



Deliverable D2.2

ANALYSIS OF GAPS AND OVERLAPS



**Funded by
the European Union**

28/02/2023

PROJECT INFORMATION

Project Number	101084007
Project Acronym	DATES
Project Full title	European Data Space for Tourism
Project Start Date	November 1 2022
Project End Date	October 31 2023
Project Duration	12 months
Funding Instrument	Digital Europe
Funding Scheme	CSA – Coordination and Support Action
Topic	DIGITAL-2021-PREPACTS-DS-01-TOURISM
Contract	101084007
Coordinator	AnySolution S.L.

DELIVERABLE INFORMATION

Deliverable No	D2.2
Deliverable Title	Analysis of gaps and overlaps
Work-Package No	2
Work-Package Title	Context analysis and agreed components
WP-Leader (Name and Short Org. Name)	INTELLERA (INTEL)
Tasks No	T2.2
Tasks Title	Analysis of existing data platforms and data sharing initiatives
Task Leader (Name and Short Org. Name)	INTELLERA
Main Authors (Name and Short Org. Name)	Giovanna Galasso – INTELLERA Marco Codastefano – INTELLERA Alessandro La Rosa - INTELLERA Beatrice Dorenti – INTELLERA Carlotta Colagrosso - INTELLERA Ginevra Bianco - INTELLERA
Other Authors (Name and Short Org. Name)	INTELLERA
Reviewers (Name and Short Org. Name)	Dolores Ordoñez- ANYSOL Jesus Herrero - TECNALIA Martin Soutschek - OUTDOORACTIVE AG Olivier Dion - ANEWGOVERNANCE
Status	Draft <input checked="" type="checkbox"/> Final <input type="checkbox"/>
Deliverable Type	Report <input checked="" type="checkbox"/> Data <input type="checkbox"/> Demonstration <input type="checkbox"/> Data <input type="checkbox"/>

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

Dissemination Level	Public (PU) <input checked="" type="checkbox"/> Sensitive (SEN) <input type="checkbox"/> EU Classified (EUC) <input type="checkbox"/> <i>PU: fully open (automatically posted online on the project Results platforms)</i> <i>SEN: limited under the conditions of the Grant Agreement</i> <i>EUC: WU-RESTRICTED, EU-CONFIDENTIAL., EU-SECRET under Decision 2015/444.</i>
Date Approved by Coordinator	28/02/2023

DOCUMENT REVIEW HISTORY

Date	Version	Reviewer Name (Short Organisation Name)	Summary of main changes
23/02/2023	V0.1	Intellera	Full report and annexes
28/02/2023	V0.2	Anysolution, Outdooractive, Tecnalía	Syntactic revision and clarification comments
28/02/2023	V1	Intellera	Final version
08/03/2023	V1.1	Intellera	Modifications under the request of the PO
13/03/2023	V2	Intellera	Final version
28/07/2023	V2.1	Intellera	Updated data, cluster analysis and taxonomy analysis according to the new inventory.

FINAL ACCEPTANCE BY COORDINATOR

DATES CONSORTIUM			
Part. No	Participant organisation name	Short Name	Country
1	ANYSOLUTION SL	ANYSOL	ES
2	FUNDACION TECNALIA RESEARCH & INNOVATION	TECN	ES
3	IDC ITALIA SRL	IDC	IT
4	MINISTERO DEL TURISMO	MITUR	IT
5	INTELLERA CONSULTING SRL	INTELLERA	IT
6	AMADEUS SAS	AMAD	FR
7	INTERNATIONAL DATA SPACES EV	IDSA	DE
8	ARCTUR RACUNALNISKI INZENIRING DOO	ARCTUR	SI
9	NETWORK OF EUROPEAN REGIONS FOR COMPETITIVE AND SUSTAINABLE TOURISM ASBL	NECS	BE
10	OUTDOORACTIVE AG	OUTD	DE
11	DIH TOURISM 4.0, ZU	DIHT	CZ
12	AVORIS RETAIL DIVISION SL	AVOR	ES

15	ANEWGOVERNANCE	ANG	BE
----	----------------	-----	----

LEGAL DISCLAIMER

The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. The information in this document is provided "as is". Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. The DATES Consortium members shall have no liability for damages of any kind, including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials, subject to any liability which is mandatory due to applicable law.

INDEX

List of Tables.....	7
List of Figures.....	7
EXECUTIVE SUMMARY.....	8
1 Introduction.....	9
1.1 Objectives of the report	9
2 Methodology.....	10
2.1 Data collection activities	10
2.1.1 Workshop methodology.....	10
2.2 Approach used for the analysis.....	12
2.2.1 Data sharing initiatives inventory analysis.....	12
2.2.2 Cluster analysis	12
2.2.3 Taxonomy analysis.....	13
3 Analysis.....	14
3.1 Analysis of the data sharing initiatives inventory (D2.1)	14
3.1.1 Geographical level and partner type.....	14
3.1.2 Challenges addressed.....	16
3.1.3 Tourism sub-sectors	17
3.1.4 Data purpose.....	18
3.1.5 Data user	19
3.1.6 Data sources.....	20
3.1.7 Data collection mode and data sharing strategy.....	21
3.2 Cluster analysis of the data sharing initiatives.....	22
3.2.1 Business	22
3.2.2 Data	23
3.2.3 Governance.....	22
3.2.4 Geographical.....	24
3.3 Taxonomy.....	25
3.3.1 Business	27
3.3.2 Data	29
3.3.3 Governance.....	30
3.3.4 Geographical.....	30
3.4 Workshop results analysis.....	32
3.5 Main tourism sector data gaps.....	34
3.5.1 Data gaps and overlaps emerging from the analysis.....	34
3.5.2 Most needed data according to priority business challenges.....	36
3.5.3 Most needed gaps of data characteristics	37
4 Conclusions.....	39
4.1 Overview of key findings.....	39

•

4.2 Next steps.....	39
Annex I - Post Workshop Report	41

List of Tables

Table 1 - Data collection activities.....	10
Table 2 - Legend of taxonomy.....	25
Table 3 - Definition of data sources.....	25
Table 4 - Challenges rating and priority.....	32
Table 5 - Next WP2 activities.....	40

List of Figures

Figure 1 - Geographical level.....	14
Figure 2 - Initiatives per country.....	15
Figure 3 - Partner type.....	16
Figure 4 - Initiatives per challenge addressed.....	17
Figure 5 - Initiatives per tourism sub-sector.....	18
Figure 6 - Initiatives per data purpose.....	19
Figure 7 - Initiatives by data users.....	19
Figure 8 - Initiatives by data source.....	20
Figure 9 - Initiatives per data collection mode.....	21
Figure 10 - Initiatives by data sharing strategy.....	21
Figure 11 - Clusters of business dimension.....	22
Figure 12 - Clusters of data dimension.....	23
Figure 13 - Clusters of geographical dimension.....	24
Figure 14 - Taxonomy of first cluster of business dimension.....	27
Figure 15 - Taxonomy of second cluster of business dimension.....	28
Figure 16 - Taxonomy of third cluster of business dimension.....	29
Figure 17 - Taxonomy of data dimension.....	30
Figure 18 - Taxonomy of first cluster of geographical dimension.....	31
Figure 19 - Taxonomy of second cluster of geographical dimension.....	31
Figure 20 - Taxonomy of third cluster of geographical dimension.....	32

EXECUTIVE SUMMARY

The present document represents the second DATES Work Package 2 deliverable – *Analysis of gaps and overlaps (D2.2)*.

The Introduction (Chapter 1) presents the structure of the report and provides an overview of its objectives.

Chapter 2 encompasses the methodology, including the data collection activities carried out and the approach used for the analysis of the gaps and overlaps among the existing tourism data sharing initiatives. Thus, the Chapter presents an incremental approach to the analysis: it starts with a descriptive analysis of the initiatives mapped in the inventory (Deliverable 2.1), and it deepens presenting the results of the cluster and the taxonomy analysis.

Moving forward, Chapter 3 focuses on confronting the results of the analyses of the inventory, the clusters, the taxonomy, and the workshop, ultimately outlining the main tourism sector data gaps and overlaps. The analysis of workshop results highlights that the most important data for the sector (i.e. the ones associated with the sector's most important challenges) are the ones regarding tourists' behaviour, mobility data and demand and offer data. Similarly, the most needed data (i.e. the ones associated with the sector's challenges to be addressed with the highest priority) are those on tourist behaviour, mobility data, and tourist flow data. At the same time, the results of the survey highlight that data on demand and offer, sustainability, mobility, and behaviour of tourists are currently missing. The comparison of such workshop and survey results underlines the importance of filling the data gaps regarding tourists' behaviour, mobility data and demand and offer data.

Moreover, results of the inventory, cluster and taxonomy analysis highlight several other data gaps present in the current data sharing initiatives, such as the limited availability of data at local level, of data regarding car and other rentals, of user generated contents, private business datasets and card transactions data.

The chapter further highlights how incomplete data, lack of interoperability and data not being updated in a timely manner are the most common shortcomings of the current available data, and how there is a lack of cooperation and coordination for sharing data in the tourism sector.

Ultimately, Chapter 4 presents an overview of the key findings, and the next project activities.

1 Introduction

The present report presents the results of the analyses carried out within Work Package 2 (WP2) *Task 2.2 Analysis of existing data platforms and data sharing initiatives*, with the aim of analysing gaps and overlaps among the tourism-related data sharing initiatives (Deliverable 2.1).

The report is structured as follows:

- Chapter 1 “Introduction” provides an overview of the report objectives;
- Chapter 2 “Methodology” presents the data collection activities and the incremental approach adopted to conduct the analysis of gaps and overlaps among the existing tourism data sharing initiatives;
- Chapter 3 “Analysis” which showcases the results of the inventory analysis, the cluster analysis, the taxonomy analysis and the workshop;
- Chapter 4 “Conclusions” provides an overview of the key findings, as well as an overview of the next activities that will be carried out within the scope of DATES Work Package 2 and Work Package 3.

The report further includes Annex I, which consists of the post workshop report *‘Towards a Data Space for Tourism, Prioritization of data needs and data purposes’*.

1.1 Objectives of the report

The objective of this report is to present the results stemming from the analysis of existing data platforms and data sharing initiatives. Accordingly, the report will showcase how the project team analysed the gaps and overlaps among the previously identified initiatives and ultimately identified a set of priority data needs to be addressed.

2 Methodology

2.1 Data collection activities

To identify the gaps and overlaps of the existing data sharing initiatives in the tourism sector, the project team relies on the evidence collected in the first four months of the project in the context of Work Package 2 “Context analysis and agreed components”. The data collection activities conducted are depicted in the following table.

Table 1 - Data collection activities

Data collection activity	Timing	Aim of the activity
Desk research	Nov 2022 - ongoing	Deepening the team’s knowledge on the topics of data use in the tourism sector; tourism stakeholders’ challenges that can be solved by enhanced data sharing; the notion of data spaces and how they could enhance the use of data in the sector.
	Nov 2022 – Jan 2023	Identifying the EU and extra-EU data sharing initiatives focused on the tourism sector. These have been included in the <i>D2.1 Data sharing initiatives inventory</i>
Survey	15 Dec 2022 – 16 Jan 2023	Exploring (i) which data are needed by tourism sector stakeholders, and (ii) which data sharing initiatives in the tourism sector are known/used by sector stakeholders. The questionnaire also aimed at creating a list of stakeholders interested in being invited in future project consultations.
Workshop	8 Feb 2023	(i) Raising awareness on the concept of data spaces and the DATES project; and (ii) setting the ground for use cases development and prioritization of the sector’s data gaps, by prioritizing a series of stakeholders’ business challenges and the related types of data.

A detailed description of the methodology used for developing the data sharing inventory and the survey has been presented in the report accompanying the *D2.1 Data sharing initiatives inventory*. The methodology used for the workshop is presented in paragraph 2.1.1.

2.1.1 Workshop methodology

The workshop, entitled “Towards a Data Space for Tourism- Prioritization of data needs and data purposes”, took place on the 8th of February, from 10:00 AM until 12:15PM.

The workshop was structured in two sections.

The objective of the first section of the workshop was to raise awareness on the concept of data spaces and on the DATES project. After debriefing the audience with the project's objectives and its benefits for the tourism sector, the policy context and the key characteristics of a data space were showcased, followed by the practical example of EONA-X, the mobility, tourism and transport data space. Afterwards, the initial findings of the project were presented and the setup of the second section of the workshop was introduced.

Accordingly, the second section of the workshop focused on prioritizing a series of business challenges that the participating stakeholders face in their day-to-day activities, which could be potentially solved by an enhanced use of data. This activity allowed to define the stakeholders' data needs, as they were asked to associate different types of data with various challenges pertaining to the tourism ecosystem. To do so, participants were split into four break-out rooms, one for each identified data purpose. The break-out rooms were:

Breakout room I: Increase of tourism sustainability & accessibility;

Breakout room II: Conduct market analysis & inform decision-making;

Breakout room III: Improvement of the interaction and engagement of the tourist;

Breakout room IV: Improvement of planning and operations of tourism services.

Each break-out room followed the same structure of activities:

- For each presented challenge, participants were asked to rate its importance (on a scale from 1 to 5, where 1 was absolutely not important and 5 very important) and to mention which types of data they would need to address it;
- Then, they were asked to rank the priority of the challenges (by positioning the one with the highest priority at the top of the list and the one with the lowest priority at the bottom);
- Finally, they were asked an open question to investigate whether there was any missing challenge, assign a rating for it and highlight the data needed to address it.

207 participants attended the event. It is worth mentioning that the workshop participants covered several categories of stakeholders, ranging from destination management organizations to consulting firms and public authorities. Accordingly, the participants were distributed in a way that the same proportion of each stakeholders category was present in each break-out room.

2.2 Approach used for the analysis

To conduct the analysis of gaps and overlaps among the existing tourism data sharing initiatives, the project team adopted an **incremental approach** to the analysis. It starts from the analysis of the inventory, and it deepens through the cluster and taxonomy analysis. Firstly, the report analyses the single dimensions of the data sharing initiatives inventory individually. Secondly, thanks to the cluster and taxonomy analysis, it compares the initiatives based on the clusters identified according to different dimensions.

The methodology used for each analysis is explained in the following sections.

2.2.1 Data sharing initiatives inventory analysis

This analysis focuses on each dimension of the data sharing initiatives inventory (Deliverable 2.1) individually. To detect the gaps and overlaps of the shared tourism-related data, the team presents descriptive statistics to point out how many of all initiatives cover a specific inventory dimension.

2.2.2 Cluster analysis

To identify and understand the characteristics of the data sharing initiatives mapped in the inventory (Deliverable 2.1), a cluster analysis was performed. Cluster analysis is a statistical method for processing data, and it works by organizing items into groups (clusters) based on how closely associated they are.

The cluster analysis allowed the grouping of data sharing initiatives in different categories, based on the following four dimensions.

1. **Business** (partner type, challenge addressed and tourism sub-sectors);
2. **Data** (data purpose, data users and data sources);
3. **Governance** (collection mode, data sharing strategy and partner type);
4. **Geographical** (geographical level and partner type).

2.2.2.1 Clustering method

Identifying groups of similar observations in datasets can be done with many clustering methods developed in statistics, machine learning, and the applied sciences. Clustering is the process of dividing a dataset into a plurality of groups and clusters composed of similar objects. The common clustering method used in the field of data mining is the K-means algorithm, which measure the similarity between data and the dataset that needs to be classified¹. However, the K-means measurement can only process numeric data and is not efficient when working with categorical data, as in the current case. To identify hidden patterns and groupings in the datasets containing the data sharing initiatives, the project team implemented the K-modes clustering algorithm, which is an extension of the K-means algorithm. The K-modes method is used for grouping categorical data, defining clusters based on the number of matching categories between data points. This algorithm uses simple overlap distance to measure the dissimilarity between objects, resulting in the best method for categorical data.

¹ Yawen Dai, Guanghui Yuan, Zhaoyuan Yang, Bin Wang, "K-Modes Clustering Algorithm Based on Weighted Overlap Distance and Its Application in Intrusion Detection", *Scientific Programming*, vol. 2021, Article ID 9972589, 9 pages, 2021. <https://doi.org/10.1155/2021/9972589>

Before proceeding with the clustering, the project team converted the categorical variable through the one-hot-encoding process, so that data can be provided to machine learning algorithms to improve classification.

2.2.3 Taxonomy analysis

To better understand and analyse possible gaps and overlaps within the data sources², a comprehensive taxonomy has been developed. Generally, a taxonomy represents a hierarchical structure of classes or types of objects within a knowledge domain by using a controlled vocabulary to make it easier to find related information. The purpose of this taxonomy is to discover connections between initiatives and data sources, highlighting the extension of the families and possible gaps in terms of data sources to be filled.

The role of a taxonomy is to determine what particular class a given entity most clearly falls into. Three processes are involved: i) organizing or grouping similar or related data sources into larger categories; ii) identifying differences between sets of subcategories and larger or overarching categories, and iii) representing the relationships among the categories and subcategories of data sources.

² A comprehensive explanation of what is meant by “data source” is provided in section 3.1.6.

3 Analysis

As mentioned above, the study adopts an incremental approach to the analysis. This chapter presents the results of the inventory analysis, cluster analysis, the taxonomy analysis and the workshop.

3.1 Analysis of the data sharing initiatives inventory (D2.1)

This section starts from an overview of all mapped dimensions of the inventory. Sections 3.1.1, 3.1.3, 3.1.4 and 3.1.6 will be at the base of the analysis of the gaps and points of convergence of data shared in the tourism sector.

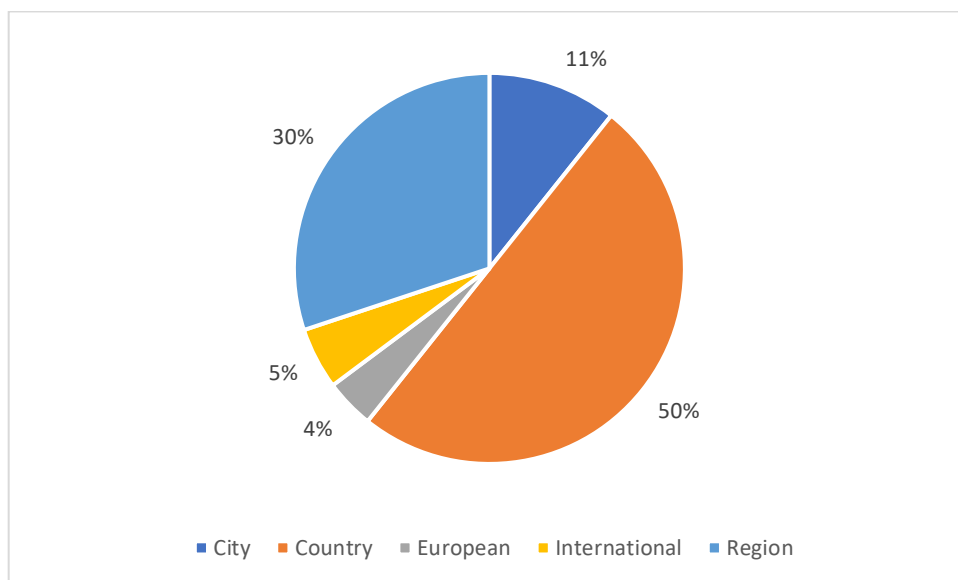
Overall, the project team mapped 194 data sharing initiatives, 36 of which were identified through the survey.

