HAL
open science

A Mobility Data Model for Web-Based Tourists Tracking

Thouraya Sakouhi, Jamal Malki, Jalel Akaichi

To cite this version:

Thouraya Sakouhi, Jamal Malki, Jalel Akaichi. A Mobility Data Model for Web-Based Tourists Tracking. The 14th International Baltic Conference on Databases and Information Systems (BalticDB&IS 2020), Jun 2020, Tallinn, Estonia. hal-02869356

HAL Id: hal-02869356

https://hal.science/hal-02869356

Submitted on 15 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Mobility Data Model for Web-Based Tourists Tracking

Thouraya Sakouhi[1], Jamal Malki[2], and Jalel Akaichi[3]

[1] Université de Tunis, ISG, LR99ES04 BESTMOD, 2000, Le Bardo, Tunisia
thouraya.sakouhi@gmail.com
[2] Université de La Rochelle, L3i Lab., 17000, La Rochelle, France
jmalki@univ-lr.fr
[3] Bisha University, College of Computer Science, 255, Bisha 67714, Saudi Arabia
jakaichi@ub.edu.sa

Abstract. Tracking tourists activities at different levels of their journeys provides an overview on their mobility and a comprehension of their behavior and preferences. Most information related to tourism services and tourists are collected and stored through web platforms. In fact, self-drive tourists access touristic information available on the web to plan for their trips. Accordingly, tourism professionals track their requirements in touristic information and then their mobility. Yet, since touristic information is managed at a territorial level, tracking tourists' movement by tourism professionals, out of their territory, is not a straightforward task. Accordingly, the latters do not have a complete overview of tourists movements. Throughout this paper authors will start by discussing mobility data capture through the web and the related challenges. Then, they'll introduce an integrated mobility data model for tracking tourists.

Keywords: Mobility data, Trajectory, Semantic modeling, Ontology, Web-based tracking.

1 Introduction

Nowadays, wireless sensor devices such as GPS, RFID and mobile phones are tied to mobile entities to trace their movements and generate what is called mobility data. Once analyzed, this data is expected to express a lot of semantics about mobile entities' activities and their dynamics. For instance in the case of tourists mobility data, this motivated large number of research works to investigate novel approaches for extracting from tourists mobility data activities performed during their travel. However, tourists mobility data generated in this way is not always available for researchers and this is due to privacy issues.

From another side, public institutions detain information about touristic offer and tourists demand and possess efficient tools for the managements of related

PLAIBDE Project. Programme Opérationnel FEDER-FSE de la région Nouvelle Aquitaine - France

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

data. These tools represent the tourism ecosystem. Such data encompasses information about touristic offer obviously but also data related to the tourist himself, and his requests according to the touristic offer. It then presents a valuable source of information about tourists mobility during their journey and present an alternative of the mobility data captured by positioning devices. Yet, due to the information sparsity within and across different destinations, tourists do not have a complete and clear overview of touristic offer neither public institutions and service providers on tourists preferences. Then integration efforts have to be made to unify touristic information related to destinations. State of the art works consider only the integration of touristic offer data in their proposed models for tourism field. In the remainder of this paper, we'll present our proposed model of tourist mobility data. This model will give a complete view of the tourism field data. We'll make use of the DATAtourisme ontology being a new platform for integrating touristic information allover France.

Accordingly, this paper is organized as follows. In the second section, authors will define mobility data capture and the related challenges. In the third section, authors will present the tourism ecosystem and the data flow within. After that, authors will overview state of the art touristic data models and the DATAtourisme ontology. Then, we will introduce our tourist mobility data model. Finally, we will discuss challenges raised by the introduction of our model.

2 Mobility Data Capture

2.1 Sensor-Based Mobility Data Capture

Nowadays, sensor devices are integrated to all kinds of moving entities: vehicles, animals and humans. However, while animals and vehicles' mobility data are readily available and accessible, this is not the case for humans mobility data, which is considered as sensitive by the law. Indeed, personal information are protected by the means of a set of norms and laws in different countries. For example, in French law, the CNIL (Commission Nationale de l'Informatique et des Libertés), a committee responsible for protecting personal data in France, and RGPD (Réglement Génénral sur la Protection de Données) for the European Union, impose regulations on expanders of personal information, thus limiting the dissemination of this data, its use and its processing to avoid violation of individuals' privacy. Consequently, this legal constraints made access to this type of data complicated despite its abundance. From the other hand, mostly data collection for research made in the context of human mobility data modeling and analysis requires the recruitment of participants, volunteering or payed, to wear positioning sensors as they go about day-to-day activities. So even the collection of mobility data is not a straightforward task.

