

Linked Data in Enterprise Information Integration

Philipp Frischmuth ^{a,*}, Jakub Klímek ^b, Sören Auer ^a, Sebastian Tramp ^a, Jörg Unbehauen ^a, Kai Holzweißig ^c, and Carl-Martin Marquardt ^c,

^a *Universität Leipzig, Institut für Informatik, AKSW, Postfach 100920, 04009 Leipzig, Germany*

E-mail: {lastname}@informatik.uni-leipzig.de

^b *XML and Web Engineering Research Group, Charles University in Prague, Malostranské náměstí 25, 118 00 Praha 1, Czech Republic*

E-mail: klimek@ksi.mff.cuni.cz

^c *Enterprise Services Portal, CMS & Search, Daimler AG, Plant 096-0191, 70546 Stuttgart, Germany*

E-mail: {firstname.lastname}@daimler.com

Abstract. Data integration in large enterprises is a crucial but at the same time costly, long lasting and challenging problem. While business-critical information is often already gathered in integrated information systems, such as ERP, CRM and SCM systems, the integration of these systems itself as well as the integration with the abundance of other information sources is still a major challenge. In the last decade, the prevalent data integration approaches were primarily based on XML, Web Services and Service Oriented Architectures (SOA). However, we become increasingly aware that these technologies are not sufficient to ultimately solve data integration challenges in large enterprises. In this article, we argue that classic SOA architectures may be well-suited for transaction processing, however more efficient technologies are available today that can be employed for enterprise data integration. In particular, the use of the Linked Data paradigm for integrating enterprise data appears to be a very promising approach. Similarly, as the data web emerged complementing the document web, data intranets can complement the intranets and SOA landscapes currently found in large enterprises. In this paper, we explore the challenges large enterprises are still facing with regard to data integration. These include, but are not limited to, the development, management and interlinking of enterprise taxonomies, domain databases, wikis and other enterprise information sources. We discuss Linked Data approaches in these areas and present some examples of successful applications of the Linked Data principles in that context.

Keywords: Linked Data, Enterprise Data Integration, Data Governance, Data Web, RDF, SKOS, WebID

1. Introduction

Data integration in large enterprises is a crucial but at the same time costly, long lasting and challenging problem. While business-critical information is often already gathered in integrated information systems such as ERP, CRM and SCM systems, the integration of these systems itself as well as the integration with the abundance of other information sources is still a major challenge. Large companies often oper-

ate hundreds or even thousands of different information systems and databases. This is especially true for large OEMs. For example, it is estimated that at Volkswagen there are approximately 5000 different information systems deployed. At Daimler – even after a decade of consolidation efforts – the number of independent IT systems still reaches 3000.

After the arrival and proliferation of IT in large enterprises, various approaches, techniques and methods have been introduced in order to solve the data integration challenge. In the last decade, the prevalent data integration approaches were primarily based on XML, Web Services and Service Oriented Archi-

*Corresponding author. E-mail: frischmuth@informatik.uni-leipzig.de.

tectures (SOA) [20]. XML defines a standard syntax for data representation, Web Services provide data exchange protocols and SOA is a holistic approach for distributed systems architecture and communication. However, we become increasingly aware that these technologies are not sufficient to ultimately solve the data integration challenge in large enterprises. In particular, the overhead associated with SOA is still too high for rapid and flexible data integration, which is a pre-requisite in the dynamic world of today's large enterprises.

We argue that classic SOA architectures are well-suited for transaction processing, but more efficient technologies are available that can be deployed for solving the data integration challenge. Recent approaches, for example, consider ontology-based data integration, where ontologies are used to describe data, queries and mappings between them [63]. The problem of ontology-based data integration are the required skills to develop the ontologies and the difficulty to model and capture the dynamics of the enterprise. A related, but slightly different approach is the use of the Linked Data paradigm for integrating enterprise data. Similarly, as the data web emerged complementing the document web, data intranets can complement the intranets and SOA landscapes currently found in large enterprises.

The acquisition of Freebase by Google and Powerset by Microsoft are the first indicators, that large enterprises will not only use the Linked Data paradigm for the integration of their thousands of distributed information systems, but they will also aim to establish Enterprise Knowledge Bases (similar to what Freebase now is for Google) as hubs and crystallization points for the vast amounts of structured data and knowledge distributed in their data intranets.