Within the mapped initiatives, 2 data spaces were included, such as Themis-X Tourism Data Space (a consortium of GAIA-X involving 60+ major public and private stakeholders in France) and EONA-X (the European Mobility, Tourism and Travel Dataspace). Additionally, several tourism data platforms and working groups (such as the GAIA-X hubs in France and Spain) were included.

3.1.1 Geographical level and partner type

According to the mapping, these 194 initiatives focus on different geographical levels: 50% (96) cover data at county level, 30% (59) at regional level, 11% (21) at city level, 5% (10) at international level³ and 4% (8) at European level⁴ (figure below).

Figure 1 - Geographical level



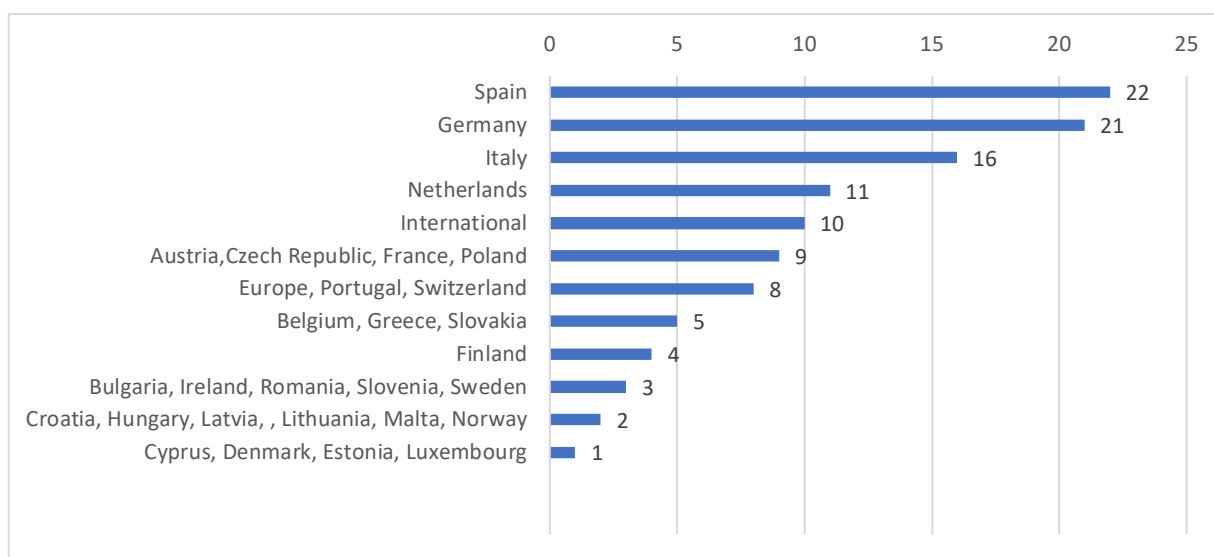
³ The data sharing initiatives labelled as “International” are the ones gathering tourism data of two or more countries worldwide.

⁴ The data sharing initiatives labeled as “European” are the ones owned/partnered by the European Commission or the European Travel Commission.

Among all geographical levels, the initiatives cover 40 countries, of which 74% are EU countries⁵, while 26% are extra-EU⁶. Moreover, 8 initiatives covering Europe and 18 covering two or more countries were identified⁷.

The following figure provides an overview of the number of initiatives identified in each EU country, and at European and International level. According to the inventory, Spain is the country with the most initiatives (22). The second and the third countries with the highest number of initiatives are respectively Germany (21) and Italy (16); while Cyprus, Denmark, Estonia and Luxembourg are the ones with the fewest initiatives (only 1 for each).

Figure 2 - Initiatives per country



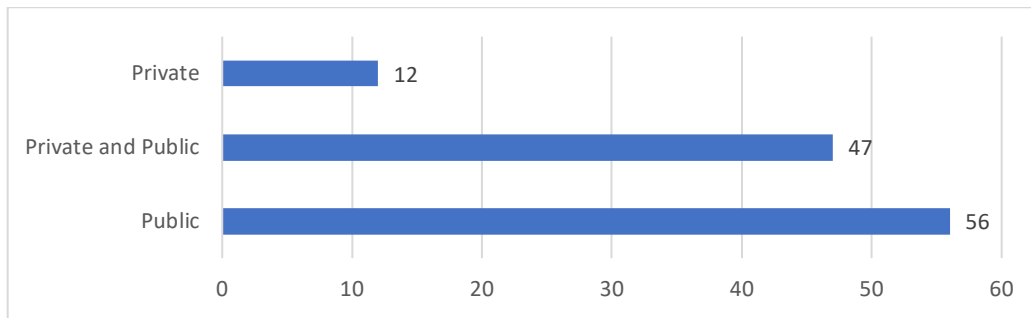
Moreover, the project team tried to identify the type of partners (public organizations or private organizations) taking part in the initiative. Of the 115 initiatives for which it was possible to identify the partner type, 49% (56) have public partners, 40% (47) have private and public partners, while only 11% (12) have exclusively private partners (see figure below).

⁵ Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden.

⁶ Argentina, Australia, Indonesia, Japan, New Zealand, Norway, Northern Ireland, Singapore, Switzerland, Tasmania, Thailand, United Kingdom, and the United States of America.

⁷ Respectively the ones marked as “European” and “International” in Figure 1.

Figure 3 - Partner type



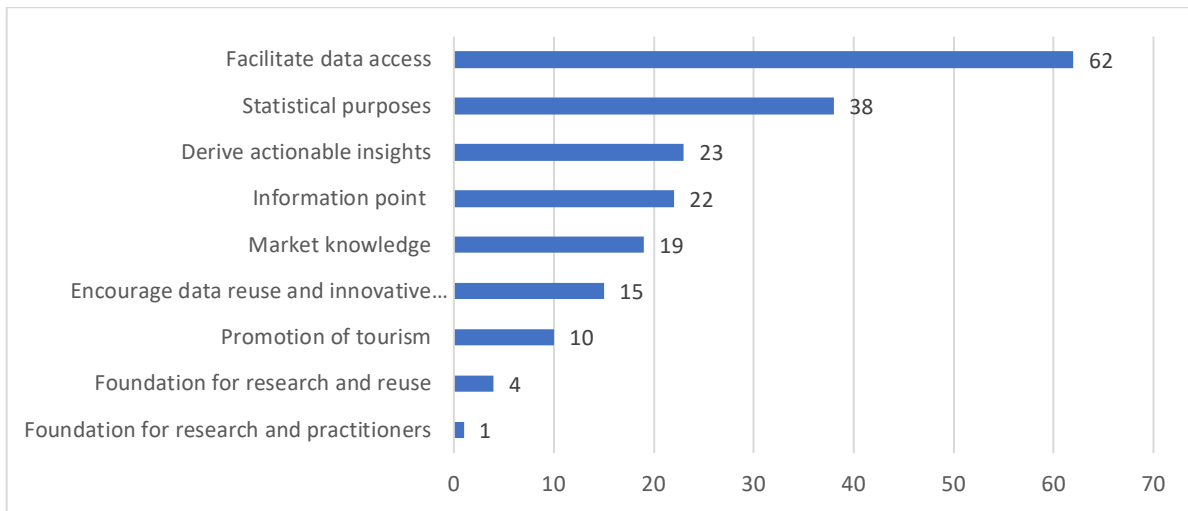
3.1.2 Challenges addressed

Exploring the inventory, the project team assigned to each initiative a label regarding the reason why the initiative was created (i.e. the challenges the initiative aims to address). The identified challenges are:

- **Facilitate data access:** the initiative aims to make data freely available to anyone under the same conditions for any purposes, increasing the transparency of tourism-related data;
- **Statistical purposes:** give access to tourism data is intended for a statistical use;
- **Derive actionable insights:** the aim of sharing data is to generate market intelligence and derive actionable insights;
- **Information point** the initiative aims to become a focal point for tourists, enabling users to find up-to-date information about the destination;
- **Market knowledge,** refers to the challenge of providing a detailed overview on the offer, trend and evolution of demand and tourist markets;
- **Encourage data reuse and innovative service/application development:** the initiative aims to promote the reuse of data by encouraging the development of innovative ideas;
- **Promotion of tourism,** refers to the challenge of promoting and fosters tourism of the destination;
- **Foundation for research and reuse:** give access to a broad information is intended to be a base for both research and reuse by the data users;
- **Foundation for research and practitioners:** Give access to a broad tourism information is intended to be a base for research and practitioners;

As shown in the figure below, the most common reason driving the set-up of the initiatives is the facilitation of data access (31%, 62), while the least common is providing data for researchers and practitioners (1%, 1).

Figure 4 - Initiatives per challenge addressed



3.1.3 Tourism sub-sectors

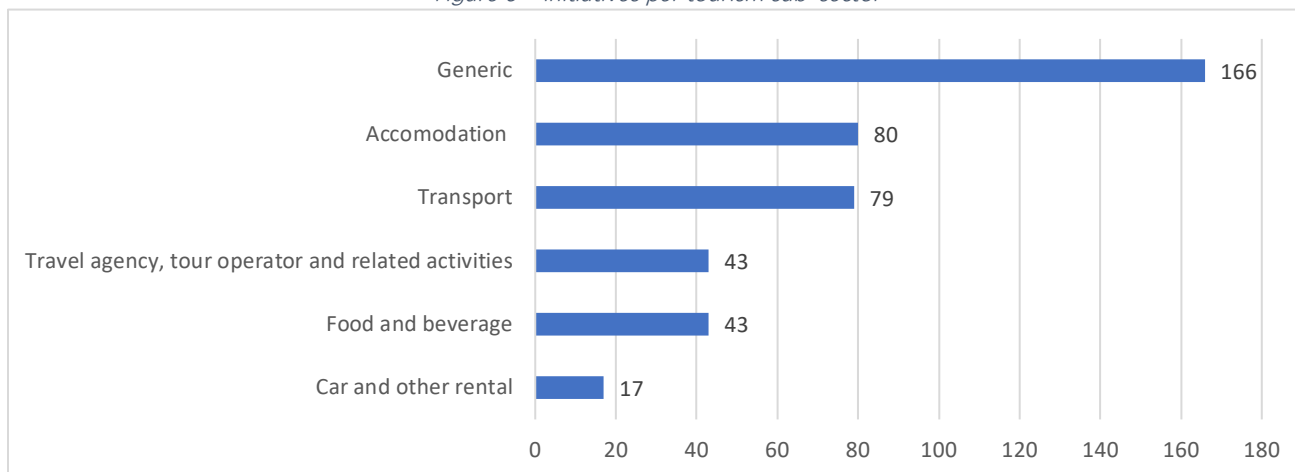
For each initiative, the project team identified the tourism sub-sectors on which it focuses⁸. The sub-sectors are defined as follows:

- **Transport:** the transportation sector aims at helping tourists to get where they need to go. This includes providing them with the means to get to and get back from their destination. It includes services related to road, rail, air and sea travel;
- **Food and beverage:** the food and beverage sector gives tourists essential refreshments at all stages of their travel experience, including during travel, and while spending time in their chosen accommodation;
- **Car and other rental:** having access to a car and other vehicles is an important part of the tourist experience, as it gives tourists the freedom to explore and travel freely. Rental services provide this kind of access and often operate in close proximity to airports, airlines or travel companies;
- **Travel agency, tour operator and related activities:** these entities are focused on connecting customers with travel services they can benefit from, as well as providing customers with important information that can assist them in their travels;
- **Accommodation:** the accommodation sector is central to the travel and hospitality industry, since people travelling to different areas require somewhere to stay, rest, sleep and unwind;
- **Generic:** the initiative covers more sub-sectors, rather than focusing on a specific one.

As depicted in the figure below, the “generic” sub-sector is the most common. This means that in most cases the data sharing initiatives do not only focus on a specific sub-sector, but they cover a variety of topics. The least frequently covered sub-sector is car and other rentals.

⁸ Eurostat- Statistics Explained, *Tourism industries – economic analysis*, March 2022, [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism industries - economic analysis#Analysis by subsectors](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_industries_-_economic_analysis#Analysis_by_subsectors)

Figure 5 - Initiatives per tourism sub-sector



3.1.4 Data purpose

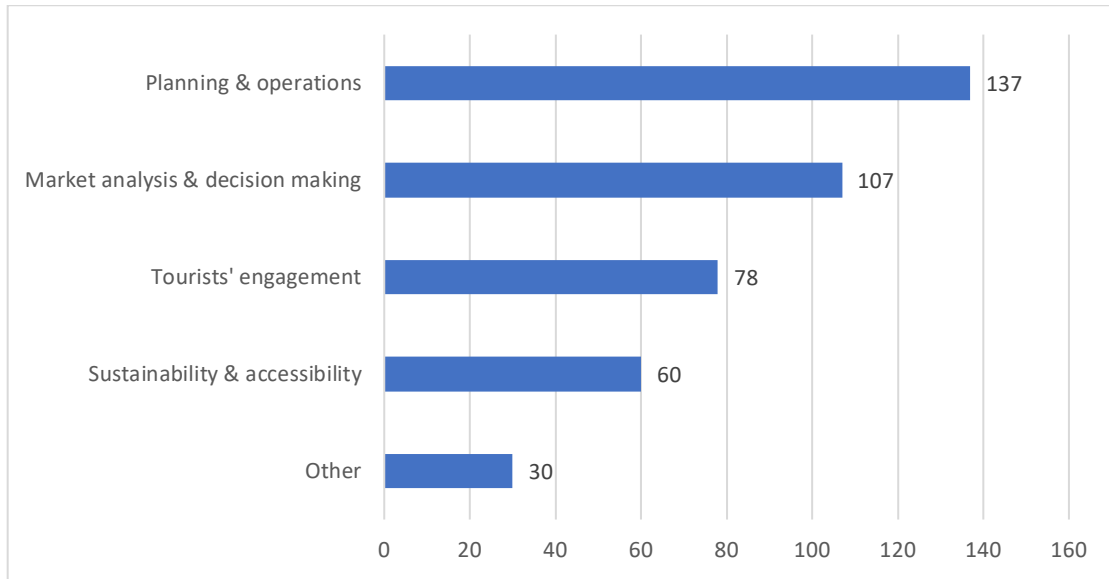
Through desk research, the project team identified the main four purposes for which data are needed in the tourism sector⁹. While compiling the inventory, one or more of the following data purposes were associated to each initiative:

- **Tourists' engagement:** improve interaction and engagement with the tourist, e.g. by developing increasingly personalised tourism services, aiming for a higher degree of interaction with the customer/end-user.
- **Planning & operations:** improve planning and operations of tourism services by understanding (and possibly predicting) tourism patterns, to improve the overall efficiency and competitiveness of the tourism ecosystem. Relevant data on tourism can help the optimisation of stakeholders' offers (e.g. improve their operations, contingency planning, re-scheduling of activities).
- **Market analysis & decision making:** conduct market analyses and inform decision-making through higher availability of data and improved data analytics capabilities. The combination and interoperability of data sources can help data users in extracting meaningful insights, improving their strategies and tailoring their offers.
- **Sustainability & accessibility:** increase sustainability and accessibility of destinations through enhanced data analysis and management, in order to produce positive impacts on society.
- **Other:** any other purpose relevant for the initiative that is not covered by the previous ones.

As shown below, among the 194 inventory initiatives, the most common data purpose is planning and operations, present in 70% (137) of the initiatives. Conversely, the least frequent data purpose is "other", associated with 15% (30) of the initiatives. Among these, the project team identified several reasons for data sharing, the most frequent ones being related to providing statistical information and making information available for research purposes.

⁹ European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, Galasso, G., Montino, C., Sidoti, A., et al., *Study on mastering data for tourism by EU destinations: main text*, Publications Office of the European Union, 2022, <https://data.europa.eu/doi/10.2873/23880>

Figure 6 - Initiatives per data purpose



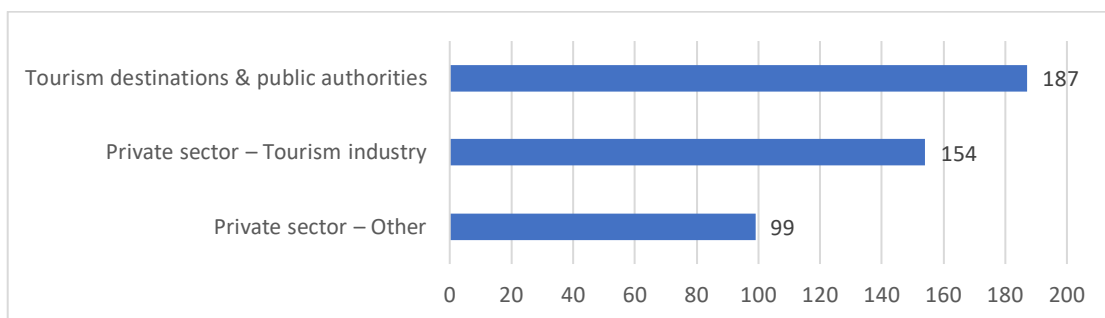
3.1.5 Data user

For each initiative, in order to understand who the end users of the data sharing initiative are, the project team mapped the dimension “data user”. The three categories of users identified are:

- **Tourism destinations & public authorities:** it includes a wide variety of different entities, from multilevel PAs, to education institutions and cultural heritage sites.
- **Private sector – Tourism industry:** a wide spectrum of private actors specialising in the provision of services for the tourism sector (e.g. big vacation rental sites and touristic metasearch engines, hotel groups, and IT and software companies offering tourism-specific data-driven and data analytics services).
- **Private sector – Other:** private companies not directly related to the tourism sector, but capable of acquiring or producing high value datasets.

