Deliverable D3.1

Reference Architecture

Funded by
the European Union

09/05/2023

PROJECT INFORMATION

Project Number	101084007
Project Acronym	DATES
Project Full title	European Data Space for Tourism
Project Start Date	1 November 2022
Project End Date	31 October 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	D3.1
Deliverable Title	Reference Architecture
Work-Package No	WP3
Work-Package Title	Design and Technical Framework
WP-Leader (Name and Short Org. Name)	FUNDACIÓN TECNALIA RESEARCH & INNOVATION - TECN
Tasks No	3.1
Tasks Title	Reference architecture for tourism data space
Task Leader (Name and Short Org. Name)	FUNDACIÓN TECNALIA RESEARCH & INNOVATION - TECN
Main Author (Name and Short Org. Name)	Valentín Sánchez (TECN)
Other Authors (Name and Short Org. Name)	Andreas Eisenrauch (AMAD), Eric Pol (ANG), Belén Martinez (TECN), Jesús Herrero (TECN), Mariano Blaya (IDSA)
Reviewers (Name and Short Org. Name)	Sonia Bilbao (TECN), Belén Martinez (TECN)
Status	Draft ☒ Final ☐
Deliverable Type	Report ☒ Data ☐ Demonstration ☐ Data ☐
Dissemination Level	Public (PU) ☒ Sensitive (SEN) ☐ EU Classified (EUC) ☐
Date Approved by Coordinator	9/05/2023

DOCUMENT REVIEW HISTORY

Date	Version	Reviewer Name (Short Organisation Name)	Summary of main changes
15/04/2023	0.1	Valentin Sánchez (TECN), Andreas Eisenrauch (AMAD), Mariano Blaya (IDSA), Jesús Herrero (TECN)	First version
16/04/2023	0.2	Andreas Eisenrauch (AMAD)	Added section 2.4
21/04/2023	0.3	Eric Pol (ANG)	Added content about Personal data
05/05/2023	1.0	Valentin Sánchez (TECN)	First complete version
09/052023	1.1	Sonia Bilbao (TECN), Belén Martinez (TECN)	Final revised version
9/05/2023	1.2	Tayrne Butler (ANYSOL)	Final formatting and addition of the list of abbreviations and acronyms

FINAL ACCEPTANCE BY COORDINATOR: 9/05/2023

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

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

Index of Figures

List of Abbreviations and Acronyms	
AISBL	Gaia-X Association for Data and Cloud
API	Application Programming Interface
BAE	Business Application Ecosystem
C2D	Compute-to-data
C2E	Compute-to-edge
CSV	Comma separated vaule
DSBA	Data Spaces Business Alliance
DGA	Data Governance Act
DID	Decentralized identifier
DMO	Destination Management Organization
DSA	Data Space Authority
DSSC	Data Spaces Support Center
DATES	Data Space for Tourism
EC	European Commission
eID	Electronic identification
eiDAS	electronic IDentification, Authentication and trust Services
EU	European Union
FAIR principles	Findable, Accessible, Interoperable and Re-usable
FIWARE	Future Internet Core Platform
GAIA-X	A Federated and Secure Data Infrastructure
GDPR	General Data Protection Regulation
GE	(Fiware) Generic Enabler
GXDCH	Gaia-X Digital Clearing House

GXFS	Gaia-X Federation Services
ICT	Information and Communication Technologies
IDM	Identity Management
IDS	International Data Space
IDSA	International Data Spaces Association
MyData	Human-centric approach to personal data management
NGSI	Next Generation Service Interfaces
NGSI-LD	Next Generation Service Interfaces – Linked Data
OPENDEI initiative	Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry
OTA	Open Travel Alliance / Open Travel Agency
PDP	Policy Decision Point
PEP	Policy Execution Point
PoC	Proof of Concept
POI	Point of Interest
REST	Representational state transfer
Simpl	EU project to design and implement the Smart middleware platform
SME	Small and Medium Enterprise
SOLID Pod	Decentralized data stores
SP	Service Provider
TA	Trust Anchor
TDS	Trusted Data Sources
UNWTO	United Nations World Tourism Organization
VP	Verifiable credential
WP	Work Package
XSLT	Extensible Stylesheet Language Transformations

EXECUTIVE SUMMARY

Data Spaces are a key instrument to promote the data sharing economy by accelerating the data economy in different verticals. Tourism, being one of the most powerful industries in Europe as worldwide tourism destination, will increase its competitiveness thanks to the creation of a sound European Tourism Data Space. DATES will work intensively with the whole tourism ecosystem to set the bases of the European Data Space of tourism based on the principles of resilience and sustainability.

The Communication and Engagement Plan provides the guidelines for effectively sharing information within the consortium and with the whole tourism ecosystem at the level of the EU and beyond as well as to define a sound strategy for transferring DATES' knowledge and results to the targeted stakeholders. This plan is designed to combine contents, target audience and means, while striving to achieve sound visibility across a multi-layered community.

1 PURPOSE OF THE DOCUMENT

This report presents the results of Task T3.1 Definition of the Tourism Data Space conceptual architecture according to European framework for data sharing. After the analysis of gaps and the inventory of data sets performed in WP2, WP3 has analysed the Technical Framework required for the implementation of the Tourism Data Space.

This work has been done in collaboration with the main European bodies and instruments for the efficient deployment of data spaces in Europe, specifically, the **Data Spaces Support Centre** and the main European initiatives in the area of data spaces, mainly GAIA-X, IDSA and FIWARE, ensuring the alignment of the tourism data space with the European Data Spaces Technical Framework and the rest of the ecosystem of data spaces.

The collaboration has been established in several ways:

- Analysing the technical documentation available.
- Installing and testing the components from different initiatives
- Participating in internal workshops and discussions
- Attending meetings other Data Spaces CSAs and DSSC

According to the proposal and grant agreement the activity of T3.1 has focused mainly on:

- Identifying the **technical requirements** for the data infrastructure of tourism data spaces (e.g. technical design, functionality, operation and governance).
- Defining the guiding design principles for the creation of tourism data spaces.
- Identifying architecture and technical data governance frameworks, including the tourism domain's specific elements.
- Identifying the **common building blocks** essential for the creation of tourism data spaces and defining technical specifications.
- Identifying common standards, including semantic standards and interoperability protocols – both tourism-specific and crosscutting.
- Identifying the potential for synergies between data spaces and coordinate related cross-cutting exchanges between data spaces.
- Identifying common toolboxes that could be used across data spaces.
- Identify existing APIs that are relevant for a Tourism Data Space

The main results have been the identification of the common building blocks, the specific requirements imposed by the tourism sector and some specific aspects of the Tourism Data Space and how they influence the data space technology to be deployed.

The report contains intermediate results of the WP. Therefore, this document must be considered as a working document that will evolve along the project's lifetime according to the ongoing work carried out by each of the initiatives that are analysed along the document, the more detailed work in task T3.2, WP2 and WP4 achievements regarding data models and standards, data governance and tourism specific use cases as well as the collaboration with the sister CSA in tourism, which is working in parallel.

1.1 Methodology

Data spaces is a hot topic in Europe. Many initiatives are running in parallel defining business cases, the scope, legal framework and technology needed to implement and deploy data spaces. Most of them are far from being stable and are producing open-source components which are not mature enough to be used in a commercial environment. Even worse, those components are evolving at a very high pace and are not interoperable with each other. Each initiative is deploying its own so-called Minimum Viable Product, built with components developed internally and highly dependent on specific versions of the components. Convergence or interoperability among the different initiatives is still an open issue.

Most of the data spaces already deployed are only prototypes, proof of concepts or minimum viable products useful for demonstration purposes, not real commercial deployments. On the contrary, data sharing among companies has been a reality for a long time by different means and technologies, so it must be clearly demonstrated what are the real benefits of using the data space approach.

In this changing and uncertain environment, it is very difficult to know in advance what will be the "winning" initiative will be, if any.

It is not the objective of DATES to define "yet another" data space reference architecture. Our objective is firstly to identify the main technical issues and challenges to be solved when implementing the data space concept. Then we will describe the main initiatives and how those issues are being tackled.

The following is a brief description of the Tourism sector, its value chain and how data is currently used and shared by the tourism stakeholders. Furthermore, some new and advanced tourism data sharing use cases are envisaged. These use cases describe specific ways of using data that the current data sharing initiatives are not covering and that would be possible using the data space approach. This information is based on the input from WP2 included in D2.1 and D2.2 and the ongoing task T2.3.

Also, an example of a current data sharing initiative and an analysis of the gaps and the path to evolve towards a data space is included: the Tourism Digital Hub by the Italian Ministry of Tourism.

Finally, the document presents some specific aspects of the Tourism data space and how these aspects influence the data space technology to be deployed.

In order to describe the current State of the Art (SOTA) of Data Space architectures and technology some data space concepts must be clearly defined, the main one being the "**Data Space**" concept itself. Also, the main stake holders and roles participating in a data space and the high-level building blocks that makeup a data space architecture must be defined, i.e., a common data spaces glossary. This common glossary will provide the basis for comparing the different approaches and technologies described in this report.

Three sources of information have been selected to set the common ground for the rest of the document: a position paper from the OPENDEI project, a glossary published by the Data Space Support Centre (DSSC), and the IDS Rulebook 2.0.

Finally, the following data spaces architectures initiatives have been included in the SOTA, since they are currently the most relevant in Europe: the International Data Spaces Association (IDSA), GAIA-X, FIWARE, The Data Spaces Business Alliance (DSBA) and the SIMPL platform.

2 OVERVIEW OF THE DATA SPACES TECHNOLOGY LANDSCAPE

This section describes the current state of the art regarding the Data Spaces technology landscape.

Three sources of information have been selected to set the common ground for the rest of the document:

- OPENDEI project: Position-paper "Design principle for data spaces" providing the technical and governance building blocks.[1]
- Data Space Support Centre (DSSC): Glossary[2]
- IDS Rulebook 2.0[3]: The aim of the rulebook is to clearly describe rules and guidelines for building and operating data spaces, taking into account functional, **technical**, operational, and legal dimensions.

Our objective is firstly to identify the main technical issues and challenges to be solved when implementing the data space concept. Then we will describe the main initiatives and how they are tackling those issues.

A description of the data spaces main technical issues and concepts has been gathered, based on the OPENDEI definition of the technical pillars and building blocks, the foundational and support concepts described in the IDS Rulebook and the Gaia-X Documentation.

The following data spaces architectures initiatives have been included in the SOTA, since they are currently the most relevant in Europe:

- International Data Spaces Association: **IDSA**[4]

[1] Design Principles for Data Spaces | Position Paper (design-principles-for-data-spaces.org)
[2] DSSC-Data-Spaces-Glossary-v1.0.pdf
[3] Front Matter - IDSA Rulebook V2 (internationaldataspaces.org)
[4] https://internationaldataspaces.org/

- **GAIA-X**[5]: A Federated and Secure Data Infrastructure
- **FIWARE**[6]: The Open-Source Platform for Smart Digital Future
- The Data Spaces Business Alliance: **DSBA**[7]
- **SIMPL**[8]: Smart Middleware Platform for Cloud-to-edge Federations and Data Spaces

2.1 Main technical issues regarding data spaces

Data is an essential building block of modern societies, and we are constantly producing more of it: citizens create data while browsing the internet, booking a trip, or buying clothes online, and public bodies generate data from weather tracking and traffic monitoring. The European Commission has developed a European data strategy[9] to help unlock its benefits. One key component of this strategy is the development of common European data spaces in strategic economic sectors and domains.

According to the DSSC glossary, a **data space** is a framework that supports data sharing within a data ecosystem. It provides a clear structure for participants to share, trade, and collaborate on data assets in a way that is compliant with relevant laws and regulations and ensures fair treatment for all involved.

In business, data spaces enable the trusted and secure sharing of commercial data assets with automated controls on legal compliance and remuneration. This can create a market among participants or facilitate collaboration among diverse, interconnected parties who rely on each other for mutual benefit. Personal data spaces, on the other hand, allow data subjects and holders to control their data and its subsequent use, ensuring that legislation surrounding the handling of personal data is followed.

According to the IDSA Rulebook, the **foundational concepts** of a data space are the following:

- Establishing of trust
- Data discoverability
- Data contract negotiation
- Data sharing & usage
- Observability
- Vocabularies and semantic models

[5] https://gaia-x.eu/
[6] https://www.fiware.org/
[7] https://data-spaces-business-alliance.eu/
[8] https://etendering.ted.europa.eu/cft/cft-display.html?cftId=12922
[9] https://digital-strategy.ec.europa.eu/en/policies/strategy-data

Additional elements that support these main functions of a data space can include these optional functional areas:

- Application and processing services
- Marketplaces
- Data trustee and escrow services
- Data incubation and service creation

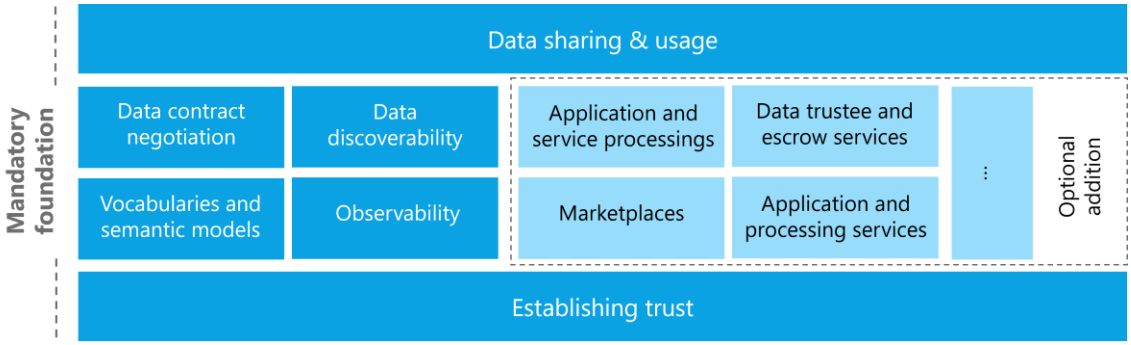

Figure 1 Data spaces foundational concept. Source: https://github.com/International-Data-Spaces-Association/IDSA-Rulebook/blob/main/documentation/media/Foundational_Concepts_in_data_spaces.png

According to the OPENDEI position paper: *"**The technical building blocks** enable the implementation of the technical architecture of a data space. They include network protocols, middleware components, (standardized) APIs, and more, facilitating the sharing of data between different parties in a secure and trustworthy fashion"*. They have defined what they call "Technical pillars" which address most of the technical concerns associated with the creation of data spaces:

- **Data interoperability**, covering aspects such as data exchange APIs, data representation formats as well as data provenance and traceability.
- **Data sovereignty**, covering aspects such as identity management, trustworthiness of participants, as well as data access and usage control.
- **Data value creation**, covering aspects such as publication of data offerings, discovery of such offerings based on metadata and data access/usage accounting, which are essential to handle data as an economic asset.

