy and community, promoting local markets and food, accessible tourism, safety and security, human rights, and helping threatened sites through donations or volunteering. The cultural dimension includes cultural valorization and decorum visiting heritage. Finally, the general dimension focuses on sustainable management and management of tourist flows. Sub-categories have been identified for each dimension to investigate and capture different aspects within the same category.

In the third step, we identified the corpus on which the dictionary is developed. We adopted a set of documents that include textual content related to sustainable tourism. In particular, we identified the following sources for the corpus creation: 1) a sustainable tourism dictionary (in Italian) previously developed by Marchi et al. (2021). It was created by adopting a web content mining approach to assess the characteristics of the online content of a sample of 20 Italian cultural tourism cities; 2) guidelines and indicators systems recognized at the international level, such as the Global Sustainable Tourism Council (2016) and the European Tourism Indicators System (ETIS) Toolkit (2017); 3) textual contents of the top 10 European tourism cities by the number of bed nights in 2019 (Statista Research Department). This content, mainly textual data, was collected using a web scraping procedure, a technique that navigates through each destination website in every internal link and extracts data without needing user interaction (Singrodia, Mitra, & Paul, 2019). A total of 6336 web pages in English were extracted in May 2021 and processed using R software and Knime Data Analytics Platform. The automated analysis supported us in analyzing patterns in text data to extract and discover knowledge for decision-making (Zhai & Massung, 2016). Then, in the fourth step, we prepared the textual contents of the 10 European urban destinations for further analysis through the pre-processing operation (such as stop words removal, reduction of all words to lowercase). Following this phase, we extracted a list of the most frequently used words (unigram and bigram) analyzed by researchers. More than 70 additional terms were classified and included in the sustainable tourism communication dictionary, such as gastronomy, wheelchair, heritage site, and circular economy.

To expand the list of keywords we integrated the dictionary-building process with the LDA model, which is a generative statistical process that explains how text documents could be generated probabilistically from a mixture of topics, where each topic has a distribution over words (Blei, 2012). LDA offers additional advantages over the semi-automatic approach adopted until now and allowed the inclusion of additional keywords resulting from processes that did not require researchers to prespecify rules or keywords (Huang, Leheavy, Zang, & Zheng, 2018). The model automatically generated topics with the most important words (weighted) for each topic which is coherent with the information used to promote and discover the European cities on official websites. We labeled each of the seven topic emerged as follows: 1. What to visit; 2. Food & Drink; 3. Shopping; 4. Culture & Leisure; 5. Enjoy the city; 6. Travel information; 7. Life & business (Fig. 1 - 6. Keywords from the topic model).

Following the creation of a list of keywords, researchers organized brainstorming sessions to analyze each keyword and the related dimensions and sub-dimensions identified. Then the seventh step includes

techniques to extend the dictionary through the identification of synonyms and the application of the stemming approach. This latter technique reduces the word, for example, the word “sustainable” may be stemmed to “sustaina*”. In this way, the word “sustainability” will also be traced.

Finally, the last step of the process included the validation of the dictionary entries. This study adopted the keyword-in-context (KWIC) method, which is an automatic system that allows the search of a particular keyword in the text and analyzes its local meaning in relation to a number of words immediately preceding and following it (Luhn, 1960). This step allowed us to identify too generalized words. For example, the word “food”, included in the sub-dimension “B2. Promotion of local markets and food”, was too generic and not necessarily related to sustainability. Researchers replaced it with the bigram “sustainable food”. We conducted different rounds for the validation of the dictionary and it was performed using the R package called ‘Quanteda’ for managing and analyzing text (Benoit et al., 2018).

The final sustainable tourism communication dictionary is composed of about 300 keywords - unigram, bigram (see Appendix). It consists of word lists and each keyword refers to each entry, dimension and sub-dimension, without term overlaps.

3.2. Sustainable tourism dictionary application

Following the creation and validation of the dictionary, an automated textual analysis was performed on the top 10 European destinations by number of bed nights in 2019 (Statista Research Department) to analyze sustainability-related contents in their websites through a set of indicators created on the basis of previous studies (Ghanem & Elgammal, 2017; Mariani & Borghi, 2021; Santos et al., 2019; Villarino & Font, 2015). A profile of the cities in the sample is provided in Table 1, which also includes the official tourism websites analyzed in this research. This sample was deemed appropriate to investigate digital sustainability communication in these urban destinations, also considering the varied profile of cities in terms of sustainability orientation based on previous rankings and research. For instance, the profile includes cities' ranking in Citizen-Centric Cities, the 2018 edition of Arcadis' Sustainable Cities Index that explores the sustainability of 100 cities from the perspective of citizens based on three pillars: People (social), Planet (environmental), and Profit (economic). Among the considered urban destinations, Lisbon was the winner of the European Green Capital Award in 2020.

The final dictionary was applied to destinations' online content through an automated analysis to deepen characteristics that the literature suggested may influence their effectiveness communication. In order to measure and monitor the actual sustainable communication of European cities, a set of indicators has been defined as an evaluation tool. This allows for the establishment of a common language and standardized units of analysis to assess city performance, supporting the decision-making process of city managers in promoting and incentivizing sustainable behaviors within cities (Huovila et al., 2019; Kourtit & Nijkamp, 2018; Nijkamp & Kourtit, 2023; Torres-Delgado & Saarinen, 2017).

In particular, the study investigated the presence of sustainable communication, which is expressed as the percentage of webpages containing sustainability messages out of the total webpages (sustainable webpages/total webpages*100) and underlines the weight of sustainability communicated online. The second indicator relates to the depth of sustainability communication, measured as the percentage of sustainable words in the dictionary out of the total words online (number of sustainable words/total words*100). This indicator provides insights into the emphasis and importance attributed to communicating the concept of sustainability within the text. It allows us to identify how extensively sustainability themes are addressed examining the degree of attention and focus given to sustainability in the overall messages. The third indicator is the dispersion of communication, calculated through