The figure below shows how many initiatives are targeting these three types of data users. Numbers show that all three categories of users are widely targeted by inventory initiatives. Indeed, tourism destinations & public authorities use 96% (187) of the mapped initiatives, private sector – tourism industry companies use 79% (154) of them and the private sector – other companies use 51% (99).

Figure 7 - Initiatives by data users



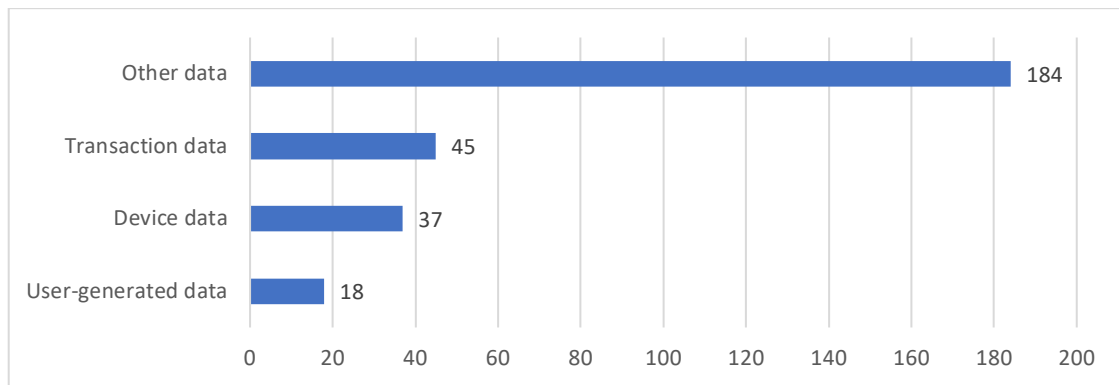
3.1.6 Data sources

An important dimension mapped in the inventory concerns the data sources, i.e. where do the data come from. The data sources could be:

- **User-generated content (UGC):** data is produced and made available by tourists themselves. UGC can be divided into two main sub-categories: textual information (such as reviews, posts, etc.) and photos and videos (usually uploaded on social media, including additional information, such as locations, time and tags);
- **Transaction data:** data is generated anytime a transaction is performed, including during the pre-visit phase. The rise in cashless payment solutions – from shops to public transports, accommodation and tourism sites – generates massive amounts of tourism-related commercial data. Transaction data include: online bookings and purchases (data generated by this kind of transactions), consumer card transactions (including credit cards data, reward cards data and payment cards data), and web search and webpage visits (data generated by web searches related to tourism related transactions).
- **Device data:** data are collected by devices and sensors that allow the tracking of movements, e.g. as part of smart city initiatives. Device data can be generated by the use of technologies, such as GPS, mobile roaming, RFID, Bluetooth, meteorological devices, Wi-Fi; and smart city sensor data (pollution, traffic, waste, etc.).
- **Other data:** high value data coming from other sources, including private businesses datasets (e.g. data on the number of airline passengers), statistics (e.g. datasets published by public authorities), and context-specific information (e.g. information on the history of a place).

As shown in the table below, “other data” is the most common data source in the inventory (94%, 184), whereas the least frequent one is user-generated data (9%, 18). Among “other data”, 59% (108) share statistics, 48% (89) context-specific data, and 17% (31) private business datasets. Among transaction data, the most frequent type of data source is online booking and purchasing data (in 69% of the initiatives, 31), followed by consumer card transactions (29%, 13) and web search and web scraping (4%, 2). Among device data, the most frequent types of data sources come from smart city devices and GPS (62%, 23), followed by mobile roaming and meteorological devices (43%, 16), Wi-Fi data (16%, 6), Bluetooth data (8%, 3) and webcams (3%,1). Finally, among the initiatives sharing user generated content, 67% (12) share textual information, while 33% (6) share photographs and videos.

Figure 8 - Initiatives by data source

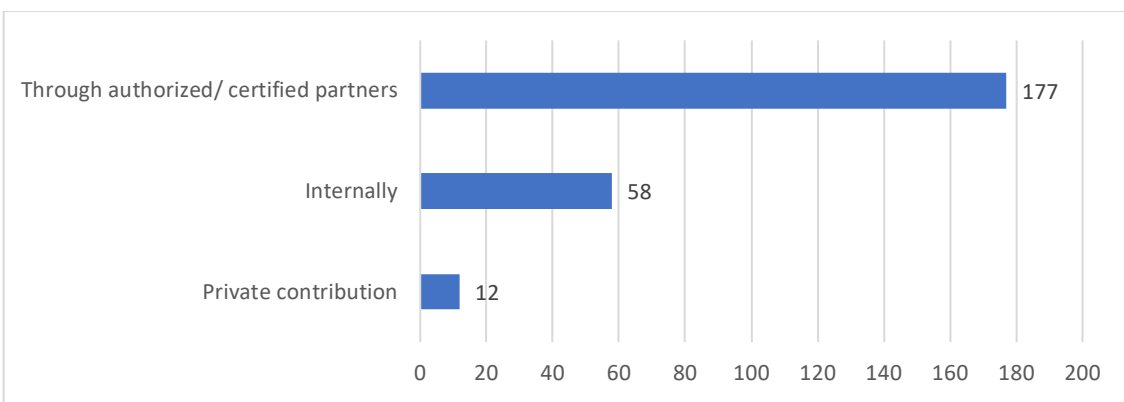


3.1.7 Data collection mode and data sharing strategy

Finally, the inventory presents the data collection mode and data sharing strategy of each initiative. The data collection mode is about how data are collected and from whom¹⁰ (e.g. whether only some specific organizations can provide data or everyone can contribute). Data sharing strategy is about how users can access the data shared through the initiative (e.g. upon payment/for free, upon registration or openly, etc...).

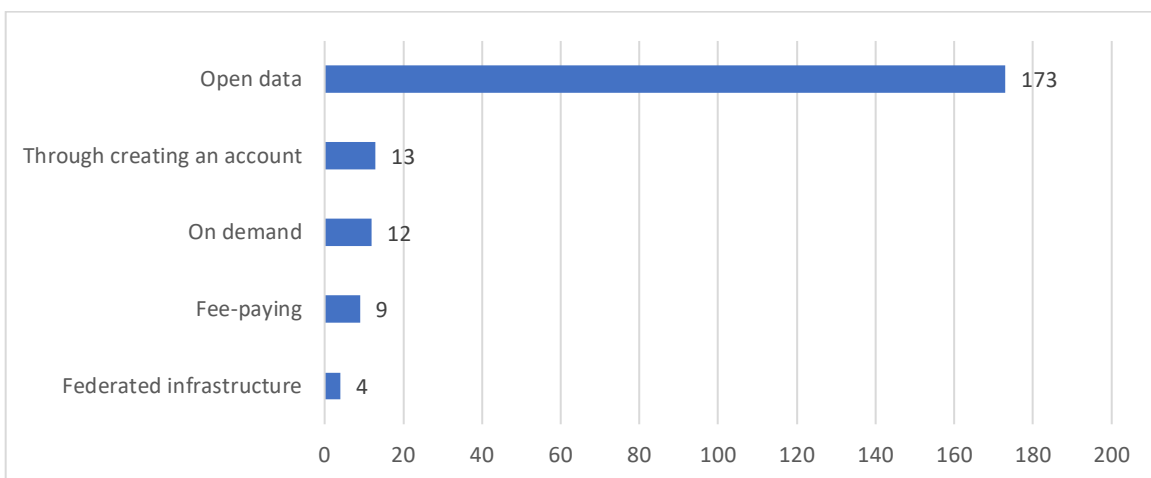
Regarding the data collection mode (figure below), the wide majority of initiatives collect data from authorized and/or certified partners. The second most common data collection mode is “internally”, meaning that data are only pooled from owners and partners taking part in the initiative. The least common collection mode is by gathering data from individuals willing to contribute.

Figure 9 - Initiatives per data collection mode



Regarding the data sharing strategy (figure below), the wide majority of initiatives offer open data, while other ways of accessing data are through an account, on demand, by paying a fee, and through a federated infrastructure (in descending frequency order). The four initiatives sharing data through a federated infrastructure are the tourism-related data spaces mapped (EONA-X, Gaia-X Hub France, Themis-X / Onecub, and Gaia-X Hub Spain).

Figure 10 - Initiatives by data sharing strategy



¹⁰ Either data owner or provider

3.2 Cluster analysis of the data sharing initiatives

The following paragraph provides an overview of the results of the cluster analysis, discussing each cluster from the business, data, geographical and data governance perspective.

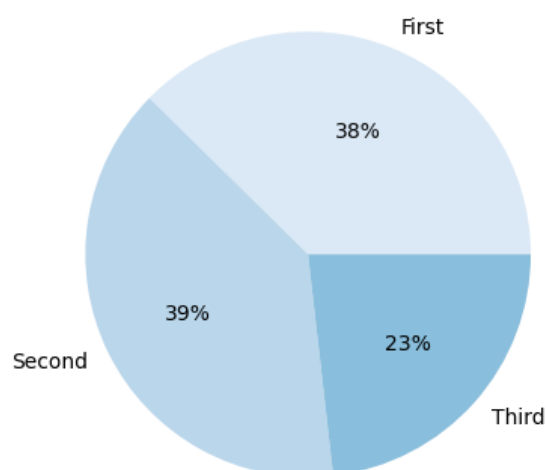
3.2.1 Governance

Regarding the data governance dimension, the cluster analysis revealed that there is not a high correlation between how initiatives collect data and how data are shared. However, predominant evidence is that initiatives with an open data strategy collect data mostly through authorized/certified partners and internally. These data sharing initiatives are mostly led by public partners. Moreover, a second insight concerns the combination of initiatives that share data on demand and collect data through private contributions and through authorized/certified partners.

3.2.2 Business

The business dimension includes aspects, such as the type of partners of the initiatives, the tourism sub-sector covered, and the challenges and issues that the initiative aims to solve. The K-modes clustering algorithm identified 3 clusters within the business dimension, distributed as depicted in Figure 11.

Figure 11 - Clusters of business dimension



3.2.2.1 First cluster – Initiatives led by public partners with the aim of data accessibility

The 73 data sharing initiatives within the first cluster are the ones led by public partners, sharing the goal of facilitating data access to everyone and delivering a more efficient and effective statistical service. Indeed, this cluster includes portals providing official statistics and open data of the country/region involved in the analysis, whose aim is generally to provide statistics as public good to everyone and for free. Given the above-described purposes, these initiatives are related to essential and fundamental sub-sectors such as generic, accommodation and transport, with the aim of providing general information to any users and, in particular, to tourism sector stakeholders.

3.2.2.2 Second cluster – Initiatives with a twofold purpose

The second cluster of the business dimension takes into consideration 76 initiatives whose key characteristics is that there is not a specified or clearly identified partner type. In terms of challenges and

purposes, we can distinguish two sub-groups. On one hand, there are data sharing initiatives that share the goal of facilitating data access to everyone and delivering a more efficient and effective statistical service. On the other hand, there are initiatives with the purpose of providing an understanding of the market and allowing to derive actionable insights from the data provided. Given the above-described purposes, these initiatives are related to essential and fundamental sub-sectors such as generic, accommodation and transport, with the aim of providing general information to any users and, in particular, to tourism sector stakeholders.

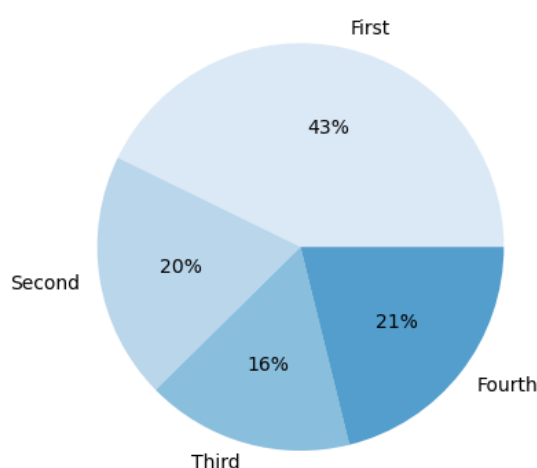
3.2.3.3 Third cluster – Public-private partnerships to foster an innovative and strategic focus

The 45 initiatives related to this cluster present a public-private partnership, with a focus on all tourism sub-sectors. Since the goal of these initiatives is to encourage the reuse of data, the creation of innovative services and derive actionable insights, it is important to have access to data about all sub-sectors of the tourism ecosystem.

3.2.3 Data

Regarding the data dimension that includes data users, data purpose and data sources, the clustering model identified four clusters of data sharing initiatives, as shown in Figure 12.

Figure 12 - Clusters of data dimension



3.2.3.1 First cluster – Planning & operations, tourists' engagement with the use of the other data

Regarding the data dimension, the first cluster involves 83 data sharing initiatives with the aim of improving the planning and operations of tourism services and increasing the tourists' engagement, using mostly data concerning principally statistics, private businesses and context-specific information. Within this cluster, the data users involved are tourism destinations, public authorities and the private sector of the tourism industry, since they are the driving force behind this process of developing and improving the tourism ecosystem.

3.2.3.2 Second cluster – Use transaction data to improve planning and operations, and conduct market analysis

This cluster refers to the 38 initiatives with the aim of improving planning and operations and conducting market analysis through the use of transaction data. Regarding the data users, those initiatives shall be

addressed to the private sector of the tourism industry and to tourism destinations & public authorities, which are the driving force behind the process of developing and improving the tourism ecosystem.

3.2.3.3 Third cluster – Increase tourists’ engagement through a combination of data sources

The 32 data sharing initiatives within the third cluster of the data dimension are related to the development of tourism services that support increasingly personalized use cases, based on a higher degree of interaction with the customer/end-user (tourists’ engagement). Given the purpose described above, the initiatives are intended mainly for tourism destinations and public authorities that leverage data about context-specific information, general statistics and transaction data, giving an overview of purchase habits of tourists.

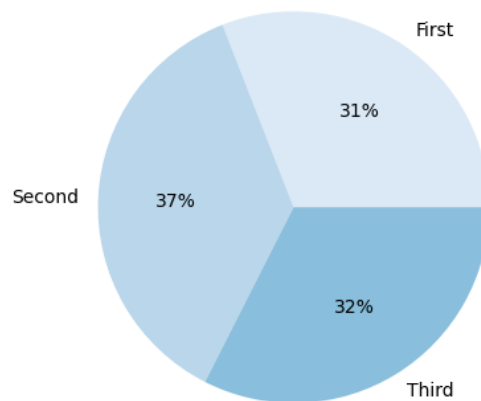
3.2.3.4 Fourth cluster – General overview of tourism ecosystem through the use of data

Within this cluster there is not a clear prevalence of the data characteristics. Indeed, the 41 initiatives engaged are the ones that cater to all the data users identified, including tourism destinations, public authorities, the private sector of the tourism industry and other industries. Moreover, the aim of giving access to these data, mainly context-specific data, statistics and private businesses data, is a combination of the data purposes identified.

3.2.4 Geographical

The geographical dimension includes variables such as the type of partners of the initiatives, and the geographical level covered by the initiative (e.g. international / country / region / city). The K-modes clustering algorithm identified 3 clusters within this dimension, distributed as depicted in Figure 13.

Figure 13 - Clusters of geographical dimension



3.2.4.1 First cluster – Initiatives at country level

The first cluster for the geographical dimension includes 60 data sharing initiatives related to the country level and concerning mostly public-private partnerships. Within this cluster, the initiatives aim to foster primarily the improvement of planning and operations and the improvement of market analysis and decision-making. The prevalence of public-private partnerships is due the fact that the establishment of synergies based on data sharing or technology implementation can be mutually beneficial and improve value propositions and increase the appeal of destinations towards prospect visitors. Moreover, these initiatives also focus on the improvement of sustainability and accessibility of tourism offers, producing

positive impacts on society at large, and on the development of tourism services based on a higher degree of interaction with the customer/end-user.

3.2.4.2 Second cluster – Initiatives at regional level

The predominant cluster for the geographical level includes 71 data sharing initiatives with a focus on the regional level. Moreover, these initiatives are mostly led by public partners. In terms of data purposes, this cluster focuses predominantly on the improvement of planning and operations of tourism services and on supporting market analysis. Also at regional level, there is a strong focus on the improvement of sustainability and accessibility of tourism offerings, producing positive impacts on society at large, and on the development of tourism services based on a higher degree of interaction with the customer/end-user. This is due the fact that data should be provided also at a more extensive and detailed level in order to ensure a complete understanding of the sector for the tourism ecosystem.

3.2.4.3 Third cluster – Improvement of market analysis at city level

The 63 data sharing initiatives within the third cluster are the ones with the key characteristic that there is not a specified or clearly identified partner type. Moreover, these initiatives are mostly related to the country level. In terms of data purposes, these initiatives mostly focus on the improvement of planning and operations of tourism services and on other objectives, such as making information available and for the purpose of statistics/research. When shifting to a more specific geographical level, like a city, there is considerable attention on the improvement of market analysis and decision-making. This is because destinations that want to extract and gain meaningful insights and improve their business strategies need to have access to data related to a higher level of detail in terms of geographic information.

3.3 Taxonomy

The following paragraph illustrates the structure for the data taxonomy. For each of the dimensions, the structure shows the hierarchy of categories and subcategories within the data sources for the clusters identified before. A colour scale was used to highlight the differences in the frequency of the data sources in the hierarchical structure, as depicted in Table 2.

Table 2 - Legend of taxonomy

Data sources frequency	Colour legend
1 to 15	
16 to 35	
36 to 55	
56 to 75	
76+	

Before delving into the taxonomy analysis, it is important to have a clear understanding of all the data sources taken into consideration in the analysis. Table 3 gives an overview on the definition of the data sources presented in the hierarchical structure.

