An Ontology for the "System Tourism" – A Call for Participation

Benedikt T. Arnold[1,3], Paul Moosmann[1,3], Anke Sange[2], Christoph Lange[1,3], Christina Gillmann[1], and Claudia Brözel[2]

[1] Fraunhofer Institute for Applied Information Technology FIT, Germany
[2] Eberswalde University for Sustainable Development, Germany
[3] RWTH Aachen University, Information Systems, Germany
benedikt.arnold@fit.fraunhofer.de

Abstract. An ontology in computer and information science formally represents domain-specific knowledge, enabling structured data integration and interoperability. In tourism, ontologies are crucial for standardizing diverse data from sources such as travel agencies, airlines, and user-generated platforms, improving semantic search, personalized recommendations, and developing intelligent systems. This paper presents a comprehensive tourism ontology built on the tourism value chain, identifying key classes, their relationships, and integrating existing ontologies. The resulting hierarchical structure supports enhanced data analysis and system intelligence. Additionally, we provide guidance on how to use and contribute to the ontology, making it accessible for broader application and development.

Keywords: Ontology, Formalization, Participatory research, Tourism value chain, Tourism system

1 Introduction

An ontology in computer and information science is a formal representation of knowledge within a specific domain, consisting of a set of concepts, categories, and their relationships. It enables structured data integration, sharing, and interoperability by providing a common vocabulary and framework for understanding and organizing information. Ontologies are crucial for applications such as semantic search, intelligent systems, and data analytics, enhancing both machine and human comprehension of complex information. Creating an ontology for the system tourism is essential for several reasons. Tourism data comes from diverse sources, including travel agencies, airlines, hotels, local attractions, and user-generated content platforms and many more. These sources often use different terminologies and data formats. Standardizing concepts such as accommodation types, transportation modes, tourist activities, and geographical locations helps in aligning disparate datasets. Traditional keyword-based search engines may not understand the context or relationships between concepts. This might be partially addressed by Generative AI, however also these systems benefit

from ontologies to avoid hallucinations [3]. By understanding the relationships between different tourism entities (e.g., hotels near a specific attraction), recommender systems can provide more accurate and contextually relevant suggestions. Ontologies underpin the development of intelligent systems that can reason about tourism data and provide automated assistance. An ontology can model user preferences and profiles, enabling personalized services such as customized travel packages, targeted marketing, and adaptive user interfaces. This way, tourists can be provided with contextual information about their destinations, including historical background, cultural significance, and local customs. Ontologies can incorporate multilingual terminologies and concepts, facilitating communication and information access for tourists from different linguistic backgrounds. Tourism ontologies can help in managing resources such as accommodations, transportation, and attractions by providing a structured representation of these entities and their relationships. Policymakers and tourism authorities can use ontologies to analyze trends, assess the impact of tourism activities, and develop informed strategies for sustainable tourism development. Researchers can use tourism ontologies to study patterns, behaviors, and trends within the tourism industry. Advanced analytics and machine learning techniques can be applied to ontology-structured data to uncover hidden patterns and predict future trends. In this paper we provide an ontology for tourism that is built upon the value chain of tourism [12]. We developed this ontology by starting with the individual phases of the value chain in tourism and derived different classes that are contained in these processes as well as their relations. In a second step, we identified classes that are contained in multiple phases of the value chain and built a combined ontology.

Tourism is a complex industry with a huge variety of different elements. The "system tourism" aims to picture all the different stakeholders and objects based on systemic theory to show their interwoven and depending relations [2]. To grasp the bigger picture and the relations of the elements it does need a holistic view. Therefore we focus on the system and explain the elements by character. So far the identified categories are Destinations, Tour Operator, Transport, Distribution, Health Resorts/Spas and travel equipment to be consumed/used while traveling, as to be seen in Figure 1. The categories of course can be divided into sub categories as for example transport includes rail, air, water and road and each of this element can be further specified, e.g. road includes cars, buses, bridges and parking areas/service stations to mention only a few. The flow between all the elements are data – that connect the complex causal relationship in the system, which we aim to explain in a universal ontology.