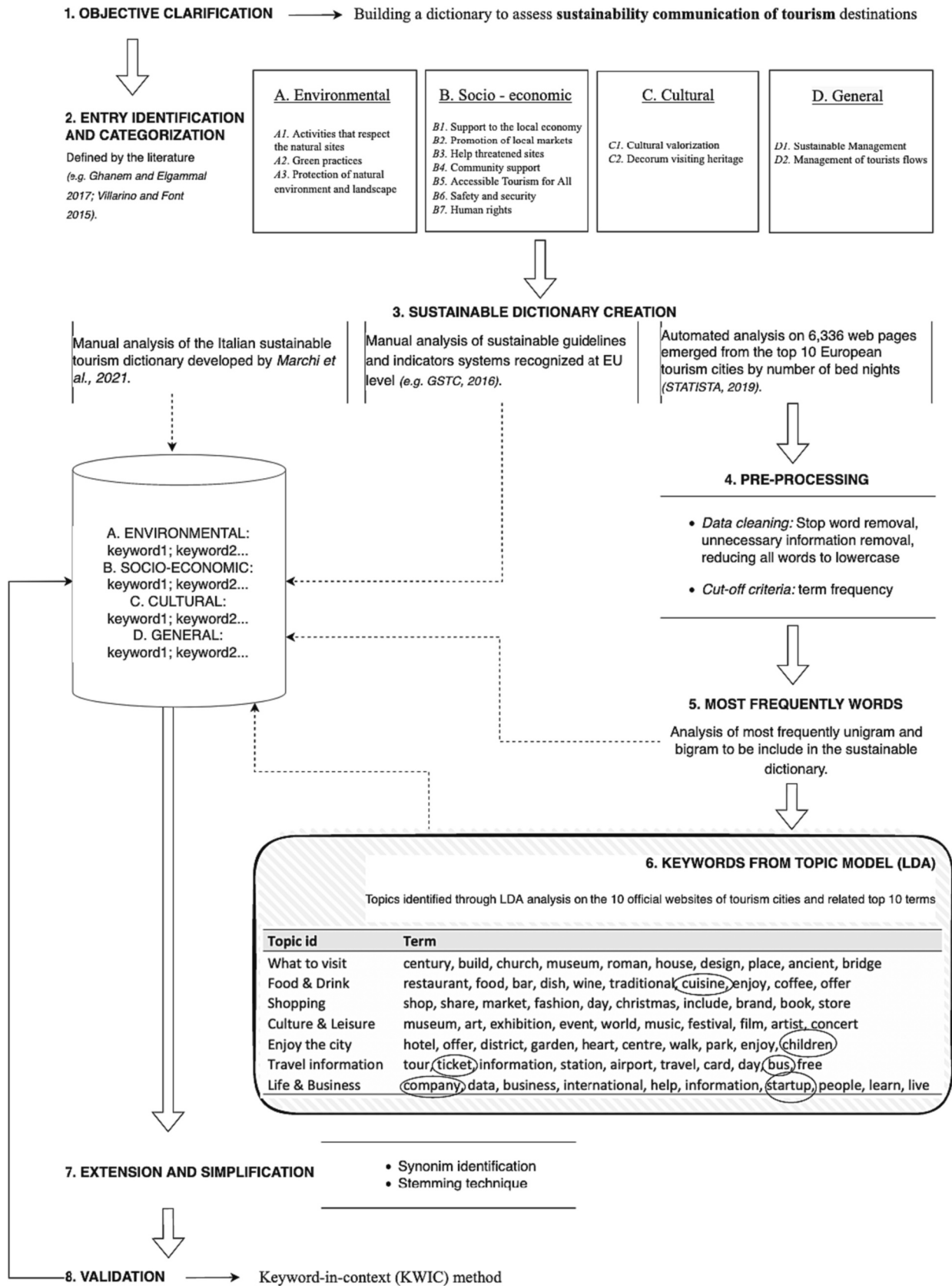


Fig. 1. The S-DBP implemented to build the sustainable tourism dictionary.

Table 1
The tourism cities in the sample.

| City | Population <i>Eurostat</i> 2018/ 2021 | Arrivals <i>Euromonitor</i> 2019 | Bed nights (mil.) <i>Statista</i> 2019 | Accommodations <i>Booking.com</i> 2021 | Main Attractions <i>Tripadvisor.com</i> 2021 | Parks, garden and nature areas <i>European Green City Index - *Other 2020</i> | Position in <i>Citizen Centric Cities</i> 2018 | Sustainability orientation of official website (<i>F. d'Angella & M. De Carlo</i>) 2016 | Total visits of the official websites <i>Similarweb.com</i> 2021 | Official Tourism Websites |
|-----------|--|--|--|--|---|--|--|---|--|--|
| Amsterdam | 854,047 | 8,835,400 | 18.38 | 1033 | 2992 | 16 | 12° | Moderate | 546.4 K | www.iamsterdam.com |
| Barcelona | 1,636,732 | 7,016,600 | 19.85 | 2375 | 4579 | 18 | 28° | Low | 112.4 K | www.barcelonaturisme.com |
| Berlin | 3,664,088 | 6,195,800 | 34.12 | 1293 | 3296 | *51 | 18° | High | 655.2 K | www.visitberlin.de |
| Lisbon | 544,851 | 3,639,900 | 13.82 | 4169 | 2958 | 10 | 62° | Low | 103.5 K | www.visitlisboa.com |
| London | 9,002,488 | 19,559,900 | 85.1 | 4405 | 8211 | 15 | 1° | High | 1.6 Mln | www.visitlondon.com |
| Madrid | 3,305,408 | 5,597,800 | 20.68 | 2948 | 3430 | *21 | 21° | Moderate | 1.2 Mln | www.esmadrid.com |
| Paris | 2,165,423 | 19,087,900 | 52.45 | 4324 | 7972 | 20 | 15° | High | 316.6 K | en.parisinfo.com |
| Prague | 1,335,084 | 9,150,900 | 18.48 | 1674 | 3932 | 18 | 23° | Moderate | 231.2 K | www.prague.eu |
| Rome | 2,844,750 | 10,317,000 | 29.07 | 8229 | 5864 | 25 | 40° | Low | 170.1 K | www.turismoroma.it |
| Vienna | 1,931,830 | 6,634,700 | 18.64 | 2001 | 1988 | 15 | 5° | Moderate | 537.8 K | www.wien.info |

the coefficient of variation (CV), which allows an analysis of the frequency distribution of sustainable words on web pages. This indicator provides insights into the concentration of sustainability across web pages giving indication of whether it is concentrated solely on a few dedicated web pages or if there is a broader approach by city managers to promote sustainability throughout the entire website. It is a standardized measure to compare European urban destinations websites, expressed as follows:

$$CV = (SD(\text{Number of sustainable words})) / (\mu (\text{Number of sustainable words})) * 100.$$

Furthermore, the study investigated the presence of sustainability labels, certifications and awards on the websites. This choice was in line with previous studies highlighting that despite the great efforts made by destinations to improve sustainable strategies, in some cases, these are not communicated and correctly perceived by tourists and residents (Cannas, 2018; Marin, Palmisani, Ivaldi, Dursi, & Fabiano, 2009; Pencarelli, Splendiani, & Fraboni, 2016). An automated content analysis was performed to investigate the communication of certifications and awards on the 10 websites of the European urban destinations. Given the high number and the variety of sustainable labels, in many cases relevant just at a national level, this study considered only certifications recognized at the European level, focusing on a set of keywords, such as 'ISO9001', 'ISO 14001' or 'ecolabel'.

4. Results

Based on the automated content analysis, a first overview can be provided on the presence of sustainability communication in the total sample. It reveals that 58.84 % of the total web pages contain content relating to sustainability-oriented practices, which are more related to specific environmental, socio-economic, and cultural practices than to sustainable tourism in general. This is relevant considering that communication focusing on benefits related to specific sustainability dimensions is likely to be more effective than generic mentions of sustainability (Villarino & Font, 2015).