Examples of public LOD data sources being highly relevant for large enterprises are OpenCorporates¹ (a knowledge base containing information about more than 50.000 corporations worldwide), LinkedGeo-Data [9] (a spatial knowledge base derived from OpenStreetMap containing precise information about all kinds of spatial features and entities) or Product Ontology² (which comprises detailed classifications and information about more than 1 million products). For enterprises, tapping this vast, crowd-sourced knowledge being freely available on the Web, is an amazing

opportunity. However, it is crucial to assess the quality of such freely available knowledge, to complement and contrast it with additional non-public information being available to the enterprise (e.g. enterprise taxonomies, domain databases etc.) and to actively manage the life-cycle of both – the public and private data – being integrated and made available in an Enterprises data intranet.

In order to make large enterprises ready for the service economy their IT infrastructure landscapes have to be made dramatically more flexible. Information and data has to be integrated with substantially reduced costs and in extremely short time intervals. Mergers and acquisitions further accelerate the need for making IT systems more interoperable, adaptive and flexible. Employing the Linked Data approach for establishing enterprise data intranets and knowledge bases will facilitate the digital innovation capabilities of large enterprises.

In this paper, we explore the challenges large enterprises are still facing with regard to data integration. These include, but are not limited to, the development, management and interlinking of enterprise taxonomies, domain databases, wikis and other enterprise information sources. We discuss Linked Data approaches in these areas and present some examples of successful applications of the Linked Data principles in that context.

Figure 1 depicts our vision of an *Enterprise Data Web* and the resulting semantically interlinked enterprise IT systems landscape. We expect existing enterprise taxonomies to be the nucleus of linking and integration hubs in large enterprises, since these taxonomies already reflect a large part of the domain terminology and corporate and organizational culture. In order to transform enterprise taxonomies into comprehensive enterprise knowledge bases, additional relevant datasets from the Linked Open Data Web have to be integrated and linked with the internal taxonomies and knowledge structures. Subsequently, the emerging enterprise knowledge base can be used (1) for interlinking and annotating content in enterprise wikis, Content Management Systems and portals; (2) as a stable set of reusable concepts and identifiers; and (3) as background knowledge for intranet, extranet and site search applications. As a result, we expect the current document-oriented intranets in large enterprises to be complemented with a data intranet, which facilitates the lightweight, semantic integration of the plethora of information systems and databases in large enterprises.

