Conceptual Normalisation of XML Data for Interoperability in Tourism

Oliver Fodor[1], Mirella Dell'Erba[1], Francesco Ricci[1], Antonella Spada[1], and Hannes Werthner[1 2]

[1] eCommerce and Tourism Research Laboratory
ITC-irst, Italy
{fodor, dellerba, ricci, spada, werthner}@itc.it
[2] University of Trento, Italy

Abstract. Currently many tourism systems and standardisation initiatives adopt XML as a standard for data representation. Together with XML, XML Schema definition language is becoming more widely used for describing the structure and constraining the contents of the documents. However neither XML itself nor XML Schema provides sufficient mechanisms for the representation of data semantics.

In this paper we discuss the issue of conceptual normalisation of XML documents. Conceptual normalisation is introduced as a pre-processing step of our ontology-mediated framework for the harmonisation of electronic tourism systems. We propose to use RDF Schema as a mechanism for conceptual modelling of data sources and investigate the relation between semantics carrying constructs of XML Schema and RDF Schema. Accordingly we elaborate on the idea of transforming XML data into a conceptual model corresponding to RDF statements and vice-versa. Finally we propose a declarative solution and a high-level architecture of a toolkit facilitating this process.

1 Introduction

Tourism in its nature is an industry strongly dependent on information exchange. In reality, however, the domain is highly fragmented mainly due to the heterogeneity of data sources.

The problem of bringing together heterogeneous and distributed systems is known as the interoperability problem. The data heterogeneity is a well-known obstacle, and the current approach to achieve the data interoperability is, mainly, to write ad-hoc data interface programs for each pair of communicating systems. Experience shows that development and maintenance of these programs is expensive in terms of both time and money. The total effort required increases with the square of the number of communicating systems. Ontology-mediated document exchange provides a solution for this scalability problem. The data model of a source document is aligned with the representation specified by the ontology maintained by mediator. The data is transformed accordingly in both ways.

Nowadays the W3C standard for semi-structured data representation XML (eXtensible Markup Language) has been adopted by many systems and domain specific standards for data exchange over the Web. XML documents are often in accordance to the schematic description provided by XML Schema. Since XML Schema describes hierarchical structure for data representation it can be considered as a logical model for XML data. Nevertheless, XML Schema does not provide a unified mechanism for representing semantics of the data. Therefore an additional pre-processing step is necessary in order to facilitate the ontology-mediated solution for interoperability. Explicit representation of concepts of the data sources enables to align them with the concepts of the mediation ontology.

Conceptual schemata can be used within the mediation task to describe the semantics of information sources and to make the content explicit. Mediation based on conceptual schemata instead of logical is preferable because conceptual schemata have more clear semantics and are not bound to specific data structures. Presence of conceptual schemata simplifies identification and association of semantically corresponding information concepts and consequent resolution of semantic conflicts.

In this paper we present a bottom-up approach to the problem of *conceptual normalisation* of XML data sources. We identify and discuss several related issues: re-engineering of conceptual schemata from existing logical schemata, associating between logical and conceptual schemata and transforming of data instances into representation corresponding to the associated conceptual schema and its notions.

The conceptual normalisation process, introduced here, is a part of the ontology-mediated interoperability platform Harmonise[3]. Harmonise aims at the development of a comprehensive solution for electronic tourism systems providing them the possibility to freely cooperate by exchanging information in a seamless manner. However we believe that the issue of conceptual normalisation is relevant for a wider variety of solutions for data integration. In Harmonise we decided to make use of the W3C standards and recommendations for the Semantic Web. We believe that relying on widely accepted standards and technologies will make the Harmonise solution compatible with related initiatives. Therefore we use RDF Schema language (RDFS) for the representation of local conceptual schemata and RDF (Resource Description Framework) metadata format for representing the data instances.

This document is structured as follows: In the following Section we will introduce the overall picture of the Harmonise platform. The Harmonisation Consortium elaborated user requirements for the interoperability task and analysed relevant tourism standards to be considered as input for the harmonisation process. An ontology-mediated solution has been proposed, distinguishing two major tasks: conceptual normalisation and semantic mapping. In Section 3 we will focus on the conceptual normalisation as a pre-processing step for the semantic mapping task. We will describe four stages of the normalisation process: the re-engineering stage supporting the creation of conceptual schemata, association discovery and association definition stages for linking be-

[3] http:// www.harmonise.org/

tween the logical and conceptual schemata and the dynamic lift stage taking care of the XML-to-RDF transformation and vice versa. In Section 4 we will propose a declarative solution relying on a set of customisable normalisation templates providing basic association facilities between XML Schema and RDF Schema. Finally, in Section 5, we will draft a high-level architecture for the normalisation toolkit as a part of the Harmonise solution.