More in detail, Table 2 reports the share of web pages containing words related to the four sustainability dimensions on the total web

Table 2
Presence of sustainability-related contents.

| | Sustainable webpages | % on total webpages |
|----------------|----------------------|---------------------|
| Environmental | 2410 | 38.04 |
| Socio-economic | 1891 | 29.84 |
| Cultural | 877 | 13.84 |
| General | 687 | 10.84 |

pages and shows that online messages are mostly related to environmental (38.04 % of total pages) and socio-economic sustainability (29.84 % of total pages). This finding was quite expected, in line with previous research that showed a stronger focus on environmental sustainability in tourism communication, reflecting the development of sustainable tourism research and practice (Tölkes, 2018). Similarly, the analysis conducted by Ghanem and Elgammal (2017) revealed a prominence of environmental and social information in destinations' online sustainability communication. It is important to underline that the number of words included in the environmental and socio-economic dimensions is greater than in the general and cultural dimensions. This stems from a lower presence of terms relating to these last two dimensions in the documents analyzed (websites, previous studies and European documents/tools), thus highlighting a lower use of communication relating to the general and cultural dimension. Despite this, the results of the analysis remain valid considering the relationship between the number of words included and the number of sustainable words that emerged.

The analysis provided a detail of the themes that are most frequently communicated in relation to the different dimensions of sustainability, thereby contributing insights into the relevance of specific themes within online content. Table 3 provides detailed information on the specific themes that are most frequently communicated in relation to the four dimensions of sustainability. For the environmental dimension, "green practices" emerges as a primary theme that is mainly present through the sub-dimension 'public transport' and 'environmental impact'. For the socio-economic dimension, the most frequently communicated themes include support for the local economy, especially through buying locally-made products and local services ('local guides'). Other highly frequent themes include accessible tourism, which is

Table 3
Sustainability dimensions, topics and sub-topics in online communication (total sample).

| A. ENVIRONMENTAL SUSTAINABILITY | | | |
|---|--|-------|-------|
| | Excursions | 28.12 | |
| A1. Tours/excursions/activities that respect natural sites, green and protected areas | Urban trekking | 0.18 | 39.56 |
| | Green areas | 6.03 | |
| | Protected areas | 1.21 | |
| | Bike path | 4.02 | |
| | Public transport | 36.14 | |
| A2. Green practices | Waste management | 0.97 | 43.09 |
| | Environmental impact | 4.43 | |
| | Public water | 0.36 | |
| | Green | 1.19 | |
| A3. Protection of natural environment and landscape | Landscape | 17.35 | 17.35 |
| B. SOCIO-ECONOMIC SUSTAINABILITY | | | |
| B1. Support to the local economy | Locally-made products | 14.04 | 37.67 |
| | Supporting local entrepreneurs | 6.65 | |
| | Local guides | 16.98 | |
| | Traditional market | 2.06 | |
| B2. Promotion of local markets and food | Slow food | 0.36 | 15.83 |
| | Kilometer zero | 0.03 | |
| | Traditional food | 13.38 | |
| B3. Help threatened sites through donations or volunteering | Threatened sites | 2.38 | 2.38 |
| B4. Community support | Involvement of local community | 17.15 | 17.15 |
| B5. Accessible tourism for All | Accessible tourism | 17.70 | 17.70 |
| B6. Safety and security | Safety | 6.60 | 6.60 |
| B7. Human rights | Preventing exploitation and discrimination | 2.67 | 2.67 |
| C. CULTURAL SUSTAINABILITY | | | |
| C1. Cultural valorization | Cultural heritage | 42.01 | 42.01 |
| C2. Decorum visiting heritage | Behave with decorum | 57.99 | 57.99 |
| D. GENERAL | | | |
| D1. Sustainable Management | Sustainable Management | 75.88 | 75.88 |
| D2. Management of tourists flows | Tourists | 24.12 | 24.12 |

present in 17.70 % of the total web pages with information relating to this dimension. Accessible tourism was found to be among the most communicated factors in relation to social sustainability also by [Ghanem and Elgammal \(2017\)](#) and this is consistent with the growing attention within research and practice on the use of digital technologies for providing information and increased access for all from the pre-visit stage of the tourist journey ([Michopoulou & Buhalis, 2013](#)). Further, ‘involvement of the local community’ emerges as another frequent sub-category in this dimension, referring in particular to ‘residents’, which appears to be particularly relevant in relation to sustainable urban destinations.

The two themes considered for the cultural dimension of sustainability relate to the valorization of cultural heritage and to promoting respectful behavior while visiting heritage. For instance, this latter is present through some recurring keywords that are related to the “don’ts” of culturally significant behavior ([Pennington-Gray & Thapa, 2004](#)), to avoid potential damage at cultural heritage sites, such as ‘touch’,

Table 4
Ranking of cities based on number of bed nights and sustainability communication indicators.

| N. of bed nights (STATISTA, 2019) | Presence | % | Depth | % | Dispersion | % |
|-----------------------------------|-----------|------|-----------|------|------------|-------|
| 1. London | Berlin | 67.1 | London | 0.64 | Berlin | 146.4 |
| 2. Paris | Madrid | 65.8 | Amsterdam | 0.62 | Lisbon | 155.7 |
| 3. Berlin | Vienna | 60.1 | Barcelona | 0.60 | Madrid | 160.0 |
| 4. Rome | Amsterdam | 59.4 | Lisbon | 0.55 | Prague | 164.3 |
| 5. Madrid | London | 58.3 | Berlin | 0.54 | Barcelona | 170.2 |
| 6. Barcelona | Paris | 56.8 | Paris | 0.50 | Amsterdam | 173.6 |
| 7. Vienna | Lisbon | 54.5 | Vienna | 0.48 | Vienna | 175.6 |
| 8. Prague | Rome | 53.8 | Madrid | 0.47 | Rome | 179.0 |
| 9. Amsterdam | Barcelona | 50.6 | Rome | 0.30 | Paris | 190.8 |
| 10. Lisbon | Prague | 38.1 | Prague | 0.27 | London | 199.9 |

‘graffiti’, ‘damage’, ‘climb’. Finally, the general dimension of sustainability is communicated in the webpages mainly through information about sustainable tourism, represented by terms including ‘sustainability’, ‘slow’, ‘ethical’, ‘sustainable development’.

The analysis of sustainability communication with regard to the European cities in the sample (Table 4) reveals that 9 out of the 10 destinations communicate sustainability on more than 50 % of their websites. More in detail, Berlin emerges as the destination with the highest presence in terms of webpages containing sustainability information of the total webpages (67.1 %), followed by Madrid (65.8 %) and Vienna (60.1 %). Based on the depth of sustainability communication, London is the destination with the highest percentage of use of sustainable words (0.64 %), followed by Amsterdam (0.62 %) and Barcelona (0.60 %). The indicator of dispersion highlights that on the website of Berlin the theme of sustainability is more concentrated on a few web pages, rather than in the total. The analysis also shows that although the city of Berlin has the highest percentage of sustainable pages out of the total, sustainable words are concentrated on a few web pages (about 40 % of sustainable words are concentrated on 60 web pages). In contrast, the website of London shows a wider distribution of sustainability-related words across web pages than the other destinations.