Therefore, this paper contributes:

- A holistic ontology of the "system tourism" based on the value chain of tourism
- An alignment of our ontology with already existing ontologies
- A guide on why and how to use this ontology
- A guide on how to contribute to the ontology

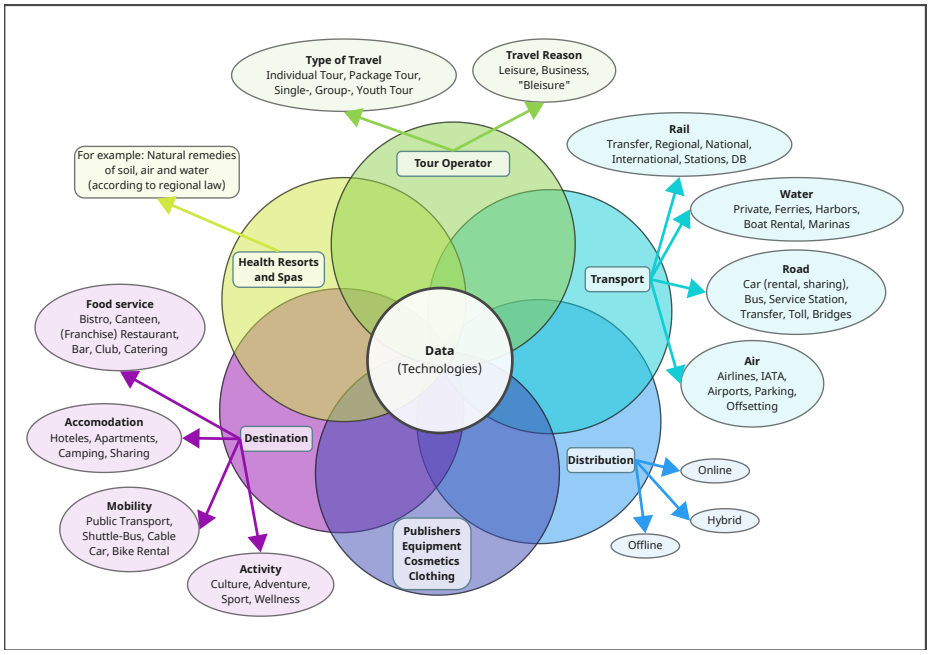

Fig. 1. The System Tourism and its Elements. Source: Authors' own.

2 Related Work

There exist a variety of attempts to structure and map terms and relations in the context of tourism. The most important will be explained in the following.

The United Nations World Tourism Organization (UN Tourism) thesaurus [13] offers a comprehensive tourism ontology covering object profiling, tourism packages, multimedia content, and archaeological object descriptions, while Mondeca's Intelligent Topic Manager software is used to maintain this travel ontology. Although it provides a comprehensive ontology, there does not exist a mechanism to contribute to the ontology.

A variety of research projects tackle ontologies in tourism. The project Open Data Germany [5] aims to collect all kinds of data from German tourism. The Harmonise project [8, 4], funded by the European Commission's 5th Framework RTD Programme, developed an ontology for the tourism industry to address data heterogeneity, covering key concepts such as travel events, accommodations, attractions, and food styles, and providing detailed information on local services, currency, languages, and event timings. DATAtourisme [1], managed by ADN Tourisme in collaboration with the French government, is a national resource that includes an ontology and platform for aggregating, standardizing, and disseminating open data in the tourism sector, structuring data on festivals, events, heritage sites, leisure activities, itineraries, accommodations, restaurants, shops, and services collected by various tourism bodies. The German Text Exploitation

and Search System (GETESS) [11], a project funded by the German Federal Ministry of Education and Research, retrieves tourism-related information from websites using natural language processing and Semantic Web methods to provide user-friendly answers to queries. These initiatives focus on specific parts of the value chain in tourism. The QUALL_ME Ontology[4], funded by the EU, is another ontology for the tourism sector and covers various aspects such as accommodation, sites, events, and transportation, and is integrated with WordNet, a large lexical database, and the SUMO foundation ontology, thereby creating a comprehensive and powerful knowledge base.

