PlanetData

Network of Excellence

FP7 – 257641

D26.1 Call 2: Semantic Modelling of Tourism Indicators

Coordinator: Marta Sabou, Adrian M.P. Braşoveanu

1st Quality reviewer: Elena Simperl

2nd Quality reviewer: Oscar Corcho

Deliverable nature:	R/O
Dissemination level: (Confidentiality)	(PU)
Contractual delivery date:	31.12.2013
Actual delivery date:	27.02.2014
Version:	0.16
Total number of pages:	32
Keywords:	Tourism indicators, TourMIS, semantic modelling, ontologies, provenance, PROV-O, statistical data, RDF Data Cube Vocabulary, linked data

Abstract

TourMIS is one of the core sources of European tourism statistics providing data about five tourism indicators, for over 150 European cities and for the last 19 years. The goal of ETIHQ is to publish the TourMIS data as high quality linked data. To ensure a high quality publication, we aimed to enable (i) high data interoperability through detailed semantic specifications and (ii) improved traceability by specifying data provenance. This deliverable describes the first step of this publishing process, namely the selection and the design of a suit of semantic models suitable for representing tourism indicators. A detailed presentation of the data available in TourMIS is followed by the description of the semantic modelling framework and its elements, which include, the RDF Data Cube Vocabulary model, various domain ontologies, models for specifying provenance information and detailed modelling of complex indicator types. These semantic models serve as a basis to the data publishing process to be performed in WP27.

Executive Summary

This deliverable describes a stack of semantic models necessary for publishing high quality linked data based on tourism indicators available in TourMIS, one of the core sources of European tourism statistics. Concretely, we worked towards the following objectives and obtained the results described next:

- *Understanding the TourMIS data*: we performed a detailed analysis of the TourMIS data and its various aspects. We concluded on some key data features to be taken into account during the publishing process such as the data's numeric nature, its heterogeneous origins and its high update frequency.
- *Exposing statistical data as linked data*: given the numeric nature of the TourMIS data, we have investigated the use of RDF Data Cube as a vocabulary for exposing it as statistical linked data. A concrete output here is a blueprint of a Data Specification Document for the arrivals indicator based on this vocabulary. We also took a number of URI design decisions for the upcoming data publishing phase.
- *Specifying provenance information*: because of the data's heterogeneous origin and our goal to ensure improved traceability, it was necessary to investigate solutions for specifying provenance information at a fine-grained, observation level as opposed to the entire dataset level as it is the current practice. The PROV-O vocabulary has been chosen to represent data provenance at micro level in terms of the providing user and the date of data creation. An example of how PROV-O will be used in conjunction with the RDF Data Cube vocabulary has also been provided.
- *Building tourism domain models by reusing existing tourism ontologies*: for achieving a high data interoperability we aimed to provide detailed semantic specifications. To that end, we have built a set of modular domain ontologies for capturing the domain knowledge available in TourMIS: a base ontology describing tourism indicators in general, an ontology for describing types of points of interest and an ontology for describing typical shopping items. In this process, we reviewed existing tourism ontologies to identify relevant parts that could be potentially reused. However, the lack of modular structure of those ontologies, their focus on tourism specific applications as opposed to modelling tourism indicators important for tourism managers as well as their lack of availability severely hampered the amount of knowledge that could be reused. To compensate for the lack of reusing tourism ontologies, we linked our ontology concepts to equivalent classes provided by three generic resources namely DBPedia, schema.org and Linked Geo Data.
- *Formally specifying the meaning of statistical indicators*: we have provided a detailed semantic specification of a variety of tourism indicators distinguishing between them in terms of the geographic region they considered as well as the type of accommodations that they took into account.

All the created semantic documents (ontologies and data structure definition) are available for download at: http://etihq.eu/wp-uploads/2014/01/ETIHQ_SemanticModels.zip.

Next steps consist in the use of the semantic models designed so far in the actual data publishing process. This will involve publishing the ontologies as linked data (i.e., by providing dereferenceble URIs for their elements) as well as potential adjustments and extensions of the selected semantic models to better fit the actual data available in TourMIS as well as the constraints of the chosen linked data publishing infrastructure.

Document Information

IST Project Number	FP7 - 257641	**Acronym**	PlanetData
Full Title	PlanetData /ETIHQ		
Project URL	http://www.planetdata.eu/		
Document URL			
EU Project Officer	Leonhard Maqua		

Deliverable	**Number**	D26.1	**Title**	Semantic Modelling of Tourism Indicators
Work Package	**Number**	WP26	**Title**	Semantic Modelling

Date of Delivery	**Contractual**	M39	**Actual**	M41
Status	version 0.16		final □	
Nature	prototype □ report □ dissemination □			
Dissemination level	public □ consortium □			

Authors (Partner)	MODUL University Vienna			
Responsible Author	**Name**	Marta Sabou	**E-mail**	marta.sabou@modul.ac.at
	Partner	MODUL University Vienna	**Phone**	

Abstract (for dissemination)	TourMIS is one of the core sources of European tourism statistics providing data about five tourism indicators, for over 150 European cities and for the last 19 years. The goal of ETIHQ is to publish the TourMIS data as high quality linked data. To ensure a high quality publication, we aimed to enable (i) high data interoperability through detailed semantic specifications and (ii) improved traceability by specifying data provenance. This deliverable describes the first step of this publishing process, namely the selection and the design of a suit of semantic models suitable for representing tourism indicators. A detailed presentation of the data available in TourMIS is followed by the description of the semantic modelling framework and its elements, which include, the RDF Data Cube Vocabulary model, various domain ontologies, models for specifying provenance information and detailed modelling of complex indicator types. These semantic models serve as a basis to the data publishing process to be performed in WP27.
Keywords	Tourism indicators, TourMIS, semantic modelling, ontologies, provenance, PROV-O, statistical data, RDF Data Cube Vocabulary, linked data

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
4.12.13	0.1	Marta Sabou	Draft Structure
9.12.13	0.2	Adrian Brasoveanu	URI Design, Ontology Reuse
10.12.13	0.3	Marta Sabou	ETIHQ Domain Ontologies
10.12.13	0.4	Adrian Brasoveanu	Cube Design, Provenance
11.12.13	0.5	Adrian Brasoveanu	Small Changes
12.12.13	0.6	Marta Sabou	Ontology Design
15.12.13	0.7	Adrian Brasoveanu	Cube Modelling, Provenance, Examples
21.12.13	0.8	Adrian Brasoveanu	Modelling Statistical Indicators
02.02.14	0.13	Marta Sabou	Finalised ontology models
03.02.14	0.14	Marta Sabou	Final complete draft.
04.02.14	0.15	Marta Sabou	First complete revision.
14.02.14	0.16	Marta Sabou	Post review changes.

Table of Contents

List of figures and/or list of tables

Abbreviations

CTO – city tourism organization

DSD – data structure definition

ECM – European Cities Marketing

ETC – European Travel Commission

LOD – linked open data

NTO – national tourism organizations

POI – point of interest

QB – RDF Data Cube Vocabulary

SDMX – statistical data and metadata exchange

VFR – visiting friends and relatives

WTO – World Tourism Organization

1 Introduction

Tourism statistics (also known as indicators) such as the number of tourists that arrive to and sleep at a destination are important for various decision making related tasks such as (i) understanding the contribution of tourism to the destination's economy [19] or (ii) promoting and marketing a destination by forecasting tourism demand, setting marketing goals and exploring potential source markets [20]. In addition, tourism planners and public agencies can use tourism statistics to decide on planning tourism related facilities and infrastructure such as airports, highways, bridges and water treatment facilities [20].

Given the importance of tourism statistics and their varied use, this project focuses on exposing as linked data the content of TourMIS, one of the core sources of European tourism statistics. Our goal is to publish high quality linked data in line with the quality dimensions defined by the PlanetData consortium [5], focusing in particular on:

- *Interoperability and understandability* - a dimension that refers to the extent to which the data is easily comprehended by the information consumer, human and machine alike. Good, machine-level interoperability is a prerequisite for building decision support systems capable of automatically making sense of and combining indicators from different sources and, therefore, it is our main focus. To achieve this aspect we focused on creating high quality semantic data models that would (i) take into account the statistic nature of the data; (ii) re-use existing vocabularies in the tourism domain or provide links to equivalent classes in well-know semantic models and (iii) provide an in-depth, machine interpretable description of the represented statistical indicators.
- *Traceability of data* is an important component of the verifiability quality dimension defined in [5], and relies on recording provenance information. This is achieved by modelling provenance information not only at dataset level but also at a more fine-grained, individual observation level.

