WORKING PAPER SERIES

MODUL VIENNA
UNIVERSITY
WKO WIEN PRIVATE UNIVERSITY

Towards Cross-domain Decision Making in Tourism:
A Linked Data based Approach

Marta Sabou, Irem Onder, Adrian M. P. Brasoveanu,
Arno Scharl

MODUL University Working Paper No. 2

March 2015

Towards Cross-domain Decision Making in Tourism:
A Linked Data based Approach

Marta Sabou[1]
Technical University of Vienna
Taubstummengasse 11, 1040, Vienna, Austria
Tel: +43-1-58801-79905450
Fax: +43-1-58801-18899
marta.sabou@ifs.tuwien.ac.at

Irem Onder[2] (corresponding author)
MODUL University Vienna
Am Kahlenberg 1, 1190, Vienna, Austria
Tel: +43 (1) 3203555 420
Fax: +43 (1) 3203555 903
irem.onder@modul.ac.at

Adrian M.P. Brasoveanu[2]
Arno Scharl[2]

[1] Vienna University of Technology
[2] MODUL University Vienna, Am Kahlenberg 1, 1190, Vienna, Austria

Abstract: The complexity of the socio-, political- and economical settings in which tourism enterprises operate, increasingly require them to make decisions that take into account data from various domains (e.g., economy, environmental sustainability). Based on a practitioners' survey that we performed, we conclude that although such cross-domain decisions are important, they are primarily performed by relying on manual data collection and aggregation, which is both time-consuming and error-prone. We propose a solution that relies on Linked Data as a technological platform for integrating data from three major tourism data sources: TourMIS, World Bank and Eurostat. Enabled by this integrated data, we developed the ETIHQ Dashboard, the first visual decision support system that supports cross-domain decisions over tourism, economic and sustainability indicators. An evaluation performed with practitioners shows that this Linked Data enabled systems brings important improvements in terms of execution times (28% faster) and answer quality when compared to current manual approaches.

Keywords: decision support, information integration, linked data, statistical data, visual dashboard

1 Introduction

In their recent research manifesto [31], Werthner and Ricci identify data analytics as one of the major influential topics for the tourism industry. Indeed, the need to make sense of large amounts of complex data is a common theme in all the five layers of their model, ranging from *decision making* needs at individual and group level, to *knowledge management* for companies and to *sector wide analytics* for the Industry level. Satisfying such data analytics requirements is increasingly challenging due to the complex network relationships at all levels as well as for the pressure of aligning tourism activities with regulatory issues on sustainability (environmental, economic, social, cultural).

In the context of this novel, complex and broad research trend, our focus is on tourism corporations (enterprises), which occupy the third layer of Werthner and Ricci's model [31]. The complexity of the socio-, political- and economical settings in which tourism enterprises operate, increasingly require them to make decisions that take into account data from various domains (e.g., economy, environmental sustainability) and usually gathered from different data providers. Examples of typical questions that support such complex decisions are: *Do changes on American GDP influence the bednights spent by American tourists in Germany? If yes, was Austria, a competitor of Germany, also affected by the economic situation in the USA?*

Answering such questions often requires gathering and combining data from various data sources. As we detailed in [25], there is no shortage of data as tourism indicators are published by a variety of international organisations along with indicators in other areas such as economy and sustainability (UNWTO, Eurostat). TourMIS [32] is one of the largest sources of European travel statistics, which, provides European-level data at a more fine-grained level than large international organizations (monthly vs. annual readings; city level vs. country level data). Common to these data sources is that, while their data is openly accessible, this data is made available using database centric technologies that require tedious manual efforts when combining data across providers. For example, a system that would allow correlating arrivals from USA to Germany with the GDP of USA, must be "told" through explicitly programmed mappings that the country "USA" in one data source (e.g., TourMIS) is actually the same as the country "United States" in another data source (e.g., World Bank). Large-scale data integration, although a key prerequisite for enabling complex, cross-domain decisions, remains challenging because (1) it requires significant manual effort to hardcode correspondences of the entities used by two data sources and (2) these manual mappings must be maintained to account for changes in the naming conventions used by the two datasets. As a direct consequence, most tourism decision support systems usually only cater for investigating tourism indicators in isolation from economic or sustainability indicators, as we detail in Section 6.

In this paper, we aim to investigate the following research questions:

How are cross-domain decisions performed in tourism? To explore the types of cross-domain decisions typical for tourism as well as the current

approaches for solving them, we performed a survey with 37 tourism practitioners, which we report in Section 2. The main conclusion was that, although cross-domain decisions are important, currently these need to be performed through manual data collection and combination approaches.

How to solve data integration challenges? As a solution to the data integration challenges in tourism, we propose the use of Linked Data technologies. Concretely, we report on using Linked Data to expose TourMIS as a linked data set. We refer to this data set as the ETIHQ data set, as it was created during a project with the same name[1]. An immediate benefit of this process was the linking of TourMIS data with data from additional data sources, such as WorldBank and Eurostat, thus overcoming data integration issues. We report on our technical solution in Section 3, by extending with details work that was published earlier on this subject in [22, 24, 25].

How to provide tool support for cross-domain decisions? To overcome the lack of tool support for cross-domain decisions, we implemented the ETIHQ Dashboard, a decision support system that relies on the multiple-integrated-views paradigm, as first reported in [25]. In this paper we provide an evaluation of this system based on a task-based evaluation and on a usability evaluation (Section 5). The evaluation showed that the current manual approaches to answering cross-domain decisions are time consuming and error-prone. In this context, the ETIHQ Dashboard led to important time savings and improvements in answer quality, thus being a first, promising approach to this problem.

We continue with describing the practitioners's survey in Section 2. Our solution in terms of Linked Data based data integration and the ETIHQ Dashboard is detailed in Sections 3 and 4 respectively. The evaluation of the ETIHQ Dashboard is described in Section 5. We then sum up with a related work section (Section 6) and a conclusion section (Section 7).

2 Practitioners' Survey

To gather information from the tourism practitioner's community about their needs for and current approaches to taking cross-domain decisions, we performed a survey collocated with the TourMIS 2014 workshop. Given the large user community of TourMIS, an annual conference is organized attracting over 100 interested participants from industry and academia alike. TourMIS 2014 therefore was a suitable venue to interact with the the target user community of our work.

---

[1] ETIHQ (Exposing Tourism Indicators as High Quality Linked Data) - http://www.etihq.eu/

Table 1 Overview of the structure and goals of the TourMIS'14 presentation and survey.

Part	Topic	Presentation	Survey
I	Introduction	Introducing the speakers and the project	Asking for the participants' occupation and country of origin
II	Multi-domain decisions in tourism	Explaining the notion of complex questions relying on multi domain statistical data	Asking participants about: how often they perform complex questions; examples of of questions; data collection method; usefulness of extending TourMIS with statistics from from other domains; most useful domains to combine with tourism data.
III	Linked Data	Explaining the LD and its use to integrate statistics from different data sources	Asking participants for their familiarity with LD and the intuitiveness of LD.
IV	Decision Support System	Explaining the decision support portal	Asking participants for the importance of such a system
V	Conclusion	Concluding the talk; Questions&Answers	Asking participants for any additional comments

2.1 Survey Design

The approach consisted of a combined presentation and survey, where participants were asked to fill in a survey during the presentation - the presenter prompted the audience to fill in the survey as she was explaining the content of her talk. The slides of the presentation are available online[2].

From approximately 50 participants to the session, a total of 37 forms have been filled in and 30 participants indicated they are interested in further information about the topic of cross-domain decision making by sharing with us their email address. Therefore, this topic is of high interest to the target user community.

2.2 Survey Results

Part I: Survey Population: The majority of the participants (24 out of 37, i.e. 65%) were tourism practitioners. Nine academics also completed the survey. Two respondents did not fill in this field of the form, while other two came from the travel industry and a non-for-profit organization respectively. In terms of geographic spread, participants originated from 19 countries, primarily European, mostly from Hungary followed by Austria and Belgium.

