Inf Technol Tourism (2015) 15:101–130
DOI 10.1007/s40558-015-0023-2

ORIGINAL RESEARCH

Business intelligence for cross-process knowledge extraction at tourism destinations

Wolfram Höpken[1] · Matthias Fuchs[2] · Dimitri Keil[2] · Maria Lexhagen[2]

Received: 13 August 2014 / Revised: 20 March 2015 / Accepted: 21 April 2015 /
Published online: 6 May 2015

Abstract Decision-relevant data stemming from various business processes within tourism destinations (e.g. booking or customer feedback) are usually extensively available in electronic form. However, these data are not typically utilized for product optimization and decision support by tourism managers. Although methods of business intelligence and knowledge extraction are employed in many travel and tourism domains, current applications usually deal with different business processes separately, which lacks a cross-process analysis approach. This study proposes a novel approach for business intelligence-based cross-process knowledge extraction and decision support for tourism destinations. The approach consists of (a) a homogeneous and comprehensive data model that serves as the basis of a central data warehouse, (b) mechanisms for extracting data from heterogeneous sources and integrating these data into the homogeneous data structures of the data warehouse, and (c) analysis methods for identifying important relationships and patterns across different business processes, thereby bringing to light new knowledge. A prototype of the proposed concepts was implemented for the leading Swedish mountain destination Åre, which demonstrates the effectiveness of the proposed business intelligence architecture and the gained business benefits for a tourism destination.

Keywords Business intelligence · Knowledge extraction · Data mining · Data warehousing · Management information systems · Tourism destinations

✉ Wolfram Höpken
wolfram.hoepken@hs-weingarten.de

[1] Business Informatics Group, University of Applied Sciences Ravensburg-Weingarten, Doggenriedstr., 88250 Weingarten, Germany

[2] European Tourism Research Institute (ETOUR), Mid-Sweden University, Kunskapens Väg 1, 83125 Östersund, Sweden

1 Introduction

The competitiveness of tourism destinations strongly depends on how information needs of stakeholders are satisfied by information and communication technologies (Buhalis 2006; Back et al. 2007). Vast amounts of data on customers, products, and competitors are electronically available in tourism destinations. For example, web-servers store tourists' website navigation, computer reservation systems (CRS) save bookings and customer profiles, property management systems (PMS), and destination management systems (DMS) store tourism offers and supplier information. However, these valuable knowledge sources are rarely utilized in tourism destinations (Pyo 2005; Höpken et al. 2011, p. 417). Consequently, managerial expertise and organisational learning in tourism destinations can be significantly enhanced by applying state-of-the-art methods of *business intelligence* (BI). BI offers reliable, up-to-date and strategically relevant information about tourists' travel motives and service expectations, channel use and related conversion rates, booking trends, and estimates about the quality of the service experience and value-added per guest segment (Min and Emam 2002; Ritchie and Ritchie 2002; Sambamurthy and Subramani 2005; Wong et al. 2006; Höpken et al. 2014; Fuchs et al. 2014).

Various methods of business intelligence have been applied in the field since the early stages of ICT adoption in tourism. Most ICT systems used in tourism today (e.g. CRS, PMS or DMS) offer business intelligence functionalities, like reporting or OLAP (online analytical processing). In addition, approaches have been developed to extract knowledge from existing data that foster knowledge-based decision making. For example, in 1988 American Airlines developed DINAMO, a yield management and dynamic pricing instrument based on demand forecasts and cancellation/no-show predictions (Smith et al. 1992). A number of subsequent approaches have been developed and implemented including TourMIS (Wöber 1998), MANOVA WEBMARK (Kepplinger 2006), Destinometer™ (Fuchs and Weiermair 2004), Destimetrics, T-stats, IIQ-Check, and TrustYou. These approaches are described in more detail in the following section.

Although the examples cited above do support knowledge extraction and decision making in tourism, their focus is typically limited to very specific aspects, or business processes, like handling reservations & bookings, tourist arrivals, tourism offers, or customer feedback. Thus, they lack a cross-process analysis approach with the capability to look at all strategically relevant business processes at the same time. Cross-process analyses intend to interlink different business processes based on some common characteristics (e.g. they are initiated by the same customer, or deal with the same product). This approach enables the identification of important patterns and trends across different processes, such as the relationship between web search and booking behaviour and customers' feedback and satisfaction. It is important to stress that within the business intelligence and data warehousing domain, business processes are typically not modelled as a flow of activities but on a more abstract level as overall processes, while their final output is modelled in the form of performance indicators (Kimball et al. 2008; Vela et al. 2012). In the case of the booking process, for example, only performance indicators,

like the turnover or the number of persons, together with characteristics, like the customer, the booked product, the booking date, etc., are relevant for analysis purposes but not the way how the booking is handled and executed by a flow of different activities. Consequently, cross-process analyses look at the patterns and relationships between those performance indicators across different processes, but do not try to interlink the flow of activities within the business processes.

This paper proposes a novel approach for cross-process knowledge extraction and decision support for tourism destinations, thus overcoming the limitations of current approaches briefly cited above. A technical architecture enables the extraction and integration of heterogeneous data from all relevant source systems or data sources at the level of tourism destinations. A newly developed homogeneous and comprehensive data model enables the integration of data from different business processes into a central, process-overarching destination data warehouse. Specific business intelligence-based analysis approaches (i.e. dashboards, OLAP analyses and data mining methods) facilitate the identification of relationships and patterns across different business processes and the extraction of previously unknown and unavailable knowledge. This approach enables managers to get responses to relevant questions like: Can web navigation behaviour be used to predict bookings or arrivals? Can customer segments be identified by a specific and meaningful relation between web navigation behaviour, booking behaviour and customer satisfaction, which may serve as input to product optimization and personalization?

The paper is structured as follows: Sect. 2 discusses past and current approaches to business intelligence in the domain of travel and tourism. Section 3 presents the proposed overall architecture for applying methods of business intelligence to tourism destinations, the *knowledge-based destination architecture*, as well as the corresponding technical framework. Section 4 describes the core aspect of the proposed approach. This description focuses on the multi-dimensional data model for the central destination data warehouse, which enables the integration of data from heterogeneous data sources and different business processes and, thus, supports powerful cross-supplier and cross-process analyses. Section 5 describes mechanisms for extracting data from different data sources and integrating them into the homogeneous data warehouse, which includes mechanisms for structured as well as unstructured data (e.g. customer reviews in social media). Section 6 presents cross-supplier as well as cross-process analyses demonstrating the power and flexibility of analyses based on the proposed multi-dimensional data warehouse model. Analysis examples are shown which are based on the knowledge-based destination architecture prototypically implemented at the leading Swedish mountain destination Åre. The paper concludes by summarizing the most important results and by providing an outlook on consecutive research activities.

2 Related work

Since the advent and widespread adoption of computerized reservation and booking systems in the 1980s, vast and comprehensive databases have been available for all types of tourism transactions related to customers' booking and consumption

behaviour, respectively [e.g. Passenger Name Record (PNR) databases of global distribution systems (GDS), the Airline On-Time Performance database of the Bureau of Transportation Statistics (http://www.transtats.bts.gov/), etc.].

In particular, airline companies began to analyse customer transaction data as input to process and product optimization. The most prominent application areas of business intelligence (BI) in the airline industry comprise demand forecasting (Subramanian et al. 1999) and the prediction of customers' cancellation behaviour and no-shows (Garrow and Koppelman 2004). A prominent example in the area of revenue and yield management in the airline industry is the DINAMO system introduced by American Airlines in 1988 (Smith et al. 1992). DINAMO builds on American Airline's GDS SABRE as data source, providing comprehensive information on all transactions, related to the areas *reservation/booking*, *cancellation* (*no-show*), and *offerings/resource management*. In order to reduce complexity, the yield management problem is broken down into the sub-problems *overbooking, discount allocation* and *traffic management*. The *overbooking* problem is solved by forecasting cancellations, no-shows, and over sale (i.e. compensation) costs, while a consecutive revenue optimization identifies the optimal overbooking level which equals the marginal revenue gained and over sale costs. The *discount allocation* problem is represented by a decision tree, based on demand predictions for multiple fare types, using exponential smoothing time-series techniques and a passenger-choice model reflecting customer reactions on schedule and price changes. Finally, *traffic management* handles the problem of single flights serving different markets based on connecting flights in a hub and spoke network, and is handled by clustering a multitude of market/fare combinations into a limited and manageable number of similar-valued groups, called *buckets* (Smith et al. 1992).

Early applications of BI can also be found in the area of tourism destinations and the hospitality industry. A common example is the Austrian tourism marketing information system TourMIS (Wöber 1998), offering market research information and decision support for tourism destinations and stakeholders. Based on a homogeneous data model for tourism arrivals, overnight stays and visits at tourism attractions, TourMIS collects data directly from destination management organisations by a manual data input process, restricting data granularity to mostly yearly, or in some cases monthly, aggregates. TourMIS supports descriptive (i.e. OLAP-like) analyses of tourism performance indicators like arrivals, overnights or visits aggregated on the level of tourism destinations, regions, countries, or customer characteristics like sending country. In addition, trend analysis techniques and prediction models are applied in order to identify seasonal or long-term trends and to predict future tourism demand or guest mix changes.

The Tyrolean (Austria) benchmarking tool Destinometer™ analyses representative survey data on customers' satisfaction with the destination offer (e.g. accommodation, gastronomy, animation, wellness, sport, shopping, etc.), thereby offering various benchmarking functions. The first analysis approach supplements and combines these data with data on customer price satisfaction, thus, showing the perceived value-for-money along the major destination value-chain areas (Fuchs 2004a). The second analysis approach utilizes Kano's (1984) factor structure model of customer satisfaction and employs Brandt's (1988) dummy-based regression

method to identify those destination activities and value-chain areas with the highest relative potential to delight the customer (Fuchs and Weiermair 2004). The third and final analysis approach adds supply-side destination data, such as output data (e.g. overnight stays, price levels for the various accommodation categories) along with input data (e.g. destination resource data such as the bed base, marketing costs, cost for energy, water and recycling, as well as aggregated wages for tourism personnel). By employing a data envelopment analysis (DEA), the relative efficiency level of the destination is determined, and optimal strategies to enhance customer satisfaction can be deduced, which in turn, also improve the aggregated level of destination efficiency (Fuchs 2004b; Fuchs and Höpken 2005; Weiermair and Fuchs 2007).