Concretely, as part of work performed in WP26, we have achieved the following results reported in this deliverable. Firstly, we have performed a detailed overview of the TourMIS data and its characteristics to inform our semantic modelling (Section 2). Then, we have investigated the use of the RDF Data Cube Vocabulary for publishing tourism statistics and we have built a blue-print dataset specification (Section 4) complemented with provenance information encoded in terms of the PROV-O ontology (Section 5). Finally, we have turned our attention to modelling tourism domain knowledge. After a thorough overview of the existing tourism ontologies, we decided to build our own ontology by extending our earlier work in this area. We adopted a modular design and created three ontologies suitable to represent indicator information, point of interest (POI) information and shopping related information (Section 6). We also investigated a more in-depth semantic specification of the tourism indicators (Section 7) .

All the created semantic documents (ontologies and data structure definition) are available for download at: http://etihq.eu/wp-uploads/2014/01/ETIHQ_SemanticModels.zip and will be shortly made public using linked data principles as part of the data publishing step in WP27.

2 The TourMIS Data

TourMIS[1] is an online database that consists of tourism market research data in various countries and cities and which aims to provide comparable data that support tourism managers in their decision-making processes [7]. This section provides an overview of the data stored in TourMIS which will be exposed as Linked Data.

2.1 Core Tourism Indicators

TourMIS collects the following core tourism indicators: *arrivals* (the number of tourists that arrive to various types of accommodations at a destination), *bednights* (the number of bednights spent at various types of accommodations), and *capacities* (the total bed capacity at accommodations at a destination). The dataset about the three indicators is large being available from 1985 onwards, in relation with 154 European destinations (i.e., cities), for 19 different markets (markets refer to the origin of the tourists) as well as recording both monthly and annual measurements. Table 1 provides some concrete TourMIS data in a tabular format.

Table 1: Example monthly tourism arrival data for 3 destinations (Vienna, London, Amsterdam) and two markets (DE, US).

	DE				**US**			
	Jan’10	**Feb’10**	**…**	**Dec’13**	**Jan’10**	**Feb’10**	**…**	**Dec’13**
Vienna	1,500	1,245	…	2,675	4,560	3,193	…	3,128
London	2,345	2,141	…	6,508	5,354	2,831	…	4,962
A’dam	1,982	1,345	…	3,405	1,274	2, 529	…	3,865

When measuring all the indicators above, variations exist along two main dimensions. Firstly, the geographic extent considered by the indicator can be the *city area only* or, more broadly, the *greater city area* which also includes establishments in the immediate vicinity of the city. Secondly, indicators can be measured for different types of accommodations. Most generically, indicators can be measured for *all accommodation establishments including VFR*. (VFR stands for visiting friends and relatives), for *all forms of paid accommodation* and, most specifically, for *hotels and similar establishments*. Combinations of variations among these directions lead to 16 different measurement types. Annex A provides a description of each measurement as specified in TourMIS, while Table 2 sums up the mapping between the various values of the dimensions and the measurement’s code names.

Table 2: Overview of the various measurement types depending on geographic extent and the type of accommodation taken into account.

Geographic Extent	Type of Accommodation	Measurement Code
Greater city area	all accommodation establishments including VFR	AZS, NZS
Greater city area	all forms of paid accommodation	AAS, NAS, KAS
Greater city area	hotels and similar establishments	AGS, NGS, KGS
City area only	all accommodation establishments including VFR	AZ, NZ
City area only	all forms of paid accommodation	AA, NA, KA
City area only	hotels and similar establishments	AG, NG, KG

[1] http://www.tourmis.info/index_e.html

2.2 Arrivals to Points of Interest

Besides destination level tourism statistics, TourMIS also collects tourism indicators (namely number of tourists) for particular tourism sights also known as points of interest (POI). TourMIS distinguishes between 19 categories of POI's (see Figure 1) and currently contains data about over 330 concrete POIs, with over 250 POIs from Austria (see Table 3).

Museums or galleries
Churches and Monasteries
Important streets or hiking paths
Castles, ruins and palaces
Adventure/amusement parks and exhibitions
Natural Parks and reserves
Cable cars, elevators and similar
Ferries and boat excursions
Company/premises exhibitions/tours
Historic train rides
Theatres
Operas
Concert Houses
Historic birth places or residential premises
Zoos and other animal attractions
Hot springs, spas and water sport sights
Mines and caves
Towers and viewing spots
Memorial and cemeteries

Figure 1: Categories of points of interest in TourMIS.

Table 3: Sample data for points of interest (for year 2012)

POI	**Location**	**Nr. Visitors**
Abbey Geras	Austria	30,151
Albertina	Vienna, Austria	620,333
Biological Museum	Turku, Finland	12, 174
Blamey Castle	Cork, Ireland Rep.	329,000
Church of our Lady	Dresden, Germany	2

An important extension of our earlier work [3] consists in publishing POI information as linked data.

2.3 Shopping Data

As of 2009, TourMIS is also collecting information about typical tourist expenses, for example:

- Taxi from airport to city centre
- Public transport (one day ticket)
- Dinner, 3-course, no drinks, in 4-star hotel restaurant in city center
- Espresso in lobby bar of a 4-star hotel

There are 16 expense items about which yearly average price data is collected. This data is useful to compare how expensive one city is with respect to others and to understand whether this has any effect on tourism arrivals.

2.4 Data Features

The TourMIS data has several important features that will have an influence on the semantic modelling and publishing process, as follows:

- **Heterogeneous origins**: TourMIS data is provided by a supporting consortium, which ensures the continued data upload into the system. The consortium includes, among others, the National Tourism Statistics Austria, which collects data from the Austrian accommodation suppliers regarding key tourism indicators as well as European Cities Marketing (ECM) and European Travel Commission (ETC), which support the collection of measurements for tourism indicators by encouraging their members, city tourism organizations (CTOs) of over 100 European cities and national tourism organizations (NTOs) of 33 nations respectively, to enter their data into TourMIS. Therefore, TourMIS maintains a detailed record of data provenance: for each individual uploaded measurement (e.g., arrivals in Vienna from Germany in Jan’10), TourMIS records the user who made this addition as well as the time-point when the upload was made. It follows that the semantic models will need to record provenance not just at the dataset level but also at the level of individual observations.
- **Dynamic**: The TourMIS data has a high update rate, with new data being added almost daily, as can be seen at the TourMIS webpage[2]. This requires a data publishing infrastructure that caters for daily updates of the linked data set. We will keep this requirement in mind while working on WP27 and D27.1 "Call 2: The ETIHQ Repository" due in M43 of the project.
- **Mostly numeric**: TourMIS is a collection of tourism indicators, and therefore contains mostly numeric values. Given that all these indicators are defined along similar dimensions (time, destination, market) it makes sense to seek a modelling approach that is suitable for statistical data.

[2] http://www.tourmis.info/index_e.html

3 Overview of the Semantic Modelling of TourMIS Data

The Semantic Models of the TourMIS data will rely on a mix of generic and domain specific ontologies (depicted in Figure 2), as follows:

- **RDF Data Cube** - this vocabulary is the emerging RDF standard for modeling data of statistical nature and therefore will be used as a main vocabulary to publish the TourMIS data (Section 4);
- **PROV-O** - is the standard vocabulary for specifying provenance information and its concepts will be reused to model provenance information in conjunction with the RDF Data Cube (Section 5);
- **ETIHQ Domain Ontologies** - we also developed a set of ontologies specific to the tourism domain that will complement the generic models above with concepts describing the content of TourMIS (Section 6). These ontologies will also contain a detailed modeling of complex tourism indicators (Section 7).