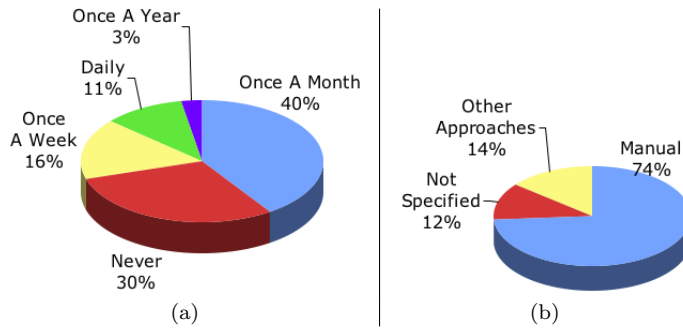[2] http://www.modul.ac.at/uploads/files/user_upload/Workshop_2014-Sabou-Onder.pdf

Fig. 1 Frequency of (a) and approaches (b) for taking cross-domain decisions.

Part II: Multi-domain decisions in Tourism: We wanted to find out how often tourism practitioners and academics need to perform tasks that involve questions relying on multi-domain indicators. 70% of respondents need to answer such questions, although not on a daily basis. In our survey population, most participants (40%) answered complex questions about once a month, 16% once a week, 11% daily and 3% once an year (see Figure 1 a). One participant commented that they performed such decisions "Everytime we see smth strange on the statistics. Once a year when we observe the markets at the moment to prepare the actions plan for next year"[3].

Figure 1 b depicts the distribution of approaches for the data collection and integration used for answering complex, cross-domain questions. In 74% of the cases, participants use a manual approach to finding, downloading and combining diverse indicator data. No answer was given to this question by 12% of the participants, while the rest of 14% relied on other approaches including: (1) using TourMIS; (2) using statistics applications; (3) inspecting citytourismbenchmark.com; (4) using warehouses internal to their company; 5) Using external tools: Smith Travel Research.

We asked participants to rate, on a scale from 1 (useless) to 10 (very important), the usefulness of extending TourMIS with capabilities to combine multi-domain statistical data. To this question 3 respondents did not provide a value, one participant provided 10+ and another 100. After excluding the void answers and changing the other two values to 10, we obtained an average rating of 8.35 with a standard deviation of 1.47, thus demonstrating a strong positive opinion towards the proposed extensions to TourMIS.

In terms of the most interesting domains that could be combined with TourMIS data, most participants were interested in economic indicators (35 votes), followed by infrastructure (22 votes), climate change (11 votes) and health (9 votes) related statistics. Additional indicators suggested were: consumer confidence (2), employment rate, exchange rate (2), demographics, airline data, sustainability indicator, visa application statistics, airline statistics, weather data, tourist preferences and trends, media coverage data.

---

[3] We reproduce all comments from the survey verbatim.

Part III: Linked Data. The notion of Linked Data technologies is not totally unknown in the tourism community. Indeed, 51% of the respondents have heard about Linked Data before, with 32% claiming to have a vague understanding of the technology, 3% actually making use of this technology and 16% having no understanding of the technology principles. Furthermore, after the talk 71% of the respondents found the principles of Linked Data easy to understand.

Part IV: Decision support system We asked participants to rate, on a scale from 1 (useless) to 10 (very important), the usefulness of a decision support system that would allow superposing multi-domain statistical indicators. Since the ETIHQ system was still under development at that time, the presentation only contained screen mockups. After removing 5 void answers, the average was 8.22 (standard deviation 1.68), thus indicating a positive perception of the usefulness of such a tool.

The most important conclusions we derive from this survey, are:

– Multi-domain questions are important, although not performed daily.
– The main approach to answering such questions is manual data collection, and therefore the data integration strengths of Linked Data technologies are highly desirable.
– Extension of TourMIS with additional domain statistics is deemed very useful, in particular regarding economic data.
– At least 50% of the participants has already heard about Linked Data and 71% claimed that the principle of this technology is easy to grasp.
– The decision support system was also considered a potentially useful tool.

3 Data Integration using Linked Data Technologies

This section provides the technical details of how Linked Data technologies were used for integrating statistical data from different sources. We start by providing some technology basics necessary to understand the technical content of this section. We then continue by describing the solution approach, which involved exposing TourMIS as linked data and aligning it with other data sources, and its main stages including semantic modelling, the creation of the dataset and of the cross-dataset links as well as the publication of this dataset on the web.

3.1 Technology Basics

Semantic Web and Linked Data technologies aim (1) to enrich raw data with semantic (that is, machine processible) information and (2) to publish it using Web based languages on the Web thus allowing links between datasets to be explicitly specified [27]. As such they are highly suitable for large-scale data integration of distributed datasets, such as those of interest for cross-domain decision making.

Expressing and Encoding Meaning. A core element of Semantic Web technologies are ontologies [10], formal domain models describing concepts in a domain and their relationships in ways that computer programs can process and reason with these descriptions. For example, a tourism ontology would describe concepts such as `City` or `Market`. Data items (e.g., Vienna) are then described in terms of ontology concepts (e.g., by associating Vienna to the concept City). Since the goal of the Semantic Web technologies is to make data public on the Web, ontology elements as well as each raw data element to be described in the Semantic Web are assigned a unique web URL, for example, `http://data.etihq.eu/etihq/page/dataset/Vienna`, for the data element Vienna or `http://data.etihq.eu/etihq/page/ontology/City` for the `City` concept of the ontology. The structure of the URLs usually indicates the name of the dataset (in our case `etihq`) as well as the type of the entity, which can be either part of the abstract data model (i.e., the ontology) or of the dataset.

Data Model. While relational databases rely on a relation (i.e., table like) data model, the Semantic Web adopts a triple (or graph based) model with all data being represented as triples. For example, to declare that Vienna is a City, a triple is created stating that `<Vienna, isA, City>`. To encode triple-based Semantic Web data, a set of languages have been developed, most notably RDF (Resource Description Format) and OWL (Web Ontology Language). Similarly to other Web specific languages, such as HTML, RDF and OWL are also based on XML (eXtendable Markup Language). OWL builds on RDF but allows expressing more complex semantics than RDF. The RDF representation of the triple above is:

```
<ebo:City rdf:ID="eds:Vienna"/>
ebo=http://data.etihq.eu/etihq/page/ontology
eds=http://data.etihq.eu/etihq/page/dataset/
```

Interlinking. One advantage of exposing datasets online using Web based standards is that **links** can be created between the elements of these datasets (similarly to hyperlinks in HTML). Most often, an `owl:sameAs` link is created between URLs from different datasets that represent the same real-life entity. For example, a single triple is sufficient to declare that two URLs in the ETIHQ and WorldBank data respectively refer to the same entity, the city of Vienna:

```
eds:Austria owl:sameAs
 <http://worldbank.270a.info/classification/country/AT.html>
```

Thanks to this link it is now possible for a computer program to understand that the term Austria as used in TourMIS is the same as the term AT used in the WorldBank dataset and that data about these terms in the two datesets can be combined.

3.2 Solution Approach

Figure 2 shows the steps followed for creating the ETIHQ Dataset as well as the main data artifacts and tools that were used.
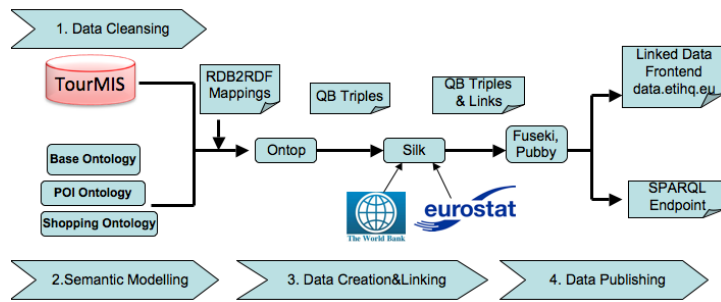
Fig. 2 Process, tools and data artifacts used for creating the ETIHQ dataset.

The process started with a multi-stage data cleansing step. Firstly, we removed tables and columns that contained redundant or unnecessary data. For smaller datasets, this step can be done with the OpenRefine tool, but we used SQL instead to cater for the complexity of this operation. Secondly, we checked the links between various tables in the database schema to make sure the previous deletions did not affect them. This was important as some of these links between tables are also reflected in the Linked Data structure. Thirdly, we changed some keys in order to obtain clean URIs later. Last, we verified that the various applications that use the database still function as expected (e.g., the web application used for data input).