2 Harmonise - Interoperability for Tourism

The Harmonise project is an initiative financed by the European Commission (EC), under the 5th Framework RTD Programme, contract number IST-2000-29329. The primary aim of the project is to establish an open international consortium - Tourism Harmonisation Network (THN) - including major tourism stakeholders, domain experts and IT professionals. The THN will serve as a forum for discussing the interoperability issues and coordinating the related activities within the tourism domain. Further, Harmonise aims to provide a solution for the interoperability problem in tourism by means of the so-called Harmonise Platform.

The Harmonise Platform is an ontology-mediated solution relying on the newest technologies for knowledge representation and following the philosophy of the Semantic Web, i. e. adding semantic annotations to the data sources. The goal is to allow the participating tourism organisations to keep their proprietary data format and simultaneously cooperate with each other, by exchanging information in a seamless manner. This will be possible through the help of a mediator module - the *Harmonisation tool (H-tool)* - providing a mediation service between different proprietary data formats and transforming the fragmented services, data and events in a common environment, where those can be easily and transparently distributed. This mediator acts as a semantic gateway between systems, permitting the receiver to view the source as an extension of its own information system, without any concern for the differences in names and representations of data.

The H-tool is based on the following three technologies as illustrated in Figure 1:

- A tourism ontology *IMHO (Interoperability Minimum Harmonisation Ontology)* modelling and storing the basic concepts used in representing the content of information exchanges in tourism transactions.
- An interchange format *HIR (Harmonise Interchange Representation)* suitable to represent the instance data used within interoperable tourism transactions.
- A set of *mapping rules* defining the transformation of data from the proprietary local format into the Harmonise Interchange Representation, and vice versa. Each system participating in Harmonise will maintain its own set of rules; the mapping rules will be defined on the base of the semantic correspondence of the local data models with respect to the IMHO.

2.1 XML as Common Input Format

The Harmonise Consortium agreed on XML as the common syntax for systems participating in the harmonisation process. This constraint was carefully decided after a wide discussion with major tourism organisations, domain experts, industry leaders and IT professionals. While the XML format was already accepted in some major industry standards (e. g. OTA[4]), several standards are currently in the preparation phase or already undertaking the evolution step towards XML. This is also true for the legacy systems of participating National Tourism Boards and their local standards (Finland, France, Portugal). Further it was agreed that logical schemata (XML Schema) of the XML documents will be also provided by the participants.

Nevertheless, it is intended to keep the Harmonise solution open for possible future extensions towards other physical level formats.

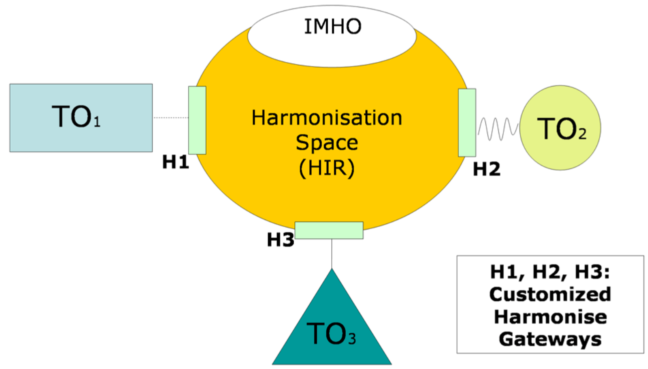

Fig. 1. Harmonise solution

2.2 Interoperable Minimal Harmonisation Ontology and Harmonise Interchange Representation

The Harmonise project will build the Interoperability Minimum Harmonisation Ontology by adopting the OPAL (Object, Process, Actor modelling Language) methodology. The core of OPAL is a business and enterprise ontology framework. The proposed framework is not a new language for ontology modelling, but intends to build, on top of an existing ontology language, domain-oriented constructs more familiar to enterprise experts than the basic modelling primitives of existing languages. The additional constructs are intended to ease the task of the tourism expert, by hiding the complexity of the underlying ontology representation formalism and supporting the construction of effective enterprise ontology.