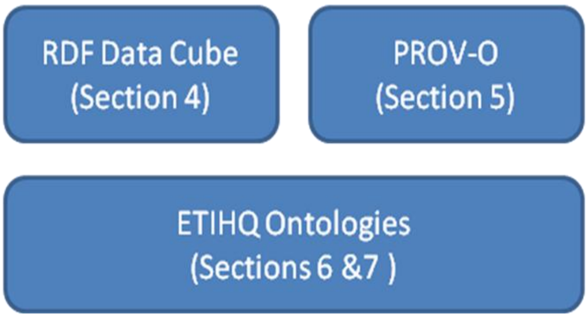


Figure 2: Overview of the ontology stack used for semantic modelling.

We will also make use of **DCAT** (http://www.w3.org/ns/dcat#) for specifying dataset level metadata as well as **XSD** (http://www.w3.org/2001/XMLSchema#) for specifying data types.

3.1 High-level URI Design

URI design plays an important role in ontology development, especially in those cases when a dataset is published as linked (open) data. While designing our URIs we made a first distinction between ontology elements and the actual dataset, which should be covered by URIs with different structure, namely, URI's that include the terms *ontology* and *dataset* respectively. Table 6 illustrates the most frequently used ontologies in our design, their namespaces and prefixes. The current {root} is http://www.etihq.eu, therefore an URI of type {root}/ontology/base would translate into http://www.etihq.eu/ontology/base.

Table 4: Namespace URIs and prefixes for the most frequently used ontologies.

Ontology	Namespace URI	Namespace prefix
RDF Data Cube	http://purl.org/linked-data/cube#	qb
PROV-O	http://www.w3.org/ns/prov#	prov
Base Ontology	{root}/ontology/base	ebo
Points Of Interest	{root}/ontology/poi	epoi
Shopping Items	{root}/ontology/shopping	esi
Dataset	{root}/dataset	eds

As we will discuss in Section 6, we adopt a modular ontology design and therefore decided to assign different URI to each ontology module. Within the dataset namespace, we will introduce a more fine-grained URI design as described in Section 4.4.

4 Modelling Statistical Tourism Data

Given the statistical nature of the TourMIS data, when exposing it as linked data we will make use of the RDF Data Cube Vocabulary (QB) [18] in order to meet the current standards in the field of statistical linked data publishing. As such, we adopt a new and more suitable approach as opposed to our earlier publishing of TourMIS, which did not consider the dataset's statistical features [3]. In this section we provide some background information about QB and detail how we have made use of it for modelling the TourMIS data.

4.1 Background - RDF Data Cube Vocabulary

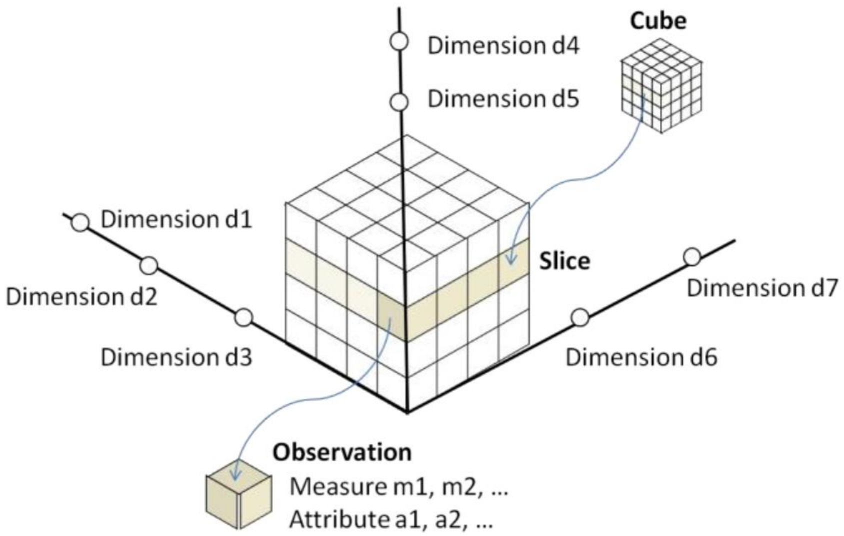
The RDF Data Cube Vocabulary is the current standard for publishing statistical data, and is a W3C Recommendation supported by industry and academia. QB has already gained acceptance by the community judging from the increasing number of statistical datasets published by using this vocabulary - see for example the datasets listed at http://wiki.planet-data.eu/web/Datasets. A further advantage of QB is that it is based on a cube model that is compatible with the SDMX standard (Statistical Data and Metadata Exchange) and designed to be general so that it enables the publishing of different types of multidimensional datasets.

The basic building blocks of the cube model are *measures*, *dimensions* and *attributes*, collectively referred to as *components*, and have the following roles (see Figure 3 for a graphical depiction of the QB elements):

- *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*.

Figure 3: The connection between cubes, dimensions, measures, slices and observations in RDF Data Cubes (adapted from [10]).



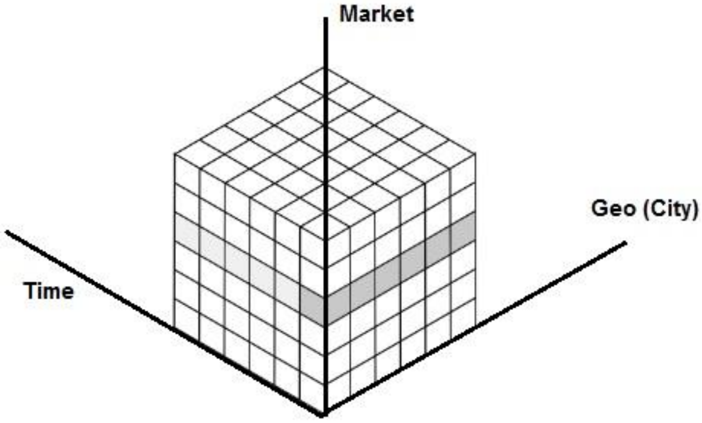


Figure 4 : Typical dimensions of the TourMIS data cubes.

4.2 Datasets

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 as depicted in Figure 4. Therefore the representation of TourMIS data using QB is quite straightforward.

The conceptual mapping between the elements of the TourMIS data and the primitives of the QB vocabulary is depicted in Table 5. 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 dataset. We believe that this modular design will allow a simpler re-use by end-users of the datasets that they are interested in.

Table 5: Datasets overview in terms of the main QB components and the key tourism concepts defined in the ETIHQ domain ontologies.

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

Note that the abstract QB notions of dimensions, measures and attributes are mapped to concrete domain concepts defined in the ETIHQ domain ontologies. In the remaining part of this report we will explain and exemplify the actual definition of the dataset.

4.3 An Example Dataset Definition

Annex B contains the listing of an example dataset definition using QB for the arrivals dataset.

The example starts with declaring a dataset (*eds:dataset_Arrivals*) and attaching to it various metadata such as comments, creators and creation time. This dataset is then assigned to a data structure definition object called *edss:dsd_Arrivals*.

