Chapter 6

The Web of Data and the Tourism Industry

Diego Berrueta
Fundación CTIC, Spain

Antonio Campos
Fundación CTIC, Spain

Emilio Rubiera
Fundación CTIC, Spain

Carlos Tejo
Fundación CTIC, Spain

José E. Labra
Universidad de Oviedo, Spain

ABSTRACT

The web of data is a new evolutionary step of the web that involves the publication, interchange and consumption of meaningful, raw data by taking full benefit of the web architecture. All the parties involved in the tourism industry should consider the opportunities offered by this new web. Entry barriers are low because existing data sources and documents can be easily leveraged to be part of this extended web. At the same time, new services and platforms that exploit the data are beginning to show the large potential for increased technological and business opportunities. A new scenario of large-scale information availability and efficient data flows is discussed in this chapter.

INTRODUCTION

The tourism sector has embraced the digital technologies at a number of levels. One of the most prominent ones is the web, which firstly was used as a world-wide showcase to publish and distribute information about tourism resources of interest, such as places, events and accommodations. Later, more advanced uses appeared, such as e-commerce (electronic ticket, room or flight availability check, direct room booking, to name a few), route planning (e.g., ticket brokers, flight
combinations), virtual visits, tourist feedback, interactive geospatial information systems, review portals, user-generated multimedia content or even encyclopaedias. All these information resources and digital services have empowered the prospective tourists to control many aspects of their trips. In this regard, the tourists have the potential to effectively become their own travel agents.

The web has been successful beyond expectations at creating an immeasurable large repository of instantaneously accessible and interlinked documents, applications and multimedia resources. However, some web researchers have realized that an even larger potential can be unleashed. The very inventor of the web in its current form, Tim Berners-Lee, champions the vision of augmenting it with data to produce a web in which data — and not just documents — are published, distributed and consumed electronically (Bizer et al., in press). This technological improvement has business consequences which will be discussed in this chapter. It is remarkable that the web of data, which is often viewed as the next generation of the web, is not intended to be a replacement of the current web of documents. The web of data does not render documents obsolete, because human beings feel more comfortable browsing documents, and therefore we will fuel the demand. Conversely, computing devices excel at processing data, but they can only exploit them when they are in their purest form, i.e., “raw” data. Up to this time, data in this degree of purity has been largely absent on the web, because textual documents contain “cooked” data woven into sentences. Even when data are isolated and structured in tables within documents, their meaning is only implicit in the context. Our brains can usually make sense of these data because we have learned to extract data from sentences and to derive the meaning from the context. Digital computing devices lack these abilities and, therefore, they cannot make use of the data embedded into a typical web document.

Constructing the web of data is, therefore, an effort to make large amounts of meaningful data available using the same technological foundations that have proven successful and scalable for the web of documents. These shared core technologies are URLs (a syntax for coining uniform identifiers), HTTP (a protocol to transfer documents over the network) and XML (a meta-language to define documents and data structures). Besides those, other pieces are specific for the web of data, such as the framework which will be described later. However, in this chapter a rather high level is maintained with respect to these technologies. Those readers interested in gaining a deeper understanding of the technical details are referred to the “Further reading” section. Similarly, the jargon of the theoretical foundations rooted in Logics (e.g., “ontology”) is also purposely omitted from this chapter for the sake of simplicity.

The web of data is more than a promising vision: for instance, it is widely used by news feeds. Most web sites are producers of news items: from traditional media portals to personal blogs, from new product announcements to releases of new multimedia content in an e-learning course. Keeping track of the updates of each one of these web sites in the traditional fashion requires the users to regularly visit them with their browsers, and to repeatedly identify the new pieces of content. Fortunately, the RSS and Atom formats allow the exchange of machine-readable information, and consequently, the creation of news “feeds”. Feed aggregation tools (aggregators) periodically gather the information from different providers. Moreover, aggregators automatically separate the new content from the old one, and organize it according to the users’ preferences. RSS and Atom are the forefront of the web of data applied to the news items, and are an evidence of how simple technologies built on top of the current web infrastructure (HTTP and XML) can dramatically improve the users’ experience by enabling a higher degree of automation in information management.

In the following pages, the application of the web of data to the tourism domain is discussed. No previous knowledge of the web of data is as-