¹<http://opencorporates.com/>

²<http://www.productontology.org/>

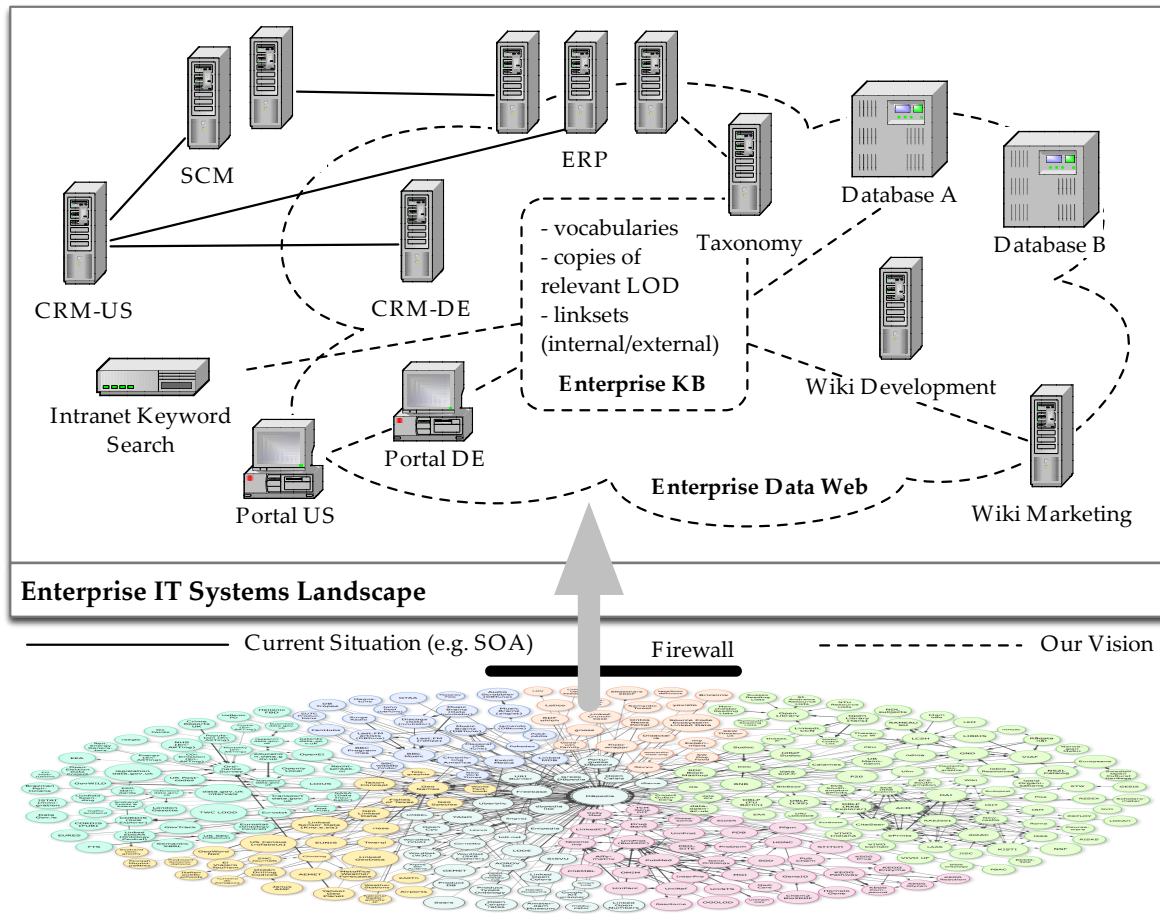


Fig. 1. The above depiction shows our vision of an Enterprise Data Web (EDW). The solid lines show how IT systems may be currently connected in a typical scenario. The dotted lines visualize how IT systems could be interlinked employing an internal data cloud. The EDW also comprises an Enterprise Knowledge Base (EKB), which consists of vocabulary definitions, copies of relevant Linked Open Data, as well as internal and external linksets between datasets. Data from the LOD cloud may be re-used inside the enterprise, but internal data is secured from external access just like in usual intranets.

2. Data Integration Challenges in the Enterprise

We identified six crucial areas (shown in table 1) where data integration challenges arise in large enterprises. In the following section we investigate those challenges in more detail, each by considering the current situation first. We then examine the benefits of employing Linked Data technologies in order to tackle the respective challenge. Finally, we describe the challenges that need to be addressed to make the transition from the current state of the art to the Linked Data approach feasible. Figure 2 shows the Linked Data lifecycle in conjunction with the aforementioned chal-

lenges. Each challenge may be related to a single or to multiple steps in the Linked Data lifecycle.

2.1. Enterprise Taxonomies

Nowadays, almost every large enterprise uses taxonomies to provide a shared linguistic model aiming at structuring the large quantities of documents, emails, product descriptions, enterprise directives, etc. which are produced on a daily basis.

2.1.1. State of the Art

It is widely agreed that taxonomies are usable, however, there are multiple challenges that must be addressed in order for taxonomies to work correctly [19].

Table 1
Overview of data integration challenges investigated in this article

Information integration domain	Current state of the art	Linked Data benefit
Enterprise Taxonomies	proprietary, centralized, no relationships between terms, multiple independent terminologies (dictionaries)	open standards (e.g. SKOS), distributed, hierarchical, multilingual, re-usable in other scenarios
XML Schema Governance	multitude of XML schemas, no integrated documentation	relationships between entities from different schemas, tracking/documentation of XML schema evolution
Wikis	text-based wikis for teams or internal-use encyclopedias	re-use of (structured) information via data wikis (by other applications), interlinking with other data sources, e.g. taxonomies
Web Portal and Intranet Search	keyword search over textual content	sophisticated search mechanisms employing implicit knowledge from different data sources
Database Integration	Data Warehouses, schema mediation, query federation	lightweight data integration through RDF layer
Enterprise Single Sign-On	consolidated user credentials, centralized SSO	no passwords, more sophisticated access control mechanisms (arbitrary metadata attached to identities)



Fig. 2. Linked data lifecycle supports four crucial data integration challenges arising in enterprise environments. Each of the challenges can relate to more than one life-cycle stage.

One obvious problem is that the creation of metadata for digital objects is additional work which is of little direct benefits to the creators of the corresponding contents. Primarily other users looking for the content benefit from a proper classification through a taxonomy. This problem may be, however, tackled by proper

explanation of the benefits to the interested parties. Ideally, the creators of metadata should be able to consume metadata of other creators so that they can see the benefits of proper metadata description directly. A bigger and long lasting problem, especially in large enterprises, is that different metadata creators use different terminologies and therefore the same object may receive different metadata descriptions by different people [18]. Another problem is that large taxonomies require certain time for the users to get their bearings so that they can start to use the taxonomies correctly and avoid creating duplicities and other errors. In [57], the author discusses whether taxonomies are really necessary and stresses the importance of establishing relations to documents via URLs, which indicates already a clear shift towards the Linked Data vision.