Based on these high-level categories and the discussions taking place now in the current data space initiatives, we have identified the list of concepts below, that we have used to compare the different approaches. The selected topics do not cover all the concepts and blocks, but they are the ones that are currently under discussion and development among the main data space initiatives.

- Identity
- Self-description
- Trust framework
- On-boarding
- Data and services discoverability: Catalogue

- Data sharing
- Data space monitoring and observability

Two more topics have been included in the analysis, since we think they are very useful in the context of the tourism data space.

- **Service orchestration**: It facilitates the SMEs with limited technical skills to compose several services to obtain the expected result.
- **Personal data management**. Personal data is the core of the tourism data space so it must be managed according to the ethical and legal constraints applicable, especially in Europe.

2.1.1 Identity

Identity plays a critical role in the context of data spaces. Data spaces are designed to enable efficient and effective sharing and management of data across different systems and applications, which requires a secure and reliable system for identifying and authenticating users.

Identification mechanisms are the basis for finding attributes of a participant in a data space and they allow the participant to exert control, to choose which data to share with whom, when and under what conditions. This ensures the participant has agency over its assets, i.e., data sovereignty.

In a data space, identity is used to establish trust between different users and systems, and to ensure that access to data is controlled and secure. For example, users may need to authenticate themselves in order to access certain data or perform certain actions within the data space. This helps to prevent unauthorized access and ensure that data is only accessed by authorized users.

In addition, identity can be used to manage data ownership and attribution within data spaces. For example, a user's identity may be associated with the data they create or contribute to, which can help to ensure that they receive proper credit and recognition for their contributions.

The design of the identity provider is the first decision for the design of the data space. If a central identity provider is chosen to manage the identities for all participants, every other service depends on this central verification, and decentralized designs are no longer fully feasible.

Which mechanism to use to identify participants is the most fundamental design decision. It impacts policies on autonomy and sovereignty as well as technical solution architectures for other components of a data space.

Regarding identity, GAIA-X proposes the use of a decentralized identity model, based on the principles of self-sovereign identity (SSI). This approach enables individuals and organizations to control their own identity information, without relying on centralized authorities or third-party providers.

In the GAIA-X model, individuals and organizations would be able to create and manage their own digital identities, which could be verified by trusted third-party providers. These identities would be based on open standards and would be interoperable across different systems and applications.

The use of decentralized identities would enable individuals and organizations to have greater control over their personal data and how it is shared. It would also help to increase security and reduce the risk of identity theft and other types of fraud.

However, identity is not only used for natural and legal persons, every dataspace component, including data sets, resources, applications, services, etc., can have an identity. Even the data space itself should have an identity since several independent data spaces can be deployed and collaborate with each other.

2.1.2 Self-Description

Participants in a data space ecosystem talk to each other in the language of Self-Descriptions. They use metadata to describe themselves, their service offerings, as well as the resources that their service offerings are composed of.

The Self-description concept was initially defined by the IDS Reference Architecture, in a restricted way, only for connectors and legal persons. GAIA-X has extended the concept to every data space component including Participants, Services and Resources. Even the data space itself has its own self-description.

In the context of data components, the term self-description refers to the ability of data objects or entities to describe themselves in a machine-readable format. It allows for easier and more efficient sharing, discovery, and integration of data across different systems and platforms.

Self-description involves providing metadata, which is information about the data itself, such as its format, structure, content, and relationships to other data elements. This metadata can be automatically extracted from the data or provided by the data producer or owner.

Self-description is an important aspect of interoperability and data integration in data spaces, as it enables systems to automatically understand and use data from different sources without the need for manual intervention or customization. It also facilitates data, participants and services discovery and enables data/services consumers to assess the quality and relevance of the data before using it.

GAIA-X is now defining the whole self-description models of the different components taking part in a data space.

2.1.3 Trust framework

Once a data space component has an associated identity and a self-description, all the information included in the self-description must be certified.

Gaia-X has defined the so-called “Trust Framework” as the process of going through and validating the set of automatically enforceable rules to achieve the minimum level of Self-Description compatibility in terms of:

- syntactic correctness
- schema validity
- cryptographic signature validation
- attribute value consistency
- attribute value verification.

On the other hand, the Gaia-X Trust Framework provides the set of rules that define the minimum baseline for being part of the Gaia-X Ecosystem. Those rules provide common governance and the basic level of interoperability between individual ecosystems, while allowing users full control of their choices.

2.1.4 Data space onboarding

A data space needs to define policies that specify what attributes an applicant must meet to become a trusted participant. This is achieved through **Data Space Self-Description (DSSD). DSSD** allows new members to provide attributes in their **Participant Self-Description (PSD**) in a format that can be understood by the **Data Space Authority (DSA)**.

Deciding which **trust anchors** and **trust frameworks**, and thus which rules and procedures of issuing and validating attributes are used, is the responsibility of the DSA and of the participants of the data space.

Participation in a data space is based on fulfilling all the policies, rules and procedures that are mandatory for membership. In its simplest form, these may be just technical or automatically verifiable policies. In more advanced cases, these can be more complex policies and rules that potentially require lengthy workflows with human interaction to verify eligibility to join a data space (e.g., a signed legal contract with a central operating company, membership in industry associations, etc).

The onboarding process defines the procedure to join a data space. It will likely include the following steps for the applicant:

1. The candidate discovers the data space and the corresponding Data Space Self-Description (DSSD)
   This can be achieved through human interaction, a website of the data space, finding the DID of the data space in some registry or through automated discovery protocol of existing participants among other things.

2. The candidate reads the DSSD and receives information about the policies and rules of the data space, as well as technical configuration information for endpoints and protocols.

3. The candidate evaluates the policies and rules and prepares additional information needed for the requirements when applying for membership in the data space.

4. When all the information and necessary proofs are collected, the candidate applies for membership through the registry function of the DSA. The technical implementation of the data space registry might vary based on the requirements.

5. The DSA requests proofs for all policies. This might include VCs and proof of technical capabilities, but also workflows including human interaction (e.g., signing a membership contract).
6. Once all policies have been satisfactorily processed the DSA issues a VC/ proof of membership and sends it to the candidate, moving them from applicant to participant.
7. The new participant sets up all the necessary technical components for participation in the data space.
8. The application process has been completed, the participant can start interacting with other participants (sharing data, browsing the catalogue(s) for data of others, negotiating data contracts).

Both GAIA-X and IDS have defined the role of the Data Space Authority.

- **IDS Data Space Authority**

The data space authority (DSA) is responsible for establishing the policies and rules of the data space. This role can be carried out by one entity, but also by multiple or even all participants. In a centralized data space, this could be the operating company. In a federated data space, this function would be performed by the federator(s) agreeing on the rules, while in a fully decentralized data space, various mechanisms are available to the participants. The mechanisms in a decentralized data space enable participants to agree on the set of policies and their enforcement, thus sharing responsibility for the data space authority function.

- **GAIA-X Clearing House**

Gaia-X Digital Clearing House[10] (GXDCH) is the one-stop place to go and get verified against the Gaia-X rules to obtain compliance in an automated way.

The Gaia-X Framework describes functional specifications, technical requirements, and SW assets necessary to be Gaia-X compliant. The GXDCH are a network of execution nodes for the compliance components that we have developed. This safeguards the distributed, decentralised ways of running the Gaia-X compliance mechanism, not operated centrally by the Association, and where anybody can benefit from the open, transparent, and secure federated digital ecosystem.

1. The **GXDCH is a node of verification** of the Gaia-X rules
2. It is the **go-to place to obtain Gaia-X compliance** and become part of the Gaia-X ecosystem

[10] GXDCH - Gaia-X: A Federated Secure Data Infrastructure

3. The GXDCH are non-exclusive, interchangeable **multiple nodes operated by market operators**, acting as a Gaia-X Federator
4. They **operate and run services of the Gaia-X Framework** (compulsory and optional), necessary to achieve compliance and support the onboarding of any Gaia-X adopter
5. They **integrate to external TA** (Trust Anchors), including CAB (Conformance Assessment Bodies) for external asseverations, Identity Verification (like eIDAS), and other TDS (Trusted Data Sources) as defined by the AISBL

2.1.5 Data and services discoverability: Catalogue

Discoverability refers to the ability to locate and access relevant data and services within a data space. Discoverability is critical for enabling efficient and effective data sharing and collaboration among users.

To achieve discoverability, data and services must be appropriately labeled, categorized, and indexed to facilitate search and retrieval. This involves the use of metadata, which provides information about the data, such as its format, content, quality, and source. Metadata also enables data and services discovery through the use of search engines and other discovery tools that allow users to browse, query, and filter resources based on different criteria.

In the case of data products, discoverability can also be enhanced through the use of data standards, ontologies, and other data management best practices that promote consistency, interoperability, and reuse of data. These practices help ensure that data can be easily discovered and understood by a wide range of users, regardless of their background or expertise.

The Catalogue is a key component of a data space and plays a critical role in facilitating data discovery, access, and sharing. It is essentially a repository or database that stores information about the data assets within the data space.

Two participants can share data by communicating directly online or offline without the need for a catalogue. But for more participants a catalogue function greatly increases the discoverability of data assets and services. If there is more than one catalogue due to a federated or decentralized design, the catalogue must allow federated searches of data assets in catalogues at multiple sites.

When choosing a target architecture for a data space, the design of the catalogue function can fall somewhere along the spectrum between a central catalogue, multiple federated catalogues, and many decentralized catalogues.

In the case of IDS, two different catalogues have been included in the reference architecture: the Meta data broker for data and the App broker for applications to be used within connectors.

GAIA-X is still defining the Federated Catalogue and the Eclipse Dataspaces Components git also includes an EDC Catalogue.

2.1.6 Data sharing

Once a participant has joined a data space and discovered available data contract offers, the mechanism of data sharing is initiated. Data sharing is the core activity to enable further data processing and value generation by using the data.

Before the actual data transfer starts, two previous steps are mandatory:

1. **Authentication:** The identities and Trust Framework are essential. Without this, you cannot connect two Participants. Identities provide general information on the Participant, and the Trust Framework appends additional claims, like verified location, or verified application of other standards or regulations.
2. **Policy negotiation and contracting** include the ability to negotiate access and usage policies between 2 parties.

Data sharing must accommodate a wide range of scenarios. From a simple file transfer between two storage providers, to API access for streaming or eventing, to quite complex implementations with secure execution environments through confidential compute enclaves, environment attestations, signed code, custom encryption algorithms, and more. The most appropriate solution depends on the data protection needs and the trust level between the participants.

2.1.7 Data space monitoring and observability

In data spaces with highly regulated data, it is necessary to make the data sharing process observable. This can be done for legal reasons to prove that data has been processed only by authorized entities, or for business reasons to provide a marketplace and billing function through a trusted third party.

Depending on the architecture of the data space, multiple solutions are possible. For a centralized architecture a central observer (clearing house, auditor or monitoring agent) can be implemented. Another option is to define a federated model of observers to distribute the information, load, and potential for error. To go a step further, a decentralized architecture can minimize the risks associated with a centralized or federated observer model.

In the case of GAIA-X, the contract negotiation can lead to both parties agreeing on a Data Transmission Logging Service which is then used by both sides to log data transmission details. The logs might also include information needed for billing and service level details.

IDSA has included in the reference architecture the “clearing house”, which receive and store the logs produced during the data transfer process.

2.1.8 Service orchestration

GAIA-X scope is not only to provide services and resources but also the possibility to compose and orchestrate different services. For example, in GAIA-X one company could use the cloud infrastructure from one provider to run an algorithm offered by other provider with the data available from a third participant and get the final result. In this scenario a workflow editor and engine should be provided to define and execute the different processes needed. This is a very complex scenario but very useful for a company without the technical skills needed to participate in a data space.

2.1.9 Personal data

The protection of personal data is a critical concern in Europe, and data spaces need to be designed to comply with the EU's General Data Protection Regulation (GDPR) and other data protection laws. Ensuring that data is processed in a way that is secure and respects individual rights is a key issue for data spaces.

2.2 EU initiatives related to Data Spaces

This section describes the main EU initiatives related to Data Spaces from a generic point of view, i.e., without considering the tourism sector specificities. The following initiatives and projects have been selected:

1. OPEN DEI project
2. DSSC: Data Space Support Centre
3. Personal data related initiatives
    - MyData
    - SOLID Pods
    - EU Wallets
4. IDSA: International Data Space Association
5. GAIA-X
6. DSBA Data Spaces Business Alliance
7. SIMPL: Smart middleware platform

Some of the information included in the following sections has been taken directly from the sources.

2.2.1 OPEN DEI: Data space principles and building blocks

The Horizon 2020 project “OPEN DEI Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry” focused on “Platforms and Pilots” to support the implementation of next generation digital platforms in four basic industrial domains. The project defined four Task Forces to address medium to long term challenges. The first Task Force was dedicated to DATA SHARING SPACES and its objective is to “define for the first time cross-sectoral and across initiatives the fundamental design principles to

build data spaces". One of the main results of TF1 has been the position paper "Design principles for Data Spaces".

The position paper underlines the importance of data spaces and though the sovereign sharing of data in creating the future data economy. It has been developed under the coordination and leadership of the International Data Spaces Association with the collaboration of more than 40 data spaces and industrial domain experts representing more than 25 organisations from 13 Horizon 2020 projects and related initiatives. This is the first approach to define the design principles for data spaces, agreements on the building blocks for a soft infrastructure and governance for data spaces.