For instance, the official tourism website of Berlin presents a specific section, labeled *Sustainable Berlin*, which promotes the sustainability of the city with information and tips, from transport to restaurants, to eco-hotels, to shopping. The tourism website of London does not include a page dedicated to sustainability and directly accessible from the homepage. These results are in part consistent with the existing sustainability-related ranking of European cities (see Table 1). For instance, Berlin and London scored as highly sustainability-oriented destinations in work by [d’Angella and De Carlo \(2016\)](#), which also included Vienna, Madrid and Amsterdam into moderate destinations based on their score. London also leads in the [Cities \(2018\)](#), which ranks Vienna, Amsterdam, Paris, and Berlin in the top 20.

The approach developed in this study also provides insights into the relative balance of communication across the different sustainability dimensions (Fig. 2). The balance among dimensions represents one of the main issues for the effectiveness of tourism sustainability communication ([Tölkes, 2018](#); [Villarino & Font, 2015](#)) and in this regard the approach developed in this study is useful for an assessment of the relative (un)balance among the different sustainability pillars.

The environmental issue is the one most dealt with by cities on their websites. In some of these, it is recorded that almost 50 % of the web pages talk about environmental sustainability. Berlin (47.8 %) and Madrid (46.8 %) are among these cities. Madrid (38.2 %), with Amsterdam (36 %), is the city that most communicates the socio-economic dimension. For the cultural dimension, Lisbon (32.7 %) is the destination with a higher percentage of web pages compared to the other cities. Finally, topics of general dimension are more present on the websites of Berlin (24 %) and Lisbon (21.2 %).

The study addressed the presence of sustainability communication also in relation to labels, certifications, and awards on the websites, as a relevant feature of online sustainability communication. The results

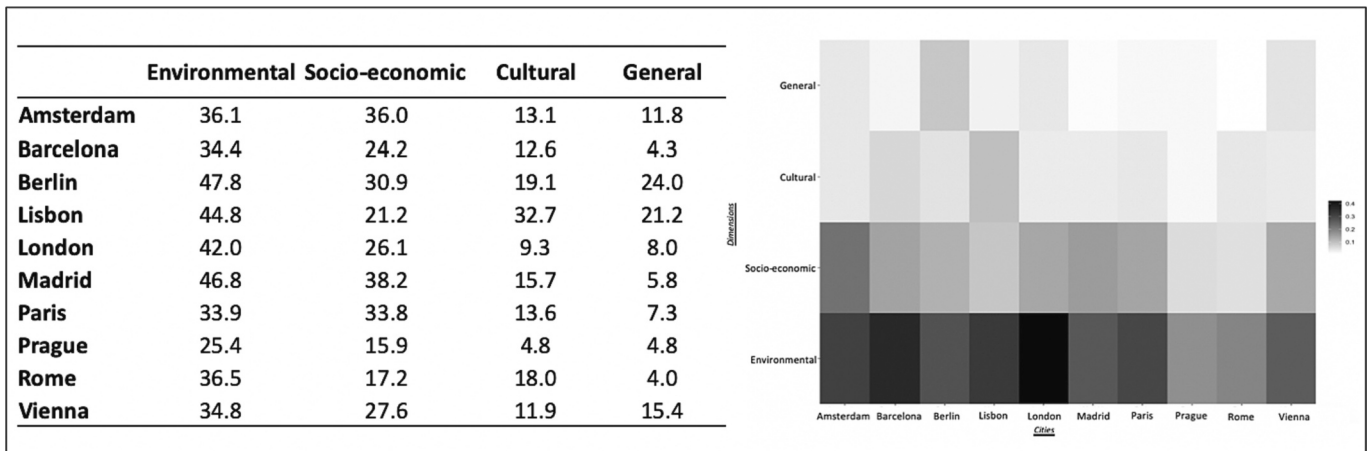


Fig. 2. Frequency of sustainability communication by dimension (%) and related heatmap. In the heatmap, color intensity depicts the ratio between the number of keywords in each dimension and the total number of words in the websites.

show that for the total sample 25 % of the words relating to labels, awards and certifications are directly related to sustainability, including ecolabels, ISO9001, ISO 14001, while the remaining concern other areas (i.e., ‘food&wine’, ‘fashion/design’, ‘cinema/theatre/music’). The majority of these certifications are related to the socio-economic dimension, with specific regard to accessible tourism (29.7 %), diversity (13.5 %), and safety & security (16.2 %). The certifications related to the environmental dimension (30.7 %) are mainly used to communicate and promote sustainable mobility to reduce traffic and pollution, while the general dimension (17.3 %) relating to sustainable management, and the cultural dimension (6.2 %) to communicate cultural valorization (i.e.

Unesco sites). Fig. 3 shows the different relevance of sustainable labels in European cities among dimensions.

Specifically, the analysis by cities partially confirms the previous ranking of urban destinations, with Paris as the city with the highest number of certifications/awards communicated on its official website (32 % of the total sample), which are particularly related to the environmental and socio-economic dimensions. Based on the analysis, the city of Prague does not promote sustainability labels or awards on its website. In other cities, such as London and Lisbon the percentage of communication is very low compared to other cities and related to a single dimension. By contrast, Rome and Paris emerge as destinations

The sustainability related labels by city and dimensions (%)