MANOVA WEBMARK (Kepplinger 2006), a management information system for Austrian tourism stakeholders, supports tourism destinations, accommodation providers, attraction providers and ski lift operators in their operative and strategic decision making process. Tourism indicators, like arrivals, overnights, visits, and passengers/transportations, as well as guest feedback and satisfaction are collected, either manually on a yearly or monthly aggregation level, or by online surveys. MANOVA WEBMARK supports the analysis of guest satisfaction (based on guests' demographic characteristics, travel motives and consumption behaviour), performance indicators and trends, benchmarking as well as strategic analyses, like SWOT analyses, or importance/performance analyses (IPA), respectively.

DestiMetrics (http://www.destimetrics.com) supports performance analyses and decision making for tourism destinations and accommodation providers in the United States and Canada. Reservation data on different accommodation types (i.e. hotel and non-hotel facilities) are imported from property management companies and vacation rental units on a monthly basis, enabling detailed analyses of past and upcoming arrivals and overnights. DestiMetrics offers performance indicators, like occupancy rate, daily average room rate, or revenue per available room (RevPAR), interlinks them with contextual factors influencing tourism demand, like holiday information, and offers benchmarking functionalities for tourism suppliers within as well as between tourism destinations.

t-stats (http://www.t-stats.co.uk), a management information system (MIS) for tourism destinations, supports descriptive analyses and benchmarking functionality in the areas of accommodation (i.e. indicators, like occupancy rates, average room rate, RevPAR, etc.), attractions (i.e. indicators, like the number of visitors, expenditures per visit, etc.), general tourism statistics (e.g. arrivals, expenditures, car parking, visitors of information centres, visits to events and festivals, weather data, exchange rates, etc.), customer feedback and satisfaction (based on customizable surveys), and website hits (i.e. web navigation behaviour). Source data are mainly entered manually by tourism stakeholders or destinations on a monthly or daily aggregation level.

To summarize, all existing major BI and data mining techniques have been applied in the tourism domain. Descriptive/explorative analyses (EDA) are used in the form of dashboards or OLAP, e.g. to visualize tourism arrivals, bookings, or visits per dimensions, like time/season, travel type, or customer origin (cf. existing systems described above). Methods of supervised learning (e.g. classification,

estimation and prediction) are used to explain tourists' booking, cancellation and consumption behaviour (Morales and Wang 2008), or to predict tourism demand (Law 1998; Chu 2004; Vlahogianni and Karlaftis 2010). As a method of unsupervised learning, clustering is one of the most heavily used data mining techniques in tourism and is typically applied to the task of customer segmentation as input to product positioning and differentiation, dynamic pricing or customer relationship management (Bloom 2004).

With growth of the World Wide Web, the topic of web data mining has gained particular attention in tourism. *Web usage mining* deals with the analysis of tourist behaviour while using online platforms or websites. Although current applications are focussing on descriptive analyses, also supervised and unsupervised learning techniques have recently been applied in the tourism domain. These include: customer segmentation for website adaptation and product recommendation (Wallace et al. 2004; Pitman et al. 2010), or sequential association rule mining for click-stream analysis (Jiang and Gruenwald 2006). *Web content mining* denotes the analysis of content from tourism online platforms and websites. On one hand, such content mining deals with the extraction of knowledge on tourism markets and offers (Walchhofer et al. 2010). On the other hand, and even more prominently, such mining deals with the analysis of user generated content (UGC), like tourists' comments in blogs or review platforms (Bronner and Hoog 2011; Lexhagen et al. 2012; Kuttainen et al. 2012). Methods of text mining, which are typically based on statistical or linguistic approaches, are applied to feedback aggregation, opinion mining or sentiment detection (Kasper and Vela 2011; Gräbner et al. 2012; Schmunk et al. 2014). Especially tourism destinations and accommodation providers can benefit from monitoring, collecting and analysing UGC. Thus, different software tools are available and already in use by tourism stakeholders. Trackur (http://www.trackur.com) and Alterian SM2 (Laine and Frühwirth 2010) are prominent examples of comprehensive social media monitoring and analysis tools. Social Mention (http://www.socialmention.com) is a social media search engine, focussing on real-time aggregation of social media content and point-in-time social media search. Tweettronics (http://www.tweettronics.com) enables the tracking of words and phrases on Twitter and to execute competitive and trend analyses of product mentions and customer sentiments. Kuttainen et al. (2012) evaluated tools and methods for collecting UGC related to the leading Swedish mountain destination of Åre and especially the current attitude of destination managers and stakeholders. The authors argue that destination stakeholders certainly make use of software tools for analysing UGC, but still lack a well-structured and efficient analysis approach.

The considerations given so far clearly show that, in general, BI techniques are used by all tourism stakeholders. However, existing BI applications, especially on the level of tourism destinations, lack a comprehensive (i.e. overarching) approach considering a customer perspective and associations between business processes. Existing approaches typically focus only on a subset of strategically relevant business processes and, thus, lack a systematic integration of data stemming from different business processes, and, consequently, do not offer cross-process analyses.

A major obstacle to integrate data stemming from different business processes and corresponding source systems is the absence of a homogeneous, process-overarching data warehouse model. Although approaches for modelling business areas, like retail and sales, procurement, accounting, human resource, E-commerce, etc., exist, an overall and general data warehouse model for the tourism domain (and for tourism destinations in specific) is still missing (Kasavana and Knutson 1999). Several initiatives for homogeneously modelling data and processes in tourism took place in the past or are still active, e.g. UN-EDIFACT TT&L (http://www.unedifact.org), ANSI ASC X12I TG08 (http://www.x12.org), IFITT RMSIG (Höpken 2004), Harmonise (Dell'Erba et al. 2005), and OpenTravel (http://www.opentravel.org). However, these initiatives focus on operative aspects and, thus, do not consider specific aspects and needs of a data warehouse and consecutive analyses as input to business intelligence-based decision support.

In contrast to existing applications, the novelty of the approach presented in this paper lies in the combination of all business processes strategically relevant for a tourism destination and in the integration of detailed transaction data from different types of data sources, thereby enabling powerful cross-process analyses. In order to reach this objective, the underlying approach is separating between process- or transaction-dependent data elements (i.e. variable data) and transaction-independent data elements (i.e. master data), following the principle of multi-dimensional modelling (MDM) (Kimball et al. 2008). Master data (i.e. dimensions of the MDM, like customer data, product data, etc.) are defined homogeneously and process-independent. Variable data (i.e. facts of the MDM, like turnover or number of persons of a certain booking) are defined process-specifically and, if necessary, on different levels of granularity (e.g. on the level of single transactions, like single bookings or clicks on a website, or on the level of aggregated transactions, like overall bookings or complete web sessions). Consequently, different processes, even with a different granularity, can be imported into a central data warehouse, independent of each other, but at the same time cross-process analyses, i.e. interlinking transactions across different processes, can effectively be executed based on homogeneously defined master data.

3 Architecture of the knowledge-based destination

The conceptual foundation of the approach presented in this paper is the knowledge-based destination architecture. This architecture consists of a knowledge generation layer and a knowledge application layer (Fig. 1; Höpken et al. 2011). Accordingly, the related activities deal with the extraction of information from different customer and supplier-based data sources as well as the generation of strategically relevant knowledge which is applied in the form of intelligent services for both customers and suppliers (i.e. destination stakeholders) (Sambamurthy and Subramani 2005).

Specifically, the *knowledge generation* layer extracts information from heterogeneous data sources and makes destination-specific knowledge available to tourists and destination suppliers. On the customer side, content is available in the form of tourists feedback (e.g. generated by guest surveys, review platforms, etc.),

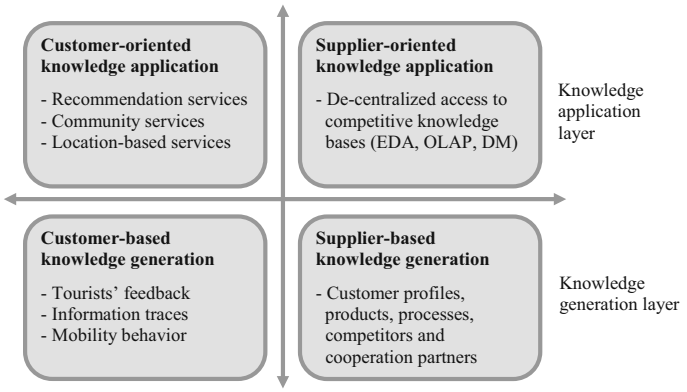


Fig. 1 The knowledge-based destination architecture (Höpken et al. 2011)

information traces (e.g. generated by booking systems, online platforms, etc.; Pitman et al. 2010) or mobility behaviour (e.g. generated by mobile applications or customer cards; Zanker et al. 2010). On the supplier side, knowledge about customers (customer profiling), products, processes, competitors and strategic partners can be extracted from existing data sources or websites (e.g. in the form of destination profiles or availability information) (Ritchie and Ritchie 2002; Gretzel and Fesenmaier 2004; Pyo 2005).

The *knowledge application* layer provides knowledge-based services for customers as well as destination suppliers and stakeholders. *Customer-oriented knowledge application* comprises, for example recommendation or community services. Recommendation services offer functionalities, like recommending products and services based on knowledge about customer preferences and consumption behaviour as well as supplier offers and market structure or automatically pushing context-sensitive messages to tourists (Höpken et al. 2008). Community services build on knowledge about customer behaviour and for example suggest tourists promising social interaction and joint activities during their destination stay. In contrast, *supplier-oriented knowledge applications* mainly fall into the category of management information and decision support systems (Cho and Leung 2002; Olmeda and Sheldon 2002). Explorative data analyses (EDA), online analytical processing (OLAP) and data mining (DM) allow for the (ad-hoc) generation, management and access to strategically relevant knowledge for the DMO as well as private and public destination suppliers (Fuchs and Höpken 2009).