As a starting point, it was decided to use RDF Schema language for representing the IMHO. This decision was taken based on an analysis of several ontology representation languages (DAML+OIL, UML, XML Schema) as a compromise between the expressive power, user friendliness and global distribution of the language. Although RDF Schema has limited expressive power we consider it as sufficient for the purpose of Harmonise whereas still allowing future extensions following new standards for representing ontologies.

Accordingly, RDF metadata format was adopted by Harmonise for representing the Harmonise Interchange Representation (HIR).

[4] OpenTravel Alliance, http://www.opentravel.org/

2.3 Analysis of the relevant standards

An analysis of relevant tourism standards (SIGRT[5], TourInFrance[6], OTA, xCBL[7], IFITT RMSIG[8], UN/EDIFACT TT&L[9], CEN/TC 329[10]) has taken place in the early phases of the Harmonise project. A comparison of overlapping concepts has shown that two different kinds of conflicts arise when an XML document has to be translated from one format to another one:

– Semantic clashes
– Structural clashes

Semantic clashes. Semantic clashes are clashes between concepts of different standards, or more precisely, between specific conceptual models or ontologies behind different standards. Typical semantic clashes are completely different concepts, different naming of concepts or different granularity. Identified semantic conflicts have been classified in eight categories. A detailed description can be seen in [13].

Table 1. Sample of semantic clashes

Different naming	PostCode vs. PostalCode
Different position	Postcode in Address rather than in ContactInfo
Different scope	TelephonePrefix and TelephoneNumber separated vs. Prefix_TelephoneNumber as single concept

Structural clashes. Structural clashes are caused by the heterogeneity of XML representation. Using XML format the same concept can be expressed in several different ways. XML Schema enables constraining of XML documents but this was designed for constraining the content of XML documents not for the conceptual representation. Within XML, structural clashes are mainly caused by the different usage of specific constructs, e.g. by a different usage of attributes rather than embedded elements or by expressing concepts in enumeration values.

Usually freely designed XML documents used for specific application purposes do not provide sufficient information about the semantics of the data. The semantics of XML elements used by Web applications is hard-coded into the applications and is typically not available in machine-processable form. This applies also to documents with available structural schemata (XML Schema), which in the most cases define the syntactical structure of XML documents without unified implicit representation of their meaning.

[5] Sistema de Informação de Gestão de Recursos Turísticos, Portugal
[6] National Tourism Board, France
[7] XML Common Business Library
[8] IFITT Reference Model Special Interest Group
[9] United Nations rules for Electronic Data Interchange for Administration, Commerce and Transport – Travel Tourism & Leisure
[10] European Committee for Standardization/ Technical Committee Tourism Services

The example in Table 2 shows three different ways of expressing the concept PostalCode in XML.

Table 2. Structural heterogeneity of XML

```
<ContactInformation>
   <Address PostalCode="X-1220">
   Wannaby Street 59, Dreamtown</Address>
</ContactInformation>
```

```
<ContactInformation>
   <Address>
      <Street>Wannaby Street 59</Street>
      <City>Dreamtown</City>
      <PostalCode>X-1220</PostalCode>
   </Address>
</ContactInformation>
```

```
<ContactInformation>
   <Address>
       Wannaby Street 59,
       <PostalCode>X-1220</PostalCode>
       Dreamtown
   </Address>
</ContactInformation>
```

In [5] the authors show that an attempt to resolve both kinds of conflicts within one transformation step causes a scalability problem. Such a solution would lead to a rather complex set of rules and consequently to possible performance problems of the processing engine.

The separation between semantic and structural clashes indicates the need of a distinction between corresponding steps in the overall transformation process. Due to these arguments the Harmonise solution introduces a pre-processing step called conceptual normalisation. This step enables a separation of semantic mapping (resolution of the semantic clashes) from the concrete physical representation of data being transformed. In case different physical representations will be used in the future, the semantic mapping definitions will still remain valid.

Figure 2 depicts the two steps of data transformation from local to the harmonised representation (HIR).