The Tourism Semantic Web Portal's OnTour prototype [10] demonstrates Semantic Web usage by providing detailed information on accommodations, activities, events, and more, allowing tourists to plan trips based on preferences and constraints, though it lacks detailed information on some concepts. We will include the concepts that have already been mapped into our ontology.

The Travel Agent Games in Agentcities (TAGA) [15], which operates on a platform compliant with the Foundation for Intelligent Physical Agents (FIPA), defines two domain ontologies covering travel concepts such as routes, services, and reservations, as well as auction protocols, to develop a heterogeneous network of agent services for various groups including travel and business services. As the ontology maps the view of agents and suppliers, it lacks concepts related to clients.

Gala et al. [6] propose an urban tourism ontology for indoor and outdoor environments, adhering to cultural heritage standards, and demonstrate its utility by developing a repository and parser that convert CSV data from a museum's data repository into RDF triples. Unfortunately, this ontology does not cover aspects such as transport.

Haridy et al. [7] propose an enhanced ontology development methodology (ON-ODM) based on ontology-driven conceptual modeling (ODCM) and natural language processing (NLP), which is applied to create a tourism ontology for Egypt useful for applications such as e-tourism. The ontology is very restricted to local concepts of Egypt.

Zhang et al. [14] focus more on the question how to create an ontology for the tourism. By analysing the documents of the Springer Encyclopedia of Tourism and 16 major international academic journals they extracted a tourism-type ontology that provides a systematic of 232 tourism types.

The OpenTravel Alliance (OTA) is developing a domain-specific ontology for the travel and hospitality industry, focusing on creating a shared vocabulary and data model that enhances interoperability between systems in the travel ecosystem. The approach already considers many individual aspects with a certain level of diversity and depth, but it does not yet fully represent the comprehensive system we aim for [9].

In summary, although there exists a variety of ontologies concerning tourism, the is no ontology yet that models the whole tourism value chain in a systemic

[4] https://qallme.sourceforge.net/

way (the "system tourism"). The existing ontologies only address sub-elements of the whole system tourism. This forms the motivation for this work.

3 Methods

3.1 Basics on Ontologies

An ontology, in the context of computer science and information science, is a formal representation of knowledge as a set of concepts within a domain and the relationships between those concepts. Ontologies are used to model domain knowledge in a structured way, facilitating data interoperability, sharing, and analysis. In the presented work we use the Web Ontology Language (OWL), which is based on the Resource Description Framework (RDF) syntax to create our ontology. OWL uses classes and properties between these classes to model ontologies. These concepts are elaborated in this section.

Classes: The primary building blocks of an ontology, representing categories or types of objects within the domain. For example, a *customer* is one of the classes in an ontology related to tourism. A defined class can be instantiated, thus representing a concrete data instance. E.g., the class *customer* might have the instance *Vasco da Gama*, depicting a specific customer.

Properties: Properties are attributes of a class. OWL has two different types of properties: datatype properties and object properties. Datatype properties describe a directed relation between two classes. The class from which the relation originates is the subject, while the class that is pointed to is the object. Together with the property, they build an RDF triple of the form *Subject - Predicate - Object*. E.g., if we want to model a customer that books his vacation at a travel agency, the triple would look like *Customer - books vacation at - travel agency*. In this example *customer* and *travel agency* are classes and *books vacation at* is an object property. Opposed to object properties, datatype properties are not relations between two classes, but between a class and a literal, which is information stored in a datatype. Examples of datatypes are *strings* to store text or *integers* to store numbers. A triple modeling the age of an instance of a customer might look like this: *Vasco da Gama - has age - 54*. Here the property *has age* points to an integer. It is also possible to model the age using other datatypes, e.g., the date datatype. This would lead to the triple *Vasco da Gama - has birthdate - 1469-01-01*. Figure 2 shows an example of a class and its properties. The green sign (C) indicates the Class, whereas the properties and their datatypes are shown below.