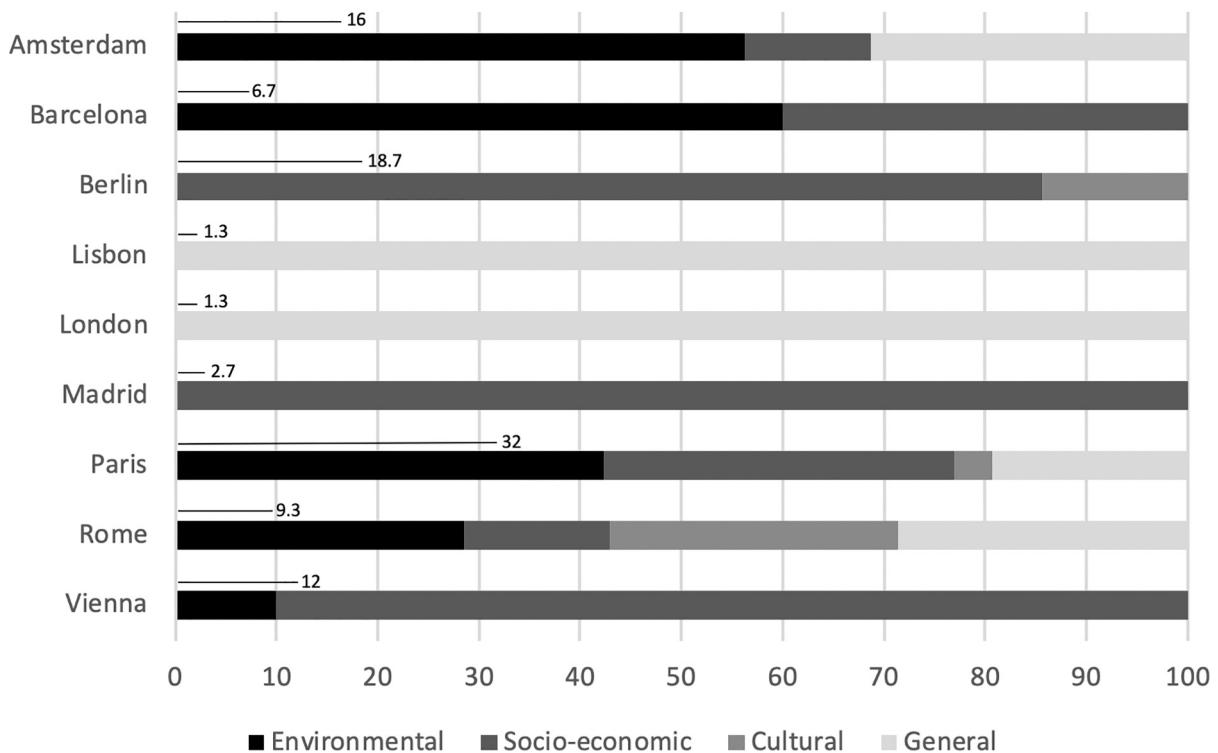


Fig. 3. The presence of sustainability labels by city and dimension (%). The solid line shows the percentage of pages that refer to at least one label for each city, while the colored bars illustrate the communication of certifications by dimensions for each website (%).

that promote certifications on all four sustainable dimensions. In other cities, some dimensions largely prevail over others, for example, the socio-economic pillar in Berlin and the environmental dimension Amsterdam.

5. Discussion

Notwithstanding the growing attention to sustainability communication in tourism, there has been only limited research into the assessment of the quality of destinations' official websites as channels of sustainable tourism content (Spinelli, 2021). This study provides a contribution to this area through an advanced approach to improve the understanding of how destinations communicate and promote sustainability through their official websites. The study advances previous research that focused only on one or a few dimensions of sustainability, by adequately taking into account the multidimensional nature of effective sustainability communication (Ghanem & Elgammal, 2017; Tölkes, 2018). In particular, the analysis provides a contribution in the direction of a holistic approach to the communication of environmental, socio-economic, and cultural sustainability in line with the recommendations of previous research (Ghanem & Elgammal, 2017; Tölkes, 2018; Villarino & Font, 2015). Through its approach, it attempts to overcome the fragmentation in the literature with regard to the analysis of web-based sustainability communication as well as the limits of non-automated methods adopted so far for content analysis. As noted by Ghanem and Elgammal (2017), the investigation of different sustainability messages and units of analysis may yield contradictory insights into the communicated contents. In this regard, this paper contributes to developing a more systematic approach to researching online sustainability communication of destinations that can provide more comparable results across a high number of units of analysis.

The results show that for some cities there is a strong imbalance of sustainable communication among the four dimensions, with high percentages for the environmental and socio-economic dimensions and almost an absence of communication of general (i.e., Rome) and cultural (i.e., Prague and London) dimensions. These results concur with those that emerged in previous studies in which destinations do not sufficiently balance their online communication in the sustainable dimensions (Ghanem & Elgammal, 2017). The prevalence of the environmental dimension in the online sustainability communication of all the cities in the sample is not surprising, given the centrality of the environment and green transition to any urban sustainable development path (Pasquinelli et al., 2022). For instance, London's strategic tourism plan *A Tourism Vision for London* (2017) recommends investments in infrastructures and amenities, among key areas of intervention, for the green transition of the city. "Championing an environmentally-sustainable city" by promoting walking, cycling and the use of clean routes is among the principles of the Visitor Destination Strategy 2019–2023 of the city. By contrast, a lower emphasis is placed on the general dimension according to the findings (Fig. 2); only two cities scored over 20 % in this dimension, notably Berlin (24 %) and Lisbon (21.2 %). In this regard, for instance, the Berlin tourism plan (*Sustainable and City-Compatible Berlin Tourism Plan 2018+*) focuses on urban compatibility and long-term sustainability to maintain the life quality of residents and harmoniously increase experience value for visitors. It replaces the emphasis on quantitative growth (more visitors) with sustainability as a maxim of action to ensure a city-compatible tourism development in the city.

This is also reflected in the presence of labels and awards, which appears to be unbalanced even in the most sustainable cities. This could be in part due to the focus of the analysis only on European labels, which thus does not consider national labels and awards. For instance, this could be the case of London, which adopts sustainable certifications that are partly different from those used by European cities and consequently the analysis showed a low percentage of communication of sustainable certifications despite its sustainable performance.

The indicators investigated through the automated approach revealed a varied picture of sustainability communication in the cities of the sample. The results highlight the performance of some urban destinations, such as Berlin and Paris, which is coherent with their position in previous sustainability rankings. For example, the cities of Berlin, London, and Madrid rank among the top 15 in the Planet sub-index of the Citizen Centric Cities Index (Cities, 2018). It is interesting to note (Table 3), however, that some of the top European city destinations for arrivals and bed nights in the sample - Paris and Rome - scored lower in communicating sustainability through their websites than less visited destinations (before Covid), such as Amsterdam and Lisbon. This finding is in line with the work by Spinelli (2021), which found no significant association between the sustainability orientation of web communication of destinations, their size (population) and tourism maturity (variation of overnight stays).

6. Conclusions, implications and future research

The main contribution of this study lies in the creation of an ad-hoc dictionary for the assessment of online sustainability communication with its multiple dimensions in tourism. This is the first attempt to apply a computer-assisted text analysis approach combined with LDA model and researchers' knowledge to establish a dictionary in tourism and explore the detailed sustainability communication adopted by 10 European cities through their official websites.

The dictionary can be used to capture the multidimensionality of sustainable tourism as communicated through the official websites of tourism cities. The approach allows to analyze big quantities of online texts in a speedy and accurate way and to reduce subjective interpretation in coding. Capturing the specific dimensions of sustainable tourism communication in a more objective way can enrich the current understanding of this phenomenon, support other researchers in effectively and efficiently investigating it, and provide city destinations with a new tool to assess and improve online communication for sustainability. Based on the ad-hoc dictionary, the study also proposes an assessment of online sustainability communication through presence, depth and dispersion indicators. These measures can be useful in providing a synthetic, representative picture of online sustainability communication that can be made available at a reasonable cost/benefit ratio, easily interpreted and used for comparisons across multiple urban destinations. This is particularly relevant given the importance of indicators as tools for supporting progress towards sustainable tourism development in consideration of the specificities of urban destinations (Diéguez-Castrillón et al., 2022).