In this study, the technical representation of the different components of the knowledge-based destination architecture described above focuses on customer-based knowledge generation and supplier-oriented knowledge application. Figure 2 illustrates all components of the technical architecture of the knowledge-based destination.

To start with, the knowledge generation layer includes:

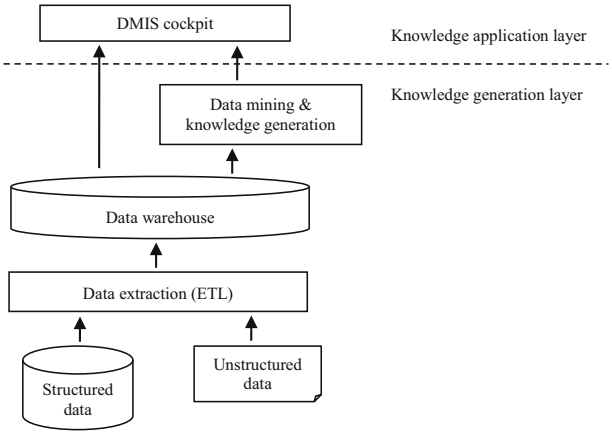


Fig. 2 Technical architecture of the knowledge-based destination

1. different types of *data sources* (e.g. reservation and booking data, web navigation data, customer feedback data) in the form of *structured* and *unstructured* data sources
2. the process of *data extraction* (ETL—extraction, transformation and load), refers to the extraction of relevant data from different data sources, transforming source data into a homogeneous (i.e. unified) data format appropriate for further analyses as well as storing/loading the data into the data warehouse
3. the *data warehouse* as a central destination data store that embraces data related to all different business processes and tourism stakeholders as basis for a destination wide and all-stakeholder and business process encompassing analysis approach
4. methods of *data mining and knowledge generation* to generate relevant knowledge for destination suppliers and the destination management by employing techniques of machine learning and artificial intelligence.

By contrast, the knowledge application layer provides the destination management information system (DMIS)—a "cockpit" that enables access to data and knowledge stored in the central data warehouse as well as the opportunity to execute specific data analyses as a means to provide decision support to destination stakeholders and managers.

4 The multi-dimensional destination data warehouse model

At the core of the knowledge-based destination architecture is a central data warehouse that embraces data related to all different business processes and tourism stakeholders (Cho and Leung 2002). Heterogeneous data from different data sources are mapped into a homogeneous data format and stored in a central data warehouse. Only through this harmonisation and integration process is it possible to carry out a

destination wide and all-stakeholder and business process encompassing analysis approach (Pyo et al. 2002). The central concept behind the data warehouse is the multi-dimensional destination data warehouse model, which is described in the following sections.

4.1 Multi-dimensional modelling

Compared to an operational database, a data warehouse is theme-oriented, time-oriented (i.e., considers periodic updates), integrated (i.e., aggregates data from different data sources) and invariant (i.e., new data are appended but existing data never changed) (Inmon 2002). Two basic approaches for modelling a data warehouse exist: multi-dimensional modelling (Kimball et al. 2008) and normalized modelling (Inmon 2002). Multi-dimensional modelling (MDM) is a business process oriented design framework, mainly differentiating between performance indicators of a business process, called *facts* (e.g. the turnover or person number of a booking), and the context of the business process execution divided into different context *dimensions* (e.g. the date and time of a booking, the booked product or the customer). A multi-dimensional model is then composed of a fact table and several dimension tables (cf. Fig. 3).

Facts, typically, show numeric and additive characteristics, which can be accumulated along a dimension (e.g. the turnover accumulated per month or per year). Thus, MDM is business process or transaction oriented. Each single MDM diagram models one single business process. Accordingly, a master MDM for a large company (or a tourism destination) may consist of 10 or even more than 20 single MDM diagrams. In turn, each MDM diagram can comprise of only a few or up to 15 or more dimensions. The advantage of the *multi-dimensional modelling* approach (Kimball et al. 2008) is its relatively simple database design that supports powerful analyses. By contrast, the advantage of the *fully normalized modelling* approach (Inmon 2002) is better support for data integration, due to reductions in redundancies and simplified processes for identifying inconsistencies, especially if inconsistencies should be solved in the original operational databases as well. Often, both approaches are combined into a two-layer data model with a normalized data structure in order to foster data integration, and a multi-dimensional data structure, generated from the normalized data structure, to support data analyses and OLAP (Chaudhuri and Dayal 1996).

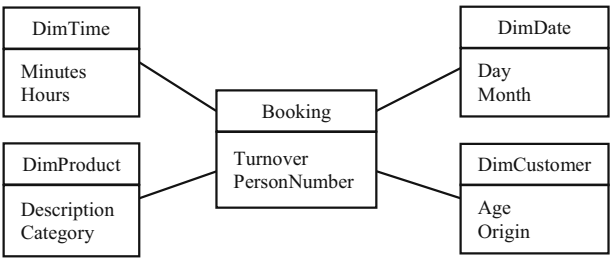


Fig. 3 Multi-dimensional model of booking process

The concept of a normalized data model is especially relevant for companies aiming to overcome inconsistencies across different operational databases as well as correcting data in operational databases as a means of master data management. However, in the case of tourism destinations, eliminating inconsistencies between source systems is not of high priority, or even unrealistic, due to the relatively strong independence of stakeholders in a tourism destination. Consequently, for tourism destinations, the direct integration of source data into a homogeneous multi-dimensional structure can be considered as the most appropriate approach.

4.2 Conformed dimensions

The integration of data from different source systems, like various CRS from different destination stakeholders, enables stakeholder or supplier overarching analyses and benchmarking (e.g. comparing booking figures of different suppliers, or web navigation behaviour on different websites). Equally important, though, is the possibility to execute business process overarching analyses. This requires that all concepts (i.e. dimension characteristics, common to several processes), are defined homogeneously and process-independent (e.g. concepts like customer or tourism product which appear in different processes, or like booking or customer feedback).

Multi-dimensional data models (MDM) are often falsely associated with negatively connotated concepts, like *data marts*, representing stove-pipe solutions by solely taking into consideration one business process or even only requirements of one department or, in the case of tourism destinations, one stakeholder. In this study, the above issue is addressed through the pivotal concept of *conformed dimensions*, which was first introduced by Kimball (1997). Thus, dimensions overlapping between several processes (e.g. time, date, product, or customer) are defined process independently. Single business processes are then modelled by using such conformed dimensions instead of defining them for each process again (cf. Fig. 4).

Making use of conformed dimensions within different processes then enables cross-process analyses (e.g. the analysis of sending-country specific correlations between web navigation behaviour, booking behaviour or customer feedback) based on the customer's origin as an overlapping (i.e. conformed) dimension characteristic. An important concept related to conformed dimensions is the hierarchical abstraction of dimension characteristics, which is used to reach the maximum degree of overlap between context characteristics among different processes. A typical example is a hierarchical product categorization, which enables matching products between different processes, at least on an abstract level (e.g., summer/winter activities or indoor/outdoor activities, instead of bicycling, hiking, skiing, etc.). Thus, dimension hierarchies do not only support OLAP analyses by enabling drill-down or drill-up along the hierarchy, but also the flexible data integration of heterogeneous source data.

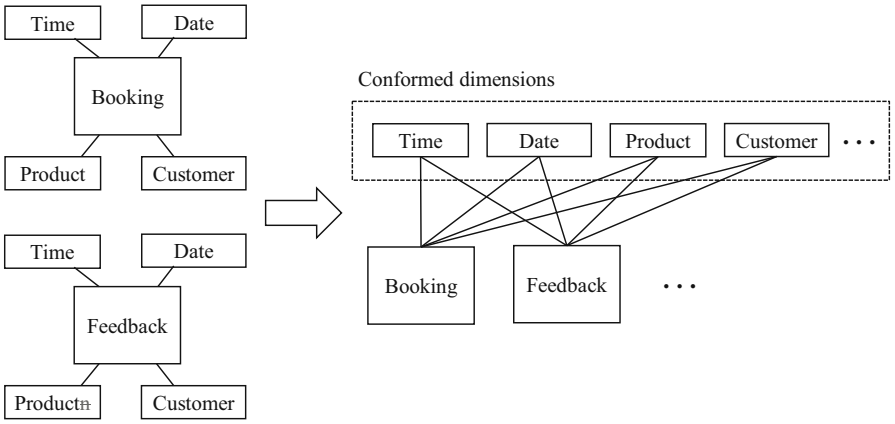


Fig. 4 Conformed dimensions within a multi-dimensional data model

4.3 The destination data warehouse model

Requirements for a comprehensive destination data warehouse model have been defined based on a literature review (Ritchie and Ritchie 2002; Pyo et al. 2002; Cho and Leung 2002; Wang and Russo 2007; Bornhorst et al. 2010; Chekalina et al. 2014; Fuchs et al. 2013, 2014). Furthermore, qualitative input from stakeholders of the Swedish mountain destination Åre was collected within a series of requirement definition workshops. The requirements (i.e. the requested business indicators) fall into three categories: *economic performance* (e.g. bookings, overnights, prices, turnover), *customer behaviour* (e.g. web page views and sessions, booking channels, conversion rates, cancellations), and *customer perception and experience* (e.g. brand awareness, guest satisfaction and loyalty). All requested indicators are next assigned to the business processes by which they are generated (i.e. measured). In addition, corresponding OLAP and data mining analyses have been defined as part of the requirement definition process. OLAP analyses define relevant context dimensions for indicators (e.g. bookings per time period, per customer, etc.) and, thus, serve as a basis for defining the multi-dimensional destination data warehouse model.