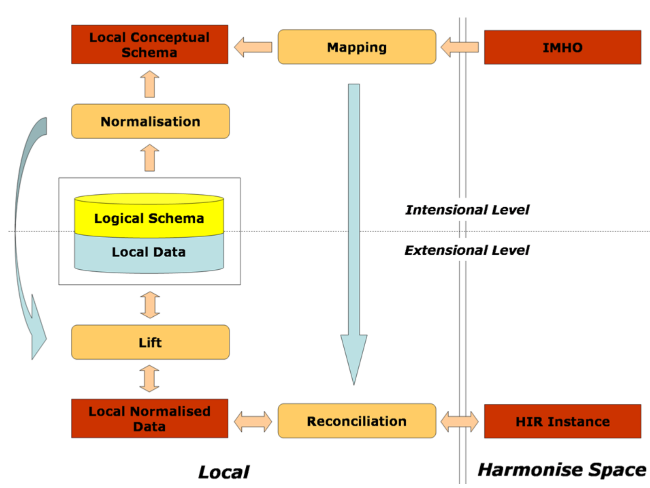

Fig. 2. Harmonise steps (forward harmonisation)

In the following Section we will discuss the basic issues related to conceptual normalisation.

3 Conceptual Normalisation

As described above, we consider conceptual normalisation as a pre-processing step facilitating the overall mediation task of the Harmonise ontology-based solution. Systems willing to participate in the harmonisation process will be encouraged to add semantics to their data by means of conceptual schemata and normalisation maps described later in this paper. We assume that there are no conceptual schemata available and a re-engineering based on existing data structures will be necessary in order to obtain them. However, we argue that available logical schemata (XML Schemata) are a promising base for the extraction of conceptual schemata. Obviously not all information about the data semantics can be extracted from the logical schemata, therefore an intervention of a system engineer is also considered. Within Harmonise we aim to develop a toolkit fully supporting the process of conceptual normalisation of XML schemata and respective data manipulation.

The process of conceptual normalisation is a bottom-up approach based on the existence of local logical schemata. We have identified four stages within the normalisation process as indicated in Figure 3. These stages will take place in a sequential step-by-step fashion.

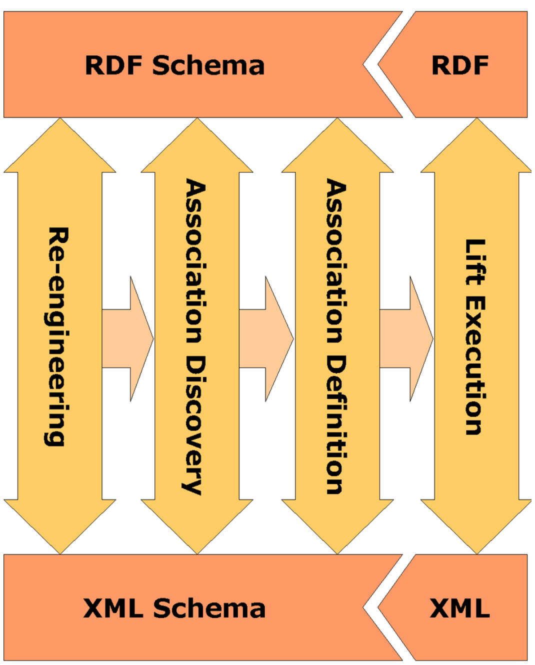

Fig. 3. Stages of conceptual normalisation process

Conceptual Level. To facilitate the semantic interoperability additional information has to be provided on the data sources. The semantics has to be expressed in an unified form (human and machine understandable) and referred to by the data to be harmonised. Therefore, Harmonise addresses the concept of *Normalized Conceptual Schemata* as explicit definition of the data semantics. Consequently, a new level of abstraction is to be considered for the representation of local data. We call this level *conceptual level*, also commonly referred to as data model level or semantic level.

The conceptual level has been introduced in the context of a three-schema framework [1] for database management systems. Three-layered approach for information representation on the Web was proposed in [2], where the semantic level was introduced above the lower object and syntax layers. Three layers for information integration (syntax, data model, ontology) were also adopted by [6] where data model layer serves for elimination of syntactical ambiguities of XML format.

Normalised Conceptual Schema. Originally, conceptual schemata played a key role in the development of a large variety of systems. In the above mentioned ANSI framework a conceptual schema is defined as “a unique central description of the information that might be in the database”. Conceptual schema is considered as a product of the conceptual modelling process.

In the Harmonise solution, the purpose of Normalised Conceptual Schema is to provide a unified human and machine understandable description of the concepts of local systems and relations among them. The important characteristic of NCS is its ability to be aligned with the formal representation of the Interoperable Minimal Harmonisation Ontology (IMHO). This intermediary product will simplify the process of semantic mapping since the mappings will be defined over representations of the concepts at the same abstraction level (conceptual level).