Table 3 - Definition of data sources

User generated data

Review	Textual information coming from reviews on websites
Social networks	Textual data coming from social media
Ratings	Data about ratings on websites
Photo	Photographs made available by tourists
Video	Videos made available by tourists
Transaction data	
Mobility	Purchases in the mobility sector
Bookings	Information about online bookings
Ticket sales	Information about ticket sales
Points of sale (Pos) data	Data coming from point of sales (Pos), including credit cards data and payment cards data
Banking data	Information about card transactions coming from banks
Web search and web scraping	Web searches tracking data about tourism related transactions
Device data	
Smart city	Data collected by smart city/data initiatives through devices and sensors
GPS	Data generated by the Global Positioning System
Mobile roaming	Data generated by the use of mobile roaming
Wi-Fi	Data generated through Wi-Fi connections
Meteorological data	Data consisting of physical parameters that are measured directly by remote sensing technologies or ground stations
Webcam	Data collected by webcams
Bluetooth	Data generated through the use of Bluetooth
Other data	
Locations data	Data about destination information, e.g. Point of interests, mobility services, cultural sites and events/activities in the area
Data of public services	Data about services in the area (e.g. list of accommodation, restaurants)
Geographical data	Data from Geographic Information Systems
Cartographic data	Data related to maps and coordinates
Governmental data	Data generated by governments
Mobility metrics	Statistics coming from mobility services
Locations metrics	Statistics about destination information
Tourism metrics	Aggregated data on tourism metrics (including occupancy and expenditure)
Travellers' information	Statistics about travellers' information
Occupancy	Data generated by private businesses on occupancy
Air traffic	Data generated by private businesses on air traffic
Expenditure	Data generated by private businesses on expenditure
Clients' information	Information on clients collected by private businesses
Business indicators	Data on indicators provided by private businesses

3.3.1 Business

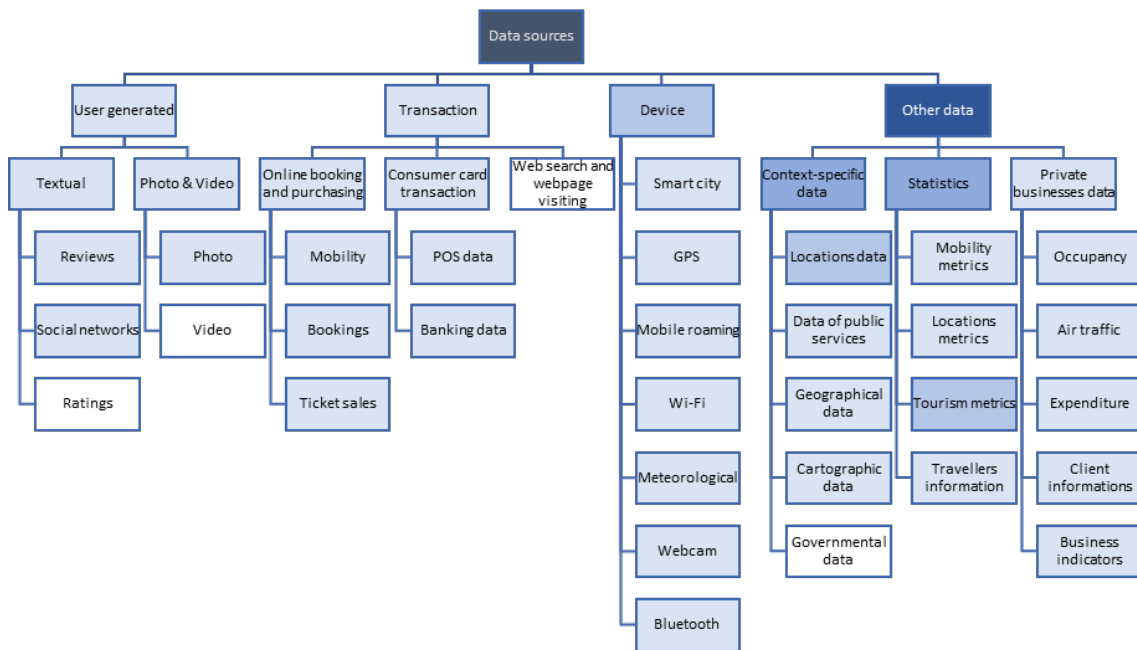
3.3.1.1 First cluster – Initiatives lead by public partners with the aim of data accessibility

The first cluster of the business dimension includes data sharing initiatives covering the generic, accommodation and transport sub-sectors with the aim of facilitating data access. Predominantly data come from other sources, focusing mainly on statistics and context-specific information. Going deeper in the hierarchical structure of those sources, it is evident that tourism metrics and locations data cover a large number of initiatives (19 and 20 data sharing initiatives, respectively).

Device data can also be considered as a frequent data source, especially data coming from smart city sensors, GPS and mobile roaming.

From a perspective of what data source is lacking and/or less frequently available, transaction and user generated data are only covered by a few initiatives. Data coming from web searches, ratings and video are also missing.

Figure 14 - Taxonomy of first cluster of business dimension



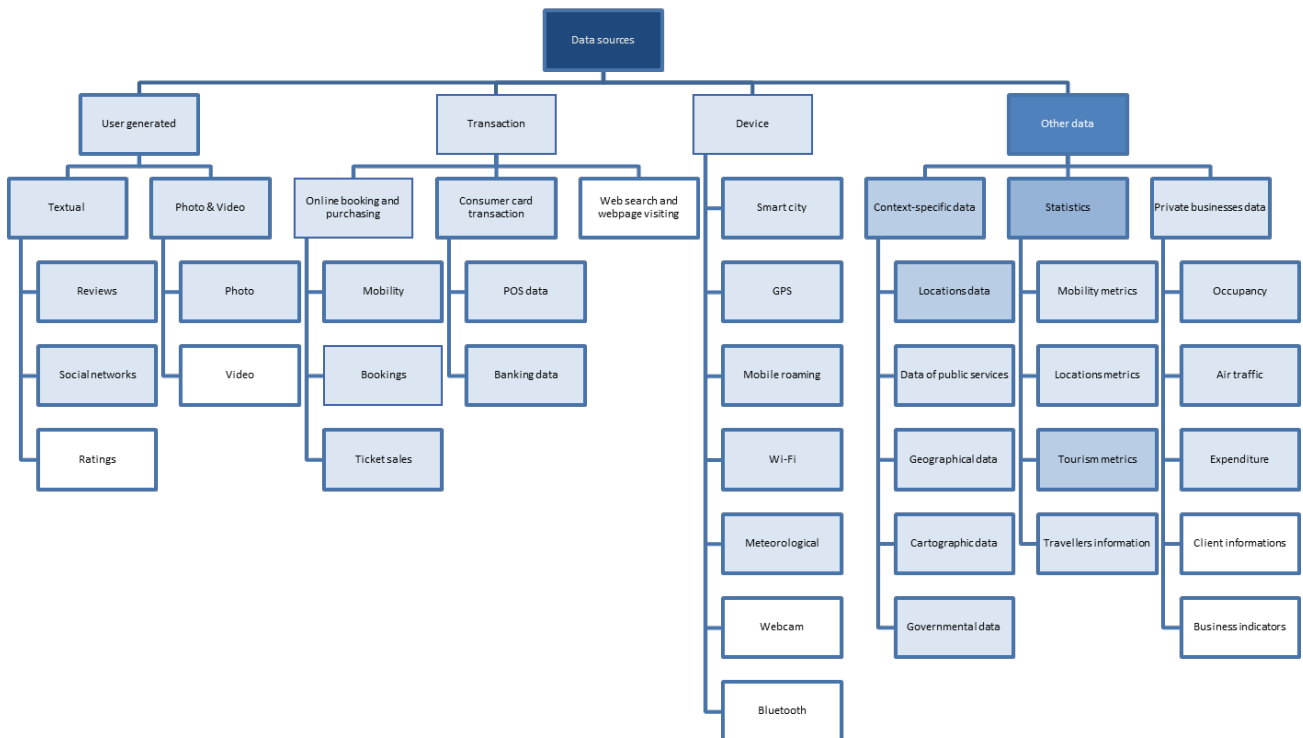
3.3.1.2 Second cluster – Initiatives with a twofold purpose

The data sharing initiatives of the second cluster, related to essential and fundamental sub-sectors (e.g. generic, accommodation and transport), cover data coming from the “other data” category, with a focus on statistics and context-specific information. Within these categories, tourism metrics and locations data appear to be the most common, with 32 and 22 data sharing initiatives respectively.

Transaction, device and user generated data are only involved to a smaller extent, with a prevalent focus on online bookings, purchasing data and consumer card transactions.

Finally, for what concerns the missing data sources, these initiatives do not rely on information coming from web searches and web scraping, webcams.

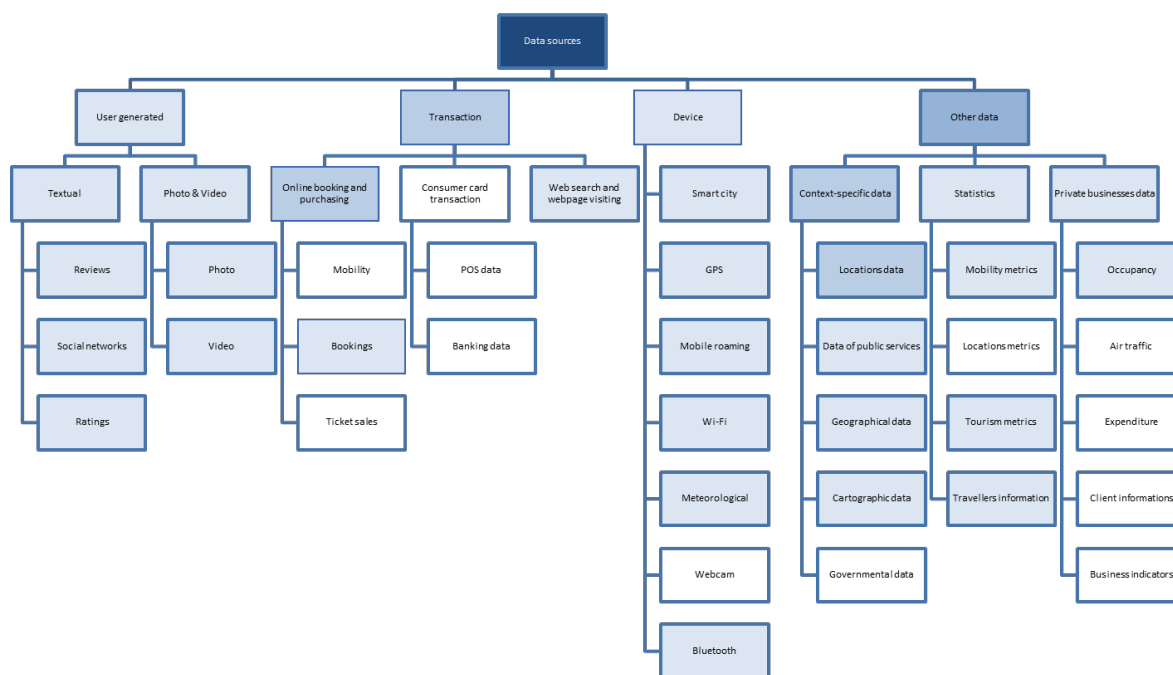
Figure 15 - Taxonomy of second cluster of business dimension



3.3.1.3 Third cluster – Public-private partnerships to foster an innovative and strategic focus

Within the third cluster of the business dimension, initiatives aim to encourage the reuse of data and the creation of innovative services. The most frequent data come from other sources, mainly tackled by context-specific data, in particular by location data. On the other hand, private business data is present in really small percentage. Shifting to transaction data, the predominant source are online bookings and purchasing data, while there is a complete lack of data coming from consumer card transactions. In addition, device data and user generated data are only present to a small extent, with a consistent frequency regarding smart city, meteorological data and textual data respectively.

Figure 16 - Taxonomy of third cluster of business dimension



3.3.2 Data

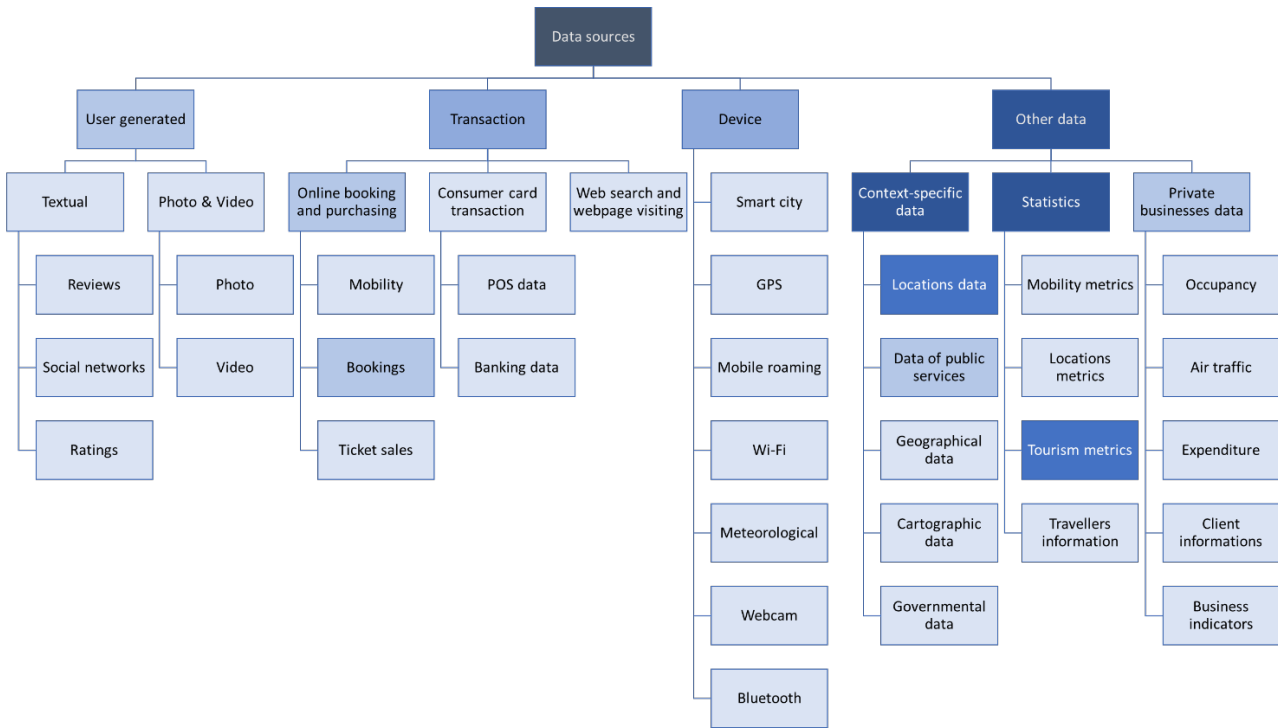
A complete and comprehensive analysis of the data sources shows that data coming from “other sources” predominate in the 194 data sharing initiatives. Regarding this source, statistics and context-specific data prevail with 106 and 91 data sharing initiatives respectively. Going deeper in the hierarchical structure of context specific data, location data is the most frequent with 71 initiatives, followed by data of public services with 27 initiatives. In contrast, geographical data, cartographic data and governmental data are the lowest in frequency. Concerning statistics, tourism metrics predominate with 61 initiatives, while mobility metrics, location metric and travellers’ information are less common. Ultimately, private business data encompasses the categories of occupancy, air traffic, expenditure, client information and business indicators that are approximately equal in frequency.

As far as transaction data are concerned, the data sharing initiatives are covered to a larger extent by online bookings and purchasing data, with 18 initiatives related to bookings, followed by data coming from consumer card transactions.

Regarding device data, 14 data sharing initiatives deal with data coming from smart city sensors and GPS. From the perspective of what is proportionally less covered, data coming from webcams and Bluetooth are less likely to be included in the overall data sharing initiatives.

Finally, for user generated data, there is a predominance of textual data, with 12 data sharing initiatives. Within textual data, reviews are the most frequent, followed by data coming from social networks and ratings.

Figure 17 - Taxonomy of data dimension



3.3.3 Governance

For the governance dimension, the cluster analysis revealed that there is not a high correlation between how initiatives collect data and how data are shared. Therefore, the taxonomy analysis exhibited the same hierarchical structure of the data dimension, previously illustrated in Figure 17.

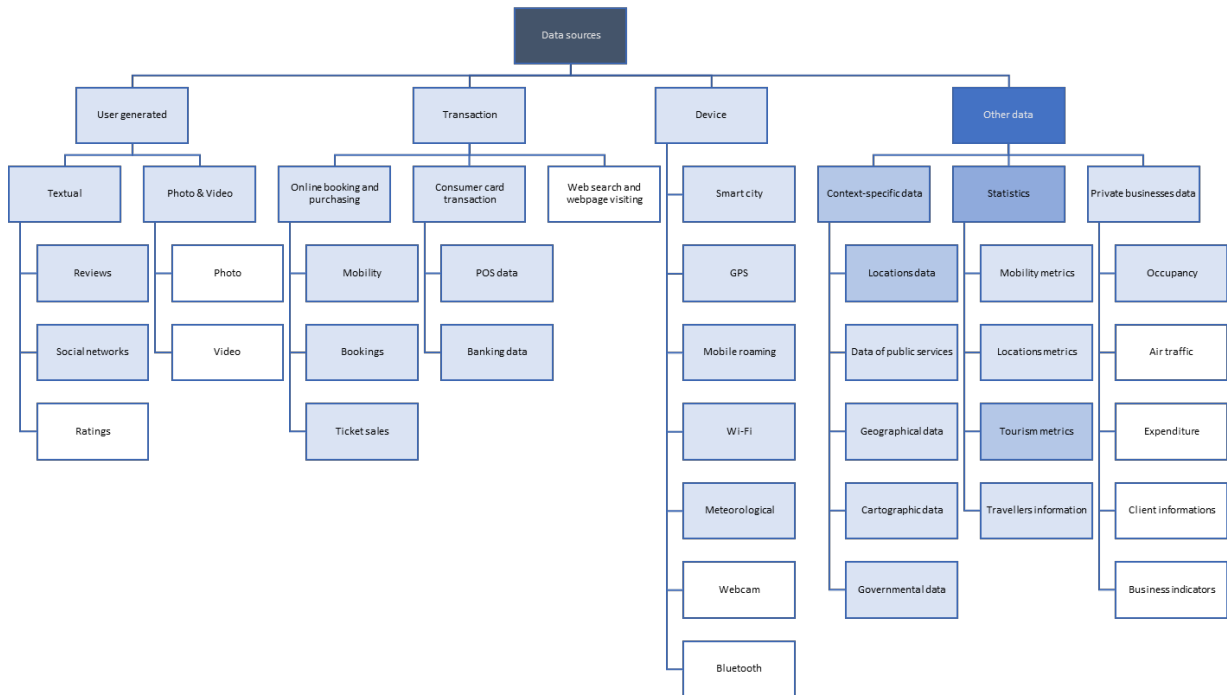
3.3.4 Geographical

3.3.4.1 First cluster – Initiatives at country level

The first cluster of the geographical dimension includes data sharing initiatives covering initiatives at the country level and concerning mostly public-private partnerships. Data coming from other sources are predominant, focusing mainly on statistics and context-specific information. Going deeper in the aforementioned sources, it appears that tourism metrics and locations data cover a major number of initiatives, with 33 and 23 data sharing initiatives respectively.