Table 1 presents business processes and corresponding dimensions defined by the multi-dimensional data model for a destination data warehouse. The fact type specifies whether the process is represented as (a) a simple transaction (e.g. an information request, or a click on a website), (b) a periodic snapshot, or (c) an accumulated snapshot, covering different phases of a process. In the case of the fact type *transaction* (T), the facts of the process are (performance) indicators recorded when the corresponding transaction took place (e.g. the feedback value given within a customer feedback). In the case of a *periodic snapshot* (PS), the facts are recorded periodically (e.g. the available capacity of an accommodation provider, which is recorded daily). In the case of an *accumulated snapshot* (AS), the facts are recorded for different steps or milestones of an overall business process. The booking

Table 1 Processes and dimensions of the destination data warehouse

Business process	Fact type	Dimensions								
		Time	Date	Customer	Cust. usage profile	Cust. demographic profile		Product	Vendor	Supplier
Information request	T	x	x	x	x	x		x		x
Web navigation	T	x	x	x	x	x		x		x
Booking	AS	x	x	x	x	x		x	x	x
Stay	AS		x	x	x	x				
Consumption	T	x	x	x	x	x		x	x	x
Location tracking	T	x	x	x	x	x				
Feedback	T	x	x	x	x	x		x	x	x
Capacity	PS		x					x	x	x
Marketing activity	T		x					x	x	x

Business process	Fact type	Dimensions						
		Channel	Location	Feedback	URI	Session	Survey	Marketing
Information request	T	x						
Web navigation	T				x	x		
Booking	AS	x						
Stay	AS		x					
Consumption	T		x					
Location tracking	T		x					
Feedback	T	x	x	x			x	
Capacity	PS							
Marketing activity	T	x						x

process, for example, consists of two steps, the booking itself and the cancellation of the booking. When the booking is executed, facts, like booking price and number of persons, are recorded. When the booking is cancelled later on, cancellation information is recorded. Thus, modelling bookings occurs as a two-step process, instead of two separate processes, thereby preserving their interdependency and simplifies corresponding analyses.

The multi-dimensional data model for tourism destinations defines the following business processes:

- *Information request*: customer information requests about the destination (e.g. products, prices, etc.) issued via a destination website, third-party websites, travel agencies, or by phone or email. Dimensions are time and date of the request, the requesting customer (together with his/her usage and demographic

profile), the corresponding product and supplier, and the used channel (e.g. web, phone, etc.).

- *Web navigation*: detailed customer online navigation behaviour on the level of single page views (URI). Dimensions are time and date of page view, customer, product (and its supplier) related to the viewed page, session and URI.
- *Booking*: customer bookings of single products, together with change or cancellation status (thus, constituting an accumulating snapshot). Dimensions are time and date of booking, customer, booked product (together with vendor and supplier) and booking channel (e.g. phone, web, etc.).
- *Stay*: individual overnights and stays in a tourism destination (on a per-day basis), covering also same-day visitors, or tourists arriving without prior booking. Dimensions are date of stay, customer and the concrete location of stay.
- *Consumption*: consumptions of single products or services of any kind (e.g. food and beverage), on a per-day and per-person basis, where appropriate. Dimensions are time and date of consumption, consuming customer, consumed product (together with vendor and supplier) and the concrete location of consumption.
- *Location tracking*: movements of customers within the destination. Dimensions are time and date, customer and the reached location (e.g. point of interest or rastered GPS coordinates).
- *Feedback*: structured and unstructured customer feedback, like ratings, comments, etc. Dimensions are time and date, customer, related product (together with vendor and supplier), channel used (e.g. web, offline survey), concrete location the feedback relates to, a description and categorization of the question or topic the feedback is given to (dimension *Feedback*) as well as the corresponding overall survey, if appropriate.
- *Capacity*: provided capacity (by number of generally existing units) of products or services on a per-day basis (periodic snapshot). Dimensions are the date and offered product (together with the supplier, i.e. service provider or producer, like hotel or ski equipment manufacturer, and vendor, i.e. seller or intermediary, like ski shop or ticket office).
- *Marketing activity*: marketing activities and corresponding investments executed by the destination. Dimensions are the time period (*date*), related product (together with vendor and supplier), marketing channel and characteristics of the marketing activity, like campaign name, type, etc.

Most dimensions in Table 1 are used by several business processes and, thus, such processes share (the same) *conformed dimensions*. One customer, for example, who typically executes several processes, like web navigation, booking, consumption and feedback, is stored only once in the customer dimension table and can then be easily identified across different processes. In these ways, *conformed dimensions* interlink different processes, thereby enabling cross-process analyses, like identifying interesting and previously unknown relationships between customers' booking/consumption, web navigation and feedback behaviour, respectively (Kimball et al. 2008).

5 Data extraction and integration

Since the uptake of CRS/GDS in the 1960s, a major part of tourism transactions are handled electronically. With the rapid growth of the WWW this portion further increased. Nowadays customers leave electronic footprints during all travel-related activities such as searching and trip planning, reservation and booking, service consumption (if based on mobile services or loyalty programmes, like electronic customer cards) and post-trip activities in community web sites and review platforms. Thus, vast amounts of data on customer transactions (e.g. customer inquiries, bookings, payment processing), customer needs and behaviour are typically available for tourism destinations. Table 2 provides a systematization of data sources considered in the proposed knowledge-based destination architecture.

The study at hand focuses on customer-based knowledge generation, using data collected mainly during the pre- and post-trip phase. Correspondingly, data integration focuses on the business processes *web navigation*, *booking* and customer *feedback*. This section discusses relevant data sources and techniques of information extraction and integration appropriate for extracting relevant information and transformation into a structure suitable for storage in the central destination data warehouse as input for consecutive OLAP analyses and data mining.

5.1 Data extraction

The most important requirement for the step of data extraction is the support of all possible data sources and data formats (cf. Table 2), which can technically be differentiated into formats for structured and unstructured data. *Structured data* comes in formats like text files (e.g. CSV-files), databases, application-specific formats (e.g. SPSS files, MS Excel files, etc.), or XML files. In this study, structured source data are available for the web navigation process as web server log files, for the booking process as databases or MS Excel files, and for the feedback process, survey data are available as SPSS files. *Unstructured data* can take different formats, like semi-structured html documents, free text or even images. Methods for extracting data from unstructured data sources vary quite widely. Data can be extracted from html documents by means of wrappers, either created manually based on static patterns or

Table 2 Potential data sources of the knowledge-based destination architecture

Intentionally provided explicit tourists' feedback	Unintentionally provided implicit tourists' information traces
Structured data: e.g. online and offline guest surveys, ratings from web 2.0 applications, user profiles from web applications and online communities, etc.	*Navigation data*: search behaviour on web sites and online portals, community sites, etc.
Unstructured data: free text from E-mails and web 2.0 applications (e.g. blogs, e-comments/reviews), rich content (e.g. YouTube.com), etc.	*Transaction data*: online requests, reservations and bookings, payments, etc.
	Tracking data: GPS/WLAN-based coverage of tourists' spatial movements
	Observation data: gathered in a laboratory context or through market observation

(semi-)automatically generated by means of (un-)supervised learning methods (Liu 2008). Free text is stored in the data warehouse as it is or transformed into structured data by means of linguistic approaches or statistical language models (e.g. word vectors with TF-IDF weights) (Manning and Schütz 2001). This study deals with unstructured data in the form of user feedback, either provided as part of customer surveys (free text answers) or as product reviews or comments on social media platforms (e.g. Tripadvisor and Booking.com). Since the number of relevant social media platforms is limited and their structure is relatively fixed, product reviews or comments are extracted using simple wrappers, based on static patterns. Free text, either stemming from customer surveys or extracted from social media platforms, is stored in the data warehouse as text. This facilitates them to be shown to the user as they are, or to further process them by text mining techniques in order to classify reviews into topics or sentiments (Schmunk et al. 2014).

5.2 Data integration

Data extracted from different data sources are integrated into the central destination data warehouse. Therefore, heterogeneous source data have to be transformed into the homogeneous data format of the data warehouse. *Webserver logfiles* typically follow a standardized structure (e.g. the common logfile format or the extended logfile format; http://www.w3.org) and, thus, data heterogeneity is not an issue. The most important and, unfortunately, most work-intensive integration task (done manually during system setup) is the mapping of single URI requests to more abstract categories, for example, the function executed by clicking on the webpage (e.g. information request, booking) and the products the webpage is related to (e.g. accommodation, food and beverage, events, sightseeing). Since single URIs are on a much too high level of granularity, such mapping is indispensable for meaningful and especially website- and supplier-overarching analyses.

In the case of *booking data* stemming from different booking systems, transforming heterogeneous data into a homogeneous structure is a comparatively complex task. Automatic or semi-automatic mapping techniques, based for example on approaches of schema matching (Liu 2008), do not reach a sufficient accuracy in the case of complex booking data. Therefore, in the presented study such mappings are defined manually for each data source at hand. Typical issues are the mapping of different age groups, customer types, booking channels, or even product types, into a homogeneous format with minimal data corruption. In the case of *survey data*, concrete questions of customer surveys are mapped to product and feedback categories (i.e. awareness, loyalty and service quality/satisfaction with sub-categories reliability, variety/choice, family friendliness, cleanliness, value for money, etc.). This enables survey- and supplier-overarching analyses although the underlying concrete questions do not fully correspond.

After integrating data from different data sources, data entries representing the same real-world entity have to be identified (typically referred to as *record linkage;* Kimball et al. 2008). In this study, duplicated entries only occur for bookings, caused by a destination-wide booking system that aggregates bookings of supplier-specific booking systems, as well as for customers executing a booking. Duplicated

bookings are successfully identified by a reference to the original booking system and the exact booking time. Single customers within a booking system are simply identified by a unique customer identifier. As global unique customer identifiers do not exist across data from different booking systems, a collection of *key attributes* is defined, uniquely identifying customers across different data sources (i.e. mainly first name, last name, birthdate and origin). Although this approach works satisfactory, duplicated customers cannot be completely avoided. More complex approaches for record linkage, like fuzzy matching, rule-based approaches, or even machine-learning techniques, have not been used within this study (Kimball et al. 2008).