As stated in Section 2.2, in Harmonise the IMHO is represented by RDF Schema language. Therefore we propose to use RDF Schema as formalism for expressing the Normalized Conceptual Schemata for the systems participating in Harmonise. The convenience of RDF Schema for conceptual modelling has been discussed in [2], [3], [5] and variety of other sources. Accordingly, we use RDF format as a mechanism for representation of the data instances.

Conceptual Schema Re-engineering. Assuming that no conceptual schemata exist at the local sites these have to be obtained by a re-engineering of the available local data structures. The available XML Schemata will serve as the basis for the re-engineering of a corresponding conceptual schema. The elementary constructs and default semantics carrying elements of the logical schemata will be analysed in order to obtain an abstract conceptual model defining them. The semantic aspects of the logical schemata, e.g. the naming or granularity, will be preserved within the process. The purpose of this step is to resolve the structural heterogeneity of XML Schema and to separate the conceptual layer from the physical representation of the data.

Association Discovery. Association discovery is an intermediary step between the re-engineering and association definition and serves as a bridge between them. The knowledge obtained in the re-engineering step is used here to propose basic associations between the underlying logical and just produced conceptual schemata. Associations are derived from the default semantic correspondences of both schemata.

We consider a possible merge of the re-engineering and discovery phase and we will investigate the benefits in our future work.

Association Definition In this stage the associations between logical and conceptual schemata will be made explicit. The linking between the building constructs will be defined based on the associations proposed in the previous step and the explicit knowledge of the data structures. Therefore this step will be mainly depending on a human intervention. The definitions produced here will allow the projection of the underlying logical schema (tree) onto the conceptual schema (graph) defined in the re-engineering step and vice-versa. The symmetric character of the definitions is necessary for the backward harmonisation step where data is translated from the harmonised to the local representation.

Execution - Lift. According to the previously defined associations the data instances must be transformed so that they correspond to the representation of the concepts. Essentially, in this step RDF statements are derived from the XML documents and vice versa. This transformation of the data instances facilitates their further processing in the mediator. The lift can be seen as wrapping and un-wrapping of the data where the content is preserved, only some meta-information about the concepts is added. As already mentioned both directions must be supported therefore we distinguish between XML-to-RDF lift and RDF-to-XML extraction.

4 Declarative Approach

To the problem of conceptual normalisation we propose a declarative approach. This is based on a set of customisable *normalisation bridges* enabling declarative definition of the associations between logical and conceptual schemata. Further we propose a set of *normalisation heuristics* based on default semantic interpretations of XML Schema constructs. These heuristics will support automated and also manual phase of the re-engineering step. Normalisation heuristics will be interrelated with the association bridges, which enables to partially automate also the association discovery step. Therefore we consider our declarative approach as semi-automatic.

It is obvious that there is only a partial overlapping with respect to the semantic expressiveness of XML Schema and RDFS. The dimensions of e.g. enumeration of property values or basic data types cannot be preserved within the normalisation process. Hence, conceptual normalisation is a lossy process. However we consider the information lost acceptable for the initial release of the Harmonise solution. In the future this drawback can be removed by an extension of RDFS or by adoption of higher level modelling formalism based on RDFS.

Normalisation Ontology. First of all, we aim to create a normalisation ontology identifying all concepts related to the conceptual normalisation process. The mission of this ontology is to specify an unambiguous human comprehensible reference model as a base for the specification of normalisation bridges. The ontology will cover all elementary constructs of XML (e. g. element, attribute), XML Schema (e.g. element declaration, type definition), RDF (e. g. object, object identity), RDFS (e. g. class, property) and mechanisms for defining associations among them (normalisation heuristics, normalisation bridges). We intend to specify this ontology in common formats, like DAML+OIL, RDFS and OPAL.

Normalisation Heuristics. A set of heuristics related to the default semantic interpretations can be applied to support the modelling process in a semi-automatic way. These heuristics address the components of XML Schema and suggest their correspondences with components of RDF Schema. Usually, there are several possibilities to interpret an XML Schema construct in RDFS. To narrow down the number of options we also look at other relevant characteristics of XML Schema components, e.g. its naming (named, anonymous) or position in the schema (top-level, nested).

Consequently, the association discovery phase introduced in Section 3 can make benefit of the normalisation heuristics and accordingly automatically derive associations between the underlying schemata.

Table 3. Sample heuristics for conceptual modelling

Given an unnamed complex type definition. Create a concept (rdfs:Class) and name it with the name of the declared parent element plus suffix "_class".
Given a local element declaration declaring the element E of a complex type. Create a relation (rdf:Property) and name it with the name of the element E. Set the domain of E to the class defined by the parent of the element declaration and the range of E to the class defined by the element's type.