The proposed approach is promising in addressing the major challenges in tourism website evaluation identified by Law et al. (2010), notably the lack of a specific tourism-oriented technique, and the personal bias of methods involving human subjects. In this respect, this study overcomes these issues that largely affect the results of tourism website assessment (Law et al., 2010) and provides a repeatable method with good potential for long-term application. The proposed dictionary can be used in combination with other methods (e.g., user judgment) for assessing the quality of online sustainability-related content of tourism destinations.

From a managerial perspective, this work lies the basis for developing a digital tool for fast and systematic processing of large amounts of online text that could be integrated as a tourism intelligence instrument to support sustainable destination management towards the digital and green twin transition. Using the tool of the sustainable tourism communication dictionary, policymakers, tourism intelligence services providers, web content strategists and researchers could efficiently and accurately investigate, compare and analyze the sustainability practices communicated by city destinations.

In particular, it could assist the assessment of online communication in relation to sustainability marketing objectives and destinations' marketing strategies to address the demands of pro-sustainable tourists

by providing useful information to target sustainability-oriented messages. This appears to be relevant also in consideration of the crucial linkage between sustainability communication and destination branding (Adamus-Matuszynska, Dzik, Michnik, & Polok, 2021; Lupu, Rodrigues, Stoleriu, & Gallarza, 2021). Given the importance of a sustainable offer in tourist choices (Adamus-Matuszynska et al., 2021), there is the opportunity to use the proposed approach to assist brand strategies towards a more sustainable perspective (Lupu et al., 2021). The possibility to assess how sustainability-related content is distributed on official websites could be useful to support the analysis of brand content used by city destinations to promote their image and positioning in terms of sustainable tourism.

The indicators will also be beneficial for destinations that want indications on how to better balance sustainability communication dimensions. Indicators can provide new inputs to destinations for starting new paths focused on undervalued sustainable dimensions both in terms of communication and the policies adopted. At the same time, indicators provide valuable information for potential customers that will be more aware and able to know sustainable practices performed by destinations.

Furthermore, the sustainable tourism communication dictionary and its indicators can be adopted as an objective tool to measure whether there are differences between how much a destination utilizes sustainability as a marketing tool and how much it actually operates in sustainable way. This could be achieved through the combination of multiple indicators, related to measuring communication and sustainable performance (such as the ETIS or the GSTC), in order to identify potential instances of greenwashing or greenhushing as well (Font et al., 2017).

The main limitations of the study concern the sample and the need to extend and further test the dictionary on a broader and varied sample of destinations. Future research could use the dictionary applied in this study as a basis for further development of the tool, especially by integrating the judgment of destination experts and the tourists' perspectives to improve its efficacy.

Future research could investigate in which web pages sustainability issues are most communicated (e.g. booking pages, accommodation facilities pages) to understand when destinations deem that sustainability themes are more attractive and effective to promote themselves online and to promote sustainable tourism practices. Further, the research could attempt to integrate this approach with specific measurements to evaluate the emotional appeal of communication as a relevant aspect of its potential persuasiveness (Villarino & Font, 2015). In this regard, the future analysis could consider, next to sentiment analysis, approaches that allow to integrate the text and images for pro-sustainable tourism communications in line with the most recent research on communication for pro-sustainable tourism behaviors (Li, Saayman, Stienmetz, & Tussyadiah, 2022). Given that techniques for text analysis differ from those used on images (Balducci & Marinova, 2018), the challenge for future research could be to bring the results back into a single framework to compare the use of sustainability issues in texts with that of images.

Finally, future research could use the indicators proposed in this study to investigate factors associated with online sustainability communication at the destination level as well as outcomes in terms of influence on sustainable tourism behaviors. For instance, the approach proposed in this study could be used to explore the alignment of online sustainability communication with destinations' strategies and performances in relation to sustainable tourism, which represents a crucial but still unexplored issue (Spinelli, 2021), also including the implementation of sustainable destination management systems in line with Cannas (2018).

The study does not address the reception of sustainability messages, but directions for further research in this regard can be envisaged. The dictionary could be also used to investigate the influence of destination web-based communication in terms of the presence and balance of different sustainability dimensions and topics on visitors' awareness and

adoption of sustainable behaviors. Research could verify the effectiveness of a balanced approach to sustainability communication across it multiple dimensions in terms of audience impact by using the proposed indicators to assess and compare multiple urban destinations.

CRediT authorship contribution statement

Valentina Marchi: Conceptualization, Methodology, Data curation, Formal analysis, Writing – review & editing. **Alessandra Marasco:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Valentina Apicerni:** Conceptualization, Methodology, Data curation, Writing – review & editing.

Declaration of competing interest

Authors have no conflicts of interest to disclose.

Data availability

Data will be made available on request.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cities.2023.104590>.

References

- Aall, C., & Koens, K. (2019). The discourse on sustainable urban tourism: The need for discussing more than overtourism. *Sustainability*, 11(15), 4228.
- Adamus-Matuszynska, A., Dzik, P., Michnik, J., & Polok, G. (2021). Visual component of destination brands as a tool for communicating sustainable tourism offers. *Sustainability*, 13, 731.
- Ali, A., & Frew, A. J. (2014). Technology innovation and applications in sustainable destination development. *Information Technology & Tourism*, 14, 265–290.
- Asmelash, A. G., & Kumar, S. (2019). Assessing progress of tourism sustainability: Developing and validating sustainability indicators. *Tourism Management*, 71, 67–83.
- Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557–590.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Bogren, M., & Sörensson, A. (2021). Tourism companies' sustainability communication—creating legitimacy and value. *Scandinavian Journal of Hospitality and Tourism*, 21(5), 475–493.
- Buckley, R. (2012). Sustainable tourism: Research and reality. *Annals of Tourism Research*, 39(2), 528–546.
- Cannas, R. (2018). Communicating actions for sustainable tourism development. The implementation of the European tourism indicator system for sustainable destinations in South Sardinia. *Almatourism*, 18, 105–128.
- Chung, N., Lee, H., Ham, J., & Koo, C. (2021, January). Smart tourism cities' competitiveness index: A conceptual model. In *Information and communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism conference, January 19–22, 2021* (pp. 433–438). Cham: Springer International Publishing.
- Cities, C. C. (2018). *The sustainable Cities index 2018*. Arcadis: Amsterdam, The Netherlands. Available: https://www.arcadis.com/campaigns/citizencentriccities/1images/%7b1d5ae7e2-a348-4b6e-b1d7-6d94fa7d7567%7dsustainable_cities_index_2018_arcadis.pdf.
- d'Angella, F., & De Carlo, M. (2016). Orientation to sustainability and strategic positioning of destinations: An analysis of international tourism websites. *Current Issues in Tourism*, 19(7), 624–633.
- Day, J. (2020). Sustainable tourism in cities. In *Routledge handbook of tourism Cities* (pp. 52–64). Routledge.
- Deng, Q., Hine, M. J., Ji, S., & Sur, S. (2019). Inside the black box of dictionary building for text analytics: A design science approach. *Journal of International Technology and Information Management*, 27(3), 119–159.
- Diéguez-Castrillón, M. I., Gueimonde-Canto, A., & Rodríguez-López, N. (2022). Sustainability indicators for tourism destinations: Bibliometric analysis and proposed research agenda. *Environment, Development and Sustainability*, 24(10), 11548–11575.
- Dodds, R., Novotny, M., & Harper, S. (2020). Shaping our perception of reality: Sustainability communication by Canadian festivals. *International Journal of Event and Festival Management*, 11(4), 473–492.
- Eichelberger, S., Heigl, M., Peters, M., & Pikkemaat, B. (2021). Exploring the role of tourists: Responsible behavior triggered by the COVID-19 pandemic. *Sustainability*, 13(11), 5774.