In the case of the business processes *web navigation* and *feedback*, necessary information to identify a single customer are missing in the available source data (i.e. webserver logfiles or customer surveys). Therefore, record linkage is impossible, or put differently, not an issue. If the available information on customers (or other objects, like products) are incomplete and unique objects cannot be identified, the concept of *conformed dimensions* with attributes on different abstraction levels is a key concept to support cross-process and cross-supplier analyses (cf. Sect. 5.2).

6 Cross-process data analyses

In general, data analyses can take place on three different levels, related to the underlying processes and data. On the *data source/supplier level,* each supplier (or destination stakeholder) can execute analyses on his own data sources (e.g. bookings stemming from their own CRS, or clicks and sessions on their own website, etc.). On the *process level,* destination stakeholders can compare their performance related to the same business process of others and identify interesting patterns between and across different destination stakeholders (*cross-supplier* analyses). Finally, on the *cross-process level,* destination stakeholders can execute process-overarching analyses and identify interesting patterns and trends across processes and multiple stakeholders. For example, the relationship between web navigation behaviour (clicks and sessions) and booking behaviour (number of bookings, used booking channel, etc.) might reveal country-specific patterns and habits or even support the prediction of bookings based on web navigation behaviour.

Independent of these different abstraction levels, data analyses can be offered to the user by different analysis techniques or interfaces, for example, *dashboards* (as a predefined collection and graphical arrangement of the most prominent and relevant analyses), *OLAP* (online analytical processing—an interactive and flexible analysis technique, based on multi-dimensional data warehouse models), or *data mining*. In turn, *data mining includes* for example *classification* to explain the cancellation behaviour or the used booking channel, *estimation* to explain the main factors influencing customer satisfaction, *prediction* of future bookings and arrivals, *association rule analysis* to identify correlations between products booked together, or *cluster analysis* to identify homogeneous customer segments.

These different BI-based analysis techniques have been prototypically implemented within the DMIS cockpit for the leading Swedish mountain destination Åre (cf. Fig. 2). The prototype focuses on the three business processes *web navigation*, *booking* and *feedback*. For the process *web navigation* logfile data from the DMO website (http://www.are360.se) as well as from several hotel websites (http://www.totthotell.se/are/, http://www.copperhill.se) is used. Booking data from the destination-wide booking platform operated by SkiStar, as well as the booking system of Tott Hotel Åre is used. Finally, customer feedback data have been included from several destination surveys, stakeholder specific surveys (Tott Hotel and Copperhill Mountain Lodge), as well as from the two major online review platforms Booking.com and Tripadvisor. The prototype has been implemented based on the open source BI toolset RapidMiner (http://www.rapid-i.com) and the database system MySQL (http://www.mysql.com).

6.1 Cross-supplier analyses

The DMIS cockpit prototype offers supplier-specific as well as cross-supplier analyses for each of the three business processes described above. Figure 5 displays a dashboard for supplier-specific booking data, showing the overall turnover (booking price) grouped by the customers' order amount, and the total number of guests further grouped by the type of travel group. By selecting from different aggregation functions (e.g. sum, average, min or max), different attributes the values should be grouped by, and from different types of visualization (e.g. bar

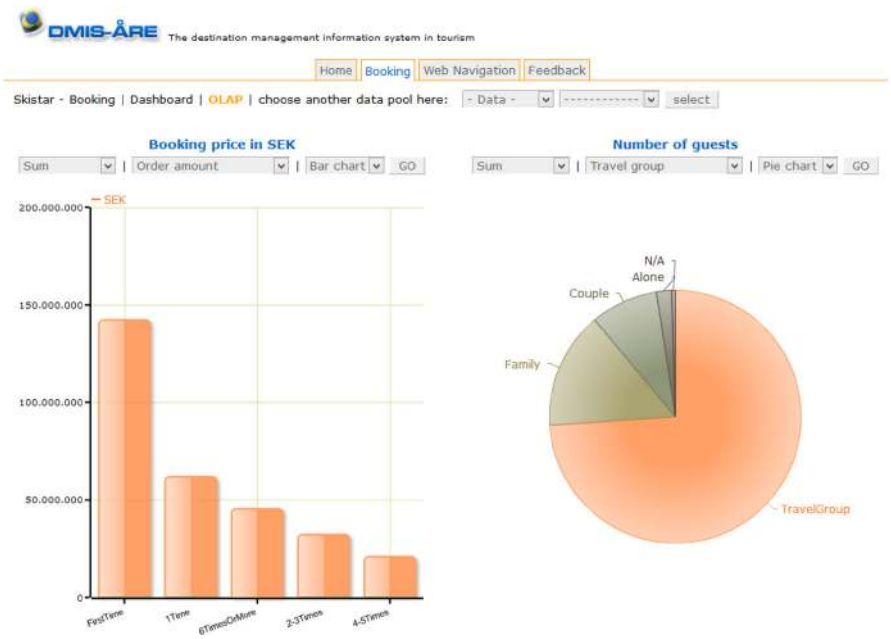


Fig. 5 Supplier-specific analyses of booking data

chart, pie chart, table), tourism managers can use explorative and OLAP-based techniques to analyse their own data in a flexible way.

Figure 6 shows a dashboard with cross-supplier analyses of booking data, offering typical benchmarking capabilities. The booking share of different accommodation providers is shown for various customer groups (based on their past order amount), stressing that most are first-time visitors to Åre and the proportion of first-time and repeat visitors significantly differs between different accommodation providers.

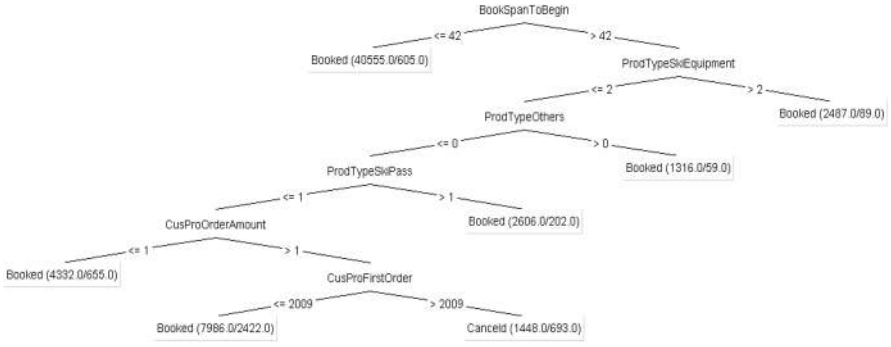
Besides descriptive analyses shown so far, decision making for destination management and stakeholders can be supported by more complex data mining techniques (Fuchs and Höpken 2009). As an example of data mining, Fig. 7 shows a decision tree which explains customers' cancellation behaviour based on destination-wide booking data. The accuracy rate of 93.2 % and the r^2 of 0.21 are reached by a C4.5 decision tree algorithm with a minimal leaf size of 2 and a confidence level for pruning of 0.25. Figure 7 shows a reduced decision tree, based on a minimal leaf size of 1200. Interestingly, the decision tree immediately enables the identification of the most important factors influencing the cancellation likelihood, such as the timespan between booking and arrival (*BookSpanToBegin*), the booked products (*ProdTypeSkiEquipment*, *ProdTypeOthers*, *ProdTypeSkiPass*), the number of previous bookings of the customer (*CusProOrderAmount*) as well as the year of the first arrival of the customer (*CusProFirstOrder*). Consequently, from a managerial standpoint a critical subgroup are customers booking more than 42 days in advance, booking two or less ski equipment items, no other products, only one or no ski passes, and have already booked before, but not earlier than in 2010 (node "Canceld (1448.0/693.0)"). With a cancellation rate of 67,6 % such customers

Fig. 6 Cross-supplier analysis of booking data

show an eight times higher cancellation rate than the average customer, representing a promising target group for specific marketing activities to prevent imminent cancellations. The explanation power of decision trees goes far beyond purely descriptive analyses shown before and constitutes a valuable input to BI-based decision support in tourism.

The business process *feedback* is supported by data from a number of different data sources, such as partner-specific customer surveys, destination-wide customer surveys, a specific online customer registration and survey platform, as well as the most relevant online platforms tripadvisor and booking.com. Within the data warehouse, each single customer feedback is stored as a single entry and assigned to an appropriate feedback and product category. Hereby, questions and responses can be looked at singularly (Fig. 8), but also aggregated by any dimension characteristics (e.g. date/time or customer characteristics), and by any product or feedback category. For example, Fig. 9 shows the average feedback value for different feedback categories, enabling the identification of strengths and weaknesses from a customer perspective.