The Semantic Modeling process focused on selecting relevant existing ontologies, creating domain specific ontologies and defining a structure for the Linked Data dataset (Section 3.3). The cleaned TourMIS dataset, the created ontologies and the dataset design served as inputs for creating the dataset itself which contained the semantic representation of the TourMIS data. Subsequently, we linked the elements of this dataset to relevant entities from other datasets therefore technologically enabling the integration of these datasets. Data creation, linking and publication are described in Section 3.4.

3.3 Semantic Modeling

In this section we briefly report on the semantic modeling performed. We focus on key elements necessary for understanding the design of the entire system. Full details of this work can be accessed in [23].

The semantic structure of the proposed solution includes a variety of ontologies. Firstly, RDF Data Cube (QB) is the emerging semantic standard for modeling data of statistical nature and therefore it was used as a main vocabulary to publish the TourMIS data. Then, PROV-O, the standard vocabulary for specifying provenance information, was reused to allow representing information about who and when contributed data to TourMIS. Finally, we also developed the ETIHQ Domain Ontologies, a set of ontologies specific to the tourism domain that complement the generic models above with con-

cepts describing the content of TourMIS. All the created semantic documents (ontologies and data structure definitions) are available for download[4].

Before creating our domain ontologies, we have conducted an overview of existing tourism ontologies that could potentially provide some concepts to be reused in our modelling. The detailed overview is reported in a project deliverable [23]. The conclusion of our overview was that existing tourism ontologies primarily support tourist-centric applications (e.g., recommendation and question answering systems to be used by tourists) and, therefore, their vocabulary is restricted by those applications' scope. Many ontologies are also no longer available online. As a result, we have focused on building a new ontology by extending our earlier work described in[22], which, to our knowledge is the only ontology concerned with modelling tourism indicators.

We have opted for a modular design, where a base ontology models the core notions related to tourism indicators and two other modules focus on points of interest and shopping items respectively. We hope that this modular design will not only make the maintaining of the ontology base easier but it will also encourage reuse of our ontologies by others: instead of importing the entire ontology base, they can import only relevant modules.
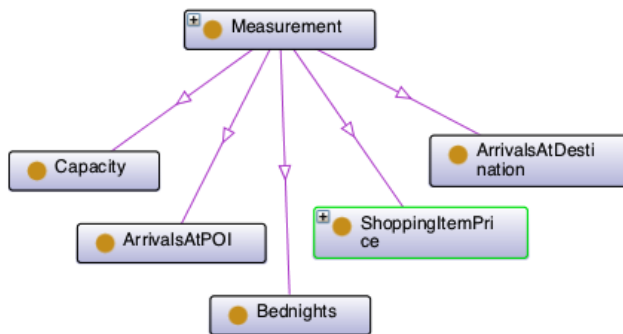

Fig. 3 Overview of the main concepts of the base ontology.

The base ontology has the `http://www.etihq.eu/ontology/` namespace (prefix `ebo` that stands for "etihq base ontology"). Core to this base ontology is the concept of `Measurement` which has five more specific concepts corresponding to all the statistical indicators in TourMIS and depicted in Figure 3. Additionally, the ontology models classes such as `Destination`, `Market` and `Currency`. Two further class hierarchies, one corresponding to `PointOfInterest` and one to `ShoppingItem`, are imported from two separately managed ontologies. The second ontology is the POI ontology declared under the `http://www.etihq.eu/ontology/poi/` namespace which models concepts corresponding to the POI categories (e.g., museums, churches, theatres) in TourMIS as specialisations of the `PointOfInterest` concept. Fi-

---

[4] http://etihq.eu/wp-uploads/2014/01/ETIHQ_SemanticModels.zip

Table 2 Datasets overview in terms of QB components and the tourism concepts defined in the ETIHQ domain ontologies.

| Dataset | Components | | |
	Dimensions	Measures	Attributes
Arrivals	Time, ebo:City, ebo:Market	ebo:ArrivalsAtDestination	unitMeasure
Bednights	Time, ebo:City, ebo:Market	ebo:Bednights	unitMeasure
Capacity	Time, ebo:City	ebo:Capacity	unitMeasure
Arrivals_At_POIs	Time, ebo:City, epoi:PointOfInterest	ebo:ArrivalsAtPOI	unitMeasure
Shopping_Items	Time, ebo:City, esi:ShoppingItem	ebo:ShoppingItemPrice	ebo:Currency

nally, the Shopping ontology (`http://www.etihq.eu/ontology/shopping/`) defines the `ShoppingItem` concept and its main subclasses as available in TourMIS and categorised in four main categories, namely, `Accommodation`, `Entertainment`, `FoodAndBeverage` and `Transport`.

Given the statistical nature of the TourMIS data, we relied on the RDF Data Cube Vocabulary (QB) for describing it. QB is the current standard for publishing statistical data, and is a W3C Recommendation supported by industry and academia. The basic building blocks of the cube model are measures, dimensions and attributes, collectively referred to as components. *Measure* components describe the things/phenomena that are observed or measured, for example, indicators such as height, weight or, in our tourism context, arrivals, bednights or capacity. *Dimension* components specify the variables that are important when defining an individual observation for a measurement. Examples of dimensions include time and space. *Attributes* help interpret the measured values by specifying the units of measurement, but also additional metadata such as the status of the observation (e.g., estimated, provisional). *Observations* are the unit elements in a dataset and they represent a concrete measurement value for a set of concrete dimension values. They correspond to a value in a statistical database. When the value of a dimension is the same in a large number of observations (for example, the geographic location) it is convenient to group these into a *slice*. A *dataset* that contains observations grouped into slices across dimensions constitutes a *cube*.

TourMIS has a structure that naturally fits the data cube formalism, with most indicators being measured along dimensions such as time, destination city and market. Therefore the representation of TourMIS data using QB is quite straightforward. As summarized in Table 2, we define five datasets corresponding to the five main indicators measured by TourMIS. The fact that these indicators are measured along different dimensions prevents us from including them all in a single QB dataset. This modular design will allow a simpler reuse by end-users of the datasets that they are interested in. Note that the abstract QB notions of dimensions, measures and attributes are mapped to concrete domain concepts defined in the ETIHQ domain ontologies.

3.4 Data Creation, Linking and Publication

The TourMIS database is powered by a Microsoft SQL Server. In order to translate it into Linked Data, we used an RDB2RDF mapping (Relational to RDF mapping, sometimes also referred to as OBDA Ontology Based Data Access mapping)[5]. The mappings file contains all the SQL queries that are used to create the RDF triples together with the rules used to create the right types and URIs for the RDF resources. These files are used by the Ontop Protege plug-in[6] to generate RDF files.

The following example mapping selects the *code*, *txe* and *etczl* columns from the region table of the TourMIS dataset and creates, for each identified market two triples: one to describe the type of the entity as $MarketForCountry$ and one to specify the label of the market.

```
mappingId map-market
target :db1/Market {code}  a :ebo/MarketForCountry;
rdfs:label {txe}.
source  SELECT  code,  txe,  etczl FROM region
WHERE etczl is not null AND LEN(code) <4
```

For each identified market, this mapping generates two triples of the form:

```
eds:market_AT  a  ebo:MarketForCountry;
rdfs:label  AT
```

We wrote 60 mappings to generate 5 different datasets from TourMIS amounting to around 16 million triples. The RDF files generated by Ontop are in the RDF Data Cube (QB) format, therefore the values for one month or one year are mapped to the values of observations. The resulting triples are uploaded to the Fuseki[7] triplestore, which a database specialized for storing RDF data.

As a final step of the dataset creation process we have specified links between the entities of ETIHQ and other datasets. We used the Silk link discovery tool [12] to create links to the DBpedia[8], Geonames[9], the WorldBank and Eurostat datasets. We compiled Geonames using a Python script, and used the dumps for DBpedia. We managed to successfully link all entities referring to countries and cities.

Once the data from TourMIS has been extracted, represented as RDF and stored in a specialized dataset, we published it on the Web. To that end, we provided a SPARQL frontend using YASGUI[10] so that users can query the data. SPARQL is a query language similar to SQL in databases but adapted to query the graph like data model of RDF.