On the other hand, accounting for the data source which is lacking/and or less frequent, user generated, transaction data and device data cover fewer initiatives. Indeed, data stemming from ratings, videos, photos, and Bluetooth are missing.

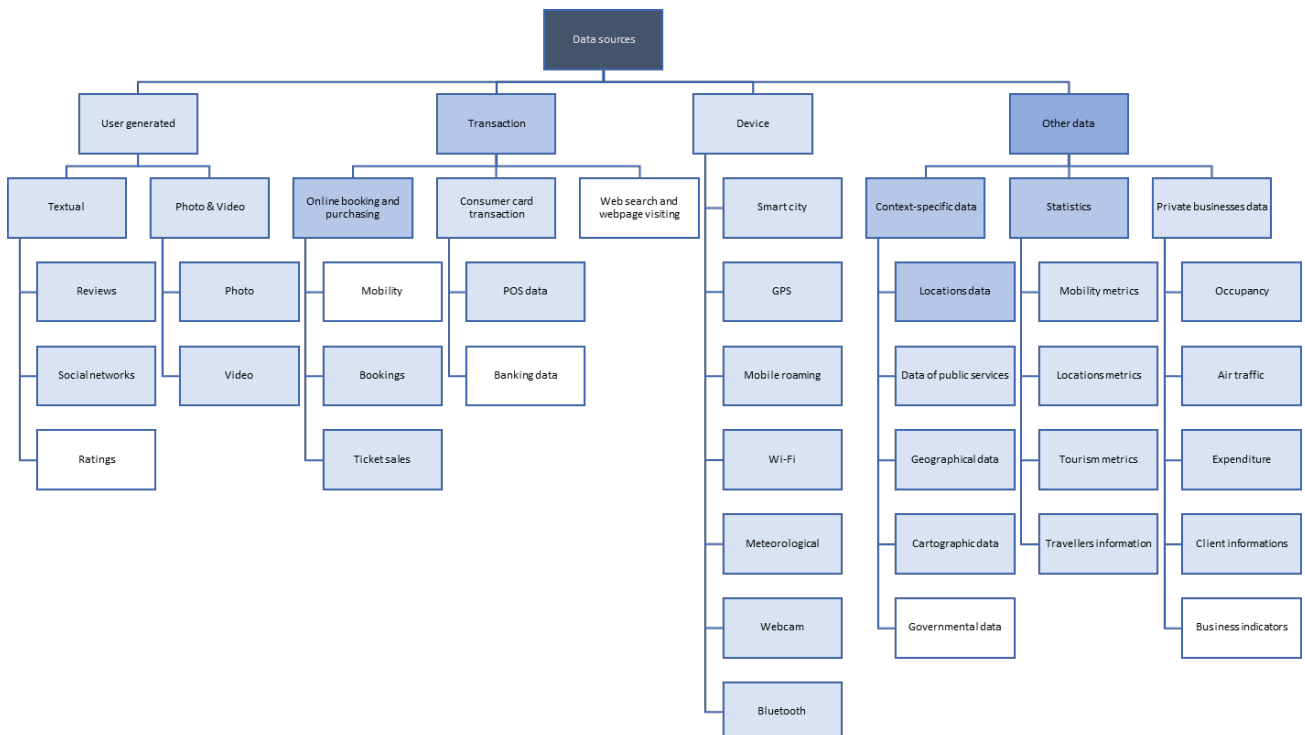
Figure 18 - Taxonomy of first cluster of geographical dimension



3.3.4.2 Second cluster – Initiatives at regional level

The data sharing initiatives of the second cluster, which pertain to the regional level and are mostly led by public partners, mainly cover data coming from other data sources, with a focus on statistics and context-specific data. Within these categories, location data are shared by a large number of initiatives. Similarly, transaction data are covered by many initiatives, with online booking and purchasing prevailing.

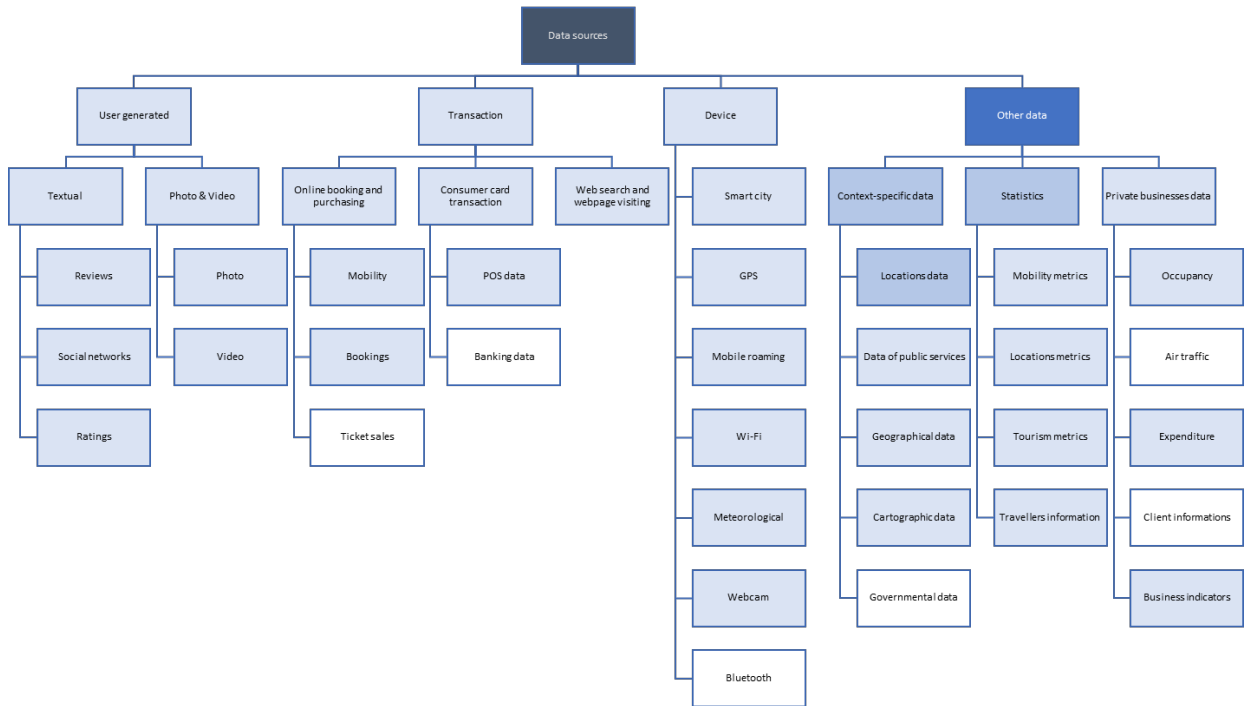
Figure 19 - Taxonomy of second cluster of geographical dimension



3.3.4.3 Third cluster – Improvement of market analysis at city level

In the third cluster, which takes into account initiatives which are characterized by not having a specified or clearly identified partner type, and are mostly related to the city level, other data is the most frequent data source, mainly comprising context specific data and specifically locations data and data of public services. Comparably, statistics data are also prevalent. Missing categories include, but are not limited to, air traffic and client information in private business data, tickets sales in online booking and purchasing, and banking data in consumer card transactions.

Figure 20 - Taxonomy of third cluster of geographical dimension



3.4 Workshop results analysis

In the workshop break-out rooms, the project team could uncover several findings pertaining to the presented challenges of each data purpose. Correspondingly, Table 4 showcases – in order of challenge priority – the rating that stakeholders gave to each challenge, on a scale from 1 to 5.

Table 4 - Challenges rating and priority

Challenge priority	Challenge Name	Challenge Rating (out of 5)
Break out room I: Increase of tourism sustainability & accessibility		
1	Support the growth of sustainable parameters awareness among the tourism industry	4.3
2	Meet the demand for more sustainable tourism	4.5

Challenge priority	Challenge Name	Challenge Rating (out of 5)
2 ¹¹	Reduce negative impacts of tourists' behaviour on local communities and environment	4.7
3	Measure and reduce the environmental footprint	4.6
4	Sustainable mobility	4.4
5	Manage tourism seasonality	4.5
6	Improve the supply and visibility of accessible tourism services	4.3
Break out room II: Conduct market analysis and inform decision making		
1	Develop effective business strategies & make investment decisions	4.3
2	Monitor the flow of visitors between districts and attractions	4.0
3	Incentivize collaboration among different types of stakeholders	4.3
4	Cultural shift regarding tourism sector data sharing	4.1
5	Assess impacts of events and infrastructure investments	3.8
6	Answer to the so-called request for "hyper personalization" of travel experiences	3.3
7	Manage the tourist-residents relationship	4.2
8	Crisis management	4.0
Break out room III: Improvement of the interaction and engagement of the tourist		
1	Improve tourists' experiences by providing information on tourism flows	4.2
2	Reach potential clients without being fully dependent on OTAS	4.4
3	Address the demand/needs of tourists of different generations	3.8
4	Create and manage a relation with tourists	3.8
5	Manage online reputation and leverage on tourists' satisfactions	4.1
6	Develop innovative immersive and digital tourism experiences	3.8
Break out room IV: Improvement of planning and operations of the tourism service		
1	Understand and better forecast the tourism flow	4.6
2	Manage seasonality in tourism sector	4.3
3	Manage and reduce overcrowding of sites and services	4.0

¹¹ Both 'meet the demand for more sustainable tourism' and 'reduce negative impacts of tourists' behavior on local communities and environment' were uncovered to have the same second ranking.

Challenge priority	Challenge Name	Challenge Rating (out of 5)
4	Manage online reputation and leverage on tourists' satisfaction	3.9
5	Multimodal ticketing-smart mobility	4.1

During the workshop, participants were also asked to express which of the following data types would be useful to address each of the aforementioned challenges. The frequencies with which each type of data is associated to each challenge are presented in Annex I - Post Workshop Report. The proposed data types are:

- **Accessibility data** are about accessibility and inclusivity of the destination/services,
- **Sustainability data** are related to the economic, social and environmental impacts of tourism,
- **Behavior of tourists** refers to their preferences related to the choice and fruition of (touristic) services while travelling,
- **Demand and offer data** allow an estimation of the “amount” of services requested and supplied,
- **Local/Residents' satisfactions and emotions** are about communities' emotions related to local tourism,
- **Tourists' satisfaction & emotions** are about customers' opinion regarding their experience,
- **Mobility data** are related to the use of transportation in and to the destination, including maritime, air and train flow data, ticketing and payment services,
- **Purchase habits** are data about how and in which occasions tourists spend their money while travelling,
- **Related industry data** concern industries related to tourism (e.g. real estate, entertainment, agriculture),
- **Typology of tourists** are about profiling of tourists (e.g. socio-demographic characteristics and personal data),
- **Tourists flow** relates to tourists' movement on a specific destination/point of attraction.

Data association with the challenges allowed the project team to identify the most important and most needed data in the tourism sector (3.5.2).

3.5 Main tourism sector data gaps

3.5.1 Data gaps and overlaps emerging from the analysis

This section presents the main findings of the analysis described above in terms of gaps and overlaps in the initiatives. In particular, the analysis focuses on the inventory dimensions relevant for understanding the data misses and convergences, namely the geographical level and country coverage, the tourism sub-sectors, the data purpose, and the data sources.

Regarding the geographical dimension, the inventory suggests that most of the initiatives provide data at country level, while the **availability of data at local level is limited**. Indeed, only 11% of the initiatives provide data at city level (Figure 1). This finding is confirmed by survey answers to the questions about

shortages and missing datasets of existing data sharing initiatives. Specifically, it was pointed out that “existing data sharing initiatives very often describe data at the national level and less at the local level” and that there are “not enough data for areas smaller than NUTS 3¹²”.

Regarding the tourism sub-sectors covered (Figure 5), an overall glance at the inventory shows that most of the initiatives cover more than one sub-sector (sub-sector “generic”). While there is **good availability of accommodation and transport data** (both covered by around 41% of the initiatives), **car and other rentals is the least covered sub-sector** (only addressed in 9% of the cases). Looking at the geographical distribution of the initiatives covering the different sub-sectors, the analysis of the three geographical clusters (3.2.4) shows that initiatives sharing transport and accommodation data are evenly distributed. Conversely, half of the initiatives covering the sub-sector car and other rentals share data at country level.

The inventory (Figure 8) shows that “other data” – covering context-specific information, statistics and private businesses datasets – represents the most common type of data sources. Conversely, **user generated data is the least frequent type of data source** mapped. Looking at the distribution of data sources across the initiatives covering the four identified data purposes, the frequency of data sources for each purpose reflects the overall distribution (Figure 8). Other data are the most frequent ones, followed by transaction data, device data and user generated data. Moreover, in the survey, when asked about the missing datasets, the open comments related to data sources most frequently pointed out the scarce availability of transaction data, device data, and private business data.

In detail, the data taxonomy (3.3.2) allows to grasp, for each data source, which data types are available. Within “other data” (the most common data source), context-specific data (89) and statistics (108) are widely available, while private-business data (31) are much less common. The **lack of private business data** is also reinforced by survey respondents¹³. Indeed, they expressed the scarcity of data typically coming from private businesses’ datasets, such as those related to hotel bookings, short-term rentals, and flights. At the same time, a comment received through the survey highlighted that private datasets are available upon payment and that they are usually expensive. Moreover, looking into initiatives covering context-specific data, almost 78% (71) focus on location data, while the other types of data, especially governmental data (1), are widely missing. Similarly, more than half (61) of the initiatives covering “statistics” provide tourism metrics, while others, especially travellers information (4), are missing.

Regarding transaction data (the second most common data source), the data are unevenly distributed among the subcategories of online bookings and purchasing, consumer card transactions, and web search and webpage visiting. The former category constitutes around 70% (31) of all initiatives sharing transaction data, while the second and the third category respectively represent around 30% (13) and 4% (2). Among the survey comments highlighting the scarcity of transaction data, several comments specifically mentioned the **lack of card transactions**. Many of these answers suggested that card

¹² The NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system for dividing up the economic territory of the EU and the UK. NUTS 3 is the smaller level for the socio-economic analyses. Source: <https://ec.europa.eu/eurostat/web/nuts/background>

¹³ Both in the question regarding the existing data sharing initiatives shortages and the one regarding the missing datasets.

transactions data are usually very expensive and difficult to be accessed, as they are shared upon payment by the private companies generating them.

As for the device data (the third most common data source), smart city and GPS data being the most available types of data (31 respectively) and webcam data being the least common (1).

Finally, for user generated data (the least frequent data source), textual (12) and photo and video (6) contents are available.

When asked about the missing datasets, many survey respondents answered focusing on the type of information that could be derived from the data. The results underline, in order of decreasing frequency, a **lack of data on demand and offer, sustainability, mobility, and behaviour of tourists**.

3.5.2 Most needed data according to priority business challenges

Through the workshop, the project team explored the priority challenges (3.4) related to each data purpose and the data needed to address each challenge (Annex I - Post Workshop Report). Therefore, the collected evidence allowed the project team to understand which are the sector's most important data and the most needed ones. The most important data are identified as the data most frequently associated with the challenges, while the most needed ones are identified based on their association to the challenges with the highest priority.

To assess the most important challenges, the project team looked at the frequency with which each data information was associated to the challenges. Initially, the project team assessed the most important data for each data purpose. The results show that the most important data for the purpose of increasing tourism sustainability and accessibility and for conducting market analysis and informing decision-making are data on tourists' behaviour, mobility data and tourists flows. For the purpose of improving the interaction and engagement of tourists, the three most important data are data on tourists' behaviour, typologies of tourists and about tourists' satisfaction and emotions. Finally, for improving planning and operations of tourism services, the three most important data are accessibility data, data on the behaviour of tourists and demand and offer data. Overall, looking at the total votes received by each type of data in all four break-out rooms, it emerged that **the three most important data to be shared in the tourism sector are the ones regarding tourists' behaviour, mobility data and demand and offer data**. This result is only partially consistent with the survey. Indeed, when asked about the most important data information, the top three data information were demand and offer data (1st), purchase habits (2nd), and data on tourists' behaviour (3rd); while mobility data only positioned as 6th out of 9¹⁴.

¹⁴ The survey question did not include all the 11 data information that were then explored in the workshop (tourists flows were excluded). The survey question "Which of the following information about tourism are the most important for you?" foresaw the following 9 options: Behavior of tourists
Benchmark with other destinations, Demand and offer data, Mobility in the destination, Nationality of the tourist, Pricing comparison, Purchase habits, Typology of tourists, and Other.

To assess the most needed challenges, the project team started by identifying the most needed data for each data purpose. Then, confronting these results, it derived the overall most needed data¹⁵. For the purpose of increasing tourism sustainability and accessibility and for conducting market analysis and informing decision-making, the three most needed data types are behaviour of tourists, mobility data, and tourism flows. For the purpose of improving the interaction and engagement of tourists, the three most needed data are data on tourists' behaviour, typologies of tourists and purchase habits. Finally, for improving planning and operations of tourism services, the three most needed data are data on tourist behaviour, tourist flow data, and demand and offer data.

As a synthesis of the four analyses, the project team compared the ranking of the most needed data for each data purpose. The results show that the overall **most needed data in the sector are data on tourist behaviour, mobility data, and tourist flow data** (the latter two equally). These results are therefore partially consistent with the analysis of the most important data. Indeed, in both analysis data on tourists' behaviour and mobility data rank in the top two positions. Differently, the demand and offer data are the third most important, but the fourth most needed. Conversely, the tourists flow data are fourth most important data, but the second most needed ones (equally with tourists flow data).

Comparing the most needed data with the data gaps stemming for the analysis, it becomes evident that **the most urgent gaps to be addressed concern data on tourists' behaviour and mobility data**. Regarding the latter gap, a possible solution can come from mobility data sharing initiatives, including data spaces such as [EONA-X](#) and the [Mobility Data Space](#).

3.5.3 Most needed gaps of data characteristics

From the survey, a multitude of shortages of existing data sharing initiatives emerged. **Incomplete data and interoperability data** were the two main gaps that were pointed out when respondents were asked about gaps in the initiatives they were aware of. Incomplete data encompassed shortages such as incomplete metadata, inaccuracy, data not being updated in a timely manner and superficiality of data in terms of granularity that do not allow for high quality analysis, such as forecasting and the use of artificial intelligence (AI) and machine learning (ML). There are also gaps in the issue of interoperability, such as insufficient data exchange possibilities and the lack of standardized data and/or definitions. The latter point is compounded by the **lack of cooperation and coordination that** does not allow for adequate standardization of data leading to difficult access to data and inefficient analysis.