The flexible assignment of single questions to product or feedback categories on different abstraction levels enables a comparison of customer feedback between different suppliers, even based on feedback data from supplier-specific and heterogeneous questionnaires. As an illustration, Fig. 10 shows the average




```
BookSpanToBegin <= 42: Booked (40555.0/605.0)
BookSpanToBegin > 42
|   ProdTypeSkiEquipment <= 2
|   |   ProdTypeOthers <= 0
|   |   |   ProdTypeSkiPass <= 1
|   |   |   |   CusProOrderAmount <= 1: Booked (4332.0/655.0)
|   |   |   |   CusProOrderAmount > 1
|   |   |   |   |   CusProFirstOrder <= 2009: Booked (7986.0/2422.0)
|   |   |   |   |   CusProFirstOrder > 2009: Canceld (1448.0/693.0)
|   |   |   ProdTypeSkiPass > 1: Booked (2606.0/202.0)
|   |   ProdTypeOthers > 0: Booked (1316.0/59.0)
|   ProdTypeSkiEquipment > 2: Booked (2487.0/89.0)
```

Fig. 7 Decision tree explaining tourists' cancellation behaviour

Respondent | guest feedback

average feedback value | by FeedQuestion select

items	average FdbFeedbackValue
The planning phase and communication with the hotel worked great	1
The check-in worked excellent	0.947
The check-out worked excellent	0.940
I am satisfied with the standard and maintenance of the facilities	0.920
The booking of my room worked well	0.904
The staff was there for me when needed	0.887
The breakfast gave me a good start of the day	0.883
The treatments at the spa lived up to my expectations	0.881
How likely is it that you would recommend Copperhill Mountain Lodge to a friend or colleague?	0.869
The spa gave me a positive experience	0.864
My visit at Copperhill Mountain Lodge made me feel noticed and the staff was courteous and conveyed a sense of caring service.	0.862
The facilities and all technical equipment worked great at the conference	0.853

Fig. 8 Average feedback values for single questions of a questionnaire

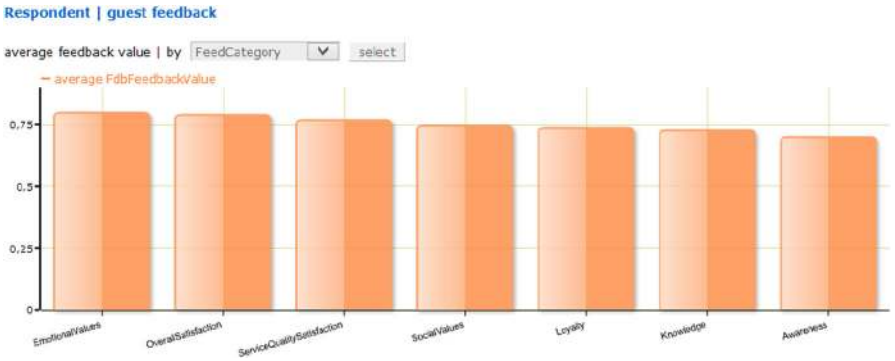


Fig. 9 Average feedback value per feedback category

feedback values per feedback category for different accommodation providers, based on feedback data from supplier-specific surveys, destination-wide surveys and a specific online customer registration and survey platform.

Analogous to structured feedback in the form of customer surveys described so far, unstructured customer feedback in the form of customer reviews provided on review platforms, like tripadvisor or booking.com, is assigned to product categories (Schmunk et al. 2014). Customer reviews are split into single statements (i.e. sentences) and classified into product categories and the sentiment of the statement by methods of text mining (cf. Sect. 3). Hereby, a comparison of customer feedback per product category across different suppliers can be executed also for customer feedback extracted from online review platforms. Figure 11 shows the average feedback value (i.e. average sentiment) of all single statements grouped by the supplier and the assigned product category, which enables suppliers to compare their strengths and weaknesses based on online customer reviews.

Additionally to the dashboard functionality, the DMIS prototype offers an OLAP interface (online analytical processing). This is a fundamental and well-known

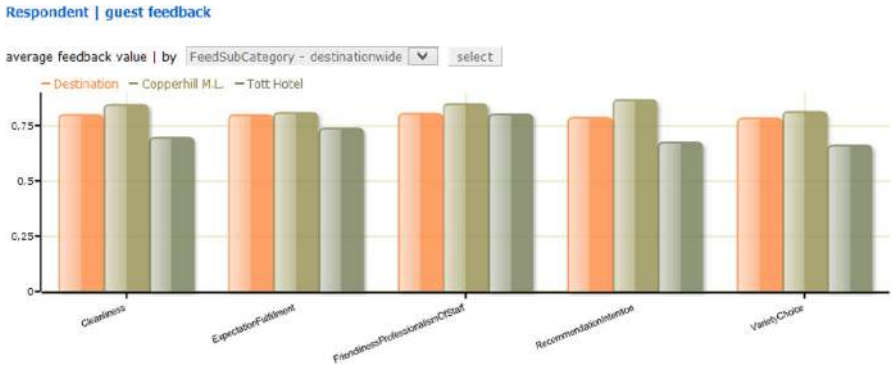


Fig. 10 Average feedback values per feedback category and supplier

Fig. 11 Average feedback values of reviews per product category and supplier

analysis technique, typically based on multi-dimensional data warehouse models. Figure 12 shows an OLAP analysis using feedback data. The user selects the facts to be analysed (e.g. the feedback value), the dimension attributes by which the facts should be aggregated (e.g. the customers' age range and the product type and area), and the aggregation function (e.g. the average). In addition, any kind of filter can be defined to restrict the analysis to relevant data fractions and to drill-down into interesting details. In the example, feedback data are filtered by the product type *indoor activities* in order to compare customers' satisfaction for different indoor activities and age ranges.

The analyses of customer feedback, as presented above, emphasize the power of the multi-dimensional data model, presented in this study. Customer feedback stemming from different and heterogeneous data sources can be compared across different suppliers and be analysed in an overall and destination-wide context.

Similar to booking data, feedback data can be analysed by more complex data mining methods. Figure 13 shows a decision tree explaining the most important factors influencing overall customer satisfaction (discretized into the values *low*, *medium* and *high*). It can easily be seen that a high agreement with at least one of

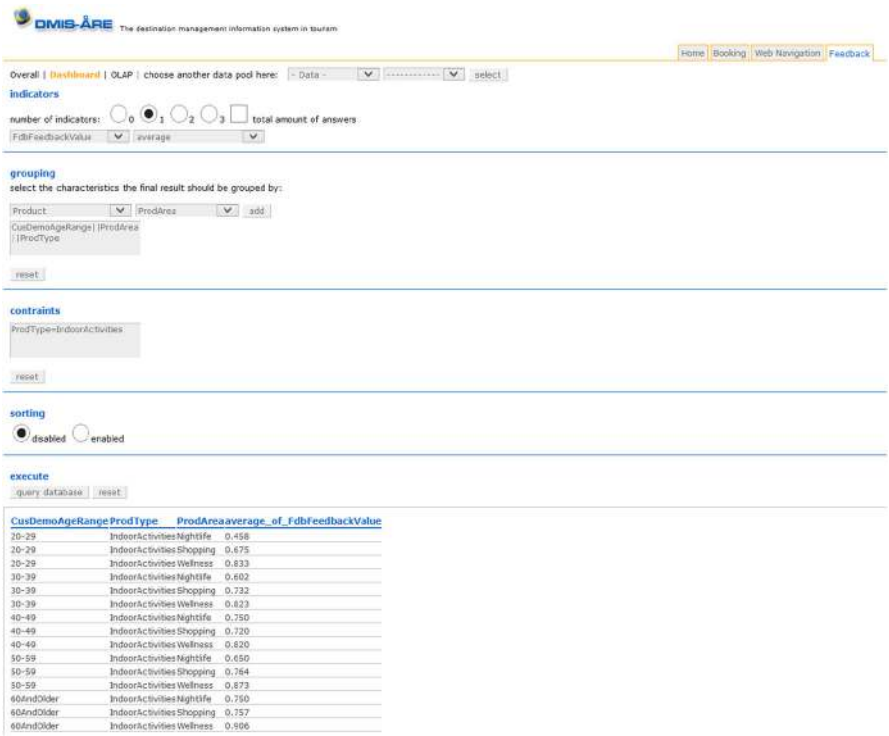


Fig. 12 OLAP analysis using feedback data

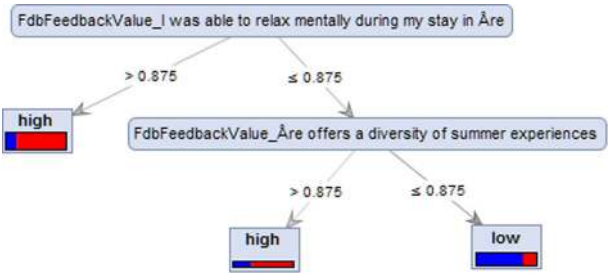


Fig. 13 Decision tree explaining factors influencing overall customer satisfaction

the two statements "I was able to relax mentally during my stay in Are" and "Are offers a diversity of summer experiences" leads to a high overall satisfaction. Otherwise, the overall customer satisfaction will be low. Such results constitute a worthwhile input to product optimization and marketing activities.

Association rules is a well-known data mining technique that can be used to identify products that are often bought together (i.e. market basket analysis), or to identify any kind of characteristics often co-occurring. Figure 14 shows association rules which identify activities often co-occurring within the travel profiles of Åre

No.	Premises	Conclusion	Support	Confid...	Lift
1	TraProHiking	TraProFishing	0.071	0.138	1.472
2	TraProHiking	TraProBicycling, TraProCableCar	0.162	0.317	1.347
3	TraProHiking	TraProCableCar	0.195	0.382	1.384
4	TraProBicycling, TraProHiking	TraProCableCar	0.162	0.410	1.488
5	TraProCableCar	TraProBicycling, TraProHiking	0.162	0.588	1.488
6	TraProBicycling, TraProCableCar	TraProHiking	0.162	0.689	1.347
7	TraProCableCar	TraProHiking	0.195	0.708	1.384
8	TraProFishing	TraProHiking	0.071	0.753	1.472

Fig. 14 Association rules showing activities often co-occurring

visitors. For example, rule three displays that Åre visitors doing *hiking* during their stay will use a *cable car* with a confidence of 38.2 %, which is 1.38 times more likely than the overall likelihood to use a cable car (this increase of likelihood is called the *lift*). Such association rules can serve as valuable input for product bundling and cross- or up-selling activities (Fuchs and Höpken 2009).

6.2 Cross-process analyses

Next to supplier-specific and cross-supplier analyses within each single business process, the DMIS cockpit prototype offers cross-process analyses that aim to find interesting patterns or trends across data from two or even more business processes. As previously discussed, cross-process analyses build on the concept of *conformed dimensions* and need at least one overlapping dimension characteristic in order to interlink corresponding transactions of different processes (e.g. bookings and feedback of the same customer or customers from the same country).