If we take a look at commercial implementations, there is Microsoft SharePoint's [39] Term Store (a.k.a. Managed Metadata), which enables enterprises using SharePoint to tag objects stored in SharePoint with terms from a taxonomy. However, there are some strong limitations to this approach. There is very restricted multilingual support – separate SharePoint language packs need to be installed for each language to be used in the taxonomy. Also, the implementation is proprietary, thus hindering the integration with taxonomies or data outside of SharePoint.

2.1.2. Linked Data Approach

We propose to represent enterprise taxonomies in RDF employing the standardized and widely used SKOS [41] vocabulary as well as publishing term definitions via the Linked Data principles. This approach entails the following main benefits:

1. Since terms are attached to globally unique identifiers (URIs), which can be dereferenced using HTTP, term definitions can be obtained without the need for additional software (a browser is sufficient).
2. For the same reason, term creation and management can be realized in a distributed scenario, where, for example, certain departments are responsible for different parts of the taxonomy. Terms can then be interlinked and re-used regardless of department boundaries.
3. By employing the SKOS vocabulary, terms can have a hierarchical order and thus the problem of different metadata granularity can be easily solved.
4. Also, since data is represented using RDF, which works with arbitrary vocabularies and properties, overlapping, coinciding or conflicting term definitions (e.g. by different departments) can be interlinked by explicitly stating the relationship between terms via links (e.g. `owl:sameAs`).
5. Terms can be assigned multiple labels³, which are represented as RDF literals. As such they can be furnished with a language tag resulting in multilingual taxonomies with very little additional effort. This is an important factor, especially in large enterprises that operate on an international scale.
6. Ultimately the result of employing the Linked Data approach for enterprise taxonomies is, that terms can be easily re-used in other scenarios as the originally intended ones. For example, a detailed term definition (with alternative labels in multiple languages) can be very useful for search applications, where terms are associated with documents which then become reachable via a variety of keywords.

Although through this approach the amount of time and effort required to build and maintain an enterprise taxonomy can be distributed on many shoulders, the initial costs for building a basic taxonomy may still

be quite high. The problem of finding initial terms (and their definitions) can be solved by integrating a service like DBpedia Spotlight [36] into tools already used (e.g. content management systems). With this service texts can be annotated with resources from DBpedia [7]. Since DBpedia is derived from Wikipedia, terms may already contain a detailed description in multiple languages in many cases. Such a description can be used as starting point for refinement.

For enterprise-specific terms, where a description is not available via DBpedia (which often will be the case in large enterprises like e.g. automotive companies with a specific terminology), a keyword extraction service like FOX [14] can be used instead. Such a service will also return a well-known URI if feasible, but will also return URIs for other found keywords (with no or little metadata attached). Thus an initial set of term URIs can be gathered, which can then be annotated in a second step.

For this manual process we propose the usage of a tool like OntoWiki (see Section 2.3). OntoWiki is a data wiki and an extensible application framework. It can be used to browse and author arbitrary RDF knowledge bases, but it can also be easily adapted to special use-cases such as taxonomy management. The advantage of employing OntoWiki for such a scenario is that all information is made available via the Linked Data principles automatically without any additional effort. Furthermore, all changes are tracked and can be reverted like in any other wiki. OntoWiki also supports the distributed maintenance of taxonomies by allowing discussions on a per-resource basis.

Figure 3 shows two screenshots, which demonstrate some of the advantages described in this section. The screenshot on the left side shows OntoWiki, which displays the definition of the term *T-Modell* contained in a taxonomy used within Daimler along with some additional information. The location bar on the top of the screen displays the URI used for this very concept, which other resources can link to. It is also possible to directly de-reference this identifier and obtain the description for this resource in a machine-readable format. The properties table for this term shows:

- the type of this resource (`skos:Concept`),
- a link to a concept that is broader (hierarchical order),
- a textual description of the meaning of this term,
- preferred labels for this term in different languages as well as
- an alternative label *Combi*.