According to the position paper, in each data space several building blocks need to be considered when setting up. They delineate areas where choices are required to enable effective and trusted sharing of data among participants that ultimately creates value and fall under two types, the technical building blocks and the governance building blocks:

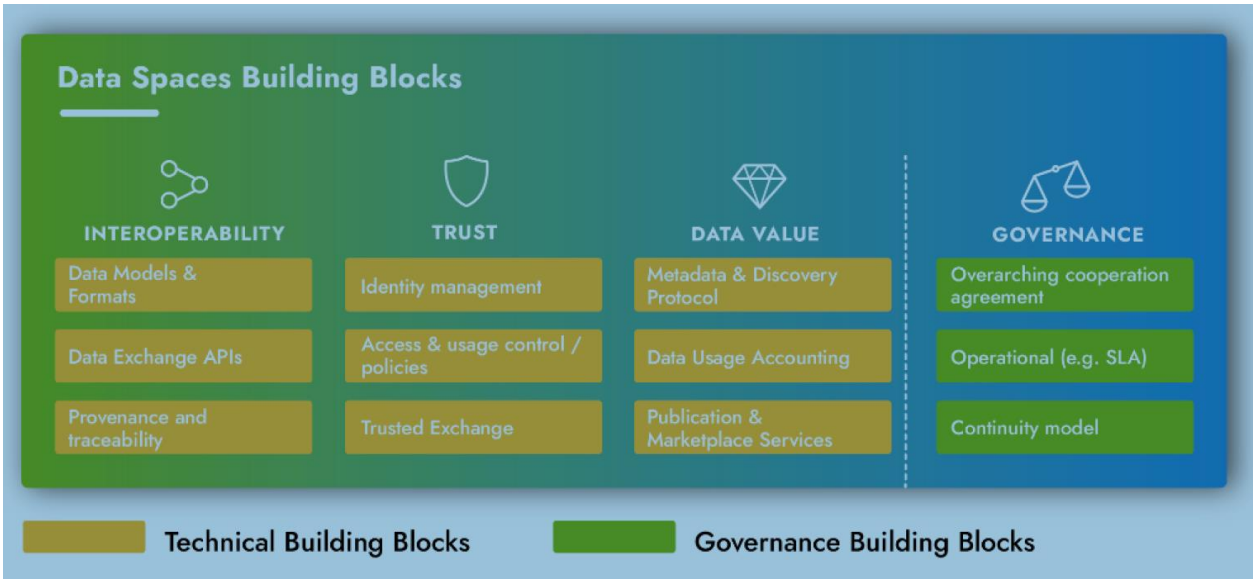

Figure 2 Data spaces building blocks

In DATES project, WP3 is focused on the **technical aspects** including technical requirements, architecture, data standards as well as common toolboxes that could be used across data Spaces, while WP4 defines the **sustainable governance framework** that will allow fair data sharing in the ecosystem. It will also identify business models that will ensure the sustainability of the tourism data space.

Building blocks and specifications defined by the OpenDEI project will save precious time for the DATES project, allowing us to focus on the technologies used by similar projects and data spaces rather than having to go through all technologies used across all industries' initiatives.

A selection of the technologies has been made available as part of the WP2 and are documented in the D2.3, chapters 2.2.2 and 3.3.

2.2.2 DSSC: Data Space Support Centre

The Data Spaces Support Centre will explore the needs of data space initiatives, define common requirements and establish best practices to accelerate the formation of sovereign data spaces as a crucial element of digital transformation in all areas.

DSSC has a plan to design, prepare and deliver a blueprint among other deliverables. At the time this document was edited, the glossary and starter kit were made public:

- Glossary[11]: The function of the glossary is to support smooth information sharing and co-development between the different data space initiatives and people involved and working with the Data Spaces Support Centre (DSSC). The secondary objective is to establish a common vocabulary for DSSC communication and publications of the DSSC.
- Starter Kit[12]: The European Data Strategy aims to build a strong and innovative data ecosystem in Europe through the development of common data spaces in strategic economic sectors and domains. Data spaces are frameworks that support data sharing within a data ecosystem and allow for the trusted and secure exchange of data assets while ensuring compliance with relevant legislation and fair treatment for all involved. The Data Spaces Support Centre has created a starter kit to help organizations navigate the challenges of creating and maintaining a successful data space. The kit includes a resource inventory that provides an overview of state-of-the-art and state-of-practice resources available for organizations interested in creating or participating in data spaces.
- DSSC Standards and specifications landscape: although it has not been made publicly available, the DATES project has made every effort to be aligned with DSSC regarding the building blocks and specifications. After some alignment meetings, we agreed with the DSSC WP5 to take the latest and stable version as the basis for our work on building blocks and specifications. If, within the lifecycle of the DATES project and new version of the spreadsheet is shared, we will consider it.

DSSC is a three-year project, and it has not even completed the first of those three years. It is work in progress and should be understood as such. Having said that, DSSC is helping all dataspaces initiatives to minimize the duplication of effort.

All projects in the dataspaces sector need to put in place mechanisms to ensure they are aligned. If, at any point of the DSSC project, a deliverable is issued, all projects should make their best effort to adopt it.

DATES project has decided to use the DSSC Glossary as the common glossary to define the data space related concepts.

[11] https://dssc.eu/download/1255/?tmstv=1679300669
[12] Navigating the challenges of creating a data space – DSSC

2.2.3 Personal data management and existing initiatives

The importance of personal data to data sharing

The EU Data Strategy's success relies heavily on sharing personal data, which is expected to drive adoption across various sectors such as health, administration, education, and mobility, and tourism.

Personal data sharing will also help ensure that the EU's values, principles, and regulations are fully implemented. As the EU digital single market is grounded in human-centricity, it aims to give back value to citizens by creating decentralized alternatives to the Big Tech platforms and their lock-in effect. By prioritizing the needs and interests of citizens, the EU can foster a more transparent and equitable digital landscape that benefits everyone.

For the tourism sector

For the tourism sector tourists will need to be able to share data across multiple organizations from different sub sectors like mobility, hospitality, and tourism activities, through a seamless user experience (tourist identity tools), while protecting their privacy (GDPR consent).

The challenges of personal data sharing

Human-centricity represents a paradigm shift in how we think about managing data and its potential. It stands in stark contrast to the prevalent "organization-centric" approach by placing the focus on the individuals involved in generating the data, rather than the organization responsible for capturing it, such as a company or government agency. Human-centricity encompasses concepts such as (self-)sovereignty, self-determination, self-governance, autonomy, and agency, which derive from the notion of human rights. At its core, a human-centric approach acknowledges that individuals have the right to determine, without coercion or compulsion, what happens to their personal data.

But human-centricity also comes with several challenges: technical issues (identity management, standardization, user experience, etc.), business issues (costs, IP strategy, etc.), legal issues (compliance with GDPR, DGA, etc.), as well as psychological factors like trust and digital resignation. Given this complexity, few existing data spaces today are actually processing personal data, despite personal data-sharing being a priority for most of them.

Addressing these challenges will require collaboration between stakeholders from a variety of levels and domains, including technical experts, legal professionals, business leaders, and psychologists. Only through such cross-disciplinary efforts can we hope to realize the full potential of a human-centric approach to data management. Given the coming widespread use of AI models in data spaces for purposes such as recommendation or personalization, it is crucial that individuals have effective means to maintain control over their personal data, which may be stored and processed by multiple organizations.

Personal data sharing and regulation

The Data Governance Act (DGA) introduces the concept of data intermediaries tools that will allow the sharing of data within a data space, and that need to be notified to competent authorities. A subset of the data intermediaries will enable the management of personal

data: the *personal data intermediaries (PDI)*. Beyond facilitating personal data-sharing, the personal data intermediaries will also provide data subjects with standard mechanisms to protect data subjects' privacy and rights (GDPR): right to be informed, right of access, right to rectification, right to erasure, right to restrict processing, right to data portability, right to object.

Tools for sharing personal data

The issue of personal data sharing is particularly complex due to specific concerns around privacy and GDPR compliance, and also given the fact that the individual is, in this case, the central point of data integration, not the organizations controlling data. The individual can potentially interact with data spaces of all locations and all sectors. Since he/she does not bear technical capabilities in itself the individual has 2 options:

- the individual authenticates to organizations controlling his/her data separately, and gives consent separately each time, which can be cumbersome.
- the individual uses dedicated tools, personal data intermediaries (PDI) for managing his/her identity and data, that aggregate data and simplify identity and consent management

There exist different approaches/paradigms to design PDI tools in Europe, here are some of the major ones:

1. **MyData operators** that are tools for GDPR consent management and personal data stores respecting the MyData declaration and label. MyData: MyData is a prominent movement advocating for human-centricity. Originally emerging from open data activism in Finland, it has since expanded into an international movement that is now run by "MyData Global," a non-profit organization. MyData provides guiding principles aimed at giving individuals greater control over the data trails they leave behind in their everyday activities. The goal is to enable individuals to see what happens with their personal data, specify who can use it, and modify those decisions over time. MyData is not solely focused on individuals' perspectives but also aims to serve commercial interests by promoting business opportunities for personal data. The MyData Principles strive to make privacy, data security, and data minimization standard practices in application design. The movement also seeks to empower individuals to understand privacy policies and to give, deny, or revoke their consent to share data based on a clear understanding of why, how, and for how long their data will be used. The Declaration of MyData Principles outlines ethical principles for personal data management and has been endorsed by over a thousand organizations and individuals worldwide.

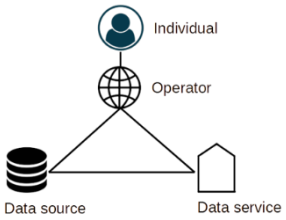

Figure 3 MyData operator concept

2. **SOLID Pods** that are tools following the SOLID specification/protocol and that let people store their data securely in decentralized data stores. The Social Linked Data project (SOLID) is a web decentralization initiative led by Sir Tim Berners-Lee, the inventor of the World Wide Web, and developed collaboratively in an open-source project consisting of multiple commercial and independent contributors. It aims at realizing Tim Berners-Lee's original vision for the Web as a medium for the secure and decentralized exchange of data. SOLID is at its core a specification/protocol that lets people store their data securely in decentralized data stores called pods (personal online data stores) (Buyle et al. 2020). Apart from a focus on separation between the application, identity provider and data storage as three interconnected entities, Solid has a strong focus on machine-readable linked data to ensure interoperability between different applications that reuse the same data source. The Linked Data architecture also helps to create (quite literally) links between one data set and another, including links between personal data and public data from various semantically related data spaces.

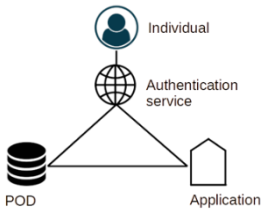

Figure 4 Solid POD concept

3. **POTENTIAL EU Wallets** allowing all European citizens to store and manage personal data in an ecosystem of standard wallets. POTENTIAL unites 148 participants from 19 EU member states from Northern, Western, Eastern, Central, and Southern Europe – representing more than 70% of the European population – and Ukraine. POTENTIAL's pilots drive European digitalization and ease numerous administrative as well as tedious identification processes in everyday situations. The aim is to vividly illustrate the possibilities, functionalities, and added value of a European Digital Identity Wallet. By involving relevant market players, POTENTIAL quickly scales solutions which build on existing market-relevant national solutions.

Within a paradigm PDI must be interoperable (the user can switch from one to another), discussions about cross-paradigms interoperability are still at a very early stage. The two main topics to address regarding interoperability for personal data in the data spaces will be about federating the individual's identity and consent. Some standardization discussions are happening now in organizations like Decentralized Identity Foundation, Kantara (consent receipt standards that inspired ISO), and MyData. At this stage, Gaia-X, IDSA, FIWARE and other related data space support organization, all explore IAA (Identity, authentication & authorization) capabilities, but they do not consider personal data apart and to not address the question in depth.

Focus on consent

Explicit consent is a cornerstone of personal data sharing, as it is the lawful basis (GDPR) of personal transfers between independent data controllers in a wide range of use-cases. When it comes to data sharing, GDPR requires that individuals are informed about the data

being shared, the purpose/finality for which it will be used, and the recipients of the data. Additionally, the individual must provide their explicit consent for their data to be shared for that specific purpose. Explicit consent means that the individual must take a clear and affirmative action, such as checking a box or signing a form, to indicate their consent. Consent must be freely given, meaning that individuals cannot be forced or coerced into giving their consent, and they must be informed of their right to withdraw their consent at any time. It's important to note that GDPR applies to any organization that collects or processes personal data of individuals residing in the EU, regardless of whether the organization is based in the EU or not. Failing to obtain explicit consent for data sharing can result in significant fines and other legal consequences. The major hurdle for consent management within the data spaces landscape now is the variety of consent formats and the lack of interoperability. In the tourism context, we can suppose that a same individual will share data across borders, which will be complex if consents are incompatible.

Complementary approaches

Some approaches help exploit data value without endangering the individual's privacy like anonymization, aggregation. More complex techniques that fall in the category of distributed processing, can drive the same results in a decentralize environment like a data space:

- **Federated learning:** type of machine learning that allows multiple devices or systems to train a model collaboratively without sharing the underlying data. The goal of federated learning is to enable machine learning models to be trained on large datasets while preserving data privacy.
- **Compute-to-Data (C2D):** lets algorithms run on a data owner's system without sharing sensitive data. Owners can control access and specify which algorithms can use their data.
- **Compute-to-Edge (C2E):** extension of C2D that uses Edge Computing to bring analytics closer to data sources, improving response times and reducing risks.

2.2.4 IDSA International Data Spaces Association

The vision of IDSA is to create the environment for trusted data exchange taking place through federated, international data spaces that are globally certified. IDS-certified products, services and systems open the door to a data economy in which businesses can share data up and down the value chain without security concerns. The aim is to create a global standard for Data Spaces, as well as fostering technologies and business models that will drive the data economy of the future in Europe and around the globe.

Two important publications of the IDSA are the IDSA Rulebook, focussing on governance of dataspaces. And the IDS-RAM V4[13] which goes into more detail on the architecture of data spaces. The Dataspace Protocol[14] is a set of specifications designed to facilitate interoperable data sharing between entities governed by usage control and based on Web technologies developed under the umbrella of IDSA. These specifications define the

[13] https://docs.internationaldataspaces.org/ids-ram-4/
[14] https://docs.internationaldataspaces.org/dataspace-protocol/overview/readme

schemas and protocols required for entities to publish data, negotiate usage agreements, and access data as part of a federation of technical systems termed a dataspace.