OWL also offers possibilities to model constraints on certain relations. E.g., it is possible to model cardinalities, such as the fact that every customer has exactly one date of birth. This is important to ensure a certain level of data quality in the data instances.

3.2 Creating the Ontology

In order to create our ontology for the tourism domain, we followed a three-step approach:

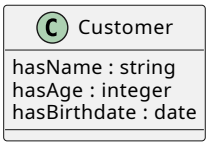

Fig. 2. An example class and its properties. This scheme will be used to visualize the provided ontology. Source: Authors' own.

1. We created a separate ontology module for each of the value chains in tourism
2. Through identifying connecting classes, we combined these separate ontologies into one merged ontology
3. We enriched the merged ontology by utilizing other existing ontologies.

We have chosen this approach to first focus on the ontology creation from the systemic perspective. The third step, however, is very critical because the more connected ontologies are to existing ones, the more useful they become.

In order to create the presented ontology and provide it in a readable, sharable and adaptable way, we utilized LinkML. LinkML is a framework designed for creating schemas, validating data, and generating code. To be able to do this, we create a YAML file using the LinkML model, which contains the domain knowledge we want to model. LinkML provides different generators which we can then use to create our ontology in different formats. we used the gen-owl command to generate the ontology as an OWL graph. Additionally, We used the gen-plantuml command to generate visual representations of the LinkML models as UML diagrams in PlantUML format. While these different generators provide a lot of flexibility, they also have some drawbacks, since different representations of the ontology (e.g., UML diagrams and OWL graphs) need to be created from the same LinkML source. This results in some modeling approaches that might not seem straightforward, but were necessary in this case. For example, we needed to introduce abstract parent classes (classes that must not be instantiated) in a few cases, where different classes shared the same properties. While this can be considered in a UML diagram, an OWL graph is not able to represent this concept natively. For this case, we added additional properties to indicate if a class is an abstract class or not in the OWL graph.

Step 1: Individual ontologies of the value chains in tourism In the first steps, we analyzed the value chain by Sutomo et al. [12] and based the system tourism ontology on this value chain. We used each stage in the value chain as a starting point (destination, transport, accomodation, food and activity) to create an individual ontology. This resulted in 5 individual ontologies. We chose this approach, since using all stages of the value chain provided a good way to think about individual objectives of tourism. Throughout the creation process of the ontologies, there has been a close collaboration between the authors who are members of the tourism community and the partners who are computer

scientists. This ensures that the resulting ontologies comprehensively represent the domain knowledge in a semantically and syntactically correct way.

Step 2: A merged ontology As a second step, we started merging the individual ontologies. Multiple concepts, such as *customers*, can be found in different stages of the ontology. We used these classes to connect the ontologies. First, duplicates have been identified and merged. In this case no conflicts were identified. In the second step, synonyms such as *customer* and *client* have been identified and merged. Here, we had to discuss which terms we use and which properties we keep.

Step 3: Enriched ontology As a last step, we utilized existing ontologies (cf. section 2) to enrich the developed system tourism ontology. We started by using a subset of the existing ontologies such as [8, 4] or [1], which can be expanded step by step in the future.

3.3 Results

As a result, we provide the developed ontology as an OWL graph using the Turtle syntax. Additionally, we provide the YAML file, which is utilized by the LinkML generator to create the OWL graph. This file can also be used by other LinkML generators to create additional assets, such as UML diagrams or documentation. Besides, another LinkML can produce SHACL shapes from the the YAML file that can be used for validating instances of the ontology. The ontology itself can be retrieved at https://github.com/Fraunhofer-FIT-DSAI/SystemTourismOntology alongside a visualization in UML that could not be included in the paper due to space constraints. To allow access to the implementation, we uploaded it to GitHub. GitHub is a web-based platform that provides version control and collaboration features for software development projects. It is built on top of Git, a distributed version control system. GitHub offers a variety of tools and features that facilitate the development, sharing, and management of code. Three key benefits of GitHub are:

- Collaboration and Version Control: GitHub provides robust tools for version control, enabling multiple developers to collaborate on projects efficiently. It tracks changes, manages code versions, and allows for seamless integration of future contributions.
- Community and Open Source Support: GitHub hosts a vast community of developers and a plethora of open-source projects. It facilitates knowledge sharing, networking, and contributions to open-source software, fostering innovation and continuous improvement in the software development ecosystem.
- Integration and Automation: GitHub integrates with various tools and services, such as continuous integration/continuous deployment (CI/CD) pipelines,

project management tools, and cloud services. This enhances workflow automation, improves deployment processes, and streamlines development practices, increasing overall productivity.