³`skos:prefLabel`, `skos:altLabel`

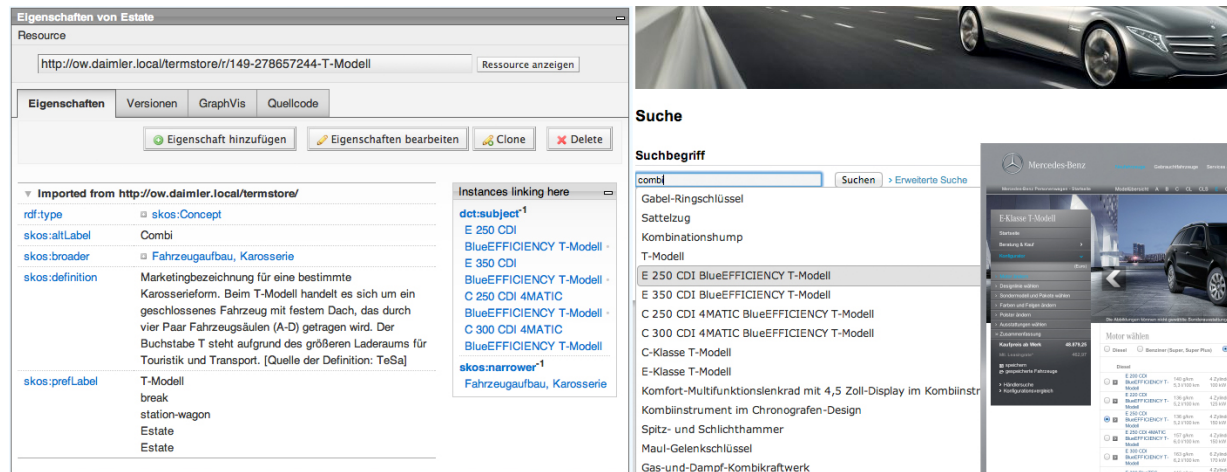


Fig. 3. The left side shows OntoWiki, which displays the definition of the term *T-Modell* from the Daimler Enterprise Taxonomy and other resources that link to this term. The right side shows a search application after searching for *combi*, which employs the term metadata as well as the links to this very concept for finding and suggesting relevant content.

Additionally, a small box on the right side of the OntoWiki screen shows other resources that link to this term. As one can imagine, the broader concept from above also contains a link to this term stating that it is a narrower term (*skos:narrower*). More interestingly the other links show that certain car models link to this concept. This circumstance is used in the search application, which is shown on the right side of Figure 3. This screenshot shows an adapted *daimler.com* website with a search field. When a user types the keyword *combi*, the knowledge base is used to obtain the fact, that this search term is a synonym for the concept *T-Modell*. Once this is done, all linked car models are retrieved and shown to the user. When a user clicks on such a link, he is directly forwarded to the car configurator for this particular model. The depicted scenario is a good example of an application of a taxonomy outside the scope of the originally intended use. One main reason for the efficient realization of this application is that data from multiple sources (term store and car configuration metadata) was made available via the Linked Data principles.

2.1.3. Challenges

Currently terminology in large enterprises is managed in a centralized manner mostly by a dedicated and independently acting department (often called *Corporate Language Management (CLM)*). CLM is in charge to standardize all corporate terms both for internal and external use. As a result they create a variety of dictionary files for different scopes that are not intercon-

nected. An employee that wants to look up a certain term, needs to know which dictionary to use in that very context, as well as where to retrieve the currently approved version of it. The latter may not always be the case, especially for new employees. The former applies to all employees, since it might be unclear, which dictionary should be used, resulting in a complicated look-up procedure or worse the abandonment of a search at all. As a result, the main challenge in the area of enterprise taxonomies is defragmentation of term definitions without centralization of taxonomy management.

Defragmentation. The proposed Linked Data approach can be implemented by keeping the centralized structure and assign CLM the task to create a knowledge base that contains the complete terminology of the company. This solves the fragmentation problem occurring with the dictionary approach, but it also keeps the barrier for participation high, since a single department still is in charge for maintaining the knowledge base. While it would be easy to assign each employee in the company the right to manipulate the taxonomy, this would very unlikely be the case, as long as a single department is responsible for it (e.g. for outwards communication).

Decentralization. On the other hand with Linked Data an entire decentralized solution can be implemented, which assigns each department in the enterprise its own taxonomy namespace. Adding new terms or refactoring existing terms becomes very easy with

this approach, due to the reduced communication overhead with other departments. Nevertheless some technical and organizational problems need to be tackled in this case, namely: Who is responsible for a given term? Which is the approved version of a term definition? Who decides, which terms are released for use in external communications? Also the problem of fragmentation arises again. It could be tackled by automatically crawling all known sub-taxonomies and providing a unified search interface.

2.2. XML Schema Governance

In most enterprises today, XML is used for message exchange, data integration, publishing and storage, often in a form of Web Services and XML databases. To be able to process XML documents efficiently, we need to know what kind of data we can expect to find in them. For this purpose, XML schemas should be present for each XML format used. XML schemas describe the allowed structure of an XML document. Currently, there are four wide-spread languages for XML schema description, the oldest and simplest DTD [11], the popular XML Schema [61], the increasingly used Relax NG [13] and the rule-based Schematron [27]. In a typical enterprise, there are hundreds or even thousands of XML schemas in use, each possibly written in a different XML schema language. Moreover, as the enterprise and its surrounding environment evolves, the schemas need to adapt. Therefore, new versions of schemas are created, resulting in a proliferation of XML schemas. *XML schema governance* now is the process of bringing order into the large number of XML schemas being generated and used within large organizations.