2.2 Web-Based Mobility Data Capture

By accessing websites and using mobile applications, users transmit indirectly different types of their data to first and third parties. The data transmission

observed between clients and servers helps parties to keep state of the user over time. Such data is historized and associated with the website visitor to be recognized later. Among the captured data, eventually data about the location of the user is collected, then his mobility according to space and time. Indeed, web/mobile applications' managers track their users online using different methods including: cookies, HTTP metadata, device fingerprinting, local storage, etc. Originally, online behavioral data collection is limited to connecting users across multiple websites on one device. However, today companies are finding ways to correlate user's behavior, by third parties, across different devices by identifying linked devices a unique individual uses. This tracking method is referred to as cross-device tracking [2].

3 Mobility Data Within the Tourism Ecosystem

As tourism and travel are known to be information intensive domains, travel-related information is available abundantly in the web [3]. The interaction between tourists and tourism professionals generates information exchange between the two stakeholders. While tourists ask for information about their journeys, tourism professionals ask for information about the tourist, his location, where he intends to go and more, to communicate him later information he requested. Website owners register users information [2] for analytical and advertising purposes. These users' informations constitute an alternative tool for tracking tourists movement implicitly. The management of territorial touristic information requires a set of tools including, mainly the Touristic Information System (TIS), Customer Relationship Management (CRM) and Reservation System (RS), which represent the tourism ecosystem illustrated in figure 1.

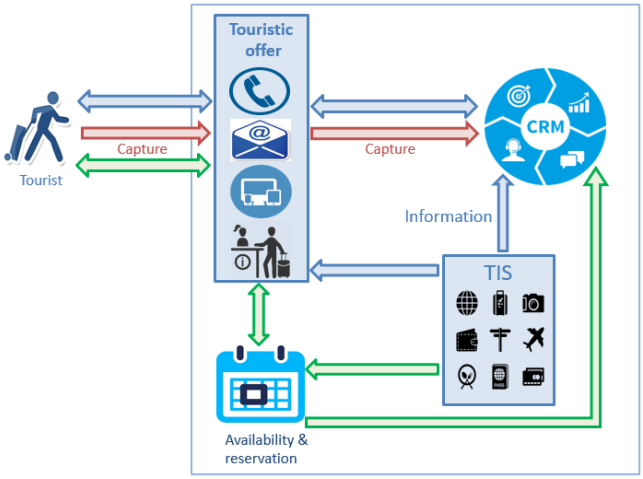

Fig. 1. Tourism ecosystem: captured data flow and processing

In France, touristic information is organized by geographical department. Actually, departmental TIS is fed by the DCT (Departemental Commitee of Tourism) [8] and its partners, the former is then the socle of all DCT publications. This pooling of resources allows a completeness of the offer and a saving of time. However, a lot of touristic service providers at the territory level still do not use the official DCT portal to share their touristic offers. Accordingly, the diversity and variety of information sources on the web related to one territory, evenmore all over France, makes a huge amount of information. And as quantity of available information is large, self-drive tourists pose a big challenge in efficiently accessing and managing the tourism related data and find out about their desired destinations in France. Consequently, the actual tourism ecosystems data in France has the following properties:

- Proprietary: it is the property of the party that generates them: the department, the service provider, etc, and are not accessible to the other sector stakeholders.
- Heterogeneous: the different TISs do not have the same data structure, nor the same semantic and syntactic representation.
- Territorial: limited to the representation of tenders at the concerned territory level.

Accordingly, efforts for the integration of territorial tourism ecosystems data throughout France are essential. On the one hand this will permit to improve the visibility online of touristic offer and increase its availability. On the other hand, this will give a complete view of tourists, their interactions with information sources and then their mobility within French departments. So, an upgrade to a unified tourism data model at a national level is highly required. In the sequel, authors review the more common proposals of tourism data integration models.

4 Tourism Data Modeling: An Overview

In the following, we concentrate in French initiatives of national tourism data integration. TourInFrance (TIF), TIFSem and DATAtourisme are from of those research thrusts that proposed solution for touristic data globalization at a national level in France. We provide a description of these models in the following.