4 Discussion

4.1 Potential of the Ontology

An ontology provides a common framework and vocabulary, facilitating the integration of data from heterogeneous sources. This ensures that information can be seamlessly shared and understood across different systems, enabling more accurate data exchange and reducing inconsistencies, which improves the overall quality of tourism information systems. Thus, the ontology may help touristic companies to provide their guests with personalized recommendations that better match their preferences and interests, enhancing their travel experience through chatbots that understand complex queries, virtual travel agents that plan itineraries, and decision support systems for tourism management.

Tailored experiences cater to tourists' specific needs and desires, increasing satisfaction and engagement while enhancing the educational and experiential value of travel, making trips more enriching and meaningful. Additionally, ontologies ensure that tourists can access information in their preferred language, breaking down language barriers and improving accessibility. The LinkML model currently does not implement multi-language support, so this is not directly reflected in the outcome of the LinkML generators. However, the generated OWL graph can be extended to incorporate multi-language support in a post-processing step. Efficient allocation and management of resources optimize operations and enhance service delivery, supporting evidence-based decision-making and strategic planning, leading to better outcomes for the tourism sector and local communities.

Ontologies provide a rich source of structured data that can be analyzed to gain insights, develop theories, and inform practice. This structured data enables more accurate forecasting and trend analysis, supporting proactive and informed decision-making. Moreover, ontologies enable semantic search capabilities, where search engines understand the meaning and context of queries. For example, understanding that "buffet" can refer to both a buffet meal and waves buffeting a ship, and retrieving relevant results accordingly.

4.2 How to use this Ontology?

We developed this ontology in order to harmonize and standardize terms and relations in tourism. Further, the improvement of interoperability is a major point that can be addressed by an ontology.

An ontology establishes a common vocabulary for the tourism sector, defining key terms and their meanings in a standardized manner. This standardization is crucial for aligning different stakeholders, such as travel agencies, hotels, and

regulatory bodies, around a shared set of definitions. It ensures that everyone uses the same language, which reduces miscommunication and fosters clearer, more effective communication across the industry.

The ontology can be used to explicitly create a knowledge graph with instances of the concepts defined in it, e.g. for some touristic destination. Besides, in practice, the defined terms can be referenced by their URIs to refer to the underlying concepts in, e.g., databases or CSV files. This way, the URI can be used to look up the term and to unambiguously get its meaning.

By defining a common framework and language, an ontology facilitates the integration and interaction of various systems and platforms within the tourism industry. This enhanced interoperability allows different information systems to exchange data seamlessly, leading to better coordination between services, improved data sharing, and the creation of more comprehensive and integrated tourism solutions. Consequently, it supports more efficient and cohesive operations across diverse tourism-related entities.

4.3 Contribution and Maintenance Concept

The provided ontology is assumed to be an open source project, licensed under CC-BY-SA 4.0 and available at https://github.com/Fraunhofer-FIT-DSAI/SystemTourismOntology. Besides the free use of this ontology, we would like to encourage an active contribution to it. Anyone can create an account on GitHub for free. Once the account is created, the GitHub repository of the system tourism ontology can be accessed. It can be downloaded for free. In case a user wants to request a change, the request can be handed in via the ticket system of GitHub. The incoming requests will be reviewed and handled by a governing change committee to be founded at the conference–consisting of both people from tourism and computer science.