2.2.1. State of the Art

There are different tools that allow us to visualize individual schemas [5,49,60] and there are tools that allow us to discover relationships among schemas [4]. Then there are integration products that allow to establish mappings between XML schemas [3] as well as server solutions that provide large libraries of XML schemas used in wide-spread enterprise standards. They enable us to execute the mappings and handle message transformations between different formats [2,37,28].

Even though XML is available since 1998, there are still some areas of XML schema management that remain unsolved or that are still undergoing research. One of the areas is conceptual modeling of XML

schemas [45] and XML schema evolution [44] implemented in an XML schema management tool *eXolution* [30]. The conceptual model can help with XML schema governance when there are lots of schemas describing the same problem domain as it is often the case in large distributed organizations. There are two basic use cases for the conceptual model. First, enterprises have to communicate with various providers or consumers of similar data. Each of the partners can work with a slightly different XML format to describe the same data and the enterprise needs to adhere to those formats. The second case is also wide-spread. When we use XML for communication among components of an enterprise IT infrastructure, parts of various messages often overlap. This can be, for example, the customer information in a message containing an order and the customer information in a message containing a list of all customers. In that case, it is useful to have a conceptual model in place that interconnects the various XML schemas and provides a visual documentation and overview of which concepts are used in which XML schemas. The interconnection is made via mapping of concepts present in the XML schemas to the concepts in the conceptual model. The conceptual model is also crucial in the above mentioned problem of evolution of multiple XML schemas. Suppose the representation of customer information changes (e.g. turnover figures should be stored additionally with each customer). This change needs to be reflected in all the XML schemas in which customer information is present. Thanks to the mappings to the conceptual model, these changes can be propagated to all affected schemas automatically. In addition, having the conceptual model can ease the process of publishing XML data as Linked Data. We can map the conceptual model to an ontology, we can then automatically create XSLT scripts to transform XML data valid against our modeled XML schemas to RDF according to the ontology and back again [31].

2.2.2. Linked Data Approach

The conceptual model approach is viable only for a set of schemas within a common problem domain. However, in large enterprises, there are often multiple problem domains in which XML solutions are deployed. In that case, we would like to have at least a light-weight overview of the existing schemas, the languages and namespaces used, the element and attribute naming conventions and the links between the schemas. For this purpose an XML schema repository would be useful. We have taken first steps toward such

a repository and created an *XML schema repository ontology*⁴, which is a Linked Data approach for describing XML schemas. Part of the description can be generated from the XML schema itself, another part (like the contact responsible for the schema) can be added manually. The advantage is that, like with other Linked Data, users are able to query the repository using standard SPARQL queries, they can collaborate on the descriptions using tools like OntoWiki and they can link to the schema (or element or attribute) descriptions from other Linked Data using their URIs. Those can be taxonomy concepts (cf. Section 2.1), development or design discussions in wikis or even emails or other text documents that can contain URIs like documentation, presentations etc. Thus, the already existing landscape of XML data, tools and technologies can be interlinked and integrated with other information and data sources in the enterprise.

2.2.3. Challenges

The sheer number of IT systems deployed in large enterprises that make use of the XML technology bear a challenge in *bootstrapping and maintaining an XML schema repository*. In order to create such a repository a bridging between XML schemata and RDF needs to be established. This requires in a first place the identification of XML schema resources and the respective entities that are defined by them. Some useful information can be extracted automatically from XML schema definitions that are available in a machine-readable format, such as XML Schema and DTDs. While this is probably given for systems that employ XML for information exchange, it may not always be the case in proprietary software systems that employ XML only for data storage. In the latter case as well as for maintaining additional metadata (such as responsible department, deployed IT systems, etc.) a substantial amount of manual work is required. In a second step the identified schema metadata needs to be represented in RDF on a fine-grained level. The challenge here is the development of an ontology, which not only allows for the annotation of xml schemas, but also enables domain experts to establish *semantic relationships* between schemas. Another important challenge is to develop methods for capturing and describing the evolution of XML schemata, since IT systems change over time and those revisions need to be aligned with the remaining schemas.

2.3. Wikis

Wikis have become increasingly common through the last years reaching from small personal wikis to the largest Internet encyclopedia Wikipedia. The same applies for the use of wikis in enterprises [34].