Certification is a major aspect of the IDS-RAM V4 to achieve Data Sovereignty. Interoperability, and compliance criteria are tested and validated based on an Open Source testing framework the IDS-Reference-Testbed[15]. The overarching view on a vibrant Data Space community as perceived by IDSA is part of the Data Space Landscape document[16]. Data Sovereignty is expressed in IDS based on human and machine readable Usage Policies, which can be enforced in technology or organizational manners. Therefore, the definition, description, specification and validation of Data Space Connectors, which implement policy negotiation and enforcement are the most important aspect of the IDSA work, including the interaction of the Data Space Connectors with other components and services in Data Spaces.

2.2.5 GAIA-X

Gaia-X aims to create a federated open data infrastructure based on European values regarding data and cloud sovereignty. The mission of Gaia-X is to design and implement a data sharing architecture that consists of common standards for data sharing, best practices, tools, and governance mechanisms.

From the technical point of view, Gaia-X aims to connect the Data and Infrastructure Ecosystems and relies on 3 conceptual pillars to achieve that:

- **Gaia-X Compliance**: Decentralized services to enable objective and measurable trust
- **Data Spaces / Federations**: Interoperable & portable (Cross-) Sector datasets and services
- **Data Exchange**: Anchored contract rules for access and data usage

2.2.6 DSBA: technical convergence document

The DSBA convergence document is a document published by the Data Spaces Business Alliance (DSBA), a coalition of four European associations that aim to promote data spaces across Europe and beyond.

The document defines a common reference technology framework for creating data spaces based on the technical convergence of existing architectures and models. The document

[15] https://github.com/International-Data-Spaces-Association/IDS-testbed
[16] https://internationaldataspaces.org/download/39041/?tmstv=1681683482

also discusses the key technology pillars for data spaces, such as data interoperability, data sovereignty and trust, and data value creation.

The aim of the convergence on dataspaces as a whole is to avoid dependencies while allowing the overlapping. If a given functionality or component is defined in more than one reference document, the documents will be updated to allow compatibility and leave it up to every initiative to choose which one to take with no impact on the overall dataspace's architecture.

Next figure shows an example of how the different initiatives could contribute to the reference technology framework.

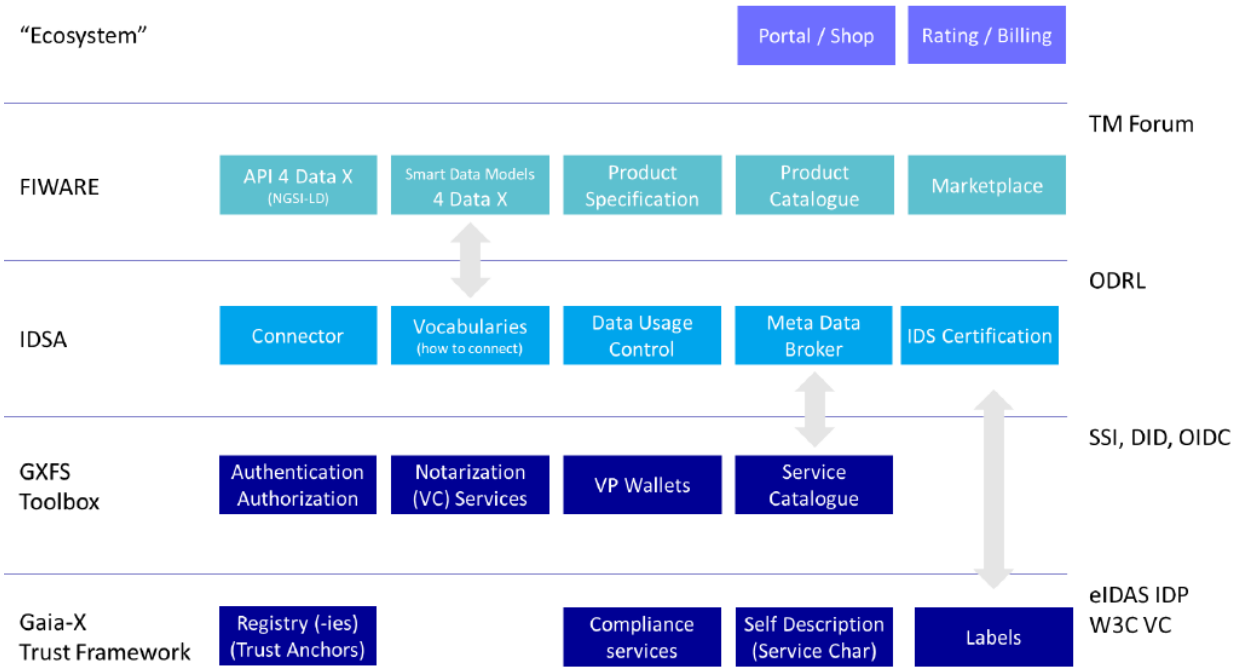

Figure 5 Technical convergence architecture example.

Recently, a second version of the convergence document has been produced.

2.2.7 Simpl

As an important element in implementing its Data Strategy, the European Commission announced its intention to develop a Smart middleware platform (so called "Simpl"). The information in this section has been taken from the tender documentation.

Simpl is a project financed by the Digital Europe program of the European Commission, endowed with 150 million euros and with an execution period of three years.

Its objective is to make software (middleware) available to society for the construction of data ecosystems and infrastructure services in the cloud that support the European values of data sovereignty, privacy and a fair market. Such a middleware would enable cloud-to-edge federations and support major data initiatives funded by the European Commission, such as common European data spaces.

Under the name of **Simpl**, the tender specifications refer to three distinct products:

- **Simpl-Open**: the core product of Simpl. It is an open-source software stack over which Simpl-Labs and Simpl-Live are built.
- **Simpl-Labs**: the provision and management of a pre-installed demonstration/playground environment where third parties (typically sectoral data spaces in their early phases of deployment) can experiment with/on the most up-to-date version of Simpl-Open before deploying it for their own needs.
- **Simpl-Live**: the provision of several instantiations of Simpl-Open in the form of customised production environments for sectoral data spaces where the European Commission itself plays an active role in their management.

The next figure shows the Simpl-Open Conceptual architecture showing the scope of the project.

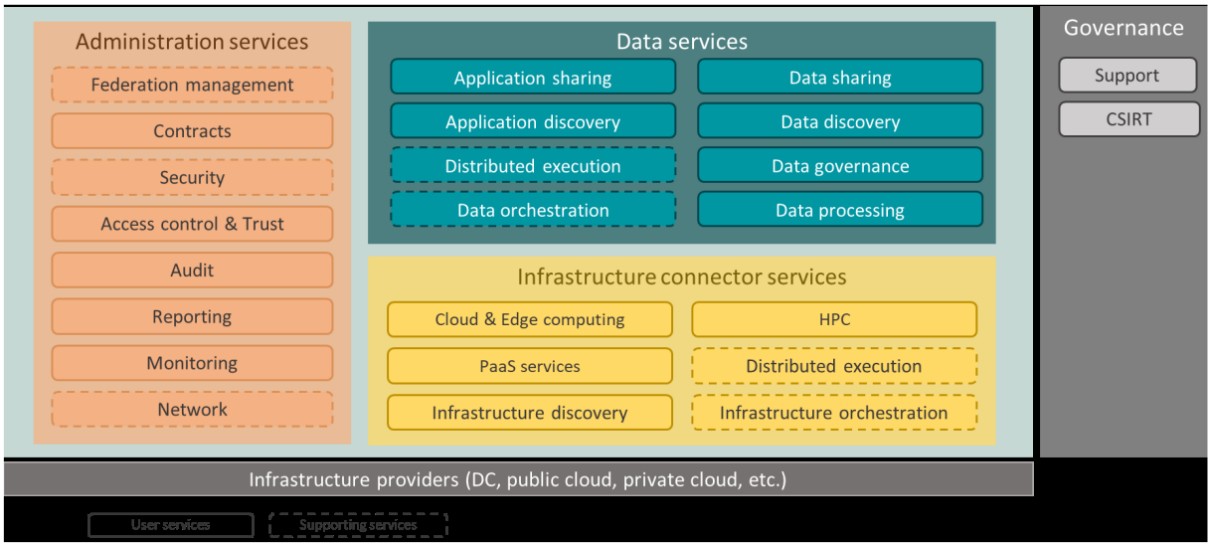

Figure 6 Simpl-Open Conceptual architecture

The architecture is based on four layers:

- **The administration services layer** is a support layer providing user and supporting administrative services that are used by the data and infrastructure service layers. Management and supervision services are included here to enable better control and interoperability among Simpl-Open's services. Issues such as security, personal data protection, access control, federation management, contract management, monitoring or reporting are integrated.
- **The data services layer** enables the exchange of data assets and applications by offering capabilities such as application sharing, data sharing, discovery (of data and applications) and data-relative functionalities such as data processing, data governance and data orchestration. More specifically, the data service layer also includes the implementation of a set of mechanisms to allow different modalities of data access (e.g., online streaming, batch download, ...), transfer (bulk transfer and

data streaming) and integration capabilities to merge data coming from various sources or data spaces.

- **The infrastructure connector services layer** manages the infrastructure–related assets and services, allowing to also connect a variety of heterogeneous third-party infrastructural services. The services provided herein allow the finding and provisioning of computing and storage resources so that application workloads can be efficiently and securely managed end-to-end. This layer presents capabilities such as the discovery of infrastructural elements from the catalogue, infrastructure orchestration, deployment and execution of applications, infrastructural services (such as containers and Virtual Machines), and PaaS-related services (e.g., Databases, AI capabilities among others).
- **Governance layer** is a transversal layer supporting the layers above, which focuses on the provisioning of security contingency and personal data protection measures and support capabilities.

The tender also includes an analysis of the relation of Simpl with other initiatives like the EU Digital Building blocks and GAIA-X.

Regarding Gaia-X, the analysis concludes that Gaia-X developments, by means of its Federation Services, are today mainly addressing Simpl-Open Administration Services Access control and trust features.

Beyond these capabilities, other Gaia-X developments such as Self-descriptions and Federated Catalogue could integrate the implementation of the Simpl-Open Infrastructure discovery building blocks along with certain aspects of contracts management, monitoring and infrastructure orchestration.

Overall, the analysis shows that Simpl-Open has a broader perimeter than the current Gaia-X developments. However, existing results have the potential to contribute to Simpl-Open implementation. Gaia-X developments and specifically its Federation Services can act as a source for open-source components which could be integrated, re-utilised and extended as part of Simpl-Open implementation.

2.3 Other Common principles / building blocks: eID (digital identity), eDelivery (secure data exchange), FIWARE's Context Broker

2.3.1 Digital Europe Building Blocks

In the course of the EU Digital Europe Programme (DIGITAL) (2021-2027) the European commission continues to provide tools and support for the digitalisation of all sectors[17].

[17] https://ec.europa.eu/digital-building-blocks/wikis/display/DIGITAL/About+us

Their implementation is especially designed to foster the creation of data exchange platforms responding the eIDAS regulation[18].

Among the modules published by Digital Europe Program[19], there are two building blocks of interest, implemented during previous periods of the program and established for more than 10 years, namely those for identity management (eID) and data delivery (eDelivery), respectively.
The main focus of the so-called Connected Europe Facilities (CEF), however, is on the digitalization of the relationships between citizens and administrations. B2B use cases are not yet fully supported. For example, marketplace functions, like those specified in the Data Space models shown above, are not yet available in the centralized metadata handling component of the eDelivery building block. Moreover, in the majority of the EU countries national electronic identification schemes (eIDs) are available only for authorities and human persons, but not for private legal entities, like enterprises.

2.3.1.1 eID - Identifying users from all across Europe.

[information shown in this subchapter is mainly taken from the DIGITAL homepage]

The eID building block can be seen as an eIDAS enabler, which provides access to free tools and support to build digital services in line with the eIDAS regulation. In detail, eID is a set of services provided by the European Commission to enable the mutual recognition of national electronic identification schemes (eID) across borders. It allows its users to authenticate themselves by using their national eIDs and connecting with their Identity Provider (IdP) from their country, when accessing online services from other European countries.

However, the eID tools cannot be used independently from the national eIDAS structures. For example, an identity provider is required to contact its Member State representative for the notification of its eID so that it can be used over the eIDAS network. And a service provider needs advice from its Member State representative to connect to the respective national eIDAS node.

Integration of eID with GAIA-X: A Verifiable Credential is *Gaia-X conformant* if the issuer of the Verifiable Credential itself has an identity coming from one of the Trust Anchors. One of the GAIA-X trust anchors is eiDAS.

eiDAS	Issuers of Qualified Certificate for Electronic Signature as defined in eIDAS Regulation (EU) No 910/2014 (homepage: https://esignature.ec.europa.eu/efda/tl-browser/#/screen/home) (machine: https://ec.europa.eu/tools/lotl/eu-lotl.xml)

[18] Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG
[19] Digital Homepage (europa.eu)

2.3.1.2 eDelivery - Exchange documents and data securely and reliably

[information shown in this subchapter is mainly taken from the DIGITAL homepage]

eDelivery is a building block providing technical specifications and standards, installable software and ancillary services to allow projects to create a network of nodes for secure digital data exchange[20]. By building with eDelivery, public and private organizations from different sectors can easily create a safe and interoperable channel to transfer documents and data among each other over a public or private network. The eDelivery "Access Point" can be implemented as an extension of a Data Space Connector, which may also connect to a federated middleware providing marketplace, contracting, and clearance features in the sense of the GAIA-X/IDSA/Open DEI concepts (see Figure XX).