---

[5] https://github.com/ontop/ontop/wiki/ObdalibObdaFile

[6] http://ontop.inf.unibz.it/

[7] http://jena.apache.org/documentation/serving_data/

[8] http://dbpedia.org

[9] http://www.geonames.org/

[10] http://yasgui.laurensrietveld.nl/

Besides allowing query-based access, the ETIHQ data (both ontologies and data triples) is exposed in a Web-page like format using the Pubby[11] application (deployed in Tomcat). Thanks to this, users can navigate the online linked data repository using any web browser. Several online access points to our dataset can be accessed at http://data.etihq.eu.

4 Visual Decision Support Dashboard

The ETIHQ dashboard is built on top of the Linked Data set that links together TourMIS, Eurostat and WorldBank statistical indicators. It allows users to select, slice, visualize and compare indicators from tourism (Arrivals and Bednights from TourMIS), economics (GDP growth (annual %); Inflation, consumer prices (annual %); Consumer price index (2005 = 100)) or sustainability (CO2 emissions (kt);Forest area (% of land area); Air transport of passengers; Rail transport of passengers; Final energy consumption by transport). As such, to the best of our knowledge, it is the first system to allow cross-domain support for decision making in tourism. The main goal of the system is to help tourism executives, decision makers and researchers to understand a broader picture (including external factors such as economics or sustainability) when examining tourism demand. A demo version of the tool is available at http://etihq.weblyzard.com/.[12]
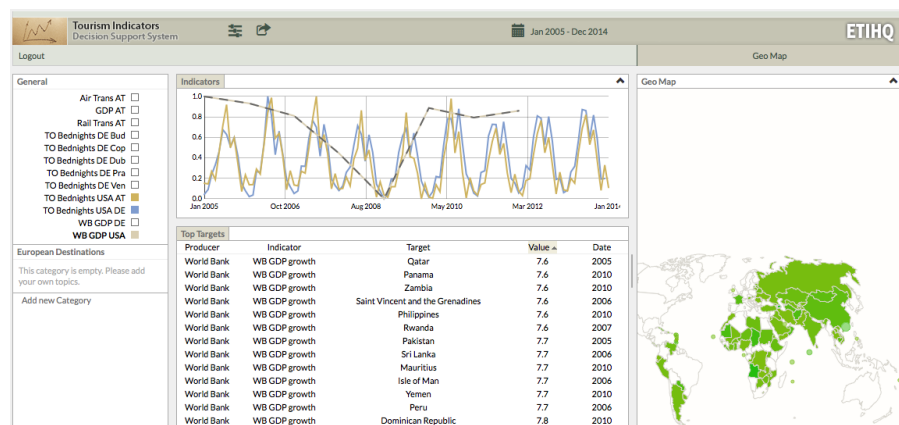

Fig. 4 The ETIHQ Dashboard interface.

Figure 4 displays a screenshot of the dashboard, which includes the following main panels. The top horizontal interface bar allows selecting the time-frame of the analysis (by clicking on the calendar icon) and offers capabilities

---

[11] https://github.com/cygri/pubby

[12] username=demo@etihq.eu; password= etihqdemo

for exporting the data visible in the dashboard and the created charts in various formats (csv or Excel for the data; svg or png for the charts). The left-most frame can be used to define new queries over statistical data, displays these queries and allows their management (e.g., (de) selection). The top part of the middle pane contains a chart based visualisation element that showcases the currently selected indicators and allows their visual comparison with other selected indicators. The lower part of the middle frame allows zooming into data details for the selected search showing the highest 250 values for the selected indicator. The right-most pane contains geographic data visualisations.

We continue with the description of the dashboard functionality in the context of the example scenario: *understanding whether, during the last decade, the behavior of American tourists in terms of their arrival to Germany and Austria was influenced by the economic situation in the USA.*
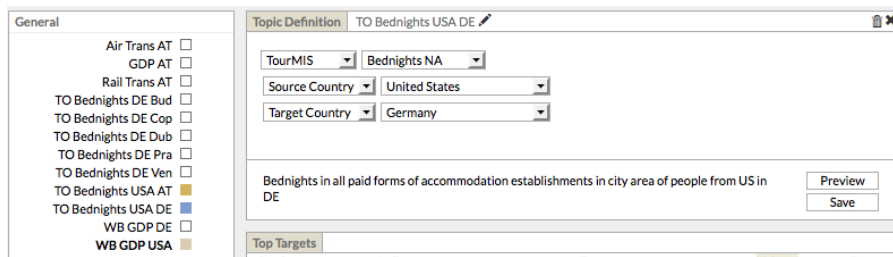

Fig. 5 Advanced Search dialog for editing the definition of a topic which represents a slice of an indicator with time as a fixed dimension.

Date selection. A click on the calendar icon allows selecting the start and end dates of the time interval of interest. TourMIS indicators are usually measured monthly, while World Bank and Eurostat indicators have different granularities (year, quarter, month). It is recommended that an interval of at least several years is inspected as values for recent months might not yet be available in the data-sources we build on. Date selection fixes an important dimension of an indicator, namely, time.

Adding a new indicator slice to the system involves creating a named representation of the indicator slice and then defining it by configuring the values for each indicator dimension.

A representation of an indicator slice is created by pushing the wheel button from the *General* menu and providing a suggestive name reflecting the definition of that particular indicator slice. We recommend the following convention: *[Source] [Short name for Indicator] [ISO country code] [Short code or Airport Code for City]*. Following such a convention, a long indicator name such as *Bednights in all paid forms of accommodation establishments in city area of Americans in Germany* will become *TO Bednights USA DE*, where TO stands for TourMIS. Other naming conventions can be used depending on the users needs.

The definition of the indicator slice is achieved by filling out details in the *Advanced Search* form that is displayed instead of the line chart. As it can be seen in Figure 5, the form requires: (1) the original *publisher* of the data (TourMIS, World Bank, or Eurostat); (2) the *indicator* to display (Bednights NA, in the picture); and (3) the geographical *sources* and *targets* of the respective indicator (source is the name of the origin from where a tourist leaves; target is the location where a tourist arrives). The Advanced Search dialogue also allows the user to perform several operations related to the new indicator: (4) rename; (5) delete (by click on the trash bin) or alert if new data about it is available (by clicking the alert button); and (6) see the entire name of the selected indicator, preview it in the interface and save it.

Statistical indicators might have multiple dimensions and typically we experienced mostly two or three dimensional indicators. In tourism, relevant dimensions include: *time, market* (mapped to the source field) and *destination* (mapped to the target field). The time dimension is fixed when the user selects the time interval. If more than two dimensions are available, another dimension can also be fixed. Eurostat and World Bank data sets have two dimensions (time; origin), whereas TourMIS data sets have three dimensions (time; origin or market; destination or city). The dashboard interface is built on the idea of fixing at least a dimension in order to display the other dimensions in various visualizations.

Visualizing an indicator slice. After closing the *Advanced Search* dialog, the selected indicator slice will be automatically displayed in the line chart with a dotted line. The *General* menu also serves as a legend for the indicators displayed in the line chart, as shown in Figure 6. If the square that follows after an indicators name is colored, the respective indicator is enabled and will be seen in the line chart. Figure 6 depicts the concurrent visualisation of the American GDP from WorldBank (WB GDP USA) as well as Bednights of American tourists to Germany and Austria as collected from TourMIS (TO Bednights USA DE, TO Bednights USA AT). To avoid visual clutter due to visualizing too many indicators at once, the maximum number of concurrently displayed indicators can be constrained at system setup time, by default 5.

Analyzing an indicator in context. An important capability of the dashboard is enabling users to inspect indicators in context and assess their performance against different types of indicators from various data sources. Multiple panels from the visual interface support this aim. The *General* menu allows selecting multiple indicator slices to be compared. The *line chart* can display multiple indicators simultaneously thus supporting the visual identification of potential correlations among indicators. For example, in Figure 6 the low American GDP is accompanied by lower numbers of American tourists occupying beds both in Germany and Austria, around 2008-2009. The *table* shows the highest 250 values for the dimensions that are not fixed (middle panel, lower part), displaying one additional tab (one extra dimension: source) for the Eurostat and World Bank data, or two tabs for the TourMIS data (source and target). A choropleth map (left panel) for 2-dimensional indicators (World Bank, Eurostat) colored according to the respective indicator's performance

(darker colors corresponding to higher indicator values) allows easy assessment of top/low performing countries from that indicators perspective. The 3-dimensional indicators from TourMIS are represented on a geographical map by depicting connections between markets and destinations. The thickness of the connecting edges is proportionate to the measurements value and therefor allows for a visual comparison of the volume of tourist traffic.