Table 3 introduces two related sample heuristics. Figure 4 shows how a RDFS conceptual schema is created based on an existing XML Schema fragment in accordance to these heuristics.

Normalisation Bridges. Normalisation bridges are a set of customisable templates implementing associations between XML Schema and RDFS. The purpose of normalisation bridges is to associate components of two concrete schemata in order to support transformation of the data instances. Normalisation bridges are customized within the association definition phase.

Each component to be associated by a normalisation bridge must be unambiguously addressed. To achieve this, we currently evaluate several techniques for naming and referencing components and structures of both schemata: XML Namespaces, XPath language and Uniform Resource Identifier (URI). We intend to rely on the URI mechanism for naming of RDF resources and XPath like naming convention (normalized uniform names) for XML Schema components. The later is described in detail in [16].

Normalisation templates must support associations in both directions. We expect that some association definitions will be symmetric whereas in some cases an explicit definition for each direction will be required. Also we expect that often several bridges will be relevant for one information item in

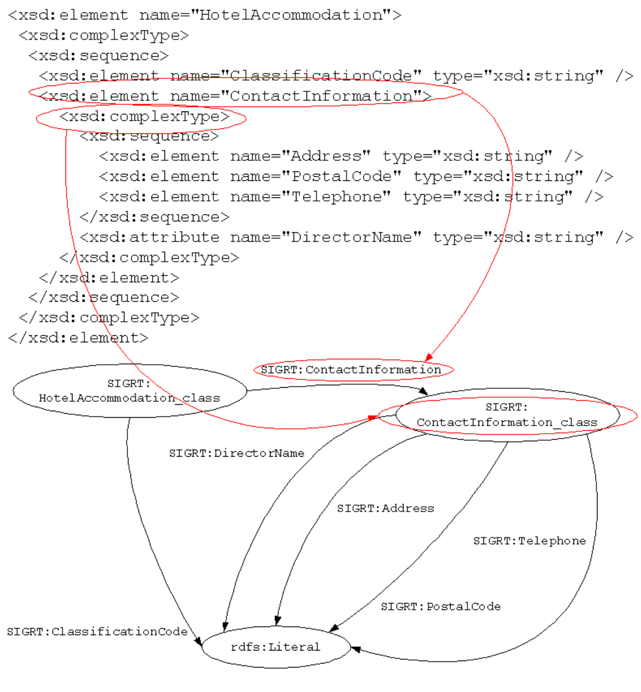

Fig. 4. Re-engineering of a conceptual schema

the data instance. E. g. following two bridges apply to one occurrence of the element `ContactInformation` in the XML instance of the schema from Figure 4:

```
<xsd2rdfs_bridges:XSD2RDFSBridge
    rdf:ID="ContactInformation_element2property"
    xsd2rdfs_bridges:XSD_NuN="HotelAccommodation/*/ContactInformation"
    xsd2rdfs_bridges:RDFS_URI="SIGRT:ContactInformation" />

<xsd2rdfs_bridges:XSD2RDFSBridge
    rdf:ID="ContactInformation_type2class"
    xsd2rdfs_bridges:XSD_NuN="HotelAccommodation/*/ContactInformation/*"
    xsd2rdfs_bridges:RDFS_URI="SIGRT:ContactInformation_class" />
```

A set of customized normalisation bridges builds a *Normalisation Map*. Its mission is to support the transformation process of underlying data instances. A Normalisation Map will be explicitly related to the schemata it applies to. An appropriate mechanism will be introduced in order to explicate relations between bridges and their composition to build the map.

Figure 5 shows how a piece of XML document is transformed into RDF with respect to the schemata from Figure 4 and the corresponding map.

5 Harmonise Normalisation Toolkit

In order to support conceptual normalisation we aim to develop a normalisation toolkit as a part of the Harmonise Platform. Figure 6 displays the high-level architecture of the toolkit including several components. In this Section we will briefly describe these components. Naturally, in order to provide a solid solution an extensive requirement study and use case analysis will take place within the next phases of the Harmonise project. Consequently, we intend to develop a prototype solution as a module based on the KAON[11] ontology infrastructure.