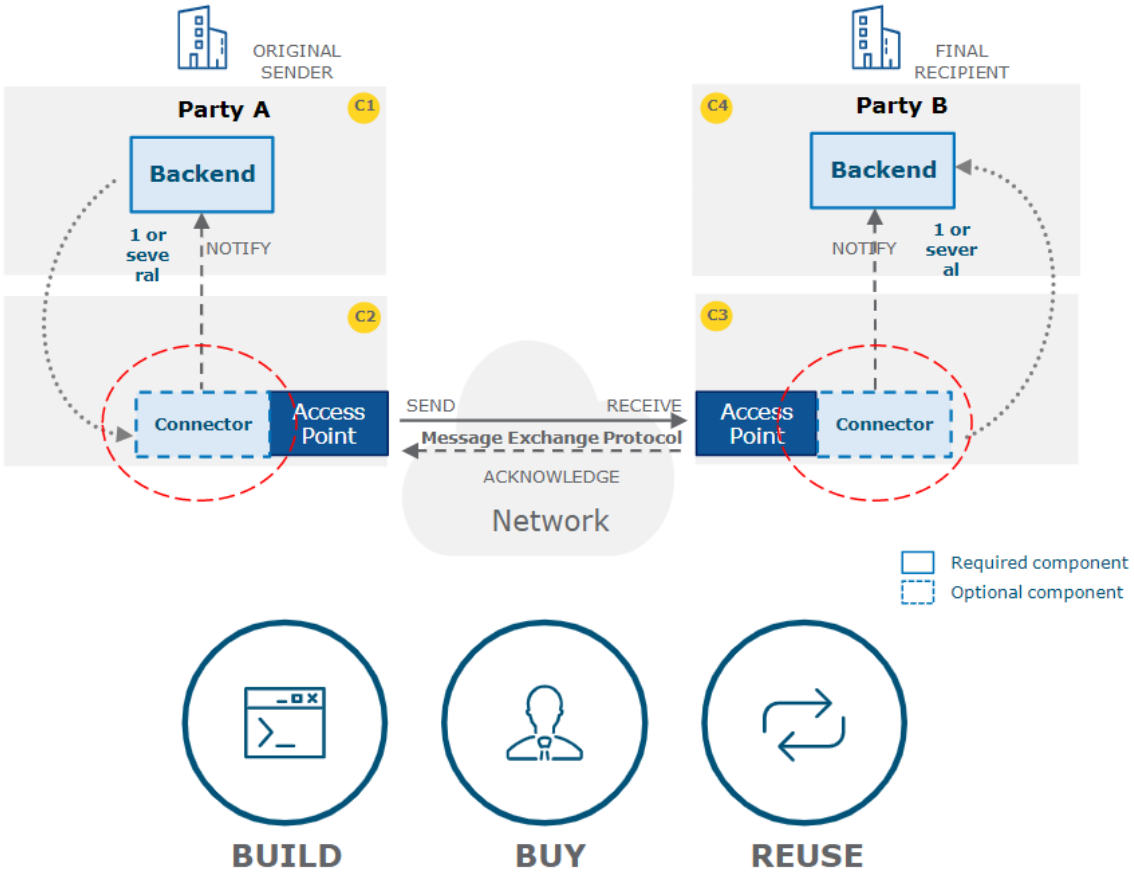

Figure 7 Integration approach of eDelivery within user's environments

As indicated in the figure above, eDelivery is designed to support the exchange of structured or unstructured information encapsulated in messages. Therefore, limitations have to be taken into account for use cases where a continuous data flow over long periods is planned.

[20] https://ec.europa.eu/digital-building-blocks/wikis/display/DIGITAL/eDelivery

2.3.2 FIWARE Building Blocks

The Fiware Foundation has developed a "curated framework of Open-Source Platform components to accelerate the development of Smart Solutions"[21]. With the Fiware Context Broker as its central and mandatory component, Fiware provides many building blocks for a complete data sharing ecosystem.

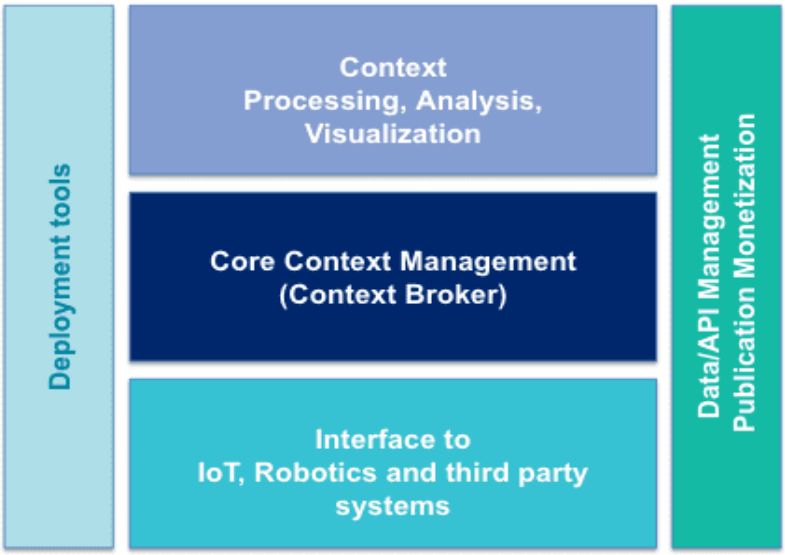

Figure 8 Fiware component schema

Since in this context it is not possible to go too much into the details of their proposal, the following subsections should point out the most remarkable elements with respect to reuse and integration into future data spaces.

2.3.2.1 FIWARE Context Broker

The FIWARE Context broker ("Orion") is a C++ implementation of the NGSIv2 REST API binding developed as the central part of the FIWARE platform[22]. Orion Context Broker allows a Data Space to manage the entire lifecycle of its context information including updates, queries, registrations and subscriptions. With the FIWARE Context Broker in centre, a rich suite of complementary open source FIWARE Generic Enablers has been made available, dealing with the following features:

- **Interfacing with the Internet of Things (IoT), Robots and third-party systems**, for capturing updates on context information and translating required actuations.
- **Context Data/API management, publication, and monetization**, bringing support to usage control and the opportunity to publish and monetize part of managed context data.

[21] https://www.fiware.org/about-us/
[22] https://github.com/telefonicaid/fiware-orion/

- **Processing, analysis, and visualization of context information** implementing the expected smart behaviour of applications and/or assisting end users in making smart decisions.

The FIWARE components are modules that can be combined with and integrated to various third party elements, to build a hybrid data-exchange platform at the user's convenience.

2.3.2.2 Smart Open Data Models

Data Models play a crucial role because they define the harmonized representation formats and semantics that will be used by applications both to consume and to publish data.

The Smart Data Models board (FIWARE Foundation, TM Forum, OASC and IUDX) are leading a joint collaboration program to support the adoption of a reference architecture and compatible common data models that underpin a digital market of interoperable and replicable smart solutions in multiple sectors. Indeed, a smart data model for tourism is available from the FIWARE GitHub repository[23]. A smart data model includes three elements: The schema, or technical representation of the model defining the technical data types and structure, the specification of a written document for human readers, and the examples of the payloads for NGSIv2 and NGSI-LD versions.

2.3.2.3 Business Application Ecosystem (BAE)

FIWARE proposes a Business Application Ecosystem (BAE). This is a marketplace component that is made up of the combination of the FIWARE Business Framework and a set of APIs provided by the TMForum. It supports the monetization of different kinds of assets during the whole service life cycle, from offering creation to its charging, accounting and revenue settlement required for billing and payment to involved participants.

See the following list of backend components and APIs associated to the FIWARE BAE Marketplace[24]:

- Backend implementing standard TM Forum APIs supporting configuration of the marketplace:
  - Catalog Management API
  - Product Ordering Management API
  - Product Inventory Management API
  - Party Management API
  - Customer Management API
  - Billing Management API

[23] https://github.com/smart-data-models/SmartDestination
[24] https://www.fiware.org/wp-content/uploads/FF_PositionPaper_FIWARE4DataSpaces.pdf

- Usage Management API

- Rating, Charging, and Billing backend;
- Revenue Settlement and Sharing System;
- Authentication, API Orchestrator, and Web portal.

2.3.2.4 FIWARE CKAN monetization architecture

FIWARE has defined and implemented an open-source architecture for data products the is able to include open and private data and static data sets along with "real time" data via APIs.

CKAN is used as the catalogue where data can be published, discovered, managed, and consumed. Within the FIWARE platform, not only static datasets (CSV, XLSX, etc) are supported, but also real (right) time context information can be published in CKAN as Context Broker queries.

In addition, using the Biz Ecosystem GE, FIWARE provides the components that allow to manage the access and monetisation of published datasets.

The following picture shows the architecture followed in FIWARE for the publication and monetization of data.

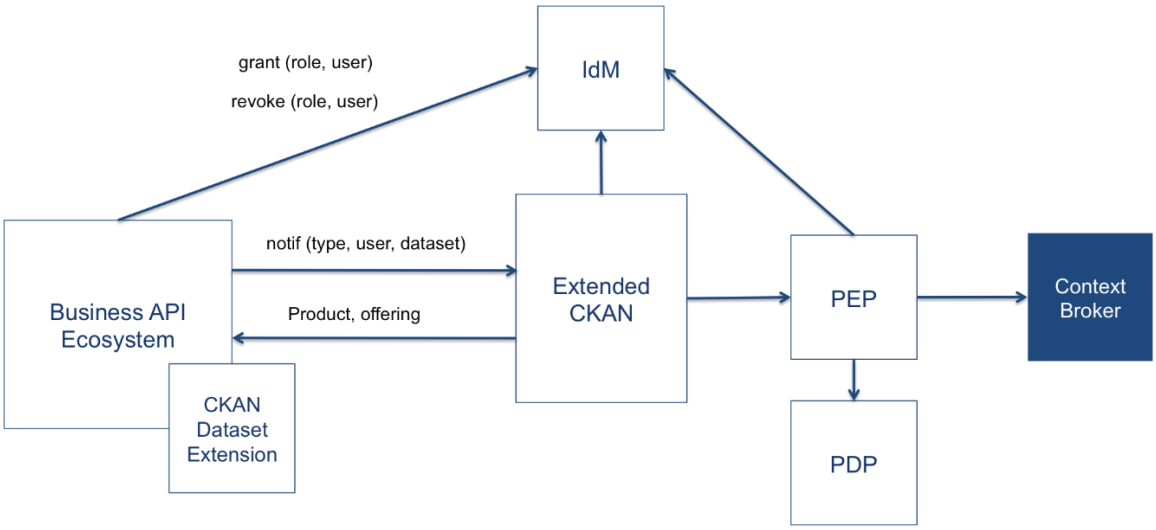

Figure 9 FIWARE Monetization architecture

The following components are integrated in order to manage and monetize data:

- **Extended CKAN**: Used for the publication of the static datasets and the Context Broker queries.
- **Biz Ecosystem**: Used for the creation and acquisition of data offerings, including customer charging, and access grant

- **IdM + PEP + PDP** (Security Framework): Used for the provision of identities, authentication, and authorization of users.

2.3.2.5 FIWARE Smart city architecture

The Smart City data space shares most of the tourism data space features. People are at the centre of the data space, a strong relationship with other data spaces and a mixture of public administration and private participants with different and sometimes conflictive goals.

FIWARE has defined the so-called Smart city architecture that provides a good overview of the smart city scope, trying to tackle some of the main problems found when deploying a smart city data space.

The Reference Architecture for Smart Cities **breaks vertical silos**, building a Context Info Management layer that provides a holistic picture of what is going on in the city. By making city data public and **merging data from multiple verticals**, city-level governance systems can be enhanced.

Due to **"de-facto" standard information models** there are no costs of adaptation to achieve full interoperability among many different systems in the city. This, in turn, enables portability of systems across sectors and cities.

Third party solution providers can benefit from right-time open data published by the city and made available through standard APIs. **They can sell their solutions to cities** across the world, targeting a larger market and boosting businesses.

Once the data gathered from all sources are made public and exchangeable, city administrators can offer higher value services to citizens, as well as improving public service delivery. This creates an innovative scenario which cities, private business, and citizens can benefit from.

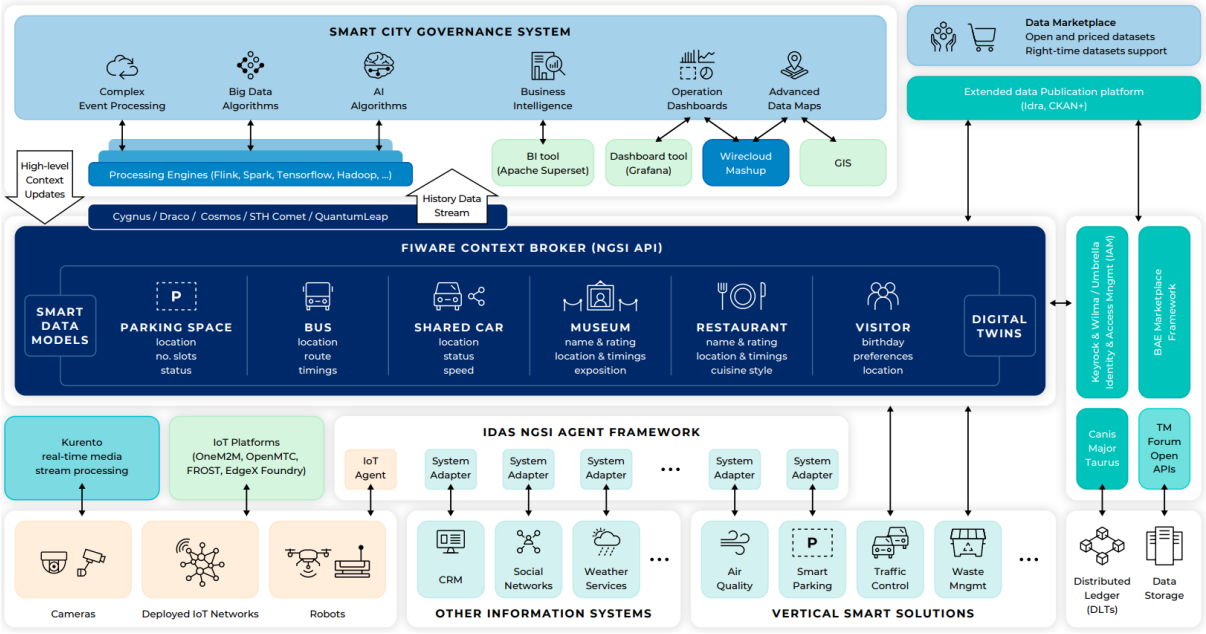

Figure 10 FIWARE Smart City architecture

3 IDENTIFICATION OF THE REQUIREMENTS OF THE TOURISM SECTOR

According to the comprehensive definition of tourism presented by Judd (2006), quoting Debbage and Danials (1998): "tourism is no single product rather a wide range of products and services that interact to provide an opportunity to fulfil a tourist experience that comprises both tangible parts (e.g., hotel, restaurant, or air carrier) and intangible parts (e.g., sunset, scenery, mood)" (Judd, 2006, p. 325).

The tourism industry is a powerful engine of the national economy and employment. However, it is affected by the **high fragmentation** of the sector, as well as by a very **heterogeneous degree of digitization**. In addition, as a transversal industry, it is **directly related to other sectors** such as mobility, health, culture, historical heritage and the environment, among others.