Hovering over dots in the line chart and map displays tooltips with additional information and summaries (the totals for a selected location in the fixed period of time, for example). Clicking a row in the table or the highlighted location (colored country or a dot that represents a city, depending on the type of indicator, 2-dimensional or 3-dimensional) triggers a new search (a new slice with the time fixed to the same interval, focused on a new indicator with a new location).
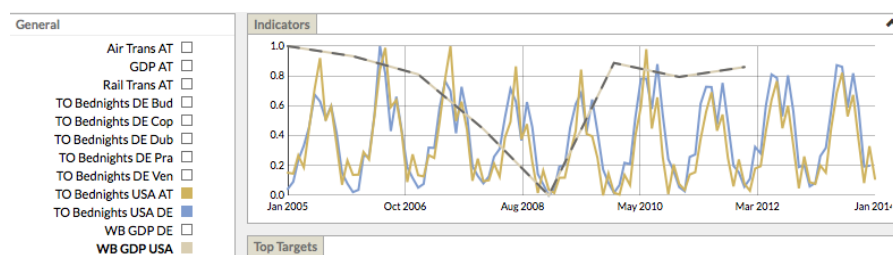

Fig. 6 Concurrent chart-based display of statistical indicators from different domain. The currently selected indicator (WB GDP USA) is displayed with dotted line.

Saving and sharing. The Export button from the horizontal upper menu, when activated, leads to a new panel for choosing the export and sharing options. Data can be saved in CSV or Excel formats, while the visualizations can be exported as pictures (SVG or PNG). This feature is important for sharing analysis data between applications or across project teams.

5 Evaluation

We performed an evaluation of the first version of the ETIHQ Dashboard. The goal of the evaluation is three-fold. Firstly, we aim to measure improvements in supporting the task of cross-domain decision making enabled by the data integration and visual analytics solutions described in the paper. Secondly, our evaluation also aims to collect additional task types that could be supported by the described infrastructure. Lastly, we wish to assess the usability of the visual dashboard and identify ideas for future improvements.

5.1 Evaluation Design

Participants. We have selected the participants to our experiments from the two major stakeholder groups that could benefit from a cross-domain decision support infrastructure: researchers in the tourism domain as well as tourism practitioners, working primarily for DMO's. From 28 contacted subjects, 17 replied positively and out of these 13 returned the evaluation forms (7 practitioners, 6 researchers). The participants were divided randomly in two groups (Group_A and Group_B), both containing an equal number and mixture of researchers and practitioners, i.e., 5 practitioners and 3 researchers per group. The final group setup is as follows: Group_A has 8 members (5 practitioners and 3 researchers), after the received evaluation forms, while Group_ B has 5 members (2 practitioners and 3 researchers).

Evaluation Setup. Prior to the evaluation itself, each participant received a tutorial consisting of an example based explanation of the tool features, as well as several hands-on exercises that ensured a basic familiarity with the tool (e.g., how to create and define indicators, how to visualise and compare their values).

We chose to perform the actual evaluation at the desk of each participant and at a time that best fitted the participant's schedule. This allowed to maintain a realistic work environment and to avoid bias potentially introduced by requesting the use of a new work environment, such as in lab-based evaluation settings.

The evaluation included four main activities, as follows:

- **Activity 1** - Participants performed three tasks using the Dashboard and recorded the time taken to perform each task and the results they have reached. See Table 3 for an overview of the tasks as well as their assignment to the two groups. Group_A performed tasks T1-T3 with the Dashboard, while Group_B focused on tasks T4-T6.
- **Activity 2** - To gather insights about potential new uses of our tool, participants were asked to create and perform two tasks of their own using the Dashboard. They noted the tasks they preformed and the insights they found.
- **Activity 3** - To collect information about how the evaluated tasks would be performed in a state of the art setting, participants performed three tasks without using the Dashboard. Since no base system exists participants were allowed to adopt their usual approach. Participants noted their findings, the time taken to perform each task as well as the approach and tools they made use of. In this activity, the role of the groups was inverted, with Group_A performing tasks T4-T6 without the Dashboard, while Group_B focused on tasks T1-T3.
- **Activity 4** - To get an insight into the usability of the tool, participants answered the ten questions that make up the System Usability Scale (SUS), the most used questionnaire for measuring perceptions of system usability [4]. Additionally, we collected their feedback about the most and least

useful features of the tool, as well as their recommendations for future extensions of the tool.

Evaluation Tasks. We specified a total of six tasks as detailed in Table 3. To measure the improvements in time savings and answer quality that the dashboard brings with respect to traditionally employed approaches, each task has been performed with the Dashboard (one group) and without it (the other group). We then measured, computed and compared group level averages for the time spent on each task as well as the correctness of the answers.

The tasks were formulated to cover two complex exploration scenarios. Task 1 to 3 investigate the influence of the American economy on bednights of American tourists to different European countries. Similarly, Tasks 4-6 cover an exploration of how Japanese economic indicators might influence arrivals of Japanese tourists to European cities. Both scenarios were investigated in the same time period, January 2005 and December 2014. The tasks covering each scenario increase in complexity and build in each other. Each scenario included one of the following types of questions:

– **Fact finding** - The first task in each scenario (i.e., Tasks 1 and 4) requested identifying a concrete fact by defining and visualising a tourism indicator, i.e., bednights and arrivals respectively.
– **Cross-domain indicator comparison** - The second task (i.e., Tasks 2 and 5) required investigating potential effects of an indicator from one domain (economy) over an indicator from another domain (tourism).
– **Benchmarking** - Finally, the last task of each scenario (i.e., Tasks 3 and 6) invited to explore whether similar conclusions about the effects of one indicator on another could be observed in the case of other entities of interest, namely, neighboring countries or cities.

5.2 Evaluation Results

We group the discussion of the evaluation results around two important themes. Firstly, we focus on a task-based evaluation of the tool, providing an insight into the performance improvement in terms of task execution time and answer quality when using the ETIHQ Dashboard, as opposed to using other approaches (Section 5.2.1). For this we contrast results obtained in Activity 1 and 3, when the same tasks were performed with and without the Dahsboard. We also report on state of the art approaches used as well as the types of tasks that participants created as part of Activity 2. Secondly, we conclude on the usability of the tool and the collected future features, based on the results from Activity 4 (Section 5.2.2).

5.2.1 Task Based Evaluation Results

Improvement of Task Execution Times. The left side of Table 4 sums up the average execution times per task and per participant group. For each

Table 3 Evaluation tasks and their assignment to the evaluator groups to be performed with and without the visual dashboard. All tasks were performed for the time period of January 2005-December 2014.

Task Nr.	Task Description	Dashboard Used? Group_A	Group_B
1	In which month was the number of bednights of tourists from the USA to Germany the highest?	Y	N
2	Explore possible similarities between the American economy (as indicated by GDP Growth) and the bednights spent by American tourists in Germany.	Y	N
3	Continuing your exploration from task 2, compare how Germany and Austria have been affected (in terms of bednights of American tourists) by the economic situation in the USA.	Y	N
4	In which month was the arrival of Japanese tourists to Vienna the lowest?	N	Y
5	Could the GDP of Japan have had any influence on the number of arrivals of Japanese tourists in Vienna?	N	Y
6	Continuing your exploration from task 5, compare how Budapest and Vienna have been affected (in terms of arrivals of Japanese tourists) by the economic situation in Japan.	N	Y

task and group we also show the completion ratio computed as the percentage of participants from a group that managed to complete a given task in less than 15 minutes. For example, task 4 was completed by 6 out of the 8 participants in Group_ A, thus leading to a completion rate of 75%. The average execution time per group and task was computed by excluding those participants who needed more than 15 minutes to complete a task. Times shown in bold are those measured when using the ETIHQ dashboard. Execution time improvement is computed as the ratio of improvement over the baseline time, which is the average execution time when not using the ETIHQ Dashboard. All numbers were rounded to the closest decimal. We conclude that the use of the ETIHQ Dashboard lead to lower execution times for all (but one) tasks when compared to the time required for answering without using the Dashboard. Positive improvements ranged from 8.82% to 46.43% and the average improvement was 34%. A negative improvement was recorded for task 4, where one of the Group_B participants needed 14 minutes to perform the task, as opposed to 2-6 minutes taken by the other group members.