The gaps related to the themes of interoperability and data not being updated in a timely manner emerged during the workshop, when participants were asked to investigate whether there were any missing challenges concerning the data purpose. The gap of interoperability, which emerged in the break-out room "conduct *market analysis and inform decision making*", concerned the fact that there

¹⁵ To understand the most needed data information for each data purpose, for each challenge, the project team assigned a weight to each data information based on the priority of the associated challenge and the frequency with which the data was associated with the challenge. Summing the weights assigned to each data information allowed the team to classify them based on their priority (the higher the sum, the more needed the data information). Similarly, to classify the priority of the data information among all four data purposes, the team assigned a weight to each of the 11 ranking positions and multiplied it by the times each data information was ranked in that position. Summing these products for each data information allowed the team to derive a ranking of the overall most needed data (the higher the sum, the more needed the data).

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

needs to be standardization on accommodations and other touristic resources. Similarly, the gap of data not being updated in a timely manner, which emerged in the break-out room “*improvement of planning and operations of the tourism service*”, related to the need of informing all visitors and local residents about temporary closures of infrastructures, before they arrive at closed locations and are forced to change plans on the spot.

4 Conclusions

4.1 Overview of key findings

The results presented in this report will constitute the ground for further research on the development of the European data space for tourism. In particular, the analysis of the tourism sector data gaps and the comparison of these results with the priority challenges of the four data purposes¹⁶ allowed to understand which are the most needed data. Overall, the **most needed data in the sector are data on tourist behaviour, mobility data and data on tourist flows**. According to the stakeholders consulted, the former two are not yet available and should therefore be made available through the new tourism data space.

At the same time, the evidence collected from stakeholders highlights that the available data has some shortcomings, the most frequent ones being **incomplete data, lack of data interoperability and data not being updated in a timely manner**. Respondents' comments highlighted data flaws that are at the core of the data space objectives. Indeed, a sector's data space aims at increasing the availability of data based on partners' involvement. Data standards will enable the exchange of data and the participation of private companies might increase the timeliness of the available data. In open comments, many respondents pointed out that there is a **lack of cooperation and coordination for sharing data in the tourism sector**. Cooperation and mutual trust are at the core of the functioning of data spaces. The deployment of a European data space for the tourism sector will surely contribute to spreading a data sharing culture within the sector.

Finally, several survey answers pointed out both a **lack of awareness regarding data sharing** and a **lack of skills needed to extract useful information from the data**. Indeed, a meaningful comment says "In my opinion, the most relevant shortage is the lack of enough professionals able to extract meaningful insights from data. In the future, we will grow the number of relevant data sources, the availability in real time, granularity, and accessibility, but we are extracting very little of the data potential which is accessible today". Both the lack of awareness and lack of skills might be among the reasons why the availability of tourism data is currently limited. Moreover, these problems are probably at the base of the lack of cooperation and coordination for sharing data in the tourism sector, as probably tourism stakeholders are not yet aware of the benefits of data sharing.

4.2 Next steps

Following the submission of deliverable D2.2, the project team will focus on determining potential common building blocks with other data spaces (Task 2.3), identifying tourism data space priority datasets and developing use cases (Task 2.4). The selection and

¹⁶ Increase of tourism sustainability & accessibility; Conduct market analysis & inform decision-making; Improvement of the interaction and engagement of the tourist; Improvement of planning and operations of tourism services

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

description of the use cases pertaining to the latter task will be executed by drawing on the results obtained from the workshop.

With respect to Task 2.4, the project team will deepen the knowledge on the initiatives already addressing the previously identified priority needs and purposes through interviews with owners and/or partners of the identified initiatives. These consultations will further help the team in collecting the contents of datasets most likely addressing the most important stakeholder's data needs. These consultation activities will also be the basis for developing an initial list of use cases, which will consequently be developed during an upcoming co-design workshop that will take place in March.

All upcoming Work Package 2 activities are summarised in Table 5.

Table 5 - Next WP2 activities

Activities	Deliverable	Deadline
Identifying potential common building blocks with other data spaces	<i>D2.3 Identification of data typology and priority lists of datasets, potential use cases and common building blocks with other data spaces</i>	30 th April 2023
Deepening on the initiatives		
Use cases definition		
Use cases workshop (March 2023)		

Furthermore, the results of WP2 will build a groundwork for Work Packages 3 and 4. Concerning Work Package 3, some of the activities of *Task 3.1 Reference architecture for tourism data space* and *Task 3.2 Common standards and interoperability* are linked to the outputs of WP2. Task 3.1 will build on the tourism data sharing initiatives inventory (Deliverable 2.1) and the analysis within the present Deliverable to define the tourism data space reference architecture and technical framework specifications. Likewise, Task 3.2 will start from the analysis of the tourism data sharing initiatives mapped in Deliverable 2.1 to identify the common standards and interoperability protocols that are essential to the development and harmonization of new data services for the tourism industry and those connected to it (e.g., transport, environment, commerce, smart cities).

Regarding Work Package 4, some of the activities of *Task 4.1 Roles and Interactions in a European Tourism Data Space* will rely on WP 2. Task 4.1 will commence from the analysis of the use cases identified in Task 2.4 to shortlist the most representative ones from a governance point of view, and to identify the tourism data space governance requirements. Moreover, Task 4.1 will analyse the outputs of WP 2 from a governance perspective, in order to be able to set a baseline of legal and regulatory aspects for the tourism data space.

Annex I - Post Workshop Report



Work Package 2: Context Analysis and Agreed Components

Towards a Data Space for Tourism –
Prioritization of data needs and data purposes

Post - workshop report

08/02/2023



**Funded by
the European Union**

CONTACT POINTS

Giovanna Galasso – Associate Partner - Intellera Consulting

giovanna.galasso@intelleraconsulting.com

Marco Codastefano – Manager - Intellera Consulting

marco.codastefano@intelleraconsulting.com

1. Workshop Overview

DATES Project Objectives

The workshop “**Towards a Data Space for Tourism – Prioritization of data needs and data purposes**” took place in the context of the project **DATES – Data Space for Tourism**, financed by the **Digital Europe Programme (DIGITAL)**. The objective of DATES is to develop a strategy roadmap for building a tourism data space, in collaboration with stakeholders of the tourism sector.

*A **data space** is defined as a decentralised and standard-based structure to enable trustworthy data sharing between the data space participants on a voluntary basis, regulated through governance, business, legal and technical combined mechanisms.*

Therefore, a data space is a particular type of data sharing initiative, characterized by a **higher level of data control** by participants. Indeed, in data spaces, data are not centrally stored and exchanges (for free or against compensation) are based on agreements. A second key aspect of data spaces is **trust among participants**, which is not only guaranteed by technological requirements, but also by the data space governance.

As a transversal industry, tourism has a great need for efficient data exchange within and across industries. A data space indeed represents an opportunity for the sector: it allows to capture value associated with data production within the industry and, at the same time, its possible interoperability with other sectors data spaces (e.g. mobility and cultural heritage data spaces) allows to also access other industries' data. Such enhanced availability of data could foster industry innovation, the digital transformation of tourism SMEs, and provide a basis for policy making.

To turn the vision of a European Tourism Data Space into reality, the overall work plan of DATES unfolds over a time span of 12 months. In this period, the project team will organize several consultation activities, if you are interested in participating please sign up at the link below.

Do you want to contribute to the European Data Space for Tourism?

Become a stakeholder, signing [here!](#)

Aim of the workshop

The workshop “**Towards a Data Space for Tourism – Prioritization of data needs and data purposes**” took place online on the 8th of February from 10:00 AM to 12:15PM CEST. The workshop agenda is presented in the following table.

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

Table 1: Workshop agenda

Time	Session	Contributor
Section I – A European data space for the tourism sector		
10:00 – 10:05	Welcome	Giovanna Galasso – Associate Partner – Intellera Consulting
10:05 – 10:15	Objectives of DATES and the benefits for the tourism sector	Dolores Ordóñez – Director - Anysolution
10:15 – 10:45	The concept of data spaces	Árpád Welker – Policy Officer – European Commission & Jean-François Cases – President – EONA X
10:45 – 10:55	First results concerning data sharing initiatives	Marco Codastefano – Manager – Intellera Consulting
10:55 – 11:00	Introduction to the breakout rooms	Daniilo Bianchini – Manager – <u>Intellera</u> Consulting
Break (11:00-11:10)		
Section II – Connecting the dots		
11:10 – 12:00	Data needs prioritization workshop	Break out room's moderators
12:00 – 12:15	Discussion and closing remarks	Break out room's moderators

As depicted in the agenda, the workshop was divided into two main sections.

The objective of the **first section** of the workshop was to raise awareness on the concept of data spaces and DATES project. Starting off, the audience was debriefed with the project's objectives and its benefits for the tourism sector. The policy context of the project and the key characteristics of a data space were accordingly showcased, followed by a practical example of a mobility, tourism, and transport data space (EONA-X). Thereafter, the initial findings of the project were presented, and the break-out rooms' setup (second section) were subsequently introduced.

The objective of the **second section** of the workshop was to prioritize a series of business challenges that stakeholders face in their day-to-day activities, and that could be potentially solved by the use of a sector data space¹⁷. At the same time, this activity allowed to define the stakeholders' data needs, as they were asked to associate different types of data with each challenge. In order to do so, participants were split into four breakout rooms, one for each of the identified data purposes, i.e. the main reasons why tourism stakeholders might be interested in accessing data. The breakout rooms were:

- Breakout room I: Increase of tourism sustainability & accessibility;
- Breakout room II: Conduct market analysis & inform decision-making;
- Breakout room III: Improvement of the interaction and engagement of the tourist;
- Breakout room IV: Improvement of planning and operations of tourism services.

¹⁷ The main sources used to identify the challenges are: European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, Galasso, G., Montino, C., Sidoti, A., et al., *Study on mastering data for tourism by EU destinations : main text*, Publications Office of the European Union, 2022, <https://data.europa.eu/doi/10.2873/23880>; and European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, *Transition pathway for tourism*, Publications Office of the European Union, 2022, <https://data.europa.eu/doi/10.2873/344425>

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

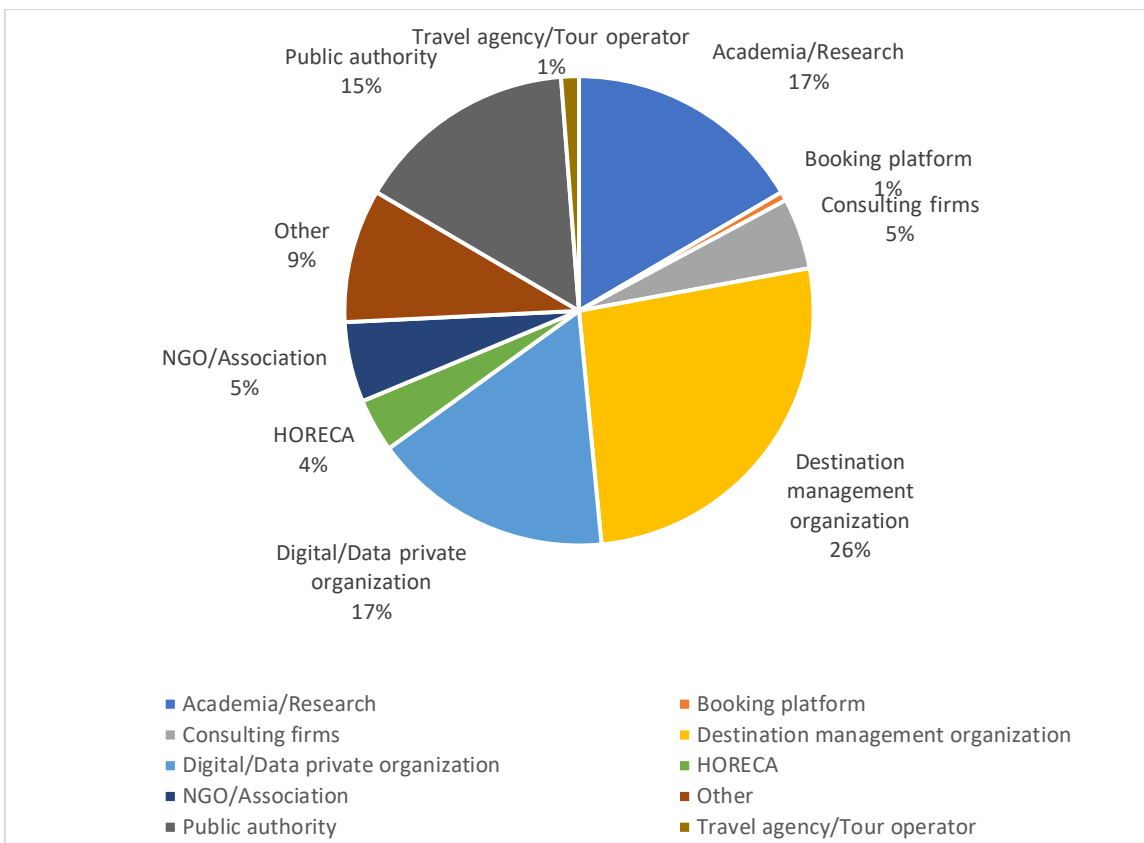
The results of such workshop will be used as the **starting point of the development of use cases**, which will help the team in defining the characteristics that a European data space for tourism should have. Moreover, workshop results will complement the analysis of the gaps and overlaps among the existing data sharing initiatives focused on the tourism sector.

The results of the breakout rooms are presented in Section 3.

2.Attendance

Overall, **207 participants** from **27 countries**¹⁸ attended the event. The attendees covered several categories of stakeholders, ranging from destination management organizations to consulting firms and public authorities. Accordingly, the graph below provides an overview of the workshop participants distribution by typology of stakeholders.

Figure 1 - Participant distribution



¹⁸ Albania, Argentina, Austria, Belgium, Bosnia and Herzegovina, Croatia, Czech Republic, Finland, France, Germany, Greece, Netherlands, Hungary, Ireland, Israel, Italy, Mexico, Norway, Poland, Portugal, Slovenia, Spain, Switzerland, Turkey, United Kingdom, United States of America.

3. Key findings

The following subsections depict the findings that were uncovered in each breakout room. The activities in each breakout room followed the same structure:

- For each presented challenge, participants were asked to rate its importance (on a scale from 1 to 5, where 1 was *absolutely not important* and 5 *very important*) and to mention which types of data they would need to address it;
- Then, they were asked to rank the priority of the challenges (by positioning the one with the highest priority at the top of the list and the one with the lowest priority at the bottom);
- Finally, they were asked an open question to investigate whether there was any missing challenge, its rate and the data needed to address it.

Correspondingly, Table 2 depicts the types of data presented when asked which data are the most useful to solve the specific challenge.

Table 2: Typologies of data needed to address a specific challenge

Type of data	Explanation
Accessibility data	Data about accessibility and inclusivity of the destination.
Behavior of tourists	Data concern to tourists' preferences related to the choice and fruition of (touristic) services while travelling.
Demand and offer data	Data allow an estimation of the "amount" of services requested and supplied (e.g. occupancy rates and/or number of overnight stays on daily basis, level of quality of tourism for region, prices for tour operators' contracts).
Local/Residents' satisfactions and emotions	Data about communities' emotions related to local tourism.
Mobility data	Data about transportation sector, including maritime, air and train flow data, ticketing and payment services.
Purchase habits	Data about how and in which occasions tourists spend their money while travelling (e.g. tourist expenditure by nationality, transaction data, consumption decisions both online and offline).
Related industry data	Data concern industries related to tourism (e.g. real estate, entertainment, agriculture...).
Sustainability data	Data on the economic, social and environmental contributions and dependencies of tourism (e.g. loading capacity requirements for geographic destinations, use of natural resources by tourism, tourism water consumption).
Typology of tourists	Data about profiling of tourists, including demographic characteristics, tourists' personal data.

Type of data	Explanation
Tourists flow	Data about tourists' movement on a specific destination/point of attraction.
Tourists' satisfaction & emotions	Data about customers' opinion regarding their experience (e.g. data about customer experience, emotional aspects of tourists, review scores, sentiment analysis).

Accordingly, the findings of each breakout room will be shown in the following paragraphs.

Guide for results interpretation

For each challenge, a table depicts three key findings: i) the average importance rating that participants gave to the challenge, implying the impact on the business to which they belonged; ii) the percentage representing the frequency of votes expressed for each rating number (from 1 to 5); and iii) the percentage representing the frequency of votes expressed for each data type, when asked which are the most useful ones to solve the challenge¹⁹.

Break out room 1: Increase of tourism sustainability & accessibility

Moderators: Tatiana Semenova and Adele Pia Villani from the Italian Ministry of Tourism

The first breakout room was dedicated to the data purpose of sustainability and accessibility, or else the improvement of sustainability and accessibility of the tourism offer producing positive impacts on society at large. For this specific data purpose, the project team uncovered 7 challenges: the respective findings of each of them are illustrated below.

Challenge I: Support the growth of sustainable parameters awareness among the tourism industry

The challenge of supporting the growth of sustainable parameters awareness among the tourism industry relates to the fact that the tourism industry is pervaded by a lack of awareness of sustainable parameters. There is a need to provide much more valuable information to users of touristic services. This implies more measurable impact indicators and monitoring systems made available for both tourists and stakeholders. Also, in this case, digital technologies can play a relevant role for both gathering data and producing valuable impact score carding systems.

As shown in the table, an average **rating of 4.3** out of 5 was given to the challenge of *supporting the growth of sustainable parameters awareness among the tourism industry*.

¹⁹ As this was investigated through a multiple-choice question, the percentages are calculated as the total number of votes received by each data type divided by the total number of votes expressed. The percentages of the results were rounded to the nearest whole number. As a result, the sum of the individual numbers does not always add up to 100%.

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

•

Correspondingly, *behavior of tourists, sustainability data and mobility data* were considered to be the three types of data that are most useful to solve this challenge.