4.1 TourInFrance

TIF is a norm that was initiated in 1999 by the French State Secretariat for Tourism to facilitate touristic data exchange between different TISs in France [5]. This norm is used by national tourism offices, departmental tourism committees and tour operators. It is represented by a metadata format describing tourist information in detail [4]. In 2004, TIF format has evolved to XML technologies in order to facilitate the uploading of data on websites and the exchange of information between different systems. However, since 2005 TIF norm has stopped

evolving. Accordingly, each of the aforementioned tourism professionals adapted the touristic content to their special requirements. Furthermore, with the recent advances of web technologies and its upgrade to the semantic web, TIF became obsolete and useless.

4.2 TIFSem

TIFSem [7] is a tourism ontology developed in the context of the TourinFlux project [6] that proposes a set of tools that allow tourism professionals to handle unified datasets and improve their display on the web. They developed for that an ontology named TIFSem that is based on the TIF norm, representing a set of concepts and relations for tourist resources. Besides, other sources are consulted to collect concepts related to the tourism domain.

4.3 DATAtourisme

DATAtourisme [1], a universal language to describe points of interest for tourists, was proposed by the French government. The latter intends to create a new platform for tourism in France in a national level, and this is by integrating data from various TISs at a territorial level. This permits then to provide touristic information requirements using Semantic Web Services. In brief, DATAtourisme is a shared system that allows territories to put their touristic data in Open Data. DATAtourisme's system has the following characteristics:

- Semantic standard since the meaning of each element of the vocabulary is formal and consensual.
- Open and simplified because available to all and easily reusable.
- Universal because generic enough to apply to any type of tourism content which made possible the harmonization of the vocabulary of each TIS.
- Scalable, this vocabulary can easily evolve to adapt to new needs.
- Language independent and can be used to describe multilingual content.
- Interoperable because it interacts with other authoritative international vocabularies, for example DATAtourisme is natively compatible with Schema.org, INSEE, DublinCore, FOAF vocabularies.

Accordingly, later on, we will make use of the DATAtourisme ontology for the construction of our tourist mobility data model.

5 Tourist Mobility Data Model

A tourist is considered as a mobile entity since he moves across different locations within his destination: museums, hotels, restaurants, etc. Such locations are considered as Points Of Interest (POI) according to the tourism context. These POIs represent also the touristic offer exposed in the web or by the way of the different available channels. While a tourist seeks touristic offer in the

web, he interacts with touristic content by sending offer requests, making reservations, etc. Consequently, this interaction, once analyzed in a global context, provides tourism professionals with valuable information about tourists, their touristic interests and then their mobility before, within and after visiting their destinations.

To do so, we start throughout this work by modeling the tourist mobility data in this context. Our objective is to represent tourist mobility data in a generic, complete and simple model. We intend to categorize the principle entities representing data within the tourism ecosystem previously described. Throughout this section, we introduce the formal model to represent data flow within the tourism ecosystem between the different stakeholders taking part in the data exchange mechanism including tourists and tourism professionals. Authors refer to this model as the Tourist Mobility Data Model (TMDM). Actually, the observation of this data flow brings out three types of data: data that represents the touristic offer, data about the set of events triggered when the tourist seeks touristic offer and data that defines the tourist. An in-depth study of the properties of each of these types of data leads us to the TMDM. Accordingly, the TMDM is mainly composed of 3 sub-models: touristic data space (TDS), event space (ES) and tourist space (TS). In the following subsections, we explain separately each components.

5.1 Touristic Data Space Model

Touristic data constitutes all data related to the touristic offer including: accommodation, entertainment activities, transportation, catering, etc. We consider DATAtourisme since it is then used by almost all governmental and private tourism institution in the country. The DATAtourisme ontology summarizes the touristic content in the POI concept as it represents the minimal definition of any touristic object. POI is defined as any touristic product that deserves to be described and valued. It is a touristic element which is managed by an agent and which could be consumed through products and services [1]. This is the minimal class to instantiate for a product to be managed in the DATAtourisme information system. A POI concept decomposes into 4 sub-concepts: Product, Tour, Entertainment and Event and Place of Interest.

5.2 Event Space Model

The event space represents data resulting from the set of interactions between tourists and touristic data resources made available to them via the different contact points (web/mobile applications, e-mail, phone, tourism offices, etc). Commonly, if an individual is using a web/mobile application, it is very likely that he is being tracked by the application developers. This actually refers to collecting and passing data about user actions when interacting with the application. Each related user action triggers an *event*. The set of collected data include data about the user: location, time, age, gender, coordinates, device

information, interests, etc, and information about his interaction with the application: information sought, time spent on every screen, clicks, ad views, etc. In the case of tourists interacting with touristic data, events are recorded within the CRM database. An event in this case represents a request from a tourist and the response to this request.