Improvement of Result Quality The right side of Table 4 sums up the quality of answers per task and per participant group computed as a precision value, i.e., the ratio of correct answers over all answers. For all the evaluated tasks, the group using ETIHQ (values shown in bold) outperformed the group that did not make use of the system. Precision values when using the

Table 4 Task execution times (in minutes) averaged per participant group and completion ratios per task (Left). Answer quality (precision) averaged per participant group (Right). Values shown in **bold** have been measured when using the Dashboard.

| Task Nr. | Execution Times | | | Answer Quality | |
| | Group Averages (mins) | | Time Improvement | Group Averages | |
	Group_A	Group_B		Group_A	Group_B
1	**4.28 (100%)**	6.5 (80%)	34.1%	**71%**	20%
2	**2.85 (100%)**	6.25 (80%)	54%	**63%**	60%
3	**3.87 (100%)**	4.25 (80%)	8.82%	**75%**	60%
4	5.5 (75%)	**6 (100%)**	-10%	63%	**100%**
5	5.6 (63%)	**3 (100%)**	46.43%	50%	**80%**
6	7 (63%)	**4.6 (100%)**	34.28%	63%	**80%**
Average Time Improvement			28%		

ETIHQ Dashboard range from 63% to 100%, while baseline approaches lead to much lower performance, with the highest precision value (63%) reaching only inferior range of the quality that is obtained when using ETIHQ. We consider this a clear indication that the tool can positively influence the quality of the answers to be obtained.

Baseline Approaches. One of the goals of our evaluation was to get an insight into the currently used approaches for solving cross-domain decision tasks. Since the respondents were familiar with TourMIS, 8 out of 12 respondents used the TourMIS system in combination with other online databases. Two of the respondents used Google to search for the GDP of Japan. The others used systems that they already knew such as World Bank, OECD, Statistics Austria, and the Database of the German Federal Statistical Office.

Five of the respondents were able to provide a correct answer/analysis for all three tasks performed without the ETIHQ Dashboard (i.e., 38% of all participants). Four of these used TourMIS and the fifth relied on World Bank. Three respondents completed 2 out of 3 tasks correctly. One of them relied on the Database of German Federal Statistical Office, one used a combination Google and TourMIS and the other one used a combination TourMIS and the OECD database. Finally, for 4 of the respondents, this activity took more than 15 minutes for each task. Only one respondent who used TourMIS and Google got the correct answer. The rest of the three respondents were not able to find all the data necessary to answer the questions such as the ones that used Statistics Austria database.

New Tasks The goal of Activity 2 was to collect tasks that participants find interesting to perform with the provided system. Each participant was asked to think of and document two such tasks. As a result, we collected 22 tasks from the 13 respondents (one participant did not complete this activity while another two provided only one example each). The collected tasks fall under the following categories:

Fact finding. Eight tasks are fact finding tasks, focusing on a single indicator and aiming to determine its minimum/maximum values (e.g., date of highest rate of rail use in Austria, most visited European cities by South Ko-

reans; from Japan, France and Germany which is the biggest source market of tourist arrivals to Vienna), or its trends over time (e.g., is the number of arrivals from China to Vienna is constantly increasing?; arrivals to France from Nigeria; German bednights in Venice).

Trend comparison. Five tasks aim to compare the trends for an indicator for different values of one of its dimensions. For example, comparing bednights of British tourists in Vienna with the bednights of British tourists in Paris, which is a benchmark partner of Vienna or evaluating the guest mix share of Berlin with Barcelona. These are typical benchmarking tasks that are often performed with TourMIS.

Multi-indicator analysis is performed in 9 cases, therefore making use of the core novelty that our system offers. Four of these tasks focus on correlations between the economic situation in a country and arrivals/bednights from that market country to a certain destination (e.g., effect of the Russian crisis on visits from Russians to Amsterdam; Dutch arrivals in Austria and Dutch GDP; has the evolution of German bednights in Stockholm been influenced by the growth rate of the German GDP?; behaviour of US bednights to Brussels related to the growth of US economy). Two further tasks investigated relations between arrivals to a country on one hand and airtransport or CO_2 emissions on the other. One task investigated relations between two transport related indicators, namely, air transport usage and rail transport usage in Austria. Two of the provided tasks were more complex than the previous ones. Firstly, one investigated economic-touristic correlation and also attempted to compare these, e.g., "Do the economic situations in Germany and France influence the number of bednights of both markets in Amsterdam?". Secondly, a comparison of three indicators is also performed in "How did the evolution of the German GDP influence the domestic tourism of the country and the internal transport usage?"

We conclude that the largest category of tasks (41%) explores the novel features of the tool and goes beyond the tasks than can be performed typically with TourMIS (e.g., fact finding or trend comparison). While some users tended to use the tool as an alternative to TourMIS, others were quick to explore the cross-domain comparisons and even to propose more complex tasks than those included in the evaluation.

5.2.2 Tool Usability Results

Based on the responses to the SUS questionnaire (first part of Activity 4), we computed an overall SUS value of 63. Therefore, the system falls in the lower category of grade C systems, just bellow systems with average usability (SUS = 68). This result indicates that further improvements need to be performed in terms of system design (especially in terms of the items indicated by our respondents, see next), however, for an alpha version system we consider these results encouraging.

The second part of Activity 4 was dedicated to the current and future features that the users find/might find useful. Due to this fact, the answers were free form texts instead of agreement scores.

Most useful tool features. As expected, the most liked feature was the ability to integrate data from multiple data sources; to easily slices for the particular periods of time or geographical entities; or to preview and visualize these slices through multiple graphics. Some users also mentioned the ease of use of the system, the visualizations and the advanced customization features. Many users appreciated our system's export functionality as a first step in their data collection strategy for future research articles or projects.

Open challenges. While the possibility to create slices was appreciated, the terminology that we used for it (topic for indicator; top targets/top sources for other dimensions) and the suggested naming conventions were not easily adopted by users. They would have preferred a traditional terminology instead (indicator; market and destination, etc.). Explaining the QB visualization philosophy proved to be difficult, even though we have prepared a short tutorial that described it. Some changes to the data aggregation features were suggested in order to ease user workflow (aggregated annual data for tourism; less data in the top targets table; the possibility to change the granularity of the data in the line charts.

Suggestions for future developments of the system include the addition of various indicators or data sources: stock exchanges, flight connections, historical events, all the GDP indicators, exchange rates. Some of the data sets indicated by users are not statistical data sets (historical events), but a next iteration of the tool will also address this issue: indeed, we are already integrating in this tool various new media data sets collected previously.

Another requested upgrade was the addition of statistical functionalities (e.g., calculation of explicit correlations; regression analysis; more data summaries) to complement the current visual comparison and to reduce the need to change to other tools for detailed mathematical analysis. Of course, some users will always prefer to do their own math, but we nevertheless plan to include some of these statistical features. Other suggested improvements referred to various user interface changes (fonts, map changes, etc.) and an interactive tutorial built directly into the tool.

6 Related Work

Related to the work presented here are Decision Support Systems (DSS) used in the tourism domain as well as DSS that explore Linked Data technologies for data integration.

6.1 Decision Support Systems in Tourism

DSS research focuses on how to improve the efficiency of decision making and making effective decisions. In general a DSS tool has the following capabilities:

a database management system that can access both internal and external data and powerful but simple user interface design that enables interactive queries and graphics [28]. Although DSS are designed to facilitate decision making, it does not automate decision making process, but it supports it.

Gretzel stresses that tourism is one of the main domains for intelligent systems, which also include decision support systems (DSS). Due to the complexity of decisions in tourism it is essential to improve tourism DSSs in a way that enriches the domain rather than just simplifying the processes [9]. One type of DSS in tourism literature includes destination recommendation systems. Some examples are case-based travel recommendations [18], travel decision styles and destination recommendations [33], and creating adaptive recommender systems using neural network approaches [16]. In addition, there are some recommendation systems that have been developed such as DieToRecs [19], TourBO [8] and MobyRek [20].