5 Conclusion

This paper introduces a comprehensive ontology for tourism as a system, grounded in the tourism value chain, addressing the need for standardization across diverse data sources in the tourism industry. By structuring key classes and their relationships, the ontology enhances the functionality of intelligent systems, enabling improved semantic search, personalized recommendations, and more efficient data management. The integration of existing ontologies further strengthens its applicability. The provided guidelines for usage and contribution encourage broader adoption and collaborative development, making the ontology a valuable tool for both researchers and practitioners in advancing data-driven tourism solutions.

As future work, we aim to promote the use of this ontology in further analytical processes such as visual analytics, explainable AI or the use in complex settings such as dataspaces. We further want to promote the participatory approach to maintain and extend the ontology.

References

1. ADN tourisme. DATAtourisme. https://gitlab.adullact.net/adntourisme/datatourisme/ontology. Accessed: 2024-07-08.
2. Claudia Brözel. *Perspektiven von Transaktionen in der Internet-Ökonomie am Beispiel der Reisebranche*. PhD thesis, Technische Universität Dresden, Dresden, 2012.
3. Diego Collarana, Moritz Busch, and Christoph Lange. Knowledge graph treatments for hallucinating large language models. *ERCIM News*, 2024(136), 2024.
4. Mirella Dell'Erba, Oliver Fodor, Wolfram Höpken, and Hannes Werthner. Exploiting semantic web technologies for harmonizing e-markets. *J. Inf. Technol. Tour.*, 7(3-4):201–219, 2004.
5. Deutsche Zentrale für Tourismus e.V. Open data destination germany. https://open-data-germany.org/open-data-germany/, 2020. Accessed: 2024-09-12.
6. Alexander Pinto-De la Gala, Yudith Cardinale, Irvin Dongo, and Regina Ticona-Herrera. Towards an ontology for urban tourism. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, SAC '21, page 1887–1890, New York, NY, USA, 2021. Association for Computing Machinery.
7. Shaimaa Haridy, Rasha Ismail, Nagwa Badr, and Mohammed Hashem. An ontology development methodology based on ontology-driven conceptual modeling and natural language processing: Tourism case study. *Big Data and Cognitive Computing*, 7:101, 05 2023.
8. Michele Missikoff, Hannes Werthner, Wolfram Höpken, Mirella Dell'Erba, Oliver Fodor, Anna Formica, and Francesco Taglino. Harmonise - towards interoperability in the tourism domain. In Andrew J. Frew, Martin Hitz, and Peter O'Connor, editors, *Information and Communication Technologies in Tourism, ENTER 2003, Helsinki, Finland, 2003*, pages 58–66. Springer, 2003.
9. OpenTravel Alliance. Opentravel alliance releases 2018a 2.0 object model publication for hospitality. *Leisure & Hospitality Monitor Worldwide*, 2018.
10. Katharina Siorpaes and Daniel Bachlechner. Ontour: Tourism information retrieval based on yars. 2006.
11. Steffen Staab, Christian Braun, Ilvio Bruder, Antje Düsterhöft, Andreas Heuer, Meike Klettke, Günter. Neumann, Bernd Prager, Jan Pretzel, Hans-Peter Schnurr, Rudi Studer, Hans Uszkoreit, and Burkhard Wrenger. Getess—searching the web exploiting german texts. In Matthias Klusch, Onn M. Shehory, and Gerhard Weiss, editors, *Cooperative Information Agents III*, pages 113–124, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.
12. Yoan Sutomo, Corinthias P. M. Sianipar, Mrittika Basu, Kenichiro Onitsuka, and Satoshi Hoshino. Tourism value chain: Synthesizing value webs to support tourism development and planning. *Tourism Review*, 77, 05 2023.
13. World Tourism Organization. *Tourism Towards 2030 / Global Overview*. UNWTO, 2023.
14. Puwei Zhang, Jia Wang, and Rui Li. Tourism-type ontology framework for tourism-type classification, naming, and knowledge organization. *Heliyon*, 9(4), 2023.
15. Youyong Zou, Tim Finin, Li Ding, Harry Chen, and Rong Pan. Taga: Travel market framework in agentcities. pages 1645–, 01 2003.