•

Table 3 – Increase of tourism sustainability & accessibility challenge 1 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0 % 0 % 7 % 57 % 36 %</p> <p>14 voters</p>	Behavior of tourists	16%
	Sustainability data	13%
	Mobility data	13%
	Tourists flow	12%
	Accessibility data	8%
	Demand and offer data	8%
	Local/Residents' satisfactions and emotions	8%
	Tourists' satisfaction & emotions	6%
	Purchase habits	6%
	Related industry data	5%
	Typology of tourists	5%
		15 voters

Challenge II: Manage tourism seasonality

The challenge of managing tourism seasonality has to do with the importance of being able to identify, address and mitigate the effects of the seasonal variation of waste, pollution and natural resources consumption generated by tourism. At the same time, crowds might also be detrimental for fragile landscapes and heritage sites. Similarly, following Covid-19 pandemic, visitors are increasingly inclined to choose destinations with lower density and able to offer safe and diversified outdoors activities.

As shown in the table, an **average rating of 4.5** out of 5 was given to this challenge. Correspondingly, **tourist flow, behavior of tourists and mobility data** were considered the most useful data to solve this challenge.

Table 4–Increase of tourism sustainability & accessibility challenge 2 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>4.5</p> <p>0 % 8 % 0 % 25 % 67 %</p> <p>12 voters</p>	Tourists flow	17%
	Behavior of tourists	14%
	Mobility data	14%
	Sustainability data	12%
	Typology of tourists	12%
	Demand and offer data	9%
	Local/Residents' satisfactions and emotions	9%
	Accessibility data	5%
	Purchase habits	5%
	Tourists' satisfaction & emotions	3%
	Related industry data	2%
		13 voters

Challenge III: Reduce negative impacts of tourists' behavior on local communities and environment

This challenge arises from the fact that it is important to ensure that tourism does not cause harm to the nature, local environment, social and cultural wellbeing of the local residents. Tourism services should not be steered and provided to visitors in ways that risk harming the local environment, culture or its people. This would reduce the attractiveness of the destination over the long term by reducing its environmental and cultural authenticity.

As shown in the table, an **average rating of 4.7** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, local/Residents' satisfactions and emotions, and tourists' satisfaction & emotions** were considered the most useful data to solve this challenge.

Table 5 – Increase of tourism sustainability & accessibility challenge 3 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>4.7</p> <p>0 % 0 % 0 % 33 % 67 %</p>	Behavior of tourists	18%
	Local/Residents' satisfactions and emotions	15%
	Tourists' satisfaction & emotions	12%
	Tourists flow	10%
	Mobility data	10%
	Sustainability data	10%
	Typology of tourists	10%
	Purchase habits	9%
	Demand and offer data	6%
	Accessibility data	0%
	Related industry data	0%
	12 voters	12 voters

Challenge IV: Meet the demand for more sustainable tourism

A Eurobarometer survey from October 2021 indicated that 82% of Europeans are willing to change their travel habits for more sustainable practices, including consuming locally sourced products, reducing waste and water consumption, travelling off-season or to less visited destinations and choosing transport options based on their ecological impact. Accordingly, this challenge pertains to the fact that consumers' choices must be empowered, through more transparent information about the sustainability and environmental footprints of the destinations/tourism services.

As shown in the table, an **average rating of 4.5** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, sustainability data, and tourists flows** were considered the most useful data to solve this challenge.

Table 6– Increase of tourism sustainability & accessibility challenge 4 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>★ 4.5 ★ ★ ★ 0 % 0 % 8 % 33 % 58 %</p>	Behavior of tourists	15%
	Sustainability data	13%
	Tourists flow	13%
	Demand and offer data	13%
	Mobility data	10%
	Purchase habits	10%
	Typology of tourists	10%
	Tourists' satisfaction & emotions	6%
	Local/Residents' satisfactions and emotions	5%
	Related industry data	5%
	Accessibility data	2%
	12 voters	12 voters

Challenge V: Measure and reduce environmental footprint

The challenge of measuring and reducing the environmental footprint arises from the fact that comprehensive data collection by industry and sectoral category rules development is required to be able to compare data within a sector or product category on ‘average’, ‘low’ or ‘good’ performance. R&I is needed to help develop sectoral category rules development for specific tourism products and services in a transparent, non-discriminatory manner (e.g. considering peripheral destinations that need to import products). Tools and technologies are also needed to make applying the environmental footprint method more user friendly for SMEs. This would make the methodology accessible to them and facilitate data collection on the environmental footprint calculation to monitor the industry’s development on sustainability.

As shown in the table, an **average rating of 4.6** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, sustainability data, and mobility data** were considered the most useful data to solve this challenge.

Table 7– Increase of tourism sustainability & accessibility challenge 5 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0 % 0 % 8 % 23 % 69 %</p> <p>13 voters</p>	Behavior of tourists	17%
	Sustainability data	16%
	Mobility data	14%
	Tourists flow	14%
	Related industry data	11%
	Purchase habits	10%
	Demand and offer data	6%
	Typology of tourists	5%
	Accessibility data	3%
	Local/Residents' satisfactions and emotions	3%
	Tourists' satisfaction & emotions	0%
		13 voters

Challenge VI: Sustainable mobility

The European Green Deal calls for a 90% reduction in greenhouse gas emissions from transport, in order for the EU to become a climate-neutral economy by 2050, while also working towards a zero-pollution ambition. To achieve this systemic change, we need to make sustainable alternatives widely available. Accordingly, this challenge relates to the fact that immediate actions to adapt mobility system are necessary to tackle climate change and reduce pollution.

As shown in the table, an **average rating of 4.4** out of 5 was given to this challenge. Correspondingly, **mobility data, behavior of tourists, and tourists flow** were considered the most useful data to solve this challenge.

Table 8– Increase of tourism sustainability & accessibility challenge 6 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>4.4</p> <p>0 % 0 % 8 % 42 % 50 %</p> <p>12 voters</p>	Mobility data	27%
	Behavior of tourists	16%
	Tourists flow	16%
	Sustainability data	13%
	Demand and offer data	9%
	Related industry data	7%
	Accessibility data	4%
	Purchase habits	4%
	Tourists' satisfaction & emotions	2%
	Typology of tourists	2%
	Local/Residents' satisfactions and emotions	0%
		12 voters

Challenge VII: Improve the supply and visibility of accessible tourism services

Development and diversification of tourism products and services aimed at ensuring equal access to tourist destinations and cultural heritage. Accessible and inclusive tourism is a key to helping all people participate fully in society. There are approximately 87 million people with some form of disability in the EU²⁰, and in 2020, 20.6% of the EU population was aged 65 or over²¹. It is therefore important to ensure the supply of accessible tourism facilities in all destinations and provide clear and accessible related information to travellers planning and reserving their stays and activities. Moreover, based on Eurostat statistics from 2019, 35% of EU residents of 15 years or older did not make overnight trips²². The share of older people not doing overnight trips was higher than younger people, and 52% of persons not doing overnight trips mentioned financial reasons for it. Access for all to tourism could be boosted by developing moderately priced off-season accommodation and travelling opportunities for unemployed, retired and people with low income.

As shown in the table, an **average rating of 4.3** out of 5 was given to this challenge. Correspondingly, **mobility data, accessibility data, and demand and offer data** were considered the most useful data to solve this challenge.

Table 9– Increase of tourism sustainability & accessibility challenge 7 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p style="text-align: center; font-size: 2em; color: #4F81BD;">4.3</p> <p style="text-align: center;"> </p> <p style="text-align: center;"> 0 % 0 % 17 % 42 % 42 % </p>	Mobility data	16%
	Accessibility data	16%
	Demand and offer data	15%
	Behavior of tourists	10%
	Tourists flow	10%
	Typology of tourists	10%
	Tourists' satisfaction & emotions	9%
	Related industry data	7%
	Sustainability data	3%
	Purchase habits	1%
	Local/Residents' satisfactions and emotions	0%
12 voters	12 voters	100%

²⁰ https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6568

²¹ <https://www.disability-europe.net/downloads/1046-edc-task-2-1-statistical-indicators-tables-eu-silc-2018>

²² https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tourism_trends_and_ageing#Nearly_half_of_the_Europeans_aged_65.2B_who_did_not_make_tourism_trips_mentioned_health_reasons

Challenges ranking

Participants were then asked to rank the seven presented challenges. Accordingly, the challenges were ranked in the following way²³:

1. Support the growth of sustainable parameters awareness among the tourism industry
2. Meet the demand for more sustainable tourism and reduce negative impacts of tourists' behaviour on local communities and environment
3. Measure and reduce environmental footprint
4. Sustainable mobility
5. Manage tourism seasonality
6. Improve the supply and visibility of accessible tourism services

Missing challenges

Participants were then asked, through an open question, to investigate whether there was any missing challenge concerning the data purpose of the break out room, with a corresponding rate out of 5 and the data needed to address it. Accordingly, the following themes emerged:

- Spreading a new culture of sustainability in tourist operators. Rating 4/5
- Importance of relationship/influence among environmental/biodiversity indicators and economic indicators in the tourism sector. Rating 4/5

Break out room 2: Conduct market analysis and inform decision making

Moderators: Dolores Ordóñez from Anysolution

The second break out room was dedicated to the data purpose of conducting market analysis and informing decision making. Indeed, higher availability of data and improved data analytics capabilities allow improvements in market analysis and decision-making. For this specific data purpose, the project team uncovered 8 challenges: the respective findings of each of them are illustrated below.

Challenge I: Develop effective business strategies and make investment decisions

The challenge of developing effective business strategies and make investment decisions arises from the fact that both private businesses and public authorities need to ground their strategic and investment decisions on reliable information. The enhancement of data availability, data quality, and quality analysis are crucial for decision making. Accordingly,

²³ Such ranking was calculated assigning a weight to each of the rank positions. Therefore, the highest weight was assigned to each vote ranking the purpose as the most important (1st) and so on. Then, the final ranking is based on the sum of all the products of the number of votes for a ranking position times the weight assigned to that rank position.

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

the combination and interoperability of the many available data sources can help tourism players in extracting meaningful insights, improving their strategies and the way priorities are defined and budget allocated.

As shown in the table, an **average rating of 4.3** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, demand and offer data, and sustainability data** were considered the most useful data to solve this challenge.

Table 10– Conduct market analysis and inform decision making challenge 1 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>5 % 0 % 5 % 40 % 50 %</p>	Behavior of tourists	14%
	Demand and offer data	12%
	Sustainability data	11%
	Typology of tourists	11%
	Mobility data	9%
	Tourists flow	8%
	Accessibility data	8%
	Local/Residents' satisfactions and emotions	7%
	Tourists' satisfaction & emotions	7%
	Purchase habits	7%
	Related industry data	5%
	20 voters	20 voters

Challenge II: Monitor the flow of visitors between districts and attractions

Often tourism destination and attractions are “overwhelmed” and suffer from “overtourism” (e.g. Venice). In the case of cities, the population of the historical city center has been constantly falling (e.g. Venice -70% from its peak in the 1950s), as residents move to escape the nuisances of tourists and price increases and to conveniently make their homes available to online vacation rentals operators, such as most notably Airbnb (“airbnbification”). Moreover, “overtourism” also generates a set of other challenges, including of the effectiveness of public services, conservation of the cultural heritage (tangible and intangible), and management of security risks. For these reasons, cities and points of interest (heritage sites, natural parks, routes etc...) need a solution to better manage the flows of tourists. The challenge of managing large tourists flows can be also linked to specific large events (e.g. festivals, expo...).

As shown in the table, an **average rating of 4.0** out of 5 was given to this challenge. Correspondingly, **mobility data, tourists flows, and behavior of tourists** were considered the most useful data to solve this challenge.

Table 11– Conduct market analysis and inform decision making challenge 2 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>5 % 5 % 14 % 36 % 41 %</p>	Mobility data	18%
	Tourists flow	17%
	Behavior of tourists	14%
	Sustainability data	9%
	Local/Residents' satisfactions and emotions	9%
	Demand and offer data	8%
	Tourists' satisfaction & emotions	8%
	Purchase habits	6%
	Typology of tourists	5%
	Accessibility data	5%
	Related industry data	1%
	22 voters	22 voters

Challenge III: Answer to the so-called request for “hyper personalization” of travel experiences

Following the Covid-19 pandemic, there is a constantly increasing request for the hyper personalization of the travelling experiences. Developing hyper-personalized strategies presents a critical challenge; due to this, optimizing hyper-personalization and designing new processes and business models takes center stage in tourism and hospitality to reach new levels of customer service and experience through the introduction and development of new solutions supported in the internet of things, software interfaces, artificial intelligence solutions, back-end and front-end management tools, and other emergent business intelligence strategies.

As shown in the table, an **average rating of 3.3** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, typology of tourists and tourists' satisfaction & emotions** were considered the most useful data to solve this challenge.

Table 12– Conduct market analysis and inform decision making challenge 3 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>4 % 4 % 52 % 30 % 8 %</p>	Behavior of tourists	19%
	Typology of tourists	17%
	Tourists' satisfaction & emotions	13%
	Purchase habits	12%
	Demand and offer data	9%
	Mobility data	8%
	Tourists flow	6%
	Sustainability data	5%
	Local/Residents' satisfactions and emotions	4%
	Accessibility data	4%
	Related industry data	4%
	23 voters	23 voters

Challenge IV: Assess impacts of events and infrastructure investments

Events attracting big amounts of attendants have positive impact on the local economy, on the visibility and reputation of the location and on the cultural environment of the destination. At the same time, they might have negative spillovers. For this reason, the challenge of assessing impacts of events and infrastructure investments arises. The challenge of assessing economic, social and environmental impact of any activity is always increasing thanks to the 2030 Sustainable Development Goals agenda and the increasing awareness and demand of tourists. Similar reasoning can be applied to investment in tourism-related infrastructure (e.g. airports, streets, railways, new bus lines etc). Forecasting impacts can help in reducing the negative spillovers.

As shown in the table, an **average rating of 3.8** out of 5 was given to this challenge. Correspondingly, **sustainability data, mobility data, and tourists flow** were considered the most useful data to solve this challenge.

Table 13– Conduct market analysis and inform decision making challenge 4 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0 % 13 % 22 % 35 % 30%</p>	Sustainability data	14%
	Mobility data	12%
	Tourists flow	12%
	Behavior of tourists	10%
	Local/Residents' satisfactions and emotions	10%
	Demand and offer data	9%
	Accessibility data	9%
	Tourists' satisfaction & emotions	6%
	Purchase habits	6%
	Typology of tourists	6%
	Related industry data	6%
	23 voters	24 voters

Challenge V: Crisis management

The challenge of crisis management stems from the fact that the tourism industry and supply chain can be suddenly disrupted if a dramatic event (e.g. pandemic, natural catastrophes, terrorists attack etc) takes place. Accordingly, having updated information on these elements can help decision making regarding risk management, disaster management and recovery. In a nutshell, reliable information can support the pathway towards resilience.

As shown in the table, an **average rating of 4.0** out of 5 was given to this challenge. Correspondingly, **mobility data, tourists flow, and related industry data** were considered the most useful data to solve this challenge.

Table 14– Conduct market analysis and inform decision making challenge 5 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0 % 8 % 17 % 39 % 34 %</p>	Mobility data	18%
	Tourists flow	12%
	Related industry data	12%
	Behavior of tourists	11%
	Tourists' satisfaction & emotions	10%
	Demand and offer data	9%
	Sustainability data	8%
	Accessibility data	8%
	Typology of tourists	5%
	Local/Residents' satisfactions and emotions	4%
	Purchase habits	3%
23 voters	23 voters	100%

Challenge VI: Incentive collaboration among different types of stakeholders

Tourism is an industry that involves a wide variety of stakeholders. For this reason, vertical and horizontal governance are key for its success. Having information on points of convergence between central/regional/local public authorities and private market players can incentivize collaboration (e.g. PPP, collaboration between central and local PA, between private market players etc) for common purposes (e.g. tourism development strategies at public level and partnership even with competitors at private level).

As shown in the table, an **average rating of 4.3** out of 5 was given to this challenge. Correspondingly, **demand and offer data, tourists flow, and tourists' satisfaction & emotions** were considered the most useful data to solve this challenge.

Table 15– Conduct market analysis and inform decision making challenge 5 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0 % 0 % 19% 30 % 50%</p>	Demand and offer data	16%
	Tourists flow	11%
	Tourists' satisfaction & emotions	10%
	Typology of tourists	10%
	Mobility data	8%
	Related industry data	8%
	Behavior of tourists	8%
	Local/Residents' satisfactions and emotions	8%
	Purchase habits	8%
	Sustainability data	7%
	Accessibility data	6%
	26 voters	26 voters

Challenge VII: Manage the tourists-residents relationship

Tourism can be an opportunity for local communities, either from an economic perspective (increase revenues for local businesses, more business opportunities) and from a social perspective (increased employment, increased services innovation, exposure to different cultures). At the same time, when tourism turns into “overtourism”, the negative sides might be more than the positive ones (e.g. increase of prices for local communities, loss of places liveliness etc…), and this is why the challenge of managing the tourists-residents relationship arises. Being aware of benefits and problems might help public authorities in making decision managing tourism in a way that maximizes the positive sides and minimize the negative ones.

As shown in the table, an **average rating of 4.2** out of 5 was given to this challenge. Correspondingly, **local/residents' satisfactions and emotions, tourists flow, and tourists' satisfaction & emotions** were considered the most useful data to solve this challenge.

Table 16– Conduct market analysis and inform decision making challenge 7 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0 % 0 % 28% 24 % 48%</p>	Local/Residents' satisfactions and emotions	17%
	Tourists flow	13%
	Tourists' satisfaction & emotions	13%
	Behavior of tourists	13%
	Typology of tourists	10%
	Mobility data	10%
	Sustainability data	7%
	Demand and offer data	7%
	Accessibility data	4%
	Purchase habits	4%
	Related industry data	2%
	25 voters	25 voters

Challenge VIII: Cultural shift regarding tourism sector data sharing

All the aforementioned challenges can be addressed through the use of data. The availability of the data is key. For this reason, it is important that tourism-related data producers (and users) understand the benefits of data sharing and commit to enhance data availability and accessibility. Indeed, at the moment, tourism industry is facing a data silos problem, meaning that data are produced and kept at company/public authority level but there is no data infrastructure enabling their exchange. At the same time, data producers/holders might be afraid that sharing some kind of data could somehow result in a damage for their organization. For this reason, they wish to keep control over their data and decide when/how/with whom to share their data. A data space for the tourism sector aims exactly at solving this dichotomy: it enhances data sharing while it ensures owners' control over the data.

As shown in the table, an **average rating of 4.1** out of 5 was given to this challenge. Correspondingly, **typology of tourists, demand and offer data, and related industry data** were considered the most useful data to solve this challenge.