For all these reasons, data spaces are optimal instruments for promoting federated business networks that serve to break down information silos and provide transparency and resilience to risks associated with disruptions in sectoral value chains. In addition, these networks are dynamic and evolving, making it possible to capitalize on the progressive digital transformation of tourism organizations, reusing investments in information systems and potentially opening new lines of business.

3.1 Main Stakeholders and value chain

Tourism is a global activity, both in its territorial and human dimensions. From one perspective, tourism can be considered as a meta-sector in which most of the traditional sectors (such as transport, mobility, energy, construction, health…) are directly or indirectly involved.

According to UNWTO[25]:

> *the **tourism value chain** is the sequence of primary and support activities which are strategically fundamental for the performance of the tourism sector. Linked processes such as policy making and integrated planning, product development and packaging, promotion and marketing, distribution and sales and destination operations and services are the key primary activities of the tourism value chain.*
>
> *Support activities involve transport and infrastructure, human resource development, technology and systems development and other complementary goods and services which may not be related to core tourism businesses but have a high impact on the value of tourism.*

In this way, the sector includes a great variety of stakeholders. In a first classification, which covers the global dimension of the sector, the following actors can be distinguished:

- **Public authorities** that manage public policies related to the territory and tourism activity.
- **Private companies** that provide tourism services and products.
- **Tourists/Users** who consume these products and services in a territory.

A specific role in this value chain is played by the Destination Management Organisation.

A **Destination Management Organization (DMO)**[26] is the leading organizational entity which may encompass the various authorities, stakeholders, professionals, and local communities. The governance structures of DMOs vary from a single public authority to a public/private partnership model. Their main purposes are implementation of tourism policies, strategic planning, product development, promotion and marketing and convention bureau activities. The functions of the DMOs may vary from national to regional and local levels depending on the current and potential needs as well as on the decentralization level of public administration.

Other main stakeholders in the tourism value chain can be identified by the role they play in these stages of tourism activity:

- **Planning and booking**: This stage involves searching and planning a trip, and making reservations for transportation, accommodation, and activities. Traditional

[25] https://www.unwto.org/glossary-tourism-terms
[26] https://www.unwto.org/glossary-tourism-terms

Travel Agencies, Online Travel Agencies (OTAs), Tour Operators can be included in this stage.

- **Transportation**: This stage involves the actual transportation of tourists to and from their destinations, whether by air, land, or sea.
- **Accommodation**: This stage involves the provision of accommodation to tourists, including hotels, resorts, hostels, and vacation rentals.
- **Activities and experiences**: This stage involves the provision of various activities and experiences to tourists, such as tours, sightseeing, cultural experiences, and recreational activities.
- **Food and beverage**: This stage involves the provision of food and beverages to tourists, including restaurants, cafes, and bars.
- **Shopping**: This stage involves the provision of shopping opportunities to tourists, such as souvenir shops and local markets.

Overall, the tourism value chain involves a complex network of activities and stakeholders, all working together to create and deliver tourism products and services to consumers.

3.2 Data types

Multiple classifications of tourism data types are widely accepted and commonly used in tourism research and analysis. In fact, different classifications can be used simultaneously to gain a more comprehensive understanding of the tourism industry and its various components.

For example, *D2.2 Analysis of Gaps and Overlaps* has provided a classification of data types in relation to the tourism sub-sector from which they originate:

- **Transport data**: data on the means used by tourists to reach their destination, including travel by road, rail, air and sea.
- **Food and beverage data**: data on food consumed at all stages of the journey, including travel and accommodation.
- **Car and other rental data**: data about rental services, including cars, motorbikes, bikes, boats, motorhomes, quads, etc.
- **Travel agency, tour operator and related activities data**: data collected by tourist agencies and related travel services.
- **Accommodation data**: data on where tourists stay and sleep.

This classification can be aligned with the findings of the EU project "Smart Tourism Destinations", as reported in [1]. This study summarizes that four main types of tourism data can be identified according to their provenance, as shown in next figure.

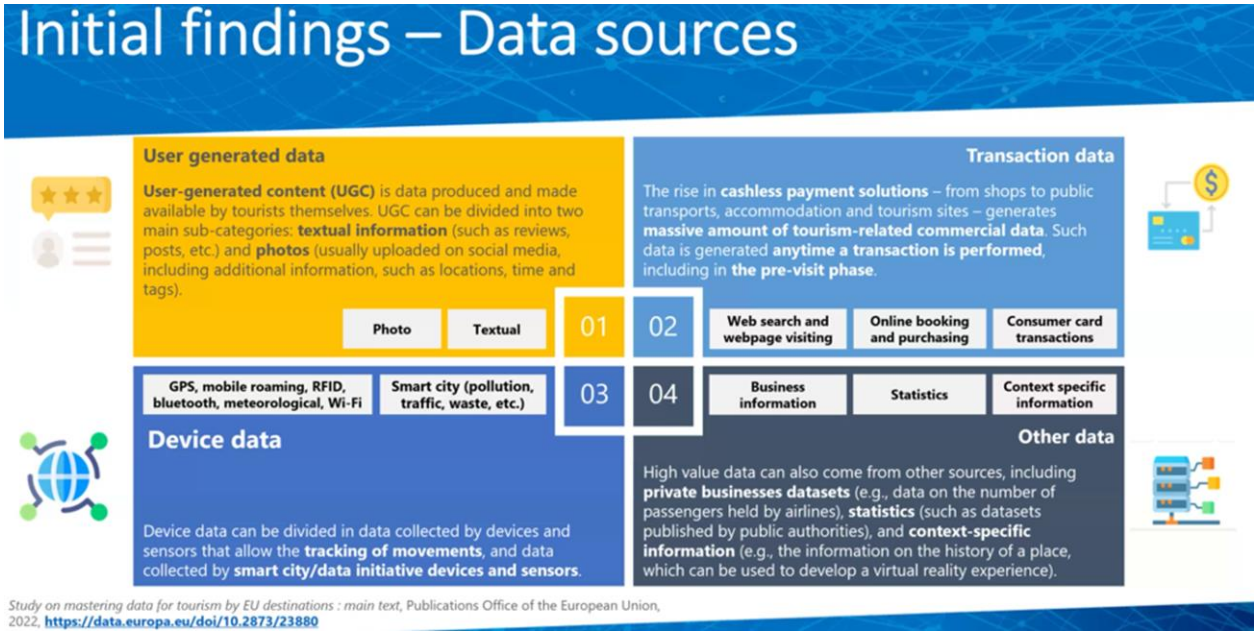

Figure 11 Study on mastering data for tourism by EU destinations

- **User generated data** can be divided into two subcategories: textual information (such as reviews, posts, surveys, etc.) and photos and videos (uploaded to social networks).
- **Transaction data** include data generated from online bookings and purchases, from internet searches and webpage visits, and from customer card transactions.
- **Device data** are generated by devices and sensors that allow the tracking of movements, or by smart city sensors that collect measurements about pollution, traffic, waste, etc.
- The **other data** category includes valuable data such as information from private companies (e.g. restaurants, hotels, airlines, ferias, etc.), statistics (e.g. hotel occupancy rate) and contextual information about the destination.

In addition to this, a different typological perspective can be obtained from [3]:

- **Traveler profile data**:
  - *Typology*: individual, couple, groups of friends, business, etc.
  - *Demographic classification*: demographic characteristics of tourists, such as age, gender, income level, education level or nationality.
  - *Trip purpose classification*: leisure / business / adventure / educational / cultural /medical tourism.
  - *Origin*: national/international, short/long distance.
  - *Time related data*: peak/off-peak season, day of the week, special season (events and specific holidays).

- *Tourism product data*: accommodation (hotels, resorts, vacation rentals), transportation (air, rail, road, water), attractions (museums, theme parks, historic sites), events (festivals, conferences, concerts), etc.
- *Transport*: air, land, sea, combined, urban transport service.

- **Traceability data:**
  - *Pre-, during-, post-trip data*: tourism data can be collected and analyzed at different stages of a traveler's journey, including before, during, and after the trip.
  - *Geographic data*: specific regions or destinations visited by tourists.
  - *Data from private/public organisations*: sales data (hotel bookings, tour packages, transportation services, tickets at tourist attractions), visitor expenditure.
  - *Tourist mobility*: data coming from IoT devices (GPS, RFID, NFC, cell phone triangulation, etc.)
- **Reputation data**:
  - *Online reviews*: tourist feedback, tourist satisfaction, complaints and suggestions.
  - *Social monitoring*: data on mentions on social media channels.
  - *Market research*: tourist satisfaction surveys.
  - *Media coverage*: data from media monitoring.
- **Marketing data**:
  - *Campaign data*: opportunity areas, target audience, number of clicks on campaign ads, demographic data, etc.
  - *Sales channels data*: success of promotions, customer feedback data, social media data, etc.
- **Public data:**
  - *Hotel data*: overnight stays, average stay, occupancy rate per place, average rate per room, hotel price index.
  - *Tourist movements at borders*: tourists according to country of residence, according to main destination region, according to main reason for travel, visitors according to type.
  - *Resident tourism survey*: number of trips, average duration of the trip, total/average/daily costs.

- *Forecast indices*: arrivals, reservations, spending, employment, tourism GDP.

- **Sustainability data**:
  - *Energy consumption data*: identification of energy consumption, usage of renewable energy resources.
  - *Carbon emissions data*: tracking of carbon emissions.
  - *Water usage*: measure water usage.
  - *Waste data*: measure the amount and types of waste, recycling efforts.
  - *Supply chain data*: data on the environmental and social impact of an organization's suppliers.
  - *Social impact data*: data on the social impact of an organization's operations, such as employee satisfaction, community engagement, and diversity and inclusion metrics.

3.3 Data purposes

Tourism data provides information on the travel and tourism industry, international tourism and domestic tourism. These data include information broken down to transaction level, such as tourism expenditure on flights, car rentals, hotel stays, restaurant visits or visits to attractions.

In a workshop conducted in WP2 in February 2023, whose results were reported in *D2.2 Analysis of Gaps and Overlaps,* some of the main uses of tourism data where analysed.

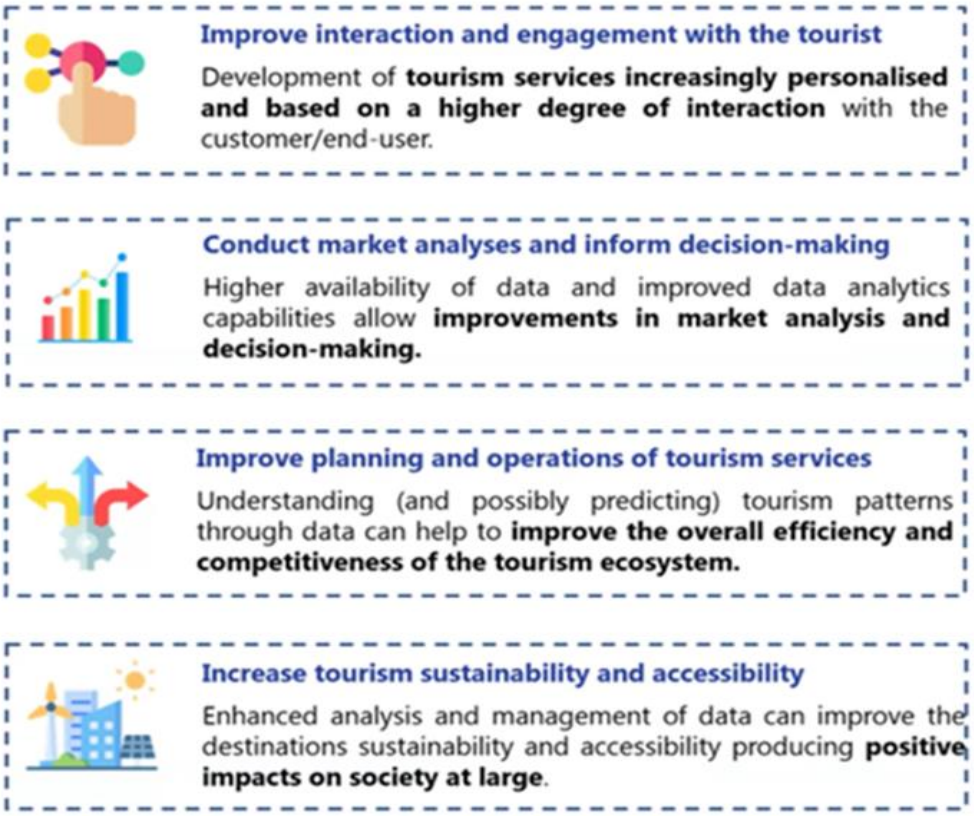

Figure 12 Main uses of tourism data

From the analysis of [2] and [3], we extract the following information on tourism data purposes:

- **Traveler profiling and personalisation of the tourism experience**: travelers can be classified in many different ways, e.g. luxury or budget, couple or kids, regional or international, business or leisure, etc. Profiling travelers can help travel companies provide better, more personalized experiences that meet the needs and preferences of their customers during their stay, increasing the propensity to tourism consumption, while also improving customer satisfaction and loyalty.
- **Analysis of tourist behavior and improvement of data traceability to optimise tourism strategy**: by collecting and tracking tourist data (e.g. before/during/after trip, flow in destiny), tourism businesses can gain valuable information about the behavior and preferences of their customers. The analysis of tourist behavior can provide insights into the factors that influence travel decisions and conducts. By understanding how tourists make decisions and behave, tourism businesses can tailor their offerings to better meet the needs and preferences of their customers. This can lead to increased customer satisfaction, profitability, and long-term success in the tourism industry.
- **Generation of public data**: by creating public repositories that make tourism data available to a wide range of stakeholders, public repositories can help promote better tourism practices. These data are useful not only for the tourism industry, but

also for public authorities to make informed decisions on tourism policy, infrastructure development, marketing strategies and sustainability initiatives. Public data repositories are useful tools that help to investigate the economic impact of tourism in a destination.

- **Creation and monitoring of marketing campaigns to improve their efficiency**: tourism data is useful for creating promotional campaigns and recommending destinations, e.g. proximity data enhances proximity marketing to engage customers in real-time. By analyzing these campaigns, tourism businesses can identify the most effective marketing strategies and tactics to attract and retain customers. In addition, monitoring these campaigns helps to measure their effectiveness, or the impact of specific events, allowing adjustments to be made to improve efficiency according to needs.
- **Management of reputation and tourist satisfaction**: this is critical in the tourism industry, as it directly affects a business's success and long-term viability. A positive reputation and high levels of customer satisfaction can give a business a competitive advantage in a crowded marketplace. Positive reviews and high customer satisfaction ratings can lead to repeat business and referrals, as satisfied customers are more likely to return and recommend a business to others. Satisfied customers are more likely to develop brand loyalty. Effective reputation management can help mitigate the impact of negative reviews and crises, ensuring that the business's reputation remains intact.