The data structure definition (DSD) plays an important role as it provides a specification of the measures, dimensions, attributes and slices used by the dataset. Accordingly, our DSD defines the three main dimensions important for all arrivals measures (time, destination, market) and the fact that an arrival is measured. The attribute specifies that we are measuring a unit value which is always required.

```
edss:dsd_Arrivals a qb:DataStructureDefinition;
  # The dimensions
   qb:component  [ qb:dimension edss:dimensionTime];
   qb:component  [ qb:dimension edss:dimensionDestination];
   qb:component  [ qb:dimension edss:dimensionMarket];

   # The measure(s)
   qb:component [ qb:measure edss:measureArrivals];

   # The attributes
   qb:component [ qb:attribute sdmx-attribute:unitMeasure;
             qb:componentRequired "true"^^xsd:boolean;
             qb:componentAttachment qb:DataSet; ] ;
   .
```

Measures and dimensions must be defined on their turn. These definitions are a good place to provide additional comments relevant for these elements but also to specify their range as well as the domain concepts that they represent. For example, the market dimension (*edss:dimensionMarket*) has been linked using *qb:concept* to the Market concept defined in our base ontology (*ebo:Market*). The other dimensions and measures are defined in a similar manner. One exception is the time dimension which was not aligned to any domain concept but rather to a well-accepted sdmx concept. Also, we will use the data.gov.uk reference time service to provide values for this dimension.

```
edss:dimensionMarket  a rdf:Property, qb:DimensionProperty;
   rdfs:label "market dimension"@en;
   rdfs:range ebo:Market;
   qb:concept ebo:Market;
 .
```

Using the DSD structure, individual observations are created. We provide one example observation for representing the arrivals of 2134 Austrian tourists to London in July 2012. The destination and market dimensions properties range over two instances of type *ebo:City* (*eds:city_London*) and *ebo:MarketForCountry* (*eds:marketAT*) respectively. These instances are also defined in Annex B. The time dimension ranges over a data.gov.uk reference interval starting on the 1st of July 2012 and lasting for a period of 1 month (see "P1M" at the end of the URI). The measurement property *edss:measureArrivals* ranges over the number of arrivals.

```
eds:obs_Arrivals_London_AT_Y2012_M07
   a qb:Observation;
   qb:dataSet  eds:dataset_Arrivals;
   edss:dimensionDestination eds:city_London;
   edss:dimensionMarket  eds:marketAT;
   edss:dimensionTime <http://reference.data.gov.uk/id/gregorian-interval/2012-07-01T00:00:00/P1M> ;
   edss:measureArrivals        2134;
   .
```

At this point we did not make use of the slicing feature of QB, but left the definition of potential slices to a later time-point during the publication of the dataset.

4.4 URI Design in the Dataset Namespace

In Section 3.1, we have defined URIs for our ontologies and for the dataset, considering the root URI to be http://www.etihq.eu/. Within the dataset namespace, we distinguish between entities that denote data set specific information (under URI eds = http://www.etihq.eu/dataset) and those that denote structuring information, for example, the DSD and its associated dimensions and measures (these have a namespace URI denoted edss = http://www.etihq.eu/dataset/structure/).

For each entity type in the dataset, we established patterns for defining their local names. These local name patterns are usually built up from a string representing the type of the entity, then a connector (_) and then the name of that entity as it appears in TourMIS. Observation entities have a more complex local name, consisting of the string "obs" and followed by the values of the measurement type, destination, market, year and month in which the observation was made. If one of these elements is missing (for example, capacity measurement do not have a market), then that element is simply not included in the local name. Table 6 lists the URI patterns for each entity type in the ETIHQ dataset.

Table 6: ETIHQ Data Cube URIs

Entity Type	**URI Pattern**
qb:DataSet	{root}/dataset/dataset_{name of dataset as per Table 5}
qb:DataStructureDefinition	{root}/dataset/structure/dsd_{name of dataset as per Table 5}
qb:DimensionProperty,	{root}/dataset/structure/dimension_{dimension name}
qb:MeasureProperty,	{root}/dataset/structure/measure_{measure name}
qb:Observation	{root}/dataset/ obs_{MeasurementType}_{Destination}_{Market}_Y{Year}_M{Month}

5 Modelling Provenance Information

Traceability of data is an important component of the verifiability quality dimension defined in [5] and relies on recording provenance information. Although this information is recommended for high quality LD, according to a recent survey of the LD sets available in TheDataHub.org, only about 35% of them specify provenance information, typically focusing on the document level [4].

For datasets that aggregate data from a variety of contributors, specifying provenance both at document and data item level is an important task. For example, TourMIS records provenance information in terms of who and when has added a set of observations to the system. Transferring this information to the linked data version of the data source is important not only in order to reflect the content of TourMIS with a high fidelity and to ensure data traceability, but also to enable the creation of a set of advanced features in the prospective applications that will be built on top of this dataset. These features could include, for example:

- *provenance based licensing* - attaching licensing information not only to the dataset as a whole, but rather to certain parts of the dataset depending on the provenance of that part. Given the heterogeneous ownership of the TourMIS data, a scenario in which different data owners have different licensing constraints is highly likely.
- *selective visualisations* - provenance metadata is a pre-requisite for building systems that allow visualising parts of a dataset based on its origin (e.g., only data uploaded by persons from Vienna) or freshness (e.g., showing only data recorded after a given time point).

The ETIHQ linked data repository will record provenance information not just at data source but also, at the more fine-grained, data item level, thus going beyond usual practice. In order to capture provenance information we will make use of PROV-O[3] [17]. PROV-O models the relations between three types of concepts: entities, activities, and agents where entities are the things about which provenance is recorded, activities are processes that lead to the creation or modification of entities and agents are involved in various activities acting upon entities.

Table 7: Important provenance concepts to be used in the ETIHQ repository.

PROV-O Concept or Property	Meaning and usage in ETIHQ
prov:Entity	The thing for which provenance is recorded, in ETIHQ these are individual observations.
prov:Person	A type of *prov:Agent* representing persons. Used to model persons that provide data to TourMIS.
prov:Organisation	A type of *prov:Agent* used to model organisations. An organisation represents a social or legal institution, a university, etc. Used to model organisations that provide data to TourMIS.
prov:generatedAtTime	The time at which an entity was completely created and is available for use (domain *prov:Entity*, range *xsd:dateTime*). Used to record the time when a certain observation was added to the TourMIS database.
prov:wasAttributedTo	Used to relate (attribute) and entity to an agent (domain *prov:Entity*, range *prov:Agent*). Used to relate the person who creates an observation in TourMIS to that observation entity.
prov:actedOnBehalfOf	Delegation of an agent to carry out a task on behalf of someone else (another agent, himself). Used to associate the persons who provide the data to the various organisations on whose behalf they act.

In TourMIS, the entities about which provenance is recorded are individual observations. Therefore each observation in our dataset is modelled as a *prov:Entity*. The activity of creating the entities is not of particular interest for the case of TourMIS and therefore we will not create any activity instances. Instead, we

[3] http://www.w3.org/TR/prov-o/

will associate observation entities directly to the persons that created them (by using *prov:wasAttributedTo*) and to the date of their creation through *prov:generatedAtTime*. Table 7 provides further details about the PROV-O concepts and properties that we will make use of to represent provenance information in the ETIHQ repository.

The following data snippet illustrates the use of these concepts at the data level by augmenting the example observation from Section 4 as follows (newly added declarations are shown in bold face):

- declaring the observation of type *prov:Entity*, therefore indicating that provenance will be assigned to this individual data element;
- attributing the observation to a creating person using *prov:wasAttributedTo*;
- specifying the time of creation for the observation with *prov:generatedAtTime*;
- defining a *prov:Person* and a *prov:Organisation* instance as well as a *prov:actedOnBahalfOf* property between these instances.

```
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix qb: <http://purl.org/linked-data/cube#>.
@prefix eds:   < http://www.etihq.eu/dataset/> .
@prefix edss:   < http://www.etihq.eu/dataset/structure/> .

@prefix ebo:   <http://www.etihq.eu/ontology/base/>.

eds:obs_Arrivals_London_AT_Y2012_M07
   a qb:Observation, prov:Entity;
   qb:dataSet  eds:dataset_Arrivals;
   edss:dimensionDestination eds:city_London;
   edss:dimensionMarket   eds:marketAT;
   edss:dimensionTime <http://reference.data.gov.uk/id/gregorian-interval/2012-07-01T00:00:00/P1M> ;
   edss:measureArrivals        2134;
   prov:wasAttributedTo eds:person_adrian;
   prov:generatedAtTime "2014-03-02"^^xsd:date;
     .

eds:person_adrian
  a prov:Person;
  foaf:givenName       "Adrian";
  prov:actedOnBehalfOf  eds:modul_university;
.

eds: modul_university
  a prov:Organisation;
  foaf:name       "MODUL University Vienna";
.
```

6 The ETIHQ Domain Ontologies

Detailed semantic specifications and links to equivalent concepts in relevant external resources contribute to data interoperability as a quality feature identified by [5]. A core component of our semantic model is a domain ontology for describing the tourism indicators covered by TourMIS. To that end, we will extend an ontology for describing tourism indicators which we developed earlier and described in [3].