Another aspect of decision support systems literature includes conceptual models of DSS and how these models can be created. The first examples of this type of research investigate how computers can help travel agents to assist in decision making for travellers [11,15]. More recently a study explains a conceptual framework of a web-based decision support system, which helps tourists visiting Nigeria to plan their trips and has the potential for tourist centres to save time, money and manpower [2]. A methodological approach to create a DSS for sustainable tourism is proposed in [1], which uses Geographic Information System (GIS) and Analytic Network Process (ANP) of Multi Criteria Evaluation (MCE) to identify the priority areas for tourism development. In a similar study, a web-based DSS is proposed that offers eco friendly trip recommendations for travellers [29].

However, most of the studies related to DSS are directed towards the consumer (i.e., the individual tourists) rather than the tourism managers therefore focusing on different content (i.e., destinations, touristic offers, events). Other types of DSSs such as yield management systems are available for tourism managers, which have been used primarily within the hospitality industry.

Furthermore there are DSS that exists for tourism managers. One example is BASTIS (www.bastis-tourism.info), which aims to give information about heritage tourism in the Baltic Sea Region. The data comes from different sources such as Eurostat, TourMIS, German Statistical Office, and etc. BASTIS can be used for benchmarking and for marketing the attractions or destinations in the specific region. However, the data maintenance is a problem since the data needs to be entered manually in most cases and the coverage of the system is only for Baltic Sea Region. PATA (mpower.pata.org/(Pacific Asian Travel Association) is another example of DSS that is useful for tourism managers and decision makers, which covers Asia Pacific countries. This system is only available to paid members. Overall, it has 80 national and regional tourism offices, airlines, hospitality groups and travel agencies as members. Some of the indicators include visitor numbers, expenditure of visitors, source markets, air passenger traffic, occupancy rates, RevPAR and forecasts or ar-

rivals and expenditure or visitors. The main issue of this DSS is that its a closed paid system and it covers one region of the world.

On the other hand there is TourMIS is an open and free tourism DSS, which includes data from over 100 European cities, 33 European nations and 19 source markets. This system is directly aimed at tourism organizations, managers, decision makers. The data in this system is entered by the official city or national tourism organization. The indicators covered include bednights, arrivals, and capacities. TOURMIS can also detect trends, has a hybrid forecasting function, and can calculate growth rates, and seasonality index among others. These different functions of TourMIS can help tourism decision makers in various ways. For instance, a DMO manager can look at the seasonality index of the destination to find out the low season months and can also benchmark its destination with its competitors to see if they have similar seasonality patterns. Based on these results the DMO manager can also investigate its competitors strategies for handling seasonality at their destinations and can apply them at his/her destination. Although TourMIS supports answering a variety of questions, it is limited to tourism specific indicators alone.

To conclude, tourism DSSs have the following drawbacks: (1) most of the tourism DSSs are designed for the demand side of the industry specifically for the tourists; (2) Some of the DSS can be used by the tourism supply side (i.e. tourism decision makers or managers) and may have data from different sources however this data is hard-coded and challenging to maintain; (3) the reviewed DSS systems focus on one domain tourism and hospitality, thus leaving other domains such as economics or sustainability, which also have an impact on tourism decisions.

6.2 Linked Data-enabled Decision Support Systems

Integrating information from multiple sources has been traditionally a difficult task before Linked Data and other NoSQL technologies [5]. Early LD visualization techniques were described in [6]. A recent survey of Semantic Web decision making tools [3] identifies several areas where these tools are lacking: a) the flexible integration and sharing of data (for example, interfaces that combine multiple modalities of exploring the data: textual, tabular, visualizations, while also enabling users to share their data); b) scalability and large-scale analytics. A number of projects and tools that use Linked Statistical Data (LSD) for visualizations and analytics were developed since the RDF Data Cube Vocabulary was introduced, but they still fall short of the requirements of Blomqvist, as we describe next.

Firstly, tools providing basic data visualizations are useful for having a quick look at the data. CubeViz [26] is an RDF Data Cube Browser which displays classical visualizations (line charts, pie charts, etc.). The OpenCube toolkit [13] also offers an OpenCube Browser for table-based views, and a tool for interactive map-based visualizations of geographical data. The automated

Visualization Wizard [17] considers vocabulary mappings, various combinations of dimensions and measures, and several visualization packages to be used for displaying the visualizations, but unfortunately these visualizations are not coordinated as in our tool (i.e., changes in one visualization pane do not reflect in related panes). Ba-Lam Do [7] developed Linked Widgets for displaying simple visualizations based on LSD on various websites.

Secondly, dashboards using multiple coordinated views and based on LSD are in their first stages of their lifecycle: they perform several functions well, but the integration of the various visualizations or features is not yet enterprise-grade. Sabol's paper [21] describes the extensions of the Visualization Wizard tool used for search and analysis over LOD, and analysis of scientific publications. LOD/VizSuite [30] only generates graph-based dashboards. Directly related to our use cases, the work on the OLAP4LD framework [14] is focused on querying and browsing LSD, and can be used as a foundation for complex dashboards. Our own tool belongs to this category, but provides a superior integration among visual elements as well as very fast response times.

Lastly, Linked Data Platforms (LDPs)[13] use REST APIs to deliver data in various formats and using various views. Some LDPs include maps or other types of visualizations (Bathing Water Quality[14] and Ordnance Survey[15], for example). Many LDPs are in fact used to publish QB datasets.

We conclude that most of the current visualization projects in the area of LSD are geared towards creating basic charts, and little effort is dedicated to complex analytic solutions that should be capable of answering a wide range of complex queries. Additionally, none of these projects have been related to tourism or specific application areas, most of them being thought of as general-purpose solutions. In this landscape, the ETIHQ Dashboard is not only the first tourism specific Linked Data enabled decision support tool, but also one of the most advanced tools that support Linked Data based analytics. The main differentiating feature from related systems is its ability to synchronize multiple visualizatios that use data from multiple datasets, in order to present complex use cases such as those described in [14] or [3].

7 Conclusions and Future Work

In this paper we focused on *cross-domain decisions in tourism*, namely on decisions that require answering questions or investigating scenarios taking into account data from diverse domains and data sources. We performed a practitioner's survey, an implementation of data integration with Linked Data and the development of a decision support prototype, which we then evaluated extensively. Based on these activities, we reached the following conclusions.

Cross-domain decisions are important tasks in the tourism domain. This has been confirmed several times during our work. Already during the

---

[13] http://www.w3.org/TR/ldp/

[14] http://www.epimorphics.com/web/projects/bathing-waterquality

[15] http://data.ordnancesurvey.co.uk/

initial practitioner's survey 70% of the respondents declared that they needed to perform such complex tasks during their work. During the evaluation of the Dashboard, in Activity 2, the most frequently proposed tasks were those involving cross-domain indicator analysis (i.e., 41%), showing an interest in this direction from the participant pool. This interest was then confirmed through comments made during Activity 4, in which respondents appreciated the most the support our tool offered in selecting indicators from different domains and concurrently visualizing those.

Tool support for cross-domain decisions is weak. Despite the increasing importance of such complex decision types, the practitioner's survey confirmed our hypothesis that, at this stage, workflows for taking such decisions involve tedious manual collection and combination of the data of interest. Such manual approaches to data collection and integration were performed by at least 74% of the survey respondents. This finding was strengthened during the tool evaluation, where the state of the art approaches reported as part of Activity 3 involved manual collection of data from TourMIS and other online data sources. As we have shown by contrasting Activity 2 and Activity 3, such approaches are both time consuming and error-prone.

Linked Data is a promising technology for solving data integration. To advance the state of the art from ad-hoc manual data integration processes, we proposed a data integration approach based on Linked Data technology. We exposed TourMIS as Linked Data and then connected it to various external datasets (WorldBank, Eurostats), thus providing a technological basis for quick and automatic integration of tourism data with statistics from other domains. We found the Linked Data technology sufficiently mature to implement our solution. Our survey also indicated that the tourism community is aware of this technology and has a good grasp of its basic principles. This indicates that a wide-spread adoption of LD in tourism could be possible, for example, by more data owners exposing and linking their datasets to the data nucleus we created in the ETIHQ project.