2.3.1. State of the Art

Since large companies have special requirements, such as fine-grained access-control, enterprise scalability, security integration and the like, a special type of wikis – enterprise wikis – emerged. Wiki-style applications are employed in different scenarios, although their purpose in the majority of cases is documentation. A survey [34] about the use of corporate wikis pointed out, that corporate users expect three benefits from using wikis, namely improved reputation, relaxation of work and helping in the advancement of processes. In [48] the authors categorize enterprise wikis into *single contributor wikis*, *group or team wikis* and *internal-use encyclopedias*. Hence, wikis are often used to either gather and organize knowledge of individual employees or small teams or manage enterprise-wide knowledge. Widely utilized wikis in the enterprise context are Confluence [6] and Jive [29]. Popular open-source wikis include Foswiki [17] and TWiki [62]. These tools differ in their provided functionality, but they are all mainly centered around textual content, although some wikis provide limited support for managing structured information (e.g. Foswiki via data forms). Consequently, the knowledge contained in those wikis can in most cases only be extracted by human reading of the documents and not by other applications used within the company.

2.3.2. Linked Data Approach

In addition to traditional wikis, there is also another category of wikis which are called semantic wikis. Those can again be divided into two categories: semantic text wikis and semantic data wikis. Wikis of this kind are not yet commonly used in enterprises, but crucial for enterprise data integration since they make (at least some of) the information contained in a wiki machine-accessible. Text-based semantic wikis are conventional wikis (where text is still the main content type), which allow users to add some semantic annotations to the texts (e.g. typed links). The semantically enriched content can then be used within the wiki itself (e.g. for dynamically created wiki pages) or can be queried, when the structured data is stored in a separate data store. An example is Semantic MediaWiki [32] and its enterprise counterpart SMW+ [46].

⁴<http://purl.org/xsr>

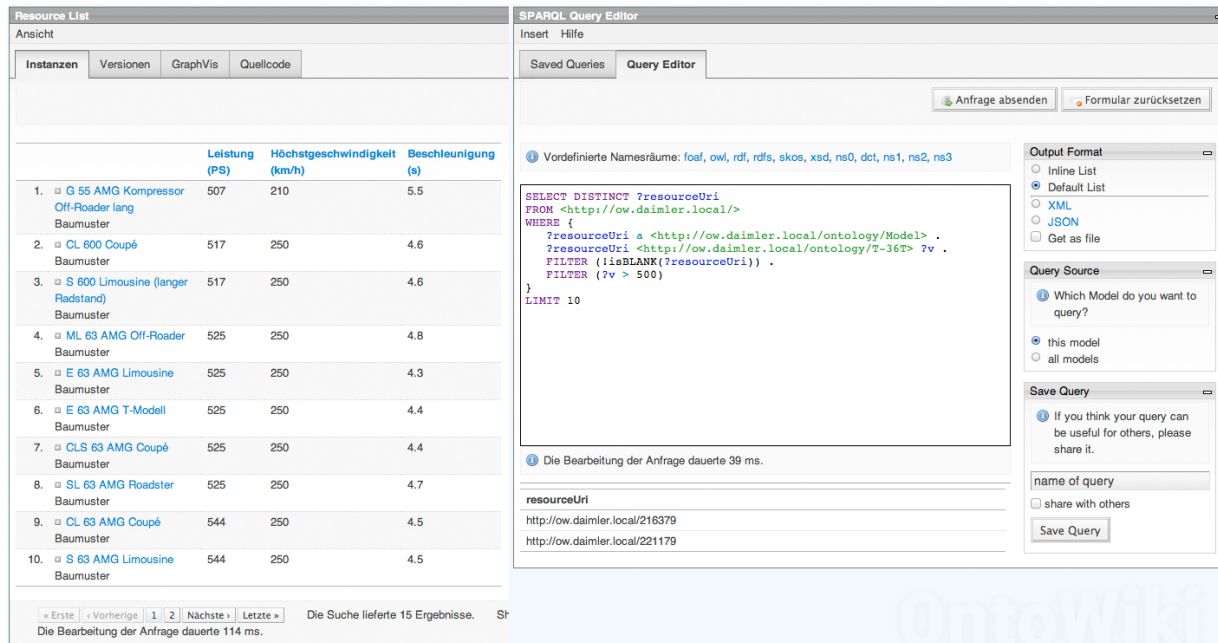


Fig. 4. The left side shows a list of car models together with some additional attributes. This list was dynamically created from the SPARQL query shown in the right screenshot.

It extends the well-known MediaWiki [35] engine (which powers Wikipedia) with syntax for typecasting links and data, classifying articles and creating dynamic pages. The knowledge in a wiki (KiWi) [54] project also developed a semantic wiki, which provides an adaptable platform for building semantic and social tools.