Figure 15 shows an example of a cross-process analysis, visualizing the correlation between bookings and web sessions (based on the date as the overlapping characteristic). The resulting graph impressively demonstrates that booking peaks typically follow web session peaks with a one or two day delay. Consequently, knowledge about the correlation between web navigation and booking behaviour constitutes a valuable input to forecasting booking behaviour and supports dynamic pricing and yield management.

Figure 16 shows the most important (i.e. key performance) indicators (KPIs) of all three business processes (defined during the requirement definition phase, cf. Sect. 4.3), using the customers' origin as the overlapping characteristic. Again, the power of such cross-process analyses is demonstrated by identifying interesting patterns across different business processes. For example, Australian customers constitute quite a small but highly satisfied customer segment with an extremely low website usage. This promising segment might be enlarged by a segment-specific adaptation of the corresponding websites. Moreover, customers from the Netherlands show the highest average booking price but a fairly low average satisfaction, which then constitute potentials for improvement.

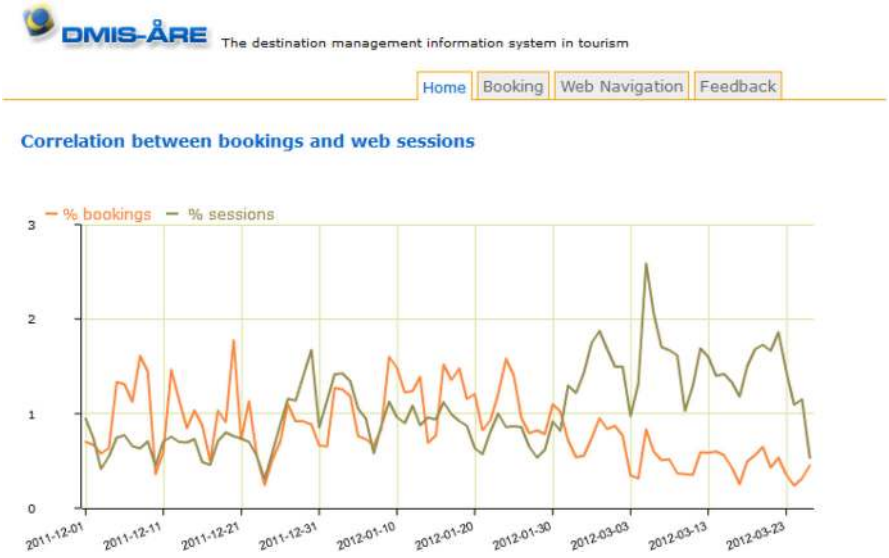


Fig. 15 DMIS cockpit Åre—correlation between bookings and web sessions

KPIs | by

Country GO ALL null values new window

Group by attribute	Total bookings	Average booking price in SEK	Average number of persons per booking	Average time between booking and arrival in days	Average stay duration per booking	Total clicks	Total sessions	Average time spent on single webpage in seconds	Average visit time on websites in seconds	Average pages visited on websites	Average feedback value	Total feedback, answers
Australia	33	4029.485	2.606	49.100	4.818	165	56	20.826	73.226	2.946	0.859	156
Belgium	33	5355.424	2.970	112	5.364	1207	361	20.692	85.960	3.344	0.702	130
China	14	3087.250	3.500	164.500	8.786	456	230	6.685	35.707	1.983	0.859	71
Denmark	1830	6235.034	4.527	115.609	6.144	2392	521	17.057	89.379	4.591	0.778	107
Estonia	1080	4837.395	5.292	85.108	5.838	161	53	24.085	94.556	3.038	0.784	16
Finland	3155	7852.046	4.386	95.536	6.160	4543	893	16.915	100.814	5.087	0.812	1039
France	25	4616	3.040	38.318	5.042	2923	739	19.187	82.978	3.955	0.715	12
Germany	105	5307.040	3.155	124.426	6.644	3268	665	13.377	82.701	4.914	0.837	150
Netherlands	279	7894.760	3.802	104.493	6.588	2168	490	13.741	87.556	4.424	0.750	150
Norway	10625	3901.213	4.911	71.890	3.297	27636	5889	16.439	87.700	4.693	0.767	9093
Sweden	56073	5023.956	3.860	83.014	5.498	162942	39139	17.407	84.981	4.163	0.773	36880
United Kingdom	1042	4929.136	3.009	59.399	6.786	48513	12843	17.259	76.811	3.777	0.765	303
United States of America	30	4719.167	4.036	39.148	4.133	4284	2307	11.236	57.168	1.857	0.539	16

Fig. 16 DMIS cockpit Åre—KPIs of bookings, web navigation and feedback

In Fig. 17 KPIs of the booking and feedback process are shown for different customer age ranges (as overlapping characteristic). The results reveal that younger customers show a relatively lower average booking price, book shorter in advance, and are less satisfied than peers. Interestingly, 40–50 year old customers by far make up the highest number of bookings with the highest average booking price and

KPIs | by

Age range GO ALL null values new window

Group by attribute	Total bookings	Average booking price in SEK	Average number of persons per booking	Average time between booking and arrival in days	Average stay duration per booking	Average feedback value	Total feedback, answers
20-29	9873	4247.327	4.626	65.800	4.304	0.749	2984
30-39	9859	4722.856	4.235	74.999	4.594	0.767	12450
40-49	17478	5738.469	4.136	95.942	5.189	0.762	19880
50-59	9370	5604.360	3.780	94.819	5.325	0.773	10090
60AndOlder	3244	5130.767	3.678	110.769	6.105	0.775	5826
UpTo19	192	4752.575	4.587	61.418	4.391	0.598	112

Fig. 17 DMIS cockpit Åre—KPIs of bookings and feedback by age ranges

a high average satisfaction. Therefore, they constitute one of the most important present customer segments of the Swedish tourism destination Åre.

These BI-based analyses examples demonstrate the power and flexibility of cross-process analyses, based on the presented multi-dimensional data model. Each characteristic of one of the conformed dimensions (cf. Table 2) can be used to interlink transactions across processes and thereby identify interesting patterns and trends among the different process indicators and KPIs. Dimension hierarchies (e.g. the location hierarchy with the levels city, state, country, continent) further improve the flexibility and power of such analyses by enabling the interlinking of processes by abstract characteristics if more concrete characteristics don't conform or are not available (e.g. the country in Fig. 16).

7 Conclusion and outlook

The paper at hand presents a knowledge-based destination architecture to enable BI-based knowledge extraction across all relevant business processes for a tourism destination. A technical architecture has been introduced to extract data from heterogeneous data sources, integrate data into a homogeneous destination data warehouse, and analyse this data using techniques of data mining and explorative data analysis (especially OLAP). The major scientific contribution is a process-overarching data model for a tourism destination data warehouse, which did not yet exist in the tourism domain and the related literature. Based on the technique of multi-dimensional data modelling and its central concept of *conformed dimensions*, a business process overarching data model was defined. This effectively enables analyses across different business processes, currently not supported by comparable systems in tourism. The presented knowledge-based destination architecture has been successfully instantiated and prototypically implemented for the leading Swedish mountain tourism destination Åre.

The present study clearly demonstrates that the concept of multi-dimensional modelling is suitable for building a tourism destination data warehouse, which enables powerful BI-based data analyses. Moreover, the concept of *conformed dimensions* proofed its ability to support process-overarching analyses.

Currently, the prototypical implementation of the knowledge-based destination architecture has been restricted to the customer-based business processes *web navigation*, *booking* and *feedback* of the knowledge generation layer and supplier-oriented knowledge application in the form of the DMIS cockpit. Consequently, new research activities will expand the scope and deal with additional customer-based processes, such as information requests, stay, consumption and location tracking. Also, supplier-based business processes (i.e. capacity and marketing activity) on the knowledge generation layer and customer-oriented knowledge applications, such as recommender systems or adaptive community services on the knowledge application layer will be considered in the course of future research.

A second vein of future research is the extension of the multi-dimensional data model in order to integrate data mining models (e.g. cluster models, association rules, decision trees) directly into the multi-dimensional structures (Meyer et al. 2015). Specific concepts will be developed to integrate even more complex data mining models, such as complete decision trees, into a multi-dimensional destination data warehouse model.

A final future research goal is the application of real-time business intelligence (http://www.gravic.com/shadowbase/, retrieved 2013; http://www.eyefortravel.com/social-media-and-marketing/savvy-data-collection-key-customer-understanding-and-personalisation, retrieved 2013) in order to gain real-time knowledge on tourists' on-site behaviour. For example, customer data can be collected through QR Code-based electronic customer cards collecting tourists' (GPS/WLAN-based) position and ad-hoc feedback (Zanker et al. 2010; Höpken et al. 2012). This valuable new knowledge can serve as input for intelligent mobile (i.e. ubiquitous) end-user applications, capable of recommending tourists the most promising matches with the actual destination offer. Consequently, this can enhance tourists' quality of experience (Wang et al. 2012). Finally, on the supply side, this newly generated knowledge input may be applied by small and medium-sized destination suppliers to react on segment specific needs in real-time (Fuchs et al. 2014, p. 208).

Acknowledgements This research was financed by the KK-Foundation project 'Engineering the Knowledge Destination' (no. 20100260; Stockholm, Sweden). The authors would like to thank the managers Lars-Börje Eriksson (Åre Destination AB), Niclas Sjögren-Berg and Anna Wersén (Ski Star Åre), Peter Nilsson and Hans Ericsson (Tott Hotel Åre), and Pernilla Gravenfors (Copperhill Mountain Lodge Åre) for their excellent cooperation.