In addition to these aspects, [4] also highlights the following aspects in which tourism data are useful:

- **Promoting the sustainable development of a destination**: by using data to understand visitor preferences and behavior, tools can be developed to provide access to environmental performance metrics that enable tourism companies to track and analyze their resource consumption, waste generation, and carbon emissions. Thus, destination managers can develop sustainable tourism products and services (e.g. to reduce waste and water usage), identify sustainable tourism opportunities, monitor tourist impact, legislate sustainability measures, and evaluate tourism policies and strategies.
- **Improving the quality of stay of visitors and the quality of life of residents**: the analysis of tourism data can help in the identification of common problems, and in the definition of possible solutions to these problems (e.g. planning city services accordingly before special events).
- **Contribute to the simplification of pre-trip organisation processes, while improving the inspiration phase**: this can be achieved by providing personalized recommendations, real-time pricing and availability, destination inspiration, and social media and user-generated content. By leveraging tourism data, travel

companies can create a seamless and engaging experience for travelers that meets their needs and exceeds their expectations.

- **Maintaining a post-travel relationship**: a post-travel relationship between the destination brand and the tourist, helps to encourage new visits based on personalised recommendations online with customer motivations.

Moreover, *D4.1 Roles and Dynamics of the Tourism industry in the EU* highlights that **establishing common semantic data standards and structures for tourism data** is a crucial objective for efficient data exchange and interoperability among all stakeholders, that will help to put the tourism sector in Europe in a leading position to address the major environmental, economic and societal challenges.

From the workshop "*Towards a data space for tourism - Use Case Co-Creation*" conducted by Intellera Consulting in March 2023 under the framework of the DATES project, we can add the following benefits of tourism data management, in addition those already identified:

- **Development and diversification of tourism products and services aimed at ensuring equal access to tourists:** provide access to data on the accessibility of a destination from different perspectives, e.g., food possibilities of restaurants and hotels (celiac, vegan, allergies), physical accessibility (availability of wheelchair ramps, elevators, accessible parking spots), visual accessibility (availability of Braille signage, audio descriptions), etc.

- **Efficient management of tourist flows to avoid "over tourism", overwhelmed tourist destinations and improve the wellbeing of local residents:** development of an observatory or dashboard to monitor the flow and profile of visitors across districts, in order to inform in real-time Destination Management Organizations and public authorities in decision making, with the aim of limiting the negative impacts of tourism and the risks of overcrowded places (e.g. impoverishment of natural sides, increased littering, residents and tourists dissatisfaction). The observatory should also provide a recommender system for tourists, to provide potential visitors with relevant information on saturation levels and suitable trip alternatives to enhance the visitor experience, as well as the quality of life of tourism regions.

- **Mitigate crisis management**: the tourism sector is highly volatile due to seasonality and other determinants. Seasonality has been extensively researched, but calamitous events and their impact on the tourism sector are still underestimated. As seen during the Covid-19 pandemic, the tourism industry can be suddenly disrupted by dramatic events, such as natural disasters, terrorist attacks, accidents, political instability and infectious diseases. Therefore, having real-time information can help to make accurate and appropriate decisions for immediate action. Integrating data relevant to crisis management with data on tourism flows can enable public authorities to plan preventive actions, interventions and also

communication. In addition, sharing this data with tour operators would help to provide real-time information on possible risks to tourists' safety.

We can conclude that tourism data can help all players in the sector to better understand their market and refine their strategies to achieve more optimal results. This information is also useful to improve the understanding of tourists' needs, to provide them with a more personalised offer and to improve their engagement. Besides, reputational information management is a crucial element that can help companies to differentiate from competitors.

It is also important to highlight that not only the information generated during the trip is useful. Pre- and post-trip information must be considered too, to improve inspiration, organisation, and even encourage return.

In addition, we cannot forget the environmental impact of tourism. A study published by Nature Climate Change in 2018[27] suggested that tourism accounts for 8% of global greenhouse gas emissions, a very high percentage if we take into account all possible sources of emissions. Tourism data should help us move towards a more sustainable tourism model that benefits both visitors and the local community.

3.4 Building blocks required to handle data categories

This chapter summarises considerations regarding the exchange of special data categories, according to the class "Other data" as depicted in the **¡Error! No se encuentra el origen de la referencia.** "Study on mastering data for tourism by EU destinations" in chapter 3.2.

These "other" data might comprise private business data offered by commercially active entities, as well as compiled data like statistics, and, for example, data which could be used to enrich the core payload data of various touristic use cases (e.g., texts, images, videos, audio files about destinations). It is not expected that the exchange of these types of data between data providers and data consumers raises remarkable technical challenges. Any standard data exchange API should be able to deal with them.

However, data providers and data consumers are supposed to have other specific requirements, which they might want to see met in a data space. Since these requirements do not necessarily depend on the data categories themselves, but predominantly on the role of data processing in the sharing process, the following sections summarise the outcome of the discussions separately for data providers, and data consumers, respectively.

3.4.1 Data providers concern

Data providers will have the need to enforce their Intellectual Property rights, at least in the case of their private data, and with respect to shared media. This requires the implementation of appropriate Metadata Models, and self-description templates,

[27] https://www.nature.com/articles/s41558-018-0141-x

respectively, enabling them to clearly express their rights and the resulting consequences for the data user.

Consequently, commercial data providers might wish to restrict the usage of the data to specific purposes or timeframes, and may limit the sharing to selected recipients (e.g., prevent access of competitors, claim for specific behaviour of the data consumer, or prevent sharing of PII data with recipients in unsecure countries). Therefore, the (smart) contracting building blocks of a given data space should enable the handling of rich data-descriptions as mentioned above, which must be able to take into account a large set of attributes defined by a data provider.

Obviously, to protect their various rights data providers will need appropriately designed building blocks implementing usage policies expression, together with substantial usage control capabilities. The state of the art to describe usage policies is set by the Open Digital Rights Language (ODRL), published by the W3C consortium. Policy enforcement can be supported on data providers side, by designing the Data Exchange APIs in a way that implements already certain restrictions, e.g., by allowing access to data only via predefined queries. In addition, data providers might wish to prevent later modification of the data (e.g., in the case of statistics). This requirement may lead to the use of building blocks which support secure data export formats, and trusted data exchange, as well as change tracking and data provenance documentation.

Finally, companies might wish to enable confidential computing, resulting in the need for features following the building block Trusted Data Exchange, e.g. providing means for anonymization and encryption, respectively.

Besides the interest in enforcing their various rights, data providers will also have an interest to implement building blocks which enhance the quality of their data, facilitating the easy and compliant usage of the data, and thus facilitating collaboration between data space participants, and also monetization of the data assets. Data quality concern is met by appropriate internal QA policies, which should define criteria for completeness, richness, compliance etc of the data offered to others. Data providers should seriously take into account implementing automated QA checks, and Privacy Enabling Technologies (PET) on their side.

3.4.2 Data consumers concern

Similar, but somehow complementary and mirror-inverted, are the needs of the data consumers. Whereas the data providers should use rich and standardised vocabularies/ontologies to comprehensively express the conditions under which they will allow data sharing, the data consumers must use the same vocabularies/ontologies to exactly know which data they can expect, and of course which usage restrictions they might encounter.

Therefore, the importance of the OpenDEI Building Blocks “Data Models & Formats” and “Metadata & Discovery Protocol” cannot be underestimated. Using an agreed semantic

base, wherever information is exchanged, is one of the most critical pillars of intra-data-space interoperability, and moreover of connection to other related data spaces, too.

With respect to the "other" data in the scope of this chapter, data consumers may wish to be informed about some additional features of the data, besides the obvious information about permissions and restrictions, commonly imposed by the data provider (e.g purpose, commercial/non-commercial, lifetime, cost, etc). In the case of statistics, they will probably need information about the parent population (size, geographical background, demographic background, business background, age (of the data), ···), and also about possible selection criteria, eventually supported by a Data Exchange API.

In the case of offered media, information about usage permissions/restrictions (e.g. according to the Creative Commons Schema), or clearly described licensing/pricing is required. Again, the metadata building block must be able to support such an information exchange with appropriate standards. Media usage may be in addition facilitated by providing information about data formats (video/audio formats, image formats), and download conditions, for example to prepare sufficient data import bandwidth. Finally, in all cases where PII data is exchanged, data consumers need to receive information about the legal ground of PII data disclosure to avoid committing data breaches, and in any case to be aware that PET is required also on the data consumer side.

3.5 Tourism Digital Hub by the Italian Ministry of Tourism: From data sharing to data spaces

As part of the National Recovery and Resilience Plan (PNRR), the Ministry of Tourism has launched the **Tourism Digital Hub** (TDH) Program, with the aim of creating a digital tourism hub, accessible through a dedicated multi-channel web platform, which, like a virtual bridge, allows the meeting between the needs of tourists and the vast Italian tourism ecosystem.

This tool, through the exploitation of data, enables the communication of digital supply and demand by offering national and international tourist services and interests tailored to different types of users in order to improve their travel experience.

The collection and aggregation of data is enabled by the interoperability platform realized within the framework of the Program, called TDH022, which enables the exchange of contents and the integration of services provided by public and private Partners through API calls. The TDH022 technology platform allows public and private partners, subject to agreements with the Ministry, to interoperate with the Tourism Digital Hub to exchange data and services.

The TDH022 Ecosystem Guidelines, identify digital standards and technologies at a national level, and guide the exchange of data and content between participants, acting as an integration interface between sector operators who wish to be part of the digital ecosystem and who are operational in Italy. Next figure shows the scope of the guidelines:

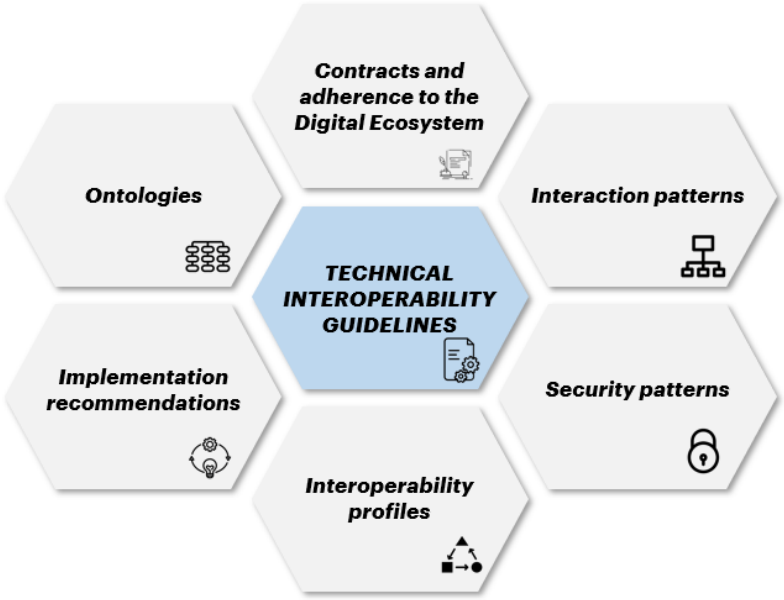

Figure 13 Guidelines and Operative Documents Structure

The Tourism Digital Hub involves the actors of the entire tourism system through an 'open' approach, capable of enhancing the existing complex of descriptive digital content present on the national territory and the services offered in this context.

Data and services integration: The TDH022 enables the integration of data and services from **public** (Regions, P.A., Central Administrations, etc.) and **private** (Hospitality, Food & Wine, Experiences, etc.) partners, allowing tourists to access a multiplicity of services through a single access point, the TDH access channels i.e. the italia.it portal and, shortly, the mobile app, capable of responding to all kinds of needs, improving and facilitating their experience before, after and during their trip.

Services for Tourists: Tourists can consult the digital platform to search for the information and/or services they need for their stay in Italy, such as accommodation, restaurants, places to visit, itineraries and activities to do. The digital hub also offers assistance services that can accompany the tourist throughout the entire journey, as well as continuously offering real-time hints and suggestions, generated on the basis of the user's preferences and interests.

Data models: The "data schemas" of the TDH022 solution follow the ontologies produced by the Ontologies Team and are validated by the Ministry. In addition, the TDH is aligned with approved national and European standardization initiatives, such as the European Interoperability Framework (EIF) document.

APIs: The Ministry provides a catalogue of APIs, available at the following website (https://developerportal.italia.it/), to ensure that parties involved in the exchange and use relationship are aware of available APIs and stated service levels. In this way, developers can easily access the APIs needed to develop new solutions. The APIs present offer the

possibility to exchange data and content with the Tourism Digital Hub, such as customer data, content of interest, events, etc.

3.5.1 Identity, Trust, and on-boarding

Following a Partner's successful completion of the manifestation of interest application process, a TDH022 system administrator proceeds to activate on the API portal the user account of the Partner's technician in question.

In the future, through the federation with the CIAM (identity provider), the Partner's technical contact person will be required to register to proceed with uploading the documentation required to pass the expression of interest. In this way, he/she will be able to automatically log in to the portal and request activation of the keys to access the exposed services.

3.5.2 Data still not available

At present, the TDH022 platform does not implement the possibility of integration with User Generated Content systems. However, it is a functionality that could be developed in the future to offer tourists more interaction and participation in content creation.

Datasets useful for the valorisation of the Program that the Ministry does not yet have at its disposal concern real-time data on tourism trends in the various Italian tourist destinations. Such data sources could help generate an even more customized offer for tourists.

The used approach is Privacy by Design, i.e., in compliance with the fundamental principles of personal data protection, at the same time appropriately declining the methods of personal data processing at the time of signing the contract with the data provider.

3.5.3 Business case

The data will be made available to users for the benefit of the community, consistent with the institutional purpose of the Ministry and no B2B use cases are envisaged.

3.5.4 Some conclusions

The Italian Tourism Digital Hub provides open-data portals and open-data based applications based on public data and some private data from companies when possible, following an “open data lake” style approach. The initiative includes a well-defined on-boarding process to include new data providers, including the definition of the necessary data models as ontologies, the SLAs and contracts and a secure way to transfer the data. Therefore, the current initiative covers several the data spaces principles, including interoperability identity management, trust, and secure data transfer, including also data-based services and application for the final user (tourist).