- European Commission, & Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. (2017). *The European tourism indicator system: ETIS toolkit for sustainable destination management*. Publications Office.
- European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. (2022). Transition pathway for tourism. Available doi: <https://doi.org/10.2873/344425>.
- European Travel Commission. (2021). *Handbook on encouraging sustainable tourism practices*. Brussels: ETC. Available: <https://etc-corporate.org/reports/handbook-on-encouraging-sustainable-tourism-practices/>.
- European Travel Commission (2022). Sustainable travel in an era of disruption: impact of Covid-19 on sustainable tourism attitudes ETC, Brussels. Available: <https://etc-corporate.org/reports/sustainable-travel-in-an-era-of-disruption-impact-of-covid-19-on-sustainable-tourism-attitudes/>.
- Font, X., Elgammal, I., & Lamond, I. (2017). Greenhushing: The deliberate under communicating of sustainability practices by tourism businesses. *Journal of Sustainable Tourism*, 25(7), 1007–1023.
- Garbelli, M., Adukaite, A., & Cantoni, L. (2017). Value perception of world heritage sites and tourism sustainability matters through content analysis of online communications. The case of Victoria Falls world heritage site. *Journal of Hospitality and Tourism Technology*, 8(3), 417–431.
- Ghanem, M., & Elgammal, I. (2017). Communicating sustainability through a destination's website: A checklist to inform, motivate, and engage stakeholders. *Journal of Travel & Tourism Marketing*, 34, 793–805.
- Global Sustainable Tourism Council (2016). *GSTC Industry Criteria*. Available: <https://www.gstccouncil.org/gstc-criteria/>.
- Hovy, D., Melumad, S., & Inman, J. J. (2021). Wordify: A tool for discovering and differentiating consumer vocabularies. *Journal of Consumer Research*, 48(3), 394–414.
- Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, 64(6), 2833–2855.
- Huovila, A., Bosch, P., & Airaksinen, M. (2019). Comparative analysis of standardized indicators for smart sustainable cities: What indicators and standards to use and when? *Cities*, 89, 141–153.
- Ivares-Baidal, J. A., Celdrán-Bernabeu, M. A., Femenia-Serra, F., Perles-Ribes, J. F., & Vera-Rebollo, J. F. (2023). Smart city and smart destination planning: Examining instruments and perceived impacts in Spain. *Cities*, 137, 104266.
- Ivares-Baidal, J. A., Vera-Rebollo, J. F., Perles-Ribes, J., Femenia-Serra, F., & Celdrán-Bernabeu, M. A. (2021). Sustainable tourism indicators: what's new within the smart city/destination approach? *Journal of Sustainable Tourism*, 1–24.
- Joseph, C., Lin, V. C. S., Nichol, E. O., & Jussem, P. M. (2014). *Sustainability disclosure on Malaysian hotel websites*. In . 29.
- Koens, K., Melissen, F., Mayer, I., & Aall, C. (2021). The Smart City hospitality framework: Creating a foundation for collaborative reflections on overtourism that support destination design. *Journal of Destination Marketing & Management*, 19, 100376.
- Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities—an experiment on Stockholm. *Land Use Policy*, 71, 24–35.
- Kourtit, K., Nijkamp, P., & Steenbruggen, J. (2017). The significance of digital data systems for smart city policy. *Socio-Economic Planning Sciences*, 58, 13–21.
- Lane, B. (2018). Will sustainable tourism research be sustainable in the future? An opinion piece. *Tourism Management Perspective*, 25, 161–164.
- Law, R., Qi, S., & Buhalis, D. (2010). Progress in tourism management: A review of website evaluation in tourism research. *Tourism Management*, 31(3), 297–313.
- Li, S., Saayman, A., Stienmetz, J., & Tussyadiah, I. (2022). Framing effects of messages and images on the willingness to pay for pro-poor tourism products. *Journal of Travel Research*, 61(8), 1791–1807.
- Luhn, H. P. (1960). Key word-in-context index for technical literature (kwic index). *American Documentation*, 11(4), 288–295.
- Lupu, C., Rodrigues, A. I., Stoleriu, O. M., & Gallarza, M. G. (2021). A textual and visual analysis of the intrinsic value dimensions of Romania: Towards a sustainable destination brand. *Sustainability*, 13, 67.
- Marchi, V., Apicerni, V., & Marasco, A. (2021). Assessing online sustainability communication of Italian cultural destinations—a web content mining approach. In *Information and communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism conference, January 19–22, 2021*(pp. 58–69). Springer international publishing.
- Mariani, M., & Borghi, M. (2021). Are environmental-related online reviews more helpful? A big data analytics approach. *International Journal of Contemporary Hospitality Management*, 33(6), 2065–2090.
- Marin, V., Palmisani, F., Ivaldi, R., Dursi, R., & Fabiano, M. (2009). Users' perception analysis for sustainable beach management in Italy. *Ocean Coastal Management*, 52, 268–277.
- Marinova, D., Singh, S. K., & Singh, J. (2018). Frontline problem-solving effectiveness: A dynamic analysis of verbal and nonverbal cues. *Journal of Marketing Research*, 55(2), 178–192.
- Michopoulou, E., & Buhalis, D. (2013). Information provision for challenging markets: The case of the accessibility requiring market in the context of tourism. *Information & Management*, 50(5), 229–239.
- Miller, D., Merrilees, B., & Coghlan, A. (2015). Sustainable urban tourism: Understanding and developing visitor pro-environmental behaviours. *Journal of Sustainable Tourism*, 23(1), 26–46.
- Mura, P., & Sharif, S. P. (2015). Exploring rural tourism and sustainability in Southeast Asia through the lenses of official tourism websites. *Worldwide Hospitality & Tourism Themes*, 7(5), 440–452.
- Nie, R. X., Tian, Z. P., Wang, J. Q., & Chin, K. S. (2020). Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory. *International Journal of Hospitality Management*, 88, 102495.
- Nijkamp, P., & Kourtit, K. (2023). *Culture and Tourism: A Sustainable Dilemma? Keeping the Focus on Sustainability: A Challenge for Governance*. 90.
- Pan, B., & Fesenmaier, D. R. (2006). Online information search: Vacation planning process. *Annals of Tourism Research*, 33(3), 809–832.
- Park, E., Kang, J., Choi, D., & Han, J. (2020). Understanding customers' hotel revisiting behaviour: A sentiment analysis of online feedback reviews. *Current Issues in Tourism*, 23(5), 605–611.
- Park, J. R., Lu, C., & Marion, L. (2009). Cataloging professionals in the digital environment: A content analysis of job descriptions. *Journal of the American Society for Information Science and Technology*, 60(4), 844–857.
- Pasquinelli, C., Trunfio, M., Bellini, N., & Rossi, S. (2022). Reimagining urban destinations: Adaptive and transformative city brand attributes and values in the pandemic crisis. *Cities*, 124, 103621.
- Pato, M. L., & Duque, A. S. (2021). Sustainability communication in rural tourism: Website content analysis in Viseu Dão Lafões region (Portugal). *Sustainability*, 13(16), 8849.
- Pencarelli, T., Splendiani, S., & Fraboni, C. (2016). Enhancement of the “blue flag” eco-label in Italy: An empirical analysis. *Anatolia*, 27, 28–37.
- Pencle, N., & Mălaescu, I. (2016). What's in the words? Development and validation of a multidimensional dictionary for CSR and application using prospectuses. *Journal of Emerging Technologies in Accounting*, 13(2), 109–127.
- Peng, N., & Chen, A. (2019). Luxury hotels going green – The antecedents and consequences of consumer hesitation. *Journal of Sustainable Tourism*, 27(9), 1374–1392.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.
- Pennington-Gray, L., & Thapa, B. (2004). DMOs and culturally responsible behaviors: An exploratory analysis. *Tourism*, 52(2), 183–194.
- Penz, E., Hofmann, E., & Hartl, B. (2017). Fostering sustainable travel behavior: Role of sustainability labels and goal-directed behavior regarding touristic services. *Sustainability*, 9(6), 1056.
- Rasoolimanesh, S. M., Ramakrishna, S., Hall, C. M., Esfandiari, K., & Seyfi, S. (2020). A systematic scoping review of sustainable tourism indicators in relation to the sustainable development goals. *Journal of Sustainable Tourism*, 1–21.
- Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The evaluative lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods*, 50(4), 1327–1344.
- Santos, M. C., Veiga, C., Águas, P., & Santos, J. A. C. (2019). Sustainability communication in hospitality in peripheral tourist destinations. *Worldwide Hospitality Tourism Themes*, 11(6), 660–676.
- Santos, M. C., Veiga, C., Santos, J. A. C., & Águas, P. (2022). Sustainability as a success factor for tourism destinations: A systematic literature review. *Worldwide Hospitality and Tourism Themes*, 14(1), 20–37.
- Šegota, T., Sigala, M., Gretzel, U., Day, J., Kokkrankal, J., Smith, M., ... Rakić, T. (2019). Future agendas in urban tourism research: Special editorial. *International Journal of Tourism Cities*, 5(2), 109–124.
- Singrodia, V., Mitra, A., & Paul, S. (2019, January). A review on web scrapping and its applications. In *2019 international conference on computer communication and informatics (ICCCI)*(pp. 1–6). IEEE.
- Smith, V. L., & Font, X. (2015). Marketing and communication of responsibility in volunteer tourism. *Worldwide Hospitality Tourism Themes*, 7(2), 159–180.
- Solís-Radilla, M. M., Hernández-Lobato, L., Callarisa-Fiol, L. J., & Pastor-Durán, H. T. (2019). The importance of sustainability in the loyalty to a tourist destination through the Management of Expectations and Experiences. *Sustainability*, 11(15).
- Spinelli, R. (2021). Communicating and promoting sustainability in tourism: An analysis of the official websites of the Italian regions. *European Journal of Tourism Research*, 29, 2916.
- STATISTA. (2019). Leading European city tourism destinations in 2019, by number of bednights. Available: <https://www.statista.com/statistics/314340/leading-european-city-tourism-destinations-by-number-of-bednights/>.
- Tölkes, C. (2018). Sustainability communication in tourism – A literature review. *Tourism Management Perspectives*, 27, 10–21.
- Tölkes, C. (2020). The role of sustainability communication in sustainable tourism's attitude-behaviour gap. *Tourism and Hospitality Research*, 20(1), 117–128.
- Torres-Delgado, A., & Saarinen, J. (2017). Using indicators to assess sustainable tourism development: A review. *New research paradigms in tourism geography*, 31–47.
- UNEP and UNWTO. (2005). *Making Tourism More Sustainable - A Guide for Policy Makers*. Available: https://wedocs.unep.org/bitstream/handle/20.500.11822/8741/Making%20Tourism%20More%20Sustainable_%20A%20Guide%20for%20Policy%20Makers-2005445.pdf?sequence=3&isAllowed=y.
- UNESCO. (2013). *The Hangzhou declaration. Placing culture at the heart of sustainable development*. International congress “culture: Key to sustainable development”, Hangzhou, China. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000221238>.
- United Cities and Local Governments. (2010). *Culture: Fourth pillar of sustainable development, world summit of local and regional leaders*. Mexico: Mexico City. Available: https://www.agenda21culture.net/sites/default/files/files/documents/en/zz_culture4pillarsd_eng.pdf.
- United Nations Conference on Environment and Development. (1992). *Rio de Janeiro, Brazil*. Available: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N92/83/6/55/PDF/N9283655.pdf?OpenElement>.
- UNWTO. (1998). *Guide for local authorities on developing sustainable tourism*. Madrid: World Tourism Organization, ISBN 92-844-0280-8.