6.1 Re-use from Existing (Tourism) Ontologies

The ontology described in [3] focuses on the Arrivals, Bednights and Capacity indicators and has no concepts suitable for modelling points of interest or shopping related information. Therefore, before extending it with this information we have conducted an overview of existing tourism ontologies that could potentially provide some concepts to be re-used in our modelling.

Table 8: Overview of Existing Tourism Domain Ontologies and their domain coverage.

Ontology	**Top Concepts**	**Covered Domains**
QALL-ME http://qallme.fbk.eu/qallme-tourism4.0.owl	Contact, Country, CreditCard, Currency, Destination, Event, EventContent, Facility, Genre, Language, Location, Period, PersonOrganization, Price, Room, Site, Transportation	Transportation Events
Hontology http://ontolp.inf.pucrs.br/Recursos/downloads/Hontology/20120417.owl	Accomodation, Facility, Room, Service, Staff, GuestType, Design, Meal, PointOfInterest, Price, Rating	Accommodation
Harmonise	Events, Attractions, Accomodations, Restaurants	Events
TGProton http://goodoldai.org/ns/tgproton.owl https://sites.google.com/site/ontotravelguides/Home/ontologies	Abstract, AccommodationRating, Adventure, Adventurer, Airline, AirplaneService, BoardService, Bunjee_jumping, BusinessAbstraction, ClassicTourist, Clubbing, Clubbing-type, Company, ContactInformation, Destination, Facility, GeneralTerm, Group, HomePage, Hotel, Language, Location, NaturalPhenomenon, Number, Offer, Roadway, Room, Service, Sightseeing, SocialAbstraction, TemporalAbstraction, Topic, TouristOffer, TouristOrganization, TransportFacility, Traveler, User, UserProfile	Users Destination
OnTour	Accommodation, Activity, Contact Data, Date Time, Event, Infrastructure, Language, Location, Room, Ticket	Period / Date Time
ACCO – Accomodation Ontology http://ontologies.sti-innsbruck.at/acco/ns.html	Room, Hotel, RoomOffer, CompoundPrice	Accomodation, Hotel Room Types and Features, Compound Prices
Hi-Touch	Documents, Objects, Publications	Tourism Documents
Hotel Search Ontology	Hotel, POI, Transportation, Features,	Hotel Search

	HotelInformation	
Task Tourism Domain Ontology	Accommodation, Attraction, Entertainment, Festival/Event, Food, History/Culture, Location, Shopping, Transportation, Weather	Tourism Attractions

Table 8 lists the tourism ontologies that we have inspected, providing examples of some of their top concepts and also indicating some of the domains that those ontologies cover. Our detailed observations about each ontology are as follows:

- QALL-ME[4] is an ontology designed to be used for question answering (QA) in the tourism domain and for benchmarking NE tagging (named entities tagging). It provides a model to describe tourism destinations, sites, events as well as transportation [2]. It is well aligned with upper ontologies such as WordNet and SUMO, but does not provide concepts reusable in our modelling.
- Hontology is a recent addition to the tourism ontology stack. It was created for multilingual usage and the paper that introduced it provides several examples of alignment with other tourism ontologies (QALL-ME) and general ontologies (DBPedia, Schema.org) [11]. It is a good ontology for modelling accommodation details: for example, the concept of Facility contains many popular hotel features ranging from Mini-Bar to Wi-Fi. Due to its focus on accommodation modelling, it is not currently useful for our modelling requirements.
- Harmonise was started more than a decade ago with the purpose of creating an interoperable tourism marketplace, but it was soon abandoned in favour of other ontologies [8]. Large tourism organizations such as WTO (World Tourism Organization), as well as national organizations (France, Spain, etc) were involved in its design. The Harmonise ontology focused on tourism events and accommodation types but unfortunately it is not available online anymore.
- TGProton (Travel Guides Proton) is an extension of the Proton Upper Ontology [12]. It was developed a decade ago at Sheffield and used for Travel Guide applications. While it has lots of concepts, and it can be used to model user behaviours in tourism, it does not contain concepts relevant for our application domain.
- OnTour, a precursor to the current eClass and GoodRelations tourism extensions, focuses primarily on Date and Time modelling [13].
- The Hi-Touch ontology models tourism destinations and their associated documentations [9]. It was in use mostly in France, being created by the Mondeca group and improved with concepts from the WTO Thesaurus. Its focus is disjoint from that of our application domain and therefore we do not make use of it.
- The Hotel Search Ontology is part of a set of ontologies designed with the use case of hotel search in mind (Person, Reisewissen, etc) [14]. Information about hotels is modelled well, but it lacks any connectors to country level statistics, and therefore also this ontology is not suitable for our needs.
- The Task Tourism Ontology focusing on modelling tourism attractions and it is not fit for modelling country level statistics [15].
- The ACCO Accommodation Ontology[5] is an extension of the Good Relations ontology and focuses on describing accommodation types and their features (e.g., hotel rooms, hotels, camping sites) as well as on modelling compound prices specific for the tourism sector (e.g., weekly cleaning fees) [16]. From the point of view of the tourism indicators collected by TourMIS, no distinctions are made between various accommodation types, that is, indicators such as *Capacity* are reported for all accommodation types at a destination. There is also no complex notion of prices, except the simple price notion related to shopping items. Therefore, at this point, we do not make use of the ACCO ontology in our modelling.

[4] http://qallme.fbk.eu/qallme-tourism4.0.owl
[5] http://purl.org/acco/ns#

The conclusion of our overview is 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 [3], 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 the relevant modules.

6.2 Base Ontology

The base ontology has the "http://www.etihq.eu/ontology/" namespace (prefix ebo that stands for "etihq base ontology") and has been developed as an extension of the TourMIS ontology[6] described in [3]. 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 5. 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 corresponding ontologies described in Sections 6.3 and 6.4 respectively.

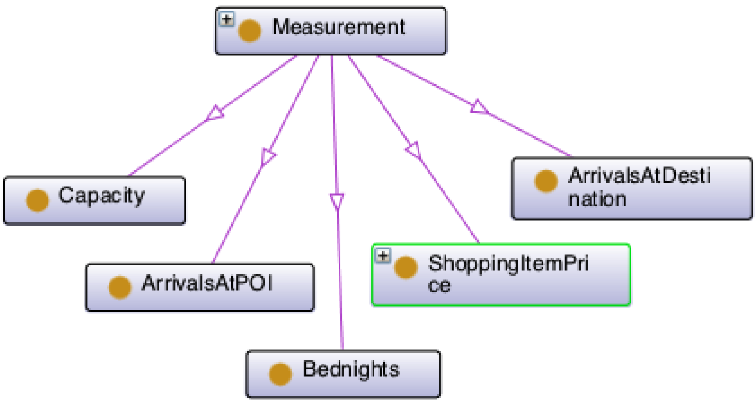


Figure 5: The Measurement concept and its subclasses in the base ontology.

Additionally, each measurement has a set of dimensions, which have been modelled as Data or Object Properties. Table 9 provides an overview of the dimensions valid for each measurement and how these were modelled with corresponding ontology properties. Accordingly, every *Measurement* has a value – and in fact we have enforced a cardinality restriction of 1 for the *hasValue* property. Two properties (*measuredForYear* and *measuredForMonth*) record the temporal dimension of a *Measurement*. Although currently only two out of the five indicators contain monthly values in TourMIS, we have defined *measuredForMonth* for all measurements – since there is no cardinality restriction on this property, it can be used only when appropriate. Furthermore, this modelling, supports any future extensions of the system that would provide monthly measurements for the other three indicators. The object property *isAboutDestinationCity* allows specifying the destination city about which a measurement has been made, and applies to all *Measurement* types. Note that TourMIS does not explicitly provide the destination city when reporting measurements about POIs, however, we plan to detect this information from external sources and provide it as part of the dataset. The property *isAboutPOI* specifies the *PointOfInterest* relevant for a given *ArrivalsAtPOI* measurement. Two measurements, *Arrivals* and *Bednights* also specify a market from which the tourists arrive. This information is recorded with the *iaAboutMarket* property. Finally, *hasCurrency*, specifies the currency for the *ShoppingItemPrice*.