References

Back A, Enkel E, Krogh G (2007) Knowledge networks for business. Springer, New York

Bloom J (2004) Tourist market segmentation with linear and non-linear techniques. Tour Manag 25(6):723–733

Bornhorst T, Ritchie J, Sheehan L (2010) Determinants for DMO and destination success: an empirical examination. Tour Manag 31(5):572–589

Brandt R (1988) How service marketers can identify value enhancing service elements. J Serv Mark 2(3):35–41

Bronner F, Hoog R (2011) Vacationers and eWOM: who posts, and why, where, and what? J Travel Res 50(1):15–26

Buhalis D (2006) The impact of ICT on tourism competition. In: Paptheodorou A (ed) Corporate rivalry and market power. IB Tauris, London, pp 143–171
Chaudhuri S, Dayal U (1996) An overview of data warehousing and OLAP technology
Chekalina T, Fuchs M, Lexhagen M (2014) A-value creation perspective on the customer-based brand equity model for tourism destinations. Finn J Tour Res 10(1):7–23
Cho V, Leung P (2002) Knowledge discovery techniques in database marketing for the tourism industry. Qual Assur Hosp Tour 3(3):109–131
Chu F (2004) Forecasting tourism demand: a cubic polynomial approach. Tour Manag 25(2):209–218
Dell'Erba M, Fodor O, Höpken W, Werthner H (2005) Exploiting semantic web technologies for harmonizing e-markets. Inf Technol Tour 7(3/4):201–220
Fuchs M (2004a) Pilot Project Destinometer™: the Tyrolean Benchmarking System (in German: "Pilotprojekt DESTINOMETER®—Benchmarkingsystem des Tiroler Tourismus"). Tour J 7(1):65–76
Fuchs M (2004b) Strategy development in tourism destinations: a data envelopment analysis approach. Poznan Econ Rev 4(1):52–73
Fuchs M, Höpken W (2005) Towards @Destination: a data envelopment analysis based decision support framework. In: Frew A (ed) Information and communication technologies in tourism. Springer, New York, pp 57–66
Fuchs M, Höpken W (2009) Data mining im tourismus. Praxis der Wirtschaftsinformatik 270(12):73–81
Fuchs M, Weiermair K (2004) Destination benchmarking—an indicator-system's potential for exploring guest satisfaction. J Travel Res 42(3):212–225
Fuchs M, Abadzhiev A, Svensson B, Höpken W, Lexhagen M (2013) Knowledge Destination framework for tourism sustainability—a business intelligence application from Sweden. Tourism 61(2):121–148
Fuchs M, Höpken W, Lexhagen M (2014) Big data analytics for knowledge generation in tourism destinations—a case from Sweden. J Destin Mark Manag 3(4):198–209
Garrow L, Koppelman F (2004) Predicting air travelers' no-show and standby behavior using passenger and directional itinerary information. J Air Transp Manag 10(6):401–411
Gräbner D, Zanker M, Fliedl G, Fuchs M (2012) Classification of customer reviews based on sentiment analysis. In: Fuchs M, Ricci F, Cantoni L (eds) Information and communication technologies in tourism. Springer, Wien, pp 460–470
Gretzel U, Fesenmaier D (2004) Implementing a knowledge-based tourism marketing information system: the Illinois tourism network. Inf Technol Tour 6:245–255
Höpken W (2004) Reference model of an electronic tourism market—version 1.3. http://www.rmsig.de/documents/referencemodel.doc
Höpken W, Scheuringer M, Linke D, Fuchs M (2008) Context-based adaptation of ubiquitous web applications in tourism. In: O'Connor P, Höpken W, Gretzel U (eds) Information and communication technologies in tourism. Springer, New York, pp 533–544
Höpken W, Fuchs M, Keil D, Lexhagen M (2011) The knowledge destination—a customer information-based destination management information system. In: Law R, Fuchs M, Ricci F (eds) Information and communication technologies in tourism. Springer, New York, pp 417–429
Höpken W, Deubele P, Höll G, Kuppe J, Schorpp D, Licones R, Fuchs M (2012) Digitalizing loyalty cards in tourism. In: Fuchs M, Ricci F, Cantoni L (eds) Information and communication technologies in tourism. Springer, New York, pp 272–283
Höpken W, Fuchs M, Lexhagen M (2014) The knowledge destination—applying methods of business intelligence to tourism applications. In: Wang J (ed) Encyclopedia of business analytics and optimization. IGI Global, Hershey, pp 2542–2556
Inmon W (2002) Building the Data Warehouse, 2nd edn. Wiley, New York
Jiang N, Gruenwald L (2006) Research issues in data stream association rule mining. SIGMOD 35(1):14–19
Kano N (1984) Attractive quality and must-be quality. J Jpn Soc Qual Control 14(2):39–48
Kasavana M, Knutson B (1999) A primer on siftware: warehousing, marting and mining hospitality data for more effective marketing decisions. J Hosp Leisure Mark 6(1):83–96
Kasper W, Vela M (2011) Sentiment analysis for hotel reviews. In: Computational linguistics-applications conference. Katowice, pp 45–52
Kepplinger D (2006) Tourismus WEBMART—Interaktive Datenerfassung und Ergebnisdarstellung durch Online-Datenbanken. In: Bachleitner R, Egger R, Herdin T (eds) Innovationen in der Tourismusforschung: Methoden und Anwendungen. Lit Verlag, Wien, pp 63–76

Kimball R (1997) A dimensional modeling manifesto. DBMS 10(9):58–70
Kimball R, Reeves L, Ross M, Thornwaite W (1998) The data warehouse lifecycle toolkit. Wiley, New York
Kimball R, Ross M, Thronthwaite W, Mundy J (2008) The data warehouse lifecycle toolkit, 2nd edn. Wiley, Indianapolis
Kuttainen C, Lexhagen M, Fuchs M, Höpken W (2012) Social media monitoring and analysis in tourism. In: Christou E, Chionis D, Gursory D, Sigala M (eds) Advances in hospitality and tourism marketing and management
Laine M, Frühwirth C (2010) Monitoring social media: tools. Characteristics and implications. Business information processing 51(2):193–198
Law R (1998) Room occupancy rate forecasting—a neural network approach. Int J Contemp Hosp Manag 10(6):234–239
Lexhagen M, Kuttainen C, Fuchs M, Höpken W (2012) Destination talk in social media: a content analysis for innovation. In: Christou E, Chionis D, Gursory D, Sigala M (eds) Advances in hospitality and tourism marketing and management. Corfu
Liu B (2008) Web data mining (2nd Ausg.). Springer, New York
Manning C, Schütz H (2001) Foundations of statistical natural language processing. MIT, Cambridge
Meyer V, Höpken W, Fuchs M, Lexhagen M (2015) Integration of data mining results into multi-dimensional data models. Information and communication technologies in tourism. Springer, Heidelberg, pp 155–168
Min H, Emam A (2002) A DM approach to develop the profile of hotel customers. Contemp Hosp Manag 14(6):274–285
Morales D, Wang J (2008) Passenger name record data mining based cancellation forecasting for revenue management. Innov Appl OR 202(2):554–562
Olmeda I, Sheldon P (2002) Data mining techniques and applications for tourism internet marketing. Travel Tour Mark 11(2/3):1–20
Pitman A, Zanker M, Fuchs M, Lexhagen M (2010) Web usage mining in tourism—a query term analysis and clustering approach. In: Gretzel U, Law R, Fuchs M (eds) Information and communication technologies in tourism. Springer, New York, pp 393–403
Pyo S (2005) Knowledge-map for tourist destinations. Tour Manag 26(4):583–594
Pyo S, Uysal M, Chang H (2002) Knowledge discovery in databases for tourist destinations. J Travel Res 40(4):396–403
Ritchie R, Ritchie J (2002) A framework for an industry supported destination marketing information system. Tour Manag 23:439–454
Sambamurthy V, Subramani M (2005) Information technologies and knowledge management. Manag Inf Syst Q 29(1):1–7
Schmunk S, Höpken W, Fuchs M, Lexhagen M (2014) Sentiment analysis—extracting decision-relevant knowledge from UGC. In: Xiang Z, Tussyadiah I (eds) Information and communication technologies in tourism. Springer, Heidelberg, pp 253–265
Smith B, Leimkuhler J, Darrow R (1992) Yield management at American Airlines. Interfaces 22(1):8–31
Subramanian J, Stidham S, Lautenbacher C (1999) Airline yield management with overbooking, cancellations, and no-shows. Transp Sci 33(2):147–167
Vela B, Blanco C, Fernández-Medina E, Marcos E (2012) A practical application of our MDD approach for modeling secure XML data warehouses. Decis Support Syst 52:899–925
Vlahogianni EI, Karlaftis MG (2010) Advanced computational approaches for predicting tourist arrivals. In: Evans T (ed) Nonlinear dynamics. InTech, Vienna, pp 309–324
Walchhofer N, Hronsky M, Pöttler M, Baumgartner R, Fröschl K (2010) Semantic online tourism market monitoring. In: Gretzel U, Law R, Fuchs M (eds) Information and communication technologies in tourism. Springer, Wien, pp 629–641
Wallace M, Maglogiannis I, Karpouzis K, Kormentzas G, Kollias S (2004) Intelligent one-stop-shop travel recommendations using an adaptive neural network. Inf Technol Tour 6(3):181–193
Wang Y, Russo S (2007) Conceptualizing and evaluating the functions of destination marketing systems. J Vacat Mark 13(3):187–203
Wang D, Park S, Fesenmaier D (2012) The role of smartphones in mediating the touristic experience. J Travel Res 51(4):371–387
Weiermair K, Fuchs M (2007) Productivity differentials across tourist destinations—a theoretical/empirical analysis. In: Keller P, Bieger T (eds) Productivity in tourism—fundamentals and concepts for achieving growth and competitiveness. Erich Schmidt Verlag, Berlin, pp 41–54

Wöber K (1998) Global statistical sources- TourMIS: an adaptive distributed marketing information system for strategic decision support in national, regional or city tourist offices. Pac Tour Rev 2(3):273–286

Wong J-Y, Chen H-J, Chung P-H, Kao N-C (2006) Identifying valuable travellers by the application of data mining. Asia Pac J Tour Res 11(4):355–373

Zanker M, Jessenitschnig M, Fuchs M (2010) Automated semantic annotation of tourism resources based on geo-spatial data. Inf Technol Tour 11(4):341–354