However, B2B and valorisation of data is out of the scope of the initiative. A data catalogue, where other companies could search for data sets or services is also missing.

4 TOURISM DATA SPACES SPECIFIC TECHNICAL REQUIREMENTS AND RELEVANT BUILDING BLOCKS

This section presents some specific issues that should be considered when setting up a tourism data space along with some initial recommendations about the strategy to address them.

4.1 Personal data

It is unclear for now to understand properly how the governance of personal data sharing will play out, since interactions, on the individual's side, will happen through interoperable personal data intermediaries (PDIs) that are, in most cases, not attached to a specific data space. At the same time certain decisions and policies will be decided at data space, national, local, or sectoral level.

Regarding the tourism sector, we recommend to start imagining a cross border, cross data space standard, that will make personal data sharing seamless across the whole industry. We propose this standard to be called Travel Connect.

We believe that in the early phase, the **Travel Connect** standard should allow full interoperability of PDIs from the 3 following approaches: MyData, SOLID, and EU Wallets.

4.2 SMEs

For most SMEs the process to on board in a data space and the technology needed to participate are too complex and very far from their usual business. They do not have the technical nor operational skills to deal with data spaces.

Some approaches both in the IDS and GAIA-X context could be very useful for SMEs: **Connector as a service** and the more ambitious **Data Space as a service**. The objective of both approaches can be summarized by the slogan used recently by Sovity[28]: "Setting up data space technology in minutes instead of months".

The strategy is to provide the SMEs with consultancy services and the software needed to share (both provide and consume) data. The software includes a user interface that hides the technical and operational complexity of data spaces showing the final user a high-level overview of the available data, the data contracts and the data transfer processes taking place, facilitating the connection with the internal applications used to produce or consume data.

Next figure shows an example of the Eclipse Data Space components data dashboard, an example dev frontend application for EDC Data Management API.

[28] https://sovity.de/about/

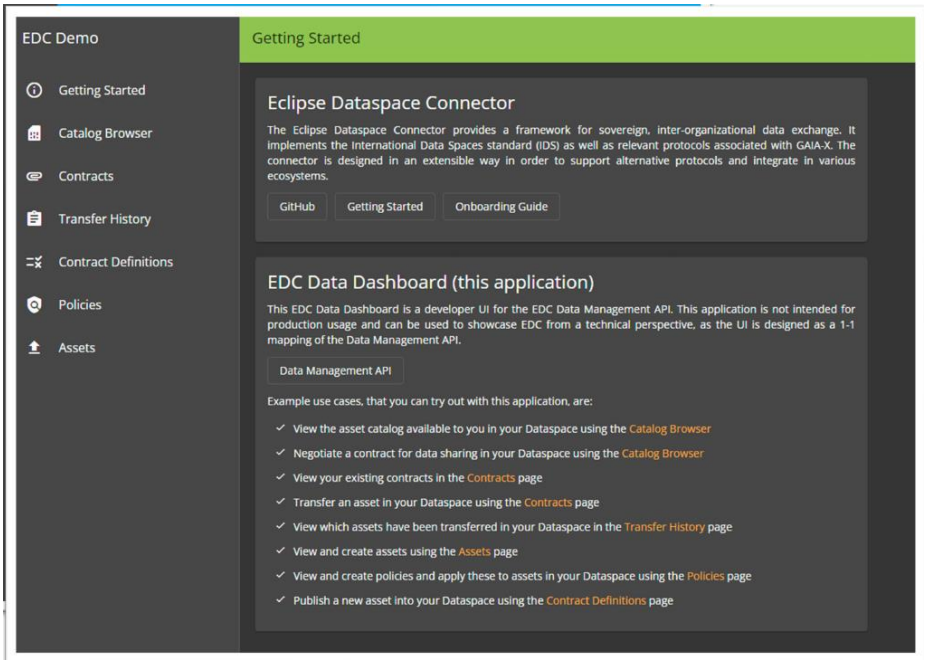

Figure 14 Eclipse Data Space components data dashboard

To better demonstrate the possibilities of dataspaces and their management, the Eclipse Dataspace components project has created under the minimum viable dataspace the so-called **Vision Demonstrator** to showcase a possible user interface that would enable end-to-end interaction - all the way from joining a dataspace to being able to publish a new data asset for others to consume.

This vision was based on 7 scenarios essential to managing a dataspace:

1. Manage my Dataspaces
2. Discover Data Shared by Others
3. Negotiate a Data Contract
4. Create a new Policy
5. Create a new Data Asset
6. Create a Data Contract
7. Review existing Data Contract and Manage Notifications

Next figure shows the main window of the vision´s mock-up.

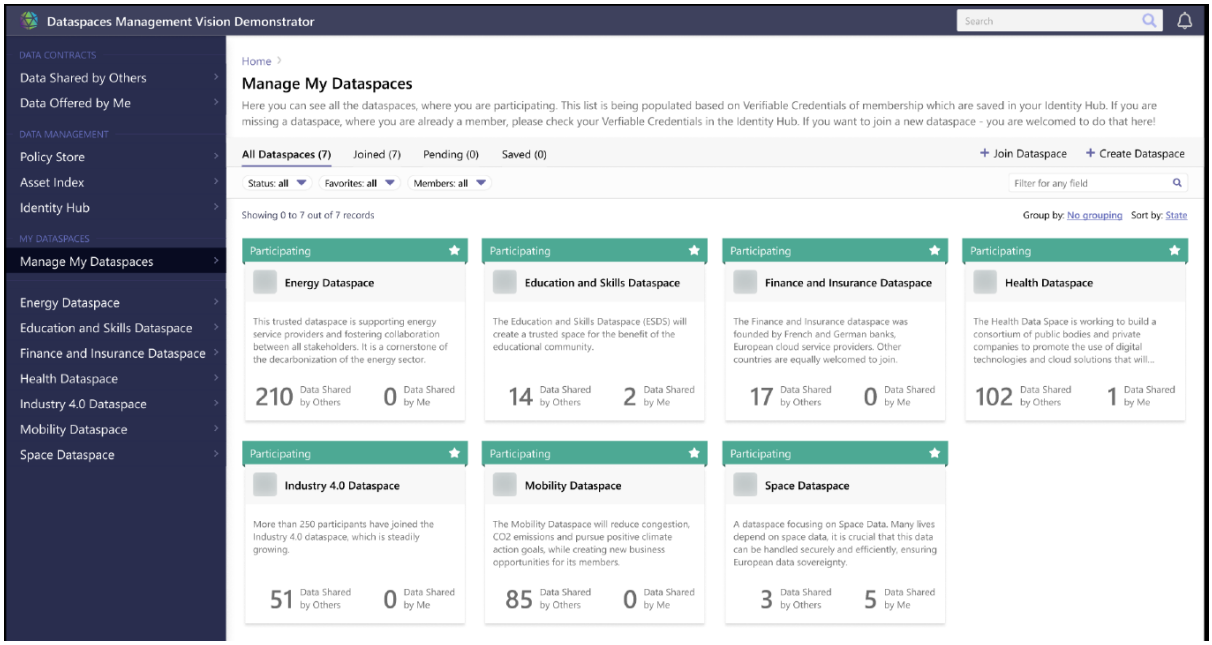

Figure 15 Data Space Vision mock-up

Gaia-X and Simpl go a step forward offering cloud infrastructure needed to run the software, data products that encapsulate data sets or access to data APIs, applications and services providing data-based applications or algorithms and the orchestration functionality needed to define specific services composition and workflows.

The Data space as a service concept includes all the steps needed to participate in a Data Space.

- Participant on boarding
- Compliance and certification (data and services)
- Catalogue registration
- Data sharing functionalities and dashboard
    - Search for data products/infrastructure/applications or services
    - Use or provide data products/infrastructure/applications or services
    - Monitor data spaces

4.3 Public and Private actors

Tourism sector stakeholders include both public administrations and private companies with different strategies and objectives regarding data sharing.

Public administrations' concern is about improving tourism destinations, improving users experience both tourists and people living in the destination and assuring tourism sustainability in the long term both from the economic and environmental points of view.

Regarding data, the main objective of public administrations is to make data as open and available as possible, so open data is the main approach. An example of this strategy is the recently approved EU Commission Implementing Act on High-Value Datasets.

The Regulation is set up under the Open Data Directive, which defines six categories of such high-value datasets: geospatial, earth observation and environment, meteorological, statistics, companies, and mobility. The datasets will be available in machine-readable format, via an Application Programming Interface and, where relevant, as bulk download. Some of this high-value datasets are very relevant for the tourism sector.

Many public administrations provide open-data portals and open-data based applications based on public data, also including private data from companies when possible. The list of current data sharing initiatives analysed in WP2 includes many examples of this kind of "open data lake" style approaches.

However, public administrations are not allowed to include commercial data or include data products for profit. Besides, public administration data sharing approaches does not facilitate B2B data sharing, which is one the main data space goals.

The public administration strategy and the technology used to implement are not aligned with two of the main characteristics of a data space: Data monetization and valorisation and data sovereignty.

- Regarding data transfer, since business is at the core of the current data space initiatives, these initiatives include contract negotiation as a mandatory step, that is not needed in an open data context.
- Open data access normally has a very low security requirements, just download files or a rest interface with no security is enough.

The only example of merging both open and proprietary data (for profit) in the same platform is the FIWARE monetization architecture.

One possible approach could be to adapt the current open data initiatives to the data space initiatives requirements, constraints, and compliance rules, becoming a participant in the data space.

The public administration managing the open data portal should follow the data space on-boarding process and all the data sets/products included should be defined and certified according to the trust framework defined in the data space.

Furthermore, another channel to get the data should be added to the current ones, a data connector compliant with the data space requirements.

4.4 (STRONG) Relation with other data spaces

The tourism sector is directly related to other sectors such as transport, mobility, energy, construction, health... In this way, **data from these sectors should also feed the ecosystem** of a tourism data space. Inter and intra data space interoperability is a critical issue for tourism data spaces.

These are some examples of the connection between tourism and these data spaces:

- **Transport:** The volume of air passengers and the reservation forecasts are very useful indicators for any tourism manager.
- **Mobility**: Data from mobile phone operators are used to measure the flow of tourists at a specific destination or point of interest.

- **Environment**: Tourism activity data is used to monitor the carbon footprint of a destination.
- **Energy:** Monitoring of energy consumption data (e.g. resort) should optimise energy efficiency systems.
- **Cultural and environmental heritage**: In protected natural resources is essential to identify the maximum carrying capacity to preserve the original conditions and usually maximum entry quotas are established.

5 CONCLUSIONS

It is not the objective of this report to define "yet another" data space reference architecture. Our objective has been firstly to identify the main technical issues and challenges to be solved when implementing the data space concept and to analyse the main initiatives and how those issues are being tackled. These main technical issues are: Identity, Self-description, Trust framework, On-boarding, Data and services discoverability, Data sharing, Data space monitoring and observability.

Several ongoing initiatives are now running in parallel to design and implement the data space concept. GAIA-X and Simpl provide by now the more comprehensive data space architectures and specifications including infrastructure, data and services in the same framework while IDSA and FIWARE provide solutions dealing with more specific aspects of the data space technology landscape.

The DSBA convergence effort and Simpl project add even more complexity to the data space scenario.

However, even with this complex scenario some solutions and approaches to specific aspects of the data spaces architecture are common or quite similar in all the initiatives, emerging as the most promising ones:

- The need for a decentralized solution for identity management. Self-sovereign identity solutions along with verifiable presentation and verifiable credentials for self-descriptions.
- The concept of Data Space Authority defining and implementing the rules to be part of a data space.
- Decoupling of control and data planes in data transfer technologies, making it possible to use any transfer protocol or technology available.
- Use common and well-established standards if available.

This report has also identified some specific issues that should be considered when setting up a tourism data space and some initial recommendations about the strategy to address them have been presented:

- **Personal data management**: People is at the centre of the tourism data space, so that data protection legislation and initiatives play a very important role.
- **SMS:** Most companies in the tourism sector are SMS and the process to on board in a data space and the technology needed to participate are too complex and very far from their usual business. They do not have the technical nor operational skills to deal with data spaces. Some approaches both in the IDS and GAIA-X context could be very useful for SMEs: **Connector as a service** and the more ambitious **Data Space as a service**.
- **Public and private actors**: Tourism sector stakeholders include both public administrations and private companies with different strategies and objectives regarding data sharing. Open data public administrations' strategy should be aligned with data protection, valorisation and monetisation strategies of private companies.
- **(STRONG) Relation with other data spaces:** Tourism sector is directly related to other sectors such as transport, mobility, energy, construction, health… In this way, data from these sectors should also feed the ecosystem of a tourism data space. Inter and intra data space interoperability is a critical issue for tourism data spaces.

The report contains intermediate results of the WP. Therefore, this document must be considered as a working document that will evolve along the project's lifetime according to the ongoing work carried out by each of the initiatives that are analysed along the document, the more detailed work in task T3.2, WP2 and WP4 achievements regarding data models and standards, data governance and tourism specific use cases as well as the collaboration with the sister CSA in tourism, which is working in parallel.

Finally, The DSSC has started the process to define the data spaces blueprint in collaboration with the CSAs in data spaces and DATES will collaborate with them bringing the tourism data space perspective.

6 REFERENCES

[1] "*Study on mastering data for tourism by EU destinations*", EU project Smart Tourism Destinations, July 2022: https://data.europa.eu/doi/10.2873/23880

[2] "*Holistic view of platforms and Intelligent Destination Platform solutions",* Segittur (Spanish Mercantile Society for the Management of Innovation and Tourist Technologies), February 2023: https://www.segittur.es/wp-content/uploads/2023/02/20230602-Vision-Holitica-Plataformas-y-soluciones-PID.pdf

[3] Results of the workshop on "*Use Cases of the Spanish Tourism Data Space*" held in October 2022 by Globaldit in collaboration with the Spanish Data Office / SEDIA: https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/noticias/2022/Infografias_WS_Turismo.pdf

[4] "*Radiography of the Spanish Tourism Data Space in Spain*", Gaia-X Hub Spain, December 2022: https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/noticias/2022/Radiografia_Espacio_Datos_Turismo_Espana.pdf