[6] http://tourmislod.modul.ac.at/TourismOntology.owl

Table 9: Overview of measurements, their dimensions and the corresponding ontology modelling.

	Measurement					Corresponding Ontology Properties		
Dimension	**Arrivals**	**Bednights**	**Capacity**	**ArrivalToPOI**	**Shopping ItemPrice**	**label**	**domain**	**range**
Value	Y	Y	Y	Y	Y	hasValue	Measurement	decimal
Year	Y	Y	Y	Y	Y	measuredForYear	Measurement	int
Month	Y	Y	-	-	-	measuredForMonth	Measurement	int
Destination	Y	Y	Y	-	Y	isAboutDestination City	Measurement	City
POI	-	-	-	Y	-	isAboutPOI	ArrivalsAtPOI	PointOfInte rest
Market	Y	Y	-	-	-	isAboutMarket	Union(Arrivals, Bednights)	Market
Currency	-	-	-	-	Y	hasCurrency	ShoppingItemPr ice	Currency

6.3 Points of Interest Ontology

Point of interest information has been incorporated in the ETIHQ semantic models in the following two ways.

Firstly, as described in the previous section, the *ArrivalsAtPOI* indicator has been added as a specific type of *Measurement*. From the properties of its superclass, *ArrivalsAtPOI* instances will make use of the *hasValue* and *measuredForYear* data properties. Specific to this indicator is the newly added object property *isAboutPOI* which ranges over a *PointOfInterest* type and allows specifying a concrete POI about which the measurement is done. The PointOfInterets concept is defined and extended with relevant specific concepts in the POI ontology, which is imported into the base ontology.

The second modelling element is the POI ontology declared under the following namespace: http://www.etihq.eu/ontology/poi/. The aim of this ontology is to model concepts corresponding to the POI categories in TourMIS, which appear as specialisations of the *PointOfInterest* concepts.

Most POI categories in TourMIS have a composite nature in the sense that they cover two or more atomic POI types. Good examples in this sense are categories such as "Churches and Monastries" or " Castles, ruins and palaces". When modelling these categories our goal was to create concepts that closely resemble them, however, creating such complex concepts made it impossible to find similar concepts in other ontologies to which we could establish links or which we could re-use. As a result, we decided that we would create a concept corresponding to each individual element of a category and then create a class based on their union to reflect the composite TourMIS category. Continuing the examples above, we created the concepts *Church, Monastry, Castle, Ruin, Palace* and then the concepts $ChurchOrMonastry \equiv Church \cup Monastry$ as well as $CastleOrRuinOrPalace \equiv Castle \cup Ruin \cup Palace$. This modelling closely reflects the TourMIS categories while explaining, to some extent, the meaning of these composite terms. Additionally, the atomic concepts have been linked to concepts from DBPedia and Schema.org using owl:equivalentClass. Table 10 provides an overview of the TourMIS POI categories, the created ETIHQ concepts (both atomic and composed) as well as corresponding equivalent classes in the three external resources.

Table 10: Overview of POI concepts and their corresponding links external resources.

TourMIS Category	Concepts in the POI Ontology	Linking		
		DBPedia	Schema.org	linkedgeodata
Museums or galleries	Museum, Gallery, MuseumOrGallery	Museum	Museum	Gallery, museum
Churches and Monasteries	Church, Monastry, ChurchOrMonastry	Church, Monastery	Church	Church, monastery
Important streets or hiking paths	ImportantStreet, HikingPath, ImportantStreetOrHikingPath	-	-	-
Castles, ruins and palaces	Castle, Ruin, Palace, CastleOrRuinOrPalace	Castle	-	Castle, palace, ruins
Adventure/ amusement parks and exhibitions	AmusementPark, Exhibition, AmusementParkOrExhibition	Amusement_park, Exhibition	AmusementPark	-
Natural Parks and reserves	NaturalPark, NaturalReserve, NaturalParkOrNaturalReserve	Park	Park	Natural_reserve, park
Cable cars, elevators and similar	CableCar, Elevator, CableCarOrElevator	Cable_car, Elevator	-	cable_car, elevator
Historic train rides	HistoricTrainRides	-	-	-
Theatres	Theatre	Theatre	-	theatre
Operas	OperaHause, OperaPerformance, OperaHauseOrOperaPerformance	-	-	-
Concert houses	ConcertHouse	-	-	Concert_hall
Historic or architectonic places or premises	HistoricPlaceOrPremise	HistoricPlace	LandmarksOrHistoricalBuildings	
Zoos or other animal attractions	Zoo	Zoo	-	zoo
Hot springs, spas and water sport sights	HotSpring, Spa, WaterSportSight, HotSpringOrSpaOrWaterSportSight	HotSpring, Spa	-	-
Mines and caves	Mine, Cave, MineOrCave	Cave	-	Mine, cave
Towers and viewing spots	Tower, ViewingSpot, TowerOrViewingSpot	-	-	tower
Memorial and cemeteries	Memorial, Cemetery, MemorialOrCemetery	Memorial, Cemetery	Cemetery	Memorial, cemetery

In terms of concept reuse, we explored the following semantic resources:

- Linked Geo Data's ontology [6], DBPedia and schema.org contain a wide range of POI concepts and we have established links from our atomic concepts to all of these resources, whenever possible, as shown in Table 10. Schema.org provides less instance information that the other two resources, however, extracting the right entity type form DBPedia and Linked Geo Data can be challenging

[21]. An example application that makes use of Linked Geo Data and where POI information is collaboratively edited is described in [22].

- GeoNames contains a large amount of primarily instance information which we will use to link our instances to in WP27. GeoNames is a primary source of geographical information in many information systems and more reliable than DBPedia in the geographic area.
- The Places Ontology[7] is a simple vocabulary of terms denoting POIs of geographic interest (e.g., Island, Gulf, Hill, Lake), which are currently not covered by TourMIS statistics that rather focus, almost entirely, on man-made structures as POIs.

6.4 Shopping Ontology

For modelling shopping information, extensions have been made to the base ontology and a new ontology has been created to model shopping related concepts (Shopping ontology). The base ontology imports the Shopping ontology.

In terms of extensions to the base ontology, the *ShoppinItemPrice* concept was added as a new type of *Measurement* (subClassOf: Measurement). This concepts inherits from its parent class a set of properties relevant for statistical indicators in the tourism domain, such as *isAboutDestination*, *hasValue*, *measuredForMonth*, *measuredForYear*. Since the value of this indicator is a monetary value measured in a certain currency we also created the *hasCurrency* property which ranges over the newly created *Currency* concept for which, for now, we have declared one instance, namely *EUR*. The *isAboutShoppingItemType* property relates this indicator to the shopping item whose price it represents and ranges over the *ShoppingItem* concept, which is defined in a dedicated ontology module, the Shopping ontology.

The Shopping ontology under the http://www.etihq.eu/ontology/shopping/ namespace defines the *ShoppingItem* concept and its main subclasses as available in TourMIS. We created a class for each shopping item defined in TourMIS and then categorised these along four main categories, namely, *Accommodation*, *Entertainment*, *FoodAndBeverage* and *Transport*. Figure 6 depicts the class hierarchy of the Shopping Ontology.