5.3 Tourist Space Model

The tourist space represents the set of tourist's intrinsic data that defines him: location, identity, coordinates, genre, interests, etc. The tourist, stands for the moving entity in the context of our application.

The association between the TDS and ES sub-models is performed between the *Request* and *Response* concepts from one side and *PointOfInterest* concept from the other side. The association between the ES and TS sub-models is performed via the *Tourist* and *Event* concepts. The TMDM is illustrated in the figure 2.

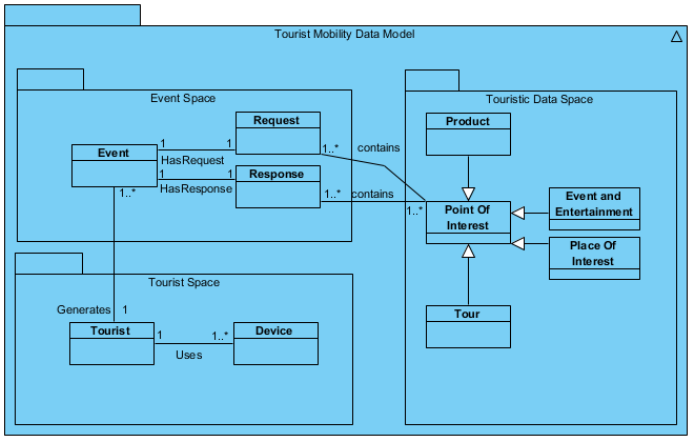

Fig. 2. Tourist Mobility Data Model

6 Discussion

The transition from the territorial to the national level after DATAtourisme will raise, mainly, scientific and technological challenges about the characterization of the touristic mobility data in the future. Actually, with TIS at the territorial level, it was not possible to link tourists behaviors across different platforms/websites they access to seek information or make reservation due to the variety of stakeholders taking part in the presentation of the touristic offer, which made difficult identifying uniquely the tourist. With DATAtourisme, since

all touristic information will be available and structured in the same vocabulary, cross-platform tracking of tourists will be affordable, and tourists will be easily recognizable at a national level. Even more, it will be possible for tourism professionals to track tourists mobility before, during and after visiting their territory. This will provide a more complete view into tourist's behavior and could be useful for a set of applications. Thereupon, this raises in turn an economical challenge about the development of tourism services and touristic offer and demand, leading to the prosperity and development of the tourism sector in the country, the latter being one of the most important economic activities, but also to the satisfaction of tourists with regards to the services offered becoming more targeted and of high quality.

Acknowledgment. This work is carried out thanks to the support of the European Union through the PLAIBDE project of the FEDER-FSE operational program for the Nouvelle-Aquitaine region. The project is supported by aYaline company, with partners: LIAS-ENSMA laboratory in Poitiers, and the L3i laboratory at La Rochelle University.

References

1. Ontologie datatourisme v1.0 documentation de référence. Technical report, 01 2017. `https://www.datatourisme.gouv.fr`
2. J. Brookman, P. Rouge, A. Alva, and C. Yeung. Cross-device tracking: Measurement and disclosures. *Proceedings on Privacy Enhancing Technologies*, 2017(2):133–148, 2017.
3. D. K. Chiu, Y. T. Yueh, H.-f. Leung, and P. C. Hung. Towards ubiquitous tourist service coordination and process integration: A collaborative travel agent system architecture with semantic web services. *Information Systems Frontiers*, 11(3):241–256, 2009.
4. T. et TERRITOIRES. Ressources documentaires tourinfrance.
5. F. Soualah-Alila, M. Coustaty, C. Faucher, and R. Wannous. Projet tourinflux: Apport des technologies du web sémantique pour la gestion des données du tourisme. In *6ème édition du colloque pluridisciplinaire AsTRES: Association Tourisme Recherche et Enseignement Supérieur*, page 12, 2016.
6. F. Soualah-Alila, M. Coustaty, N. Rempulski, and A. Doucet. Datatourism: designing an architecture to process tourism data. In *Information and communication technologies in tourism 2016*, pages 751–763. Springer, 2016.
7. F. Soualah-Alila, C. Faucher, F. Bertrand, M. Coustaty, and A. Doucet. Applying semantic web technologies for improving the visibility of tourism data. In *Proceedings of the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval*, pages 5–10. ACM, 2015.
8. CDT : Comité Départemental du Tourisme. `https://fr.wikipedia.org/wiki/Comité_départemental_du_tourisme`