Table 17– Conduct market analysis and inform decision making challenge 8 results

Average rating	Type of data that are most useful to solve the challenge		Percentage
<p>4 % 0 % 25 % 25 % 45 %</p>	Typology of tourists		12%
	Demand and offer data		12%
	Related industry data		12%
	Mobility data		11%
	Purchase habits		9%
	Tourists flow		9%
	Behavior of tourists		9%
	Sustainability data		9%
	Tourists' satisfaction & emotions		7%
	Accessibility data		7%
	Local/Residents' satisfactions and emotions		5%
	24 voters	24 voters	

Challenges ranking

Participants were then asked to rank the six presented challenges. Accordingly, the challenges were ranked in the following way²⁴:

1. Develop effective business strategies & make investment decisions
2. Monitor the flow of visitors between districts and attractions
3. Incentivize collaboration among different types of stakeholders
4. Cultural shift regarding tourism sector data sharing
5. Assess impacts of events and infrastructure investments
6. Answer to the so-called “hyper personalization” of travel experiences
7. Manage the tourists-residents relationship
8. Crisis management

Missing challenges

Participants were then asked, through an open question, to investigate whether there was any missing challenge concerning the data purpose of the breakout room, with a corresponding rate out of 5 and the data needed to address it. Accordingly, the following themes emerged:

²⁴ Such ranking was calculated assigning a weight to each of the rank positions. Therefore, the highest weight was assigned to each vote ranking the purpose as the most important (1st) and so on. Then, the final ranking is based on the sum of all the products of the number of votes for a ranking position times the weight assigned to that rank position.

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

-
- Standardization on accommodations and other touristic resources. Ranking 4/5
- Measurement of the impact of implemented policies, using long term key performance indicators
- Challenges regarding energy and resources savings (e.g. water, food...)
- Challenges related to relation between tourism and health

Break out room 3: Improvement of the interaction and engagement of the tourist

Moderators: Danilo Bianchini from Intellera Consulting

The third breakout room was dedicated to the data purpose of improving the interaction and engagement of the tourist, which implies the development of tourism services that are increasingly personalized and based on a higher degree of interaction with the customer/end user. For this specific data purpose, the project team uncovered 6 challenges: the respective findings of each of them are illustrated below.

Challenge I: Address the demand/needs of tourists of different generations

Due to sociodemographic trends, the EU tourist sector will see demand changes in the short, medium and long term. For example, Baby Boomers will become the oldest target group in the coming decades, with the highest spending capacity, but with lower digital skills than the other groups of interest. They will progressively require a set of personalized touristic services based on a combination of health and cultural tourism. Similarly, millennials, generations Z and Y are/will be highly digitally skilled tourists, but with lower spending capacity than their parents. Moreover, generations Y and Z have the propensity to remain online, and are more likely to ask for more inclusive and interconnected digital solutions and adopt a more ethical lifestyle.

As shown in the table, an **average rating of 3.8** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, typology of tourists, and demand and offer data** were considered the most useful data to solve this challenge.

Table 18– Improvement of the interaction and engagement of the tourist challenge 1 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>6 % 0 % 25 % 44 % 25 %</p>	Behavior of tourists	17%
	Typology of tourists	17%
	Demand and offer data	11%
	Tourists' satisfaction & emotions	11%
	Accessibility data	9%
	Purchase habits	9%
	Mobility data	8%
	Tourists flow	7%
	Sustainability data	5%
	Local/Residents' satisfactions and emotions	4%
	Related industry data	1%
16 voters	16 voters	100%

Challenge II: Develop innovative immersive and digital tourism experiences

Consistently with the sociodemographic trends described in the previous challenge, tourists are and will always be asking more immersive and digital experiences. For example, virtual and augmented reality services enable real-like visitor experiences and might extend the participation to tourism services before and after the travel experience. R&I in this area could provide new innovative, sustainable and accessible forms of tourism services, with innovative technologies that can be used to provide new ways to help preserve natural and cultural resources at risk.

As shown in the table, an **average rating of 3.8** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, typology of tourists, and tourists' satisfaction & emotions** were considered the most useful data to solve this challenge.

Table 19 - Improvement of the interaction and engagement of the tourist challenge 2 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0% 14 % 7% 64 % 14%</p>	Behavior of tourists	24%
	Typology of tourists	13%
	Tourists' satisfaction & emotions	13%
	Demand and offer data	9%
	Tourists flow	9%
	Purchase habits	7%
	Accessibility data	6%
	Mobility data	6%
	Sustainability data	6%
	Local/Residents' satisfactions and emotions	4%
	Related industry data	4%
14 voters	14 voters	100%

Challenge III: Improve tourists' experience by providing information on tourism flows

This challenge pertains to the fact that information regarding the amount of people actually visiting a specific point of interest and information regarding the most crowded moments will surely allow residents and tourists to make their visits safer and more comfortable. For this purpose, some cities have already installed a network of devices that collect anonymous real time data from a crowd stream by detecting Bluetooth devices. Moreover, such devices can also transmit messages to the users, who can receive information and live suggestions regarding discounts, activities, events, or new routes for visitors, directly on their smartphones.

As shown in the table, an **average rating of 4.2** out of 5 was given to this challenge. Correspondingly, **mobility data, behavior of tourists, and tourists flow** were considered the most useful data to solve this challenge.

Table 20– Improvement of the interaction and engagement of the tourist challenge 3 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0% 6% 18% 25% 50%</p> <p>16 voters</p>	Mobility data	18%
	Behavior of tourists	17%
	Tourists flow	17%
	Accessibility data	10%
	Local/Residents' satisfactions and emotions	10%
	Tourists' satisfaction & emotions	7%
	Typology of tourists	6%
	Sustainability data	6%
	Purchase habits	5%
	Demand and offer data	2%
	Related industry data	2%
	16 voters	16 voters

Challenge IV: Manage online reputation and leverage on tourists' satisfactions

In the last decades, the impact of traditional “word of mouth” has been amplified by the tourists’ active use of Internet. During and following their travelling experience, tourists often publish online information regarding their satisfaction and emotions (“sentiment”) which can of course improve or damage tourism stakeholders. The sentiment analysis - based on the Natural language processing (NLP) - permits to measure the level of satisfaction expressed online regarding a service, a product or a brand. Being aware of one’s own reputation can help in designing strategies on how to engage the customers and improve their satisfaction.

As shown in the table, an **average rating of 4.1** out of 5 was given to this challenge. Correspondingly, **tourists’ satisfaction and emotions, local/residents’ satisfactions and emotions, and behavior of tourists** were considered the most useful data to solve this challenge.

Table 21 - Improvement of the interaction and engagement of the tourist challenge 4 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
	Tourists' satisfaction & emotions	21%
	Local/Residents' satisfactions and emotions	18%
	Behavior of tourists	12%
	Typology of tourists	12%
	Demand and offer data	11%
	Accessibility data	8%
	Purchase habits	6%
	Tourists flow	5%
	Sustainability data	5%
	Mobility data	3%
	Related industry data	0%
	16 voters	17 voters

Challenge V: Reach potential clients without being fully dependent on OTAs

Nowadays tourists are able to choose, configure and assemble a highly customized journey, instead of asking for a package from a travel agent or a website. They reduced the use of intermediation channels if they do not provide a higher added value than an online website. Together with the raise of cashless payments, this process incentivized the use of online intermediaries, highlighting the need to compare or aggregate tourism supply (e.g. online travel agencies – or OTAs – search engines, and price comparison platforms). On one side, these intermediaries give visibility to tourism players' offers and help them in reaching potential clients, on the other side, they are so widespread that it is extremely difficult for players to reach clients without referring to such platforms. Additionally, these platforms generate large amounts of data that put them in an always privileged position. Data availability could help tourism stakeholders (HORECA in particular) in finding ways to reach (potential) clients without using OTAs.

As shown in the table, an **average rating of 4.4** out of 5 was given to this challenge. Correspondingly, **demand and offer data, behavior of tourists, and typology of tourists** were considered the most useful data to solve this challenge.

Table 22– Improvement of the interaction and engagement of the tourist challenge 5 results


Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0% 0% 18% 29% 53%</p> <p>17 voters</p>	Demand and offer data	19%
	Behavior of tourists	17%
	Typology of tourists	17%
	Purchase habits	17%
	Related industry data	9%
	Tourists' satisfaction & emotions	7%
	Accessibility data	4%
	Sustainability data	4%
	Tourists flow	3%
	Mobility data	3%
	Local/Residents' satisfactions and emotions	0%
		17 voters

Challenge VI: Create and manage a relation with tourists

A lot of effort is required especially by small tourism businesses to attract new customers. It makes sense then to consider the potential for stimulating repeat visits and/or referrals from them, because it can a more efficient use of scarce resources to stay in touch with previous customers than to spend on advertising to attract a continual stream of new ones. CRM loyalty programs that provide economic incentives have a positive effect on customer retention and market share. A small increase in the number of loyal customers can result in reduced marketing costs, increased sales, and higher profits. While traditional promotional activities are necessary to attract new customers, there are opportunities for (especially) small tourism businesses to supplement these with initiatives to increase loyalty.

As shown in the table, an **average rating of 3.8** out of 5 was given to this challenge. Correspondingly, **purchase habits, behavior of tourists, and typology of tourists** were considered the most useful data to solve this challenge.

Table 23– Improvement of the interaction and engagement of the tourist challenge 6 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
 <p>0% 0% 38% 38% 23%</p>	Purchase habits	20%
	Behavior of tourists	19%
	Typology of tourists	17%
	Tourists' satisfaction & emotions	16%
	Demand and offer data	11%
	Accessibility data	5%
	Local/Residents' satisfactions and emotions	5%
	Mobility data	3%
	Tourists flow	2%
	Sustainability data	2%
	Related industry data	2%
	13 voters	13 voters

Challenges ranking

Participants were then asked to rank the six presented challenges. Accordingly, the challenges were ranked in the following way²⁵:

1. Improve tourists' experiences by providing information on tourism flows
2. Reach potential clients without being fully dependent on OTAs
3. Address the demand/needs of tourists of different generations
4. Create and manage a relation with tourists
5. Manage online reputation and leverage on tourists' satisfaction
6. Develop innovative immersive and digital tourism experiences

Missing challenges

Participants were then asked, through an open question, to investigate whether there was any missing challenge concerning the data purpose of the breakout room, with a corresponding rate out of 5 and the data needed to address it. Accordingly, the following themes emerged:

- Balance tourists-related companies/organizations' interest to engage with tourists' with their (possible) desire not to be overwhelmed by lot of information

Break out room 4: Improvement of planning and operations of the tourism service

Moderators: Marco Codastefano from Intellera Consulting

²⁵ Such ranking was calculated assigning a weight to each of the rank positions. Therefore, the highest weight was assigned to each vote ranking the purpose as the most important (1st) and so on. Then, the final ranking is based on the sum of all the products of the number of votes for a ranking position times the weight assigned to that rank position.

The fourth break out room was dedicated to the data purpose of improving the planning and operations of the tourism service. This implies that understanding, and possibly predicting tourism patterns through data can help to improve the overall efficiency and competitiveness of the tourism ecosystem. For this specific data purpose, the project team uncovered 5 challenges: the respective findings of each of them are illustrated below.

Challenge I: Understand and better forecast the tourism flow

The challenge of understanding and better forecasting the tourism flow relates to the fact that tourism enterprises are facing growing uncertainty. They will need to collect as much data as possible to cover the gaps in the decision-making process. A key factor to boost competitiveness is to make better and more innovative use of data, so that operators are able to anticipate demand for their services, analyze customer profiles and business trends, and ultimately provide a better customer experience.

As shown in the table, an **average rating of 4.6** out of 5 was given to this challenge. Correspondingly, **tourist flow, behavior of tourists, and mobility data** were considered the most useful data to solve this challenge.

Table 24– Improvement of planning and operations of the tourism service challenge 1 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>4.6</p> <p>0% 0% 0% 35% 65%</p>	Tourists flow	14%
	Behavior of tourists	12%
	Mobility data	12%
	Demand and offer data	10%
	Tourists' satisfaction & emotions	10%
	Purchase habits	10%
	Local/Residents' satisfactions and emotions	9%
	Accessibility data	8%
	Sustainability data	7%
	Typology of tourists	7%
	Related industry data	4%
17 voters	17 voters	100%

Challenge II: Manage seasonality in the tourism sector

Seasonality of demand is generally considered one of the major challenges in the tourism business. Destinations with high fluctuations in seasonality often face various challenges, such as overcrowding, high prices, inadequate infrastructure in peak seasons, as well as a lack of services and job opportunities in shoulder and low seasons. Accordingly, seasonality is a measurable feature with significant economic and social impacts. Understanding the main characteristics can help to modify its occurrence.

DATES: European Data Space for Tourism

Deliverable D2.2 Analysis of gaps and overlaps

As shown in the table, an **average rating of 4.3** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, demand and offer data, and typology of tourists** were considered the most useful data to solve this challenge.

Table 25– Improvement of planning and operations of the tourism service challenge 2 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0% 0% 15% 42% 42%</p>	Behavior of tourists	16%
	Demand and offer data	14%
	Typology of tourists	12%
	Tourists flow	11%
	Tourists' satisfaction & emotions	10%
	Sustainability data	10%
	Purchase habits	9%
	Local/Residents' satisfactions and emotions	8%
	Mobility data	5%
	Accessibility data	2%
	Related industry data	2%
	19 voters	19 voters

Challenge III: Manage and reduce overcrowding of sites and services

Depending on the type of destination and characteristics of its attractions, tourism flows can present more or less clear or recurring patterns – such in relation to seasonal attractions or specific landmark/sites. It is important to create innovative tourism services that manage tourism flows based on real-time data (e.g. crowd management at attractions). Data driven destination management tools, practices and technologies are needed to boost the sustainability of tourist destinations and reduce overcrowding of sites and services.

As shown in the table, an **average rating of 4.0** out of 5 was given to this challenge. Correspondingly, **behavior of tourists, tourist flow, and local/residents' satisfaction and emotions** were considered the most useful data to solve this challenge.

Table 26– Improvement of planning and operations of the tourism service challenge 3 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>0% 5% 22% 38% 33%</p>	Behavior of tourists	14%
	Tourists flow	14%
	Local/Residents' satisfactions and emotions	12%
	Mobility data	12%
	Demand and offer data	11%
	Typology of tourists	9%
	Tourists' satisfaction & emotions	9%
	Sustainability data	9%
	Purchase habits	4%
	Accessibility data	4%
	Related industry data	2%
	18 voters	18 voters

Challenge IV: Multimodal ticketing – smart mobility

This challenge arises from the fact that in the transportation sector, tracking and identifying tourists will become increasingly important. The technologies for controlling mobility and access will be key in this aspect. Digitalization of the tourism ecosystem makes it increasingly easy to find and book journey tickets online. However, this is not easy if the door-to-door trip requires multimodal transport, for example combining air, train and local transport. The digital transformation of the transport and mobility sector requires further efforts related to data availability, access and exchange.

Table 27– Improvement of planning and operations of the tourism service challenge 4 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p>4.1</p> <p>0% 11% 16% 22% 50%</p>	Mobility data	14%
	Behavior of tourists	12%
	Demand and offer data	12%
	Accessibility data	12%
	Tourists flow	11%
	Typology of tourists	9%
	Purchase habits	9%
	Tourists' satisfaction & emotions	7%
	Sustainability data	6%
	Local/Residents' satisfactions and emotions	4%
	Related industry data	4%
	18 voters	18 voters

As shown in the table, an **average rating of 4.1** out of 5 was given to this challenge. Correspondingly, **mobility data, behavior of tourists, and demand and offer data** were considered the most useful data to solve this challenge.

Challenge V: Manage online reputation and leverage on tourists' satisfaction

The challenge of managing online reputation and leverage on tourists' satisfaction concerns the fact that user-generated content is an important data source to explore tourist satisfaction and understand how to improve guest experience, competitive positioning and marketing activities. The analysis of these data allows to understand tourist behavior in a view to developing travel recommendation systems and improve marketing activities, with a view of improving overall efficiency and competitiveness of the tourism ecosystem.

As shown in the table, an **average rating of 3.9** out of 5 was given to this challenge. Correspondingly, **tourists' satisfaction & emotions, behavior of tourists, and local/residents' satisfactions and emotions** were considered the most useful data to solve this challenge.

Table 28– Improvement of planning and operations of the tourism service challenge 5 results

Average rating	Type of data that are most useful to solve the challenge	Percentage
<p style="text-align: center; font-size: 2em; color: #007bff;">3.9</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">★ 5%</div> <div style="text-align: center;">★ 11 %</div> <div style="text-align: center;">★ 5%</div> <div style="text-align: center;">★ 38 %</div> <div style="text-align: center;">☆ 38%</div> </div> <p style="text-align: center;">18 voters</p>	Tourists' satisfaction & emotions	22%
	Behavior of tourists	20%
	Local/Residents' satisfactions and emotions	14%
	Typology of tourists	14%
	Demand and offer data	12%
	Accessibility data	7%
	Tourists flow	5%
	Purchase habits	5%
	Related industry data	2%
	Mobility data	0%
	Sustainability data	0%
		17 voters

Challenges ranking

Participants were then asked to rank the five presented challenges. Accordingly, the challenges were ranked in the following way²⁶:

1. Understand and better forecast the tourism flow
2. Manage seasonality in tourism sector
3. Manage and reduce overcrowding of sites and services

²⁶ Such ranking was calculated assigning a weight to each of the rank positions. Therefore, the highest weight was assigned to each vote ranking the purpose as the most important (1st) and so on. Then, the final ranking is based on the sum of all the products of the number of votes for a ranking position times the weight assigned to that rank position.

-
- 4. Manage online reputation and leverage on tourists' satisfaction
- 5. Multimodal ticketing – smart mobility

Missing challenges

Participants were then asked, through an open question, to investigate whether there was any missing challenge concerning the data purpose of the breakout room, with a corresponding rate out of 5 and the data needed to address it. Accordingly, the following themes emerged:

- Need to inform all visitors and local residents about (temporary) closures of infrastructure (e.g. sights, hiking tracks, cycling routes, wildlife areas, avalanche danger zones...) before they come to closed locations and need to change plans on the spot. Rating 5/5
- Need to consider weather data for the improvement of planning and operations of tourism services

4. Next steps

As mentioned above, the team will use the evidence collected from the workshop for defining the use cases of the European data space for tourism. More specifically, starting from the challenges with the highest priority, the Team will carry out desk research and will organize a workshop (in **March**) to develop and discuss the use cases.

Such use cases will also be at the base of defining the data space technical and governance requirements. Consultations related to these topics will take place in the **upcoming months**.