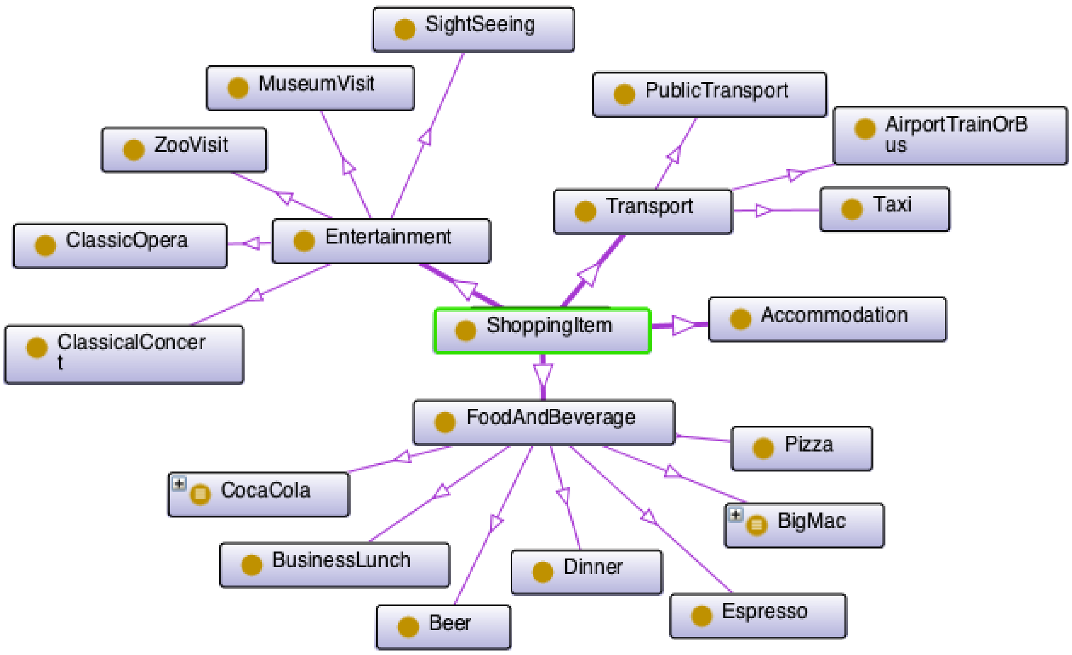


Figure 6: The Shopping ontology. Created with Protege's OntoGraf plugin (note: the type of links is *hasSubclass* with the arrows pointing from the more generic concept towards the more specific one).

[7] http://vocab.org/places/schema.html

7 Detailed Modelling of Indicator Types

As described in Section 2.1, a broad space of indicator varieties exist derived from the basic indicators and varying along several dimensions, primarily in terms of the extent of the geographic region in which they are measured as well as the types of accommodations that they take into account.

The detailed semantic modelling of such indicator varieties is important in order to automatically assess how various indicators compare to each other and therefore to be able to decide whether they are the same or at least comparable. This kinds of comparisons are pre-requisites for building (semi-)automatic solutions for combining tourism indicators from various data sources.

In the TourMIS system, indicator types are simply specified by adding a comment field to the output of the system, which specifies through a textual message the exact meaning of the indicator. See Annex A for an overview of such text messages. In our earlier modelling [3], we adopted a similar approach by simply stating the type of the indicator as an “Exception Definition”, modelled as a data type property of the *Measurement* concept ranging over a text field containing the definition of the indicator.

As part of ETIHQ we advanced earlier work by a more detailed modelling of the meaning of indicators. In particular, for the Measurement concept we defined two properties:

- *measuredForGeographicReach*, with range the newly introduced concept *GeographicReach*. We also defined two subconcepts for *GeographicReach*, namely *CityArea* ⊂ *GreaterCityArea* ⊂ *GeographicReach*.
- *measuredForAccommodationType*, with range the newly introduced concept *Accommodation*. The following hierarchy of *Accommodation* has been defined: *HotelAndSimilarEstablishment* ⊂ *PaidAccommodation* ⊂ *AccommodationIncludingVFR* ⊂ *Accommodation*.

Using these properties and classes, new indicator types can be defined by constraining the range of these properties. For example, the AA type arrivals indicator defined as “Arrivals in all paid forms of accommodation establishments in city area only” is defined as (Manchester Syntax):

```
Class: ebo:Arrivals_AA
  EquivalentTo:
    ebo:measuredForAccommodationType only ebo:PaidAccommodation,
    ebo:measuredForGeographicReach only ebo:CityArea
  SubClassOf:
    ebo:ArrivalsAtDestination
```

Similarly, the broader arrivals indicator AAZ, defined as "Arrivals in all accommodation establishments incl. VFR in greater city area" is defined by setting property restrictions on broader classes than AA, namely:

```
Class: ebo:Arrivals_AZS
  EquivalentTo:
    ebo:measuredForGeographicReach only ebo:GreaterCityArea,
    ebo:measuredForAccommodationType only  ebo:AccommodationIncludingVFR
  SubClassOf:
    ebo:ArrivalsAtDestination
```

With these definitions in place, a reasoning mechanism is able to detect that Arrivals_AA ⊂ Arrivals_AZS, and therefore that the values of AA for a given observation setting should be always smaller or equal to the values of the AZS indicator for the same setting.

8 Conclusions and Future Work

We hereby described the semantic models that provide the necessary foundation for the upcoming data publishing and linking step in WP27. The concrete outcomes from this work are:

- a blueprint dataset specification based on QB and PROV-O which illustrates how these two vocabularies will be used in conjunction with domain concepts from our own ontologies. Available in the *etihq.ttl* file;
- three tourism domain ontologies for specifying tourism statistical measures (base ontology), POIs (poi ontology) and shopping item types (shopping ontology). Available as three owl files for the base ontology (EtihqOntology.owl), the POI ontology (POIOntology.owl) and the Shopping ontology (ShoppingOntology.owl);
- a potential modelling of indicator varieties based on their different intent. Available in the base ontology.

The four semantic models are available at: http://etihq.eu/wp-uploads/2014/01/ETIHQ_SemanticModels.zip.

We can draw several conclusions from our work. Firstly, as expected, mapping the TourMIS data into the QB formalism was straightforward, although it took more effort than expected to understand all the technicalities of the QB vocabulary. Making use of the PROV-O ontology was, however, very intuitive. Secondly, although several tourism ontologies exist, none of these were found particularly relevant for our domain since they primarily focused on powering end-user applications as opposed to tools aimed for tourism managers. Some of the ontologies that we overviewed cannot be obtained online anymore. Others have a monolithic structure and, although they contained a reduced number of useful terms, it was not feasible to import them entirely into our model for the purpose of only making use of a handful of their concepts. As a result of the above issues, we developed our own ontologies which we linked to general purpose resources such as DBPedia, Linked Geo Data and schema.org.

Future work items include the publication of the created resources as linked data (not just as downloads from the project web-site) and their use in the data publishing (WP27) and application development (WP28) processes. These processes might require revisiting (and extending or changing) the hereby provided semantic models. For example, we might provide a slice structure in the generated data to better support certain visualisations. Additionally, we might be able to simplify the base ontology as many of its properties can be better modelled with the data cube elements.

References

[1] Barta, R., Feilmayr, C., Pröll, B., Grun, C., and Werthner, H. 2009. Covering the Semantic Space of Tourism: An Approach based on Modularized Ontologies. In Proc. of the 1st WS. on Context, Information and Ontologies, CIAO ’09, pages 1–8. ACM.

[2] Ou, S., Pekar, V., Orasan, C., Spurk, C. and Negri, M. 2008. Development and Alignment of a Domain-Specific Ontology for Question Answering. In Proc. of the Sixth International Language Resources and Evaluation Conf. (LREC).

[3] Sabou, M., Arsal, I. and Brasoveanu, A.M.P. 2013. TourMISLOD: a Tourism Linked Data Set. Semantic Web Journal 4(3): 271-276.

[4] Mendes, P. N., Stadtmüller, S., and Bizer, C. 2011. D4.1 PlanetData data sets, vocabularies and provisioning tools catalogue and access portal. PlanetData Deliverable.

[5] Mendes, P. N., Bizer, C., Miklos, Z., Calbimonte, J., Moraru, A., and Flouris, G. 2012. D2.1 Conceptual model and best practices for high-quality metadata publishing. PlanetData Deliverable.

[6] Stadler, C., Lehmann, J., Höffner, K., and Auer, S. 2012. LinkedGeoData: A Core for a Web of Spatial Open Data, Semantic Web Journal.

[7] Wöber, K. 2003. Information supply in tourism management by marketing decision support systems. Tourism Management. 24(3):241-255.

[8] Fodor, O., and Werthner, H. 2005. Harmonise: A Step Toward an Interoperable E-Tourism Marketplace. Int. J. Electron. Commerce, 9(2):11–39.

[9] Mondeca. 2004. Semantic Web Methodologies and Tools for Intra-European Sustainable Tourism. White Paper.