The normalisation toolkit described in this paper supports only manual process of semantic normalisation. The pro-

[11] http://kaon.semanticweb.org/

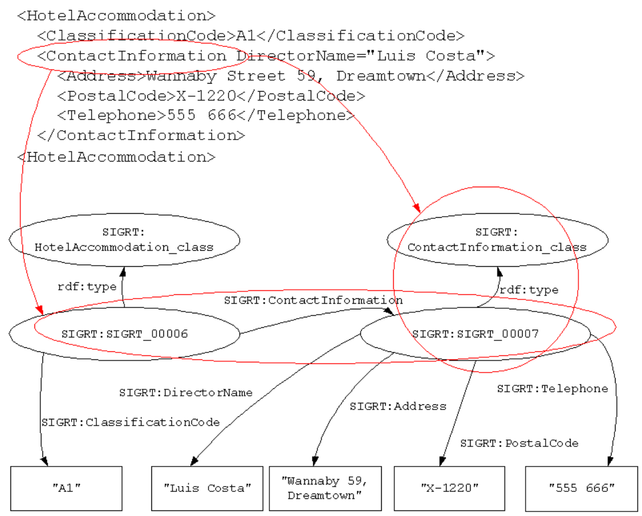

Fig. 5. XML data transformation (lift)

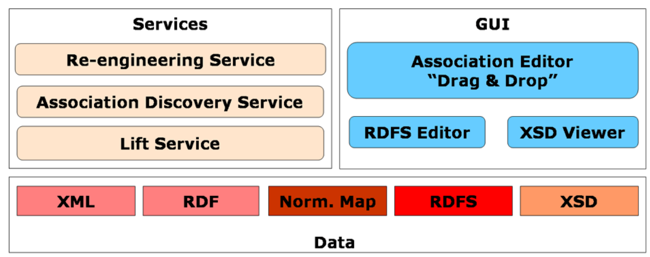

Fig. 6. Normalisation toolkit

posed solution can be extended with automated engines, supporting the re-engineering process of conceptual schemata and the association discovery process.

RDFS Editor & Viewer. In the first release only manual creation of conceptual schemata will be supported. This requires an intervention of a system engineer who is familiar with the local data structures. This person manually extracts the concepts and relations from the existing XML Schemata and creates an appropriate RDFS model. The basic component supporting this process will be an editor tool providing an easy-to-use graphical user interface for editing RDFS concepts. Naturally, this tool will also serve as a viewer for already existing models. KAON provides an ontology-management component called SOEP which we consider as a candidate to re-use.

XML Schema Viewer. We agreed that XML Schema is the initial input for the normalisation process. Therefore, a component enabling viewing of schema documents is also necessary. A convenient visualization method for the XSD format will be proposed considering both efficiency and human readability. The functionality of this component can be limited to passive browsing.

Association Editor. Once we can browse through the normalized conceptual schemata (RDFS) and the underlying logical schemata (XSD) we need to provide a support for the definition of the associations among their elementary building blocks. For this purpose a GUI editor is foreseen. The tool will cooperate with both schema viewers and provide a support for intuitive definition of the mappings, e.g. in a "drag-and-drop" fashion. The editor will implement the normalisation templates described earlier in this document. The output of this editor will be the Normalisation Map including a set of association definitions (normalisation bridges).

Lift Engine. This component will process a Normalisation Map at input and make use of it in order to transform the data to be harmonised. The lift engine will support both directions: XML-to-RDF lift and RDF-to-XML extraction. We expect to compile the normalisation map into a lower-level transformation language suitable for transforming XML, or RDF documents. XSLT is a good candidate for the XML-to-RDF whereas a RDF query language might be more appropriate for the RDF-to-XML direction.

6 Related Work

There is a rather big number of approaches to ontology-mediated integration of heterogeneous data sources. Most of these approaches, however, focus on integrated view over relational databases. Nevertheless, there are approaches [8] for mapping between XML documents although these mostly focus on direct transformation without considering the semantic level.

In [6] a three-level solution for B2B catalogue integration is presented. The authors introduce an RDF mapping language RDF-T [7], which allows to map between two RDF Schemata and translate RDF instance documents accordingly. Recently, RDF-T has been extended with basic support for mapping XML tags to RDFS resources. We follow the evolution of RDF-T and consider the language as a possible candidate for our normalisation bridges.

A rule-based approach to DTD to conceptual schema conversion is presented in [4]. This proposal bases on a set of conversion rules that takes into account DTD elements and attributes, syntactical constructs and heuristics related to their default semantic interpretations to generate a conceptual schema. This approach doesn't introduce any mechanism for an explicit definition of the associations between DTDs and corresponding conceptual schemata.