- UNWTO. (2020). Framework Convention on Tourism Ethics, UNWTO, Madrid. Available: doi:10.18111/9789284421671.
- Villarino, J., & Font, X. (2015). Sustainability marketing myopia: The lack of persuasiveness in sustainability communication. *Journal of Vacation Marketing*, 21 (4), 326–335.
- VisitEngland. (2014). *Keep it real for destinations: A guide to incorporating wise growth in your communications with visitors, residents and businesses*. London: VisitEngland.
- WCED. (1987). Our Common future: Report of the World commission of environment and development United Nations. Available: <https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf>.
- Wehrli, R., Priskin, J., Demarmels, S., Schaffner, D., Schwarz, J., Truniger, F., & Stettler, J. (2014). How to communicate sustainable tourism products to customers: Results from a choice experiment. *Current Issues in Tourism*, 1–20.
- Xiang, Z. (2018). From digitization to the age of acceleration: On information technology and tourism. *Tourism Management Perspectives*, 25, 147–150.
- Zhai, C., & Massung, S. (2016). *Text data management and analysis: A practical introduction to information retrieval and text mining* (Morgan & Claypool).
- Zolfani, S. H., Sedaghat, M., Maknoon, R., & Zavadskas, E. K. (2015). Sustainable tourism: A comprehensive literature review on frameworks and applications. *Economic Research-Ekonomska Istraživanja*, 28(1), 1–30.