[10] Leroux, H., and Lefort, L. 2012. Using CDISC ODM and the RDF Data Cube for the Semantic Enrichment of Longitudinal Clinical Trial Data. SWAT4LS.

[11] Silveira Chaves, M., de Freitas, L. A., and Vieira, R. 2012. Hontology: A Multilingual Ontology for the Accommodation Sector in the Tourism Industry. KEOD 2012: 149-154

[12] Damljanovic, D., and Devedžic, V. 2009. Applying semantic web to e-tourism. In: The Semantic Web for Knowledge and Data Management: Technologies and Practices, Information Science Reference, 243-265.

[13] Siorpaes, K., and Bachlechner, D. 2006. OnTour: Tourism information retrieval based on YARS. Demos and Posters of the 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro.

[14] Niemann, M., Mochol, M., and Tolksdorf, R. 2008. Enhancing Hotel Search with Semantic Web Technologies. JTAER 3(2): 82-96.

[15] H. Park, A. Yoon and H. -C. Kwon. 2012, Task Model and Task Ontology for Intelligent Tourist Information Service, International Journal of u - and e - Service, Science and Technology, vol. 5, no. 2, June, pp. 43-58.

[16] Martin Hepp. Accommodation Ontology Language Reference, Accommodation Ontology Metadata, V 1.0, Release 2013-03-25, http://ontologies.sti-innsbruck.at/acco/ns.html

[17] Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. 2013. PROVO: The PROV Ontology. W3C Recommendation. Available at http://www.w3.org/TR/prov-o/.

[18] Cyganiak, R., Reynolds, D., Tennison, J. (25 June 2013). The RDF Data Cube Vocabulary. W3C Reccomendation. Available at: http://www.w3.org/TR/vocab-data-cube/.

[19] Stronge, W. B. 1993. Statistical Measurements in Tourism. In VNR's Encycl. of Hospitality and Tourism, 735-745.

[20] Frechtling, D., C. 2001. Forecasting Tourism Demand: Methods and Strategies. Butterworth Heinemann.

[21] Paulheim, H. and Bizer, C. 2013. Type Inference on Noisy RDF Data. ISWC, pp: 510-525.

[22] Braun, M., Scherp, A., and Staab, S. 2010. Collaborative semantic points of interests. In Proceedings of the 7th international conference on The Semantic Web: research and Applications - Volume Part II (ESWC'10), Lora Aroyo, Grigoris Antoniou, Eero Hyvönen, Annette Teije, and Heiner Stuckenschmidt (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 365-369.

Annex A Overview of TourMIS Indicators

Arrivals Indicators:

AA - Arrivals in all paid forms of accommodation establishments in city area only

AAS - Arrivals in all paid forms of accommodation in greater city area

AG - Arrivals in hotels and similar establishments in city area only

AGS - Arrivals in hotels and similar establishments in greater city area

AZ - Arrivals in all accommodation establishments incl. VFR in city area only

AZS - Arrivals in all accommodation establishments incl. VFR in greater city area

Capacity Indicators:

KA - Number of bedspaces in all forms of paid accommodation establishments in city area only

KAS - Number of bedspaces in all forms of paid accommodation establishments in greater city area

KG - Number of bedspaces in hotels and similar establishments in city area only

KGS - Number of bedspaces in hotels and similar establishments in greater city area

Bednights Indicators:

NA - Bednights in all paid forms of accommodation establishments in city area

NAS - Bednights in all paid forms of accommodation establishments in greater city area

NG - Bednights in hotels and similar establishments in city area only

NGS - Bednights in hotels and similar establishments in greater city area

NZ - Bednights in all accommodation establishments incl. VFR in city area only

NZS - Bednights in all accommodation establishments incl. VFR in greater city area

Annex B Cube Definition Example for Arrivals Data

```
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl:      <http://www.w3.org/2002/07/owl#> .
@prefix xsd:      <http://www.w3.org/2001/XMLSchema#> .
@prefix dct:      <http://purl.org/dc/terms/> .
@prefix foaf:     <http://xmlns.com/foaf/0.1/> .
@prefix interval: <http://reference.data.gov.uk/def/intervals/> .
@prefix prov:     <http://www.w3.org/ns/prov#> .
@prefix qb:       <http://purl.org/linked-data/cube#> .

@prefix sdmx-attribute:  <http://purl.org/linked-data/sdmx/2009/attribute#> .
@prefix sdmx-concept:    <http://purl.org/linked-data/sdmx/2009/concept#> .
@prefix sdmx-dimension:  <http://purl.org/linked-data/sdmx/2009/dimension#> .

@prefix eds:      <http://www.etihq.eu/dataset/> .
@prefix edss:     <http://www.etihq.eu/dataset/structure/> .
@prefix ebo:      <http://www.etihq.eu/ontology/base/> .


# -- Data Set --------------------------------------------

eds:dataset_Arrivals a qb:DataSet;
   dct:title       "Arrivals Dataset"@en;
   rdfs:label      "Arrivals Dataset"@en;
   rdfs:comment    "Arrivals Dataset"@en;
   dct:description "Arrivals Dataset"@en;
   dct:publisher   eds:modul_university;
   dct:issued      "2014-03-02"^^xsd:date;
   qb:structure edss:dsd_Arrivals;
   .

# -- Data structure definition ----------------------------

edss:dsd_Arrivals a qb:DataStructureDefinition;

  # The dimensions
   qb:component  [ qb:dimension edss:dimensionTime];
   qb:component  [ qb:dimension edss:dimensionDestination];
   qb:component  [ qb:dimension edss:dimensionMarket];

   # The measure(s)
   qb:component [ qb:measure edss:measureArrivals];

   # The attributes
   qb:component [ qb:attribute sdmx-attribute:unitMeasure;
                qb:componentRequired "true"^^xsd:boolean;
                qb:componentAttachment qb:DataSet; ] ;
   .

# -- Dimensions and measures  ----------------------------

edss:dimensionTime  a rdf:Property, qb:DimensionProperty;
   rdfs:label "time dimension"@en;
   rdfs:subPropertyOf sdmx-dimension:refPeriod;
```

```
    rdfs:range interval:Interval;
    qb:concept sdmx-concept:refPeriod;
    .

edss:dimensionDestination  a rdf:Property, qb:DimensionProperty;
    rdfs:label "destination dimension"@en;
    rdfs:range ebo:City;
    qb:concept ebo:City;
    .

edss:dimensionMarket  a rdf:Property, qb:DimensionProperty;
    rdfs:label "market dimension"@en;
    rdfs:range ebo:Market;
    qb:concept ebo:Market;
    .

edss:measureArrivals  a rdf:Property, qb:MeasureProperty;
    rdfs:label "arrivals measure"@en;
    rdfs:range xsd:decimal;
    qb:concept ebo:ArrivalsAtDestination;
    .

# -- Observations -----------------------------------------

eds:obs_Arrivals_London_AT_Y2012_M07 a qb:Observation, prov:Entity;
    qb:dataSet  eds:dataset_Arrivals;
    edss:dimensionDestination eds:city_London;
    edss:dimensionMarket  eds:marketAT;
    edss:dimensionTime <http://reference.data.gov.uk/id/gregorian-interval/2012-07-01T00:00:00/P1M> ;
    edss:measureArrivals        2134;
    prov:wasAttributedTo eds:person_adrian;
    prov:generatedAtTime "2014-03-02"^^xsd:date;
    .

# -- Additional data instances  -----------------------------------------

eds:modul_university a prov:Organization;
    foaf:name "MODUL University Vienna"@en;
    rdfs:label "MODUL University Vienna"@en;
    owl:sameAs <http://dbpedia.org/resource/MODUL_University_Vienna>;
 .

eds:person_adrian a prov:Person;
    foaf:givenName "Adrian"@en;
    prov:actedOnBehalfOf eds:modul_university;
 .

eds:marketAT a ebo:MarketForCountry;
    rdfs:label "AT"@en;
    ebo:forCountry <http://dbpedia.org/resource/Austria>;
 .

eds:city_London a ebo:City;
    rdfs:label "LONDON"@en;
    owl:sameAs <http://dbpedia.org/resource/London>;
 .
```

[end of document]