In [15] the authors propose to combine XML Schema with RDF Schema to enhance interoperability. Since this approach is relevant for proprietary designed schemata within applications it is not suitable for existing schemata defined by already adopted standards.

7 Conclusions

In this paper we introduced the process of semantic normalisation as a part of ontology-based harmonisation platform for tourism. Since the Harmonise solution builds on common standards for data representation and especially on the fundamental technologies for the Semantic Web, this approach can be also relevant for applications in other domains. Currently there are many systems and standards based on the XML technology and more are to be expected. Therefore our solution could be considered as a generic application for the Semantic Web.

From the perspective of electronic tourism systems participating in the harmonisation process we consider the semantic normalisation as an approximation step towards modern sophisticated solutions for interoperability. We believe that tourism is a very specific domain and its evolution is tightly coupled with the evolution of the information and communication technologies, the Internet in particular. Within Harmonise we intend to disseminate new promising technologies in order to preserve the tourism domain as one of the leading domains in the World Wide Web.

Currently we are working on a high-level ontology for the conceptual normalisation (heuristics and bridges) and its formal description. Within the next phases of the Harmonise project we intend to develop a prototype of the normalisation toolkit as a module for the KAON ontology framework.

8 Acknowledgement

Hereby we would like to thank the partners of Harmonise project for their contribution to the realisation of this paper. These are as follows:

– T6 (Italy)
– ICEP (Portugal)
– IFITT (Austria, IRL)
– CNR-IASI (Italy)
– LINK (Portugal)
– EC3 (Austria)

References

1. Interim Report: ANSI/X3/SPARC Study Group on Data Base Management Systems 75-02-08. FDT - Bulletin of ACM SIGMOD Vol. 7, No. 2 (1975) 1-140
2. Melnik, S., Decker, S.: A Layered Approach to Information Modelling and Interoperability on the Web. Proceedings of the Workshop on the Semantic Web at the Fourth European Conference on Research and Advanced Technology for Digital Libraries ECDL-2000, Lisbon, Portugal (2000)
3. Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., Horrocks, I.: The Semantic Web: the Roles of XML and RDF. IEEE Internet Computing, Sept./Oct. 2000, Vol. 4, No. 5 (2000) 63-74
4. dos Santos Mello, R., Heuser, C. A.: A Rule-Based Conversion of DTD to a Conceptual Schema. Proceedings of the 20th International Conference on Conceptual Modeling ER 2001, Yokohama, Japan (2001) 133-148
5. Omelayenko, B., Fensel, D.: Analysis of B2B Catalogue Integration Problems. Content and Document Integration, In: Filipe, J., Sharp, B., Miranda, P. (eds.): Enterprise Information Systems III, Kluwer Academic Publishers, Dordrecht (2002) 270-277
6. Omelayenko B., Fensel D.: Scalable Document Integration for B2B Electronic Commerce. Submited (2001)
7. Omelayenko, B., Fensel, D., Klein, M.: RDF-T: An RDF Transformation Toolkit. Submitted (2001)
8. Su, H., Kuno, H., Rundensteiner, E. A.: Automating the Translation of XML Documents. Technical Report WPI-CS-TR-01-13, Computer Science Department, Worcester Polytechnic Institute (2001)
9. Wiederhold, G.: Mediators in the Architecture of Future Information Systems. IEEE Computer, March 1992, Vol. 25, No. 3 (1992) 38-49

10. Visser, U., Stuckenschmidt, H., Vögele, T., Wache, H.: Enabling Technologies for Interoperability. Submitted to Transactions in GIS (2001)
11. Harmonise technical committee: Harmonise Project Technical Summary. http://www.harmonise.org (2002)
12. Harmonise technical committee: D.1.1: Electronic Classification of Tourism standards. http://www.harmonise.org (2002)
13. Harmonise technical committee: D.1.2: Report on comparison on standard methodologies. http://www.harmonise.org (2002)
14. Dell'Erba, M., Fodor, O., Ricci, F., Spada, A., Werthner, H.: Harmonise: A Solution for Data Interoperability. Submitted (2002)
15. Hunter, J., Lagoze, C.: Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles. Proceedings of the Tenth International World Wide Web Conference, WWW 10, Hong Kong, China (2001) 457-466
16. Brown, A., Fuchs, M., Robie, J., Wadler, P.: XML Schema: Formal Description. W3C Working Draft, http://www.w3.org/TR/xmlschema-formal/ (25 September 2001)