The ETIHQ Dashboard improves current tool support for complex decision making. Besides providing a solution to data integration, we also implemented a Decision Support System prototype through the ETIHQ Dashboard, which leveraged the integrated data. Our evaluation of this tool shows that it provides considerable improvements over manual approaches to answering a range of complex questions. Indeed, we obtained an average time improvement of 28%, while answer quality, in terms of precision, was clearly inferior when using manual approaches (under 63%) as opposed to using ETIHQ (over 63%). These positive results were obtained despite the fact that this was the first public use of the system and its usability is still under average as indicated by the results of the SUS questionnaire.

Future work will focus on including more indicators from additional data sources, thus moving towards a big data setting. We also wish to investigate more complex decision scenarios that can be supported with our solution. In terms of the Dashboard, we will improve its usability by addressing the

comments of the respondents and experiment with including additional visual analytics.

References

1. Aminu, M., Matori, A.N., Yusof, K.: A spatial decision support system (SDSS) for sustainable tourism planning in Cameron Highlands, Malaysia. In: 8th International Symposium of the Digital Earth (ISDE8) (2014)
2. Asefe, Y.N., Bolaji, A., Enaholo, A.E., Olubukola, O.: Web-based expert decision support system for tourism destination management in nigeria. International Journal of Advanced Research in Artificial Intelligence **2**(4), 59–63 (2013)
3. Blomqvist, E.: The use of semantic web technologies for decision support - a survey. Semantic Web **5**(3), 177–201 (2014). DOI 10.3233/SW-2012-0084. URL http://dx.doi.org/10.3233/SW-2012-0084
4. Brooke, J.: SUS-A quick and dirty usability scale. Usability Evaluation in Industry **189**, 194 (1996)
5. Cattell, R.: Scalable sql and nosql data stores. ACM SIGMOD Record **39**(4), 12–27 (2011)
6. Dadzie, A., Rowe, M.: Approaches to visualising linked data: A survey. Semantic Web **2**(2), 89–124 (2011). DOI 10.3233/SW-2011-0037. URL http://dx.doi.org/10.3233/SW-2011-0037
7. Do, B., Trinh, T., Wetz, P., Anjomshoaa, A., Kiesling, E., Tjoa, A.M.: Widget-based exploration of linked statistical data spaces. In: M. Helfert, A. Holzinger, O. Belo, C. Francalanci (eds.) DATA 2014 - Proceedings of 3rd International Conference on Data Management Technologies and Applications, Vienna, Austria, 29-31 August, 2014. SciTePress (2014)
8. Franke, T.: Tourbo: A prototype of a regional tourism advising system in germany. In: D.R. Fesenmaier, K.W. Wöber (eds.) Destination Recommendation Systems: Behavioural Foundations and Applications, pp. 257–280. CABI (2006)
9. Gretzel, U.: Intelligent systems in tourism. Annals of Tourism Research **38**(3), 757–779 (2011)
10. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993)
11. Hruschka, H., Mazanec, J.: Computer-assisted travel counselling. Annals of Tourism Research **17**(2), 208–227 (1990)
12. Isele, R., Jentzsch, A., Bizer, C.: Adding missing Links while consuming Linked Data. In: COLD Workshop (2010)
13. Kalampokis, E., Karamanou, A., Nikolov, A., Haase, P., Cyganiak, R., Roberts, B., Hermans, P., Tambouris, E., Tarabanis, K.: Creating and utilizing linked open statistical data for the development of advanced analytics services. In: S.C. et al (ed.) Second International Workshop for Semantic Statistics SemStats 2014. CEUR-WS (2014)
14. Kämpgen, B., Stadtmüller, S., Harth, A.: Querying the global cube: Integration of multidimensional datasets from the web. In: K.J. et al. (ed.) Knowledge Engineering and Knowledge Management - 19th International Conference, EKAW 2014, Linköping, Sweden, November 24-28, 2014. Proceedings, *Lecture Notes in Computer Science*, vol. 8876, pp. 250–265. Springer (2014)
15. Loban, S.C.: A framework for computer-assisted travel counselling. Annals of Tourism Research **24**(4), 831–834 (1997)
16. Mazanec, J.A.: Building adaptive systems: A neural net approach. In: D.R. Fesenmaier, K.W. Wöber (eds.) Destination Recommendation Systems: Behavioural Foundations and Applications, pp. 137–170. CABI (2006)

17. Mutlu, B., Höfler, P., Sabol, V., Tschinkel, G., Granitzer, M.: Automated visualization support for linked research data. In: S. Lohmann (ed.) Proceedings of the I-SEMANTICS 2013 Posters & Demonstrations Track, Graz, Austria, September 4-6, 2013, *CEUR Workshop Proceedings*, vol. 1026, pp. 40–44. CEUR-WS.org (2013)
18. Rici, F., Cavada, D., Mirzadeh, N., Venturini, A.: Case-based travel recommendations. In: D.R. Fesenmaier, K.W. Wöber (eds.) Destination Recommendation Systems: Behavioural Foundations and Applications, pp. 67–93. CABI (2006)
19. Rici, F., Fesenmaier, D.R., Mirzadeh, N., Rumetshofer, H., Schaumlechner, E., Venturini, A., Wöber, K.W., Zins, A.H.: Dietorecs: A case-based travel advisory system. In: D.R. Fesenmaier, K.W. Wöber (eds.) Destination Recommendation Systems: Behavioural Foundations and Applications, pp. 227–239. CABI (2006)
20. Rici, F., Nguyen, Q.N.: Mobyrek: A conversational recommender system for on-the-move travellers. In: D.R. Fesenmaier, K.W. Wöber (eds.) Destination Recommendation Systems: Behavioural Foundations and Applications, pp. 281–294. CABI (2006)
21. Sabol, V., Tschinkel, G., Veas, E.E., Höfler, P., Mutlu, B., Granitzer, M.: Discovery and visual analysis of linked data for humans. In: P.M. et al. (ed.) The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I, *Lecture Notes in Computer Science*, vol. 8796, pp. 309–324. Springer (2014)
22. Sabou, M., Arsal, I., Brasoveanu, A.: TourMISLOD: a Tourism Linked Data Set. Semantic Web - Interoperability, Usability, Applicability **4**(3), 271–276 (2013)
23. Sabou, M., Brasoveanu, A.: D26.1 Semantic Modelling of Tourism Indicators. Planet-Data Project Deliverable (2014)
24. Sabou, M., Brasoveanu, A., Onder, I: Supporting Tourism Decision Making with Linked Data. In: Proc. of the of the 8th Int. Conf. on Semantic Systems (I-SEMANTICS) (2012)
25. Sabou, M., Brasoveanu, A., Onder, I: Linked Data for Cross-Domain Decision-making in Tourism. In: Proc. of the of the 22nd ENTER Conference (2015)
26. Salas, P.E., Martin, M., Mota, F.M.D., Breitman, K., Auer, S., Casanova, M.A.: Publishing statistical data on the web. In: Proceedings of 6th International IEEE Conference on Semantic Computing, IEEE 2012. IEEE (2012)
27. Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intelligent Systems **21**(3), 96–101 (2006)
28. Shim, J.P., Warkentin, M., Courtney, J., Power, D., Sharda, R., Carlsson, C.: Past, present, and future of decision support technology. Decision Support Systems **33**(2), 111–126 (2002)
29. Sinleunam, P., Poolsawad, K., Tritilanunt, S.: A Web - Based Decision Support System For Eco Tourism Planning: A Case Study of National Park In Chiangmai Province. In: Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE) (2011)
30. Vocht, L.D., Dimou, A., Breuer, J., Compernolle, M.V., Verborgh, R., Mannens, E., Mechant, P., de Walle, R.V.: A visual exploration workflow as enabler for the exploitation of linked open data. In: D.T. et al. (ed.) Proceedings of the 3rd International Workshop on Intelligent Exploration of Semantic Data (IESD 2014, *CEUR Workshop Proceedings*, vol. 1279. CEUR-WS.org (2014)
31. Werthner, H., Ricci, F.: Future research issues in IT and tourism. Information Technology and Tourism (2015). DOI 10.1007/s40558-014-0021-9
32. Wöber, K.: Information supply in tourism management by marketing decision support systems. Tourism Management **24**(3), 241 – 255 (2003)
33. Zins, A., Grabler, K.: Destination recommendations based on travel decision styles. In: D.R. Fesenmaier, K.W. Wöber (eds.) Destination Recommendation Systems: Behavioural Foundations and Applications, pp. 94–120. CABI (2006)