We propose the usage of semantic data wikis such as OntoWiki [8,22] in enterprises for the following main reasons:

1. Data wikis focus on structured information, which is kept as such and thus can be easily reused by other applications consuming the data.
2. Since OntoWiki is solely based on RDF, all information is automatically published via the Linked Data principles, making it trivial for other parties to consume the data.
3. Information fragments can be interlinked with other resources within an enterprise (e.g. taxonomies, XML schemas, databases, Web services), which leads to a better reuse of information and thus to better maintained data.
4. Since textual information can also be represented in RDF (via literals), text wikis can be emulated and thus (additional) human-friendly information can be added. Such annotations and the

structured information can then be used to create customized views on the data.

Figure 4 shows an example of structured information that is maintained within a data wiki. The figure contains two screenshots. The picture on the left shows OntoWiki displaying a list of car models that was dynamically created using the SPARQL query shown on the right. In a traditional wiki such a list needs to be created and maintained manually and the resulting table would only be useful for humans, who are able to understand the meaning. Furthermore, since this is a wiki, if a user detects an error in the data of this automatically generated list, he can fix it right in place and all users (including other applications that access the information) benefit from this change automatically.

2.3.3. Challenges

Since wikis in large enterprises are still a quite new phenomena, the deployment of data wikis instead of or in addition to text wikis will be relatively easy to tackle. A challenge, however, is to train the users of such wikis to actually create semantically enriched information. For example, the value of a fact can be either represented as a plain literal, or as a relation to another information resource. In the latter case the target of the relation can be identified either by a newly gen-

erated URI or one that was introduced before (eventually already attached with some metadata). The more users are urged to reuse information wherever appropriate, the more all participants can benefit from the data. It should be part of the design of the wiki application (especially the user interface), to make it easy for users to build quality knowledge bases (e.g. through auto-suggestion of URIs within authoring widgets). Since data in RDF is represented in the form of simple statements, information that naturally is intended to be stored in conjunction (e.g. geographic coordinates) is not visible as such per se. The same applies for information which users are accustomed to edit in a certain order (e.g. address data). A non-rational editing workflow, where the end-users are confronted with a random list of property values may result in invalid or incomplete information. The challenge here is to develop a *choreography of authoring widgets* in order to provide users with a more logical editing workflow.

Another defiance to tackle is to make the deployed wiki systems available to as many stakeholders as possible (i.e. cross department boundaries) to allow for an improved information re-use. Once Linked Data resources and potentially attached information are reused (e.g. by importing such data), it becomes crucial to keep them in sync with the original source. Therefore mechanisms for *syndication* (i.e. propagation of changes) and *synchronization* need to be developed, both for intra- and extranet semantic wiki resources.

Finally, it is also necessary to consider access control in this context. Semantic representations contain implicit information, which can be revealed by inferencing and reasoning. A challenge is to develop and deploy scalable access control mechanisms, which are aligned with existing access control policies in the enterprise and which are safe with regard to the hijacking of ontologies [24].

2.4. Web Portal and Intranet Search

The biggest problem with enterprise intranets today is the huge difference in user experience when compared to the Internet [42]. When using the Internet, the user is spoiled by modern technologies from, for example, Google or Facebook, which provide very comfortable environments, precise search results, auto-complete text boxes etc. These technologies are made possible through large amounts of resources invested in providing comfort for the millions of users, customers, beta testers and by their large development team and also by the huge number of documents avail-

able, which increases the chances that a user will find what he is looking for. In contrast, in most enterprises, the intranet experience is often poor because the intranet uses technologies from the previous millennium.

2.4.1. State of the Art

Putting the money issue aside and assume that a company has a state-of-the-art intranet data management system with proper full text search and a comfortable user interface. The current search engines provide federated search, i.e. for a single user query multiple data sources are searched, possibly including the entire Internet as a supplementary source and the results are merged, ranked and presented to the user. They include Microsoft's FAST Search [38], SAP's Netweaver Enterprise Search [53] or Autonomy's IDOL Universal Search [10]. These search engines are based on full-text search, taxonomy support, custom ranking and context awareness. Even though the search engines are quite sophisticated, there is still a lot of room for improvement that can be tackled by publishing the data as Linked Data and allowing it to be queried as such [19]. In [55], the authors present a vision of enriching enterprise search with the use of business domain, technical and user information ontologies and their mappings. In [21], the author identifies several challenges in enterprise search, one of them being internal multi-source search, which is when a user has a precisely formulated question but only a keyword search is available. The author uses an example where a manager in an oil company wants to identify all wells previously drilled by the company in the Black Gold field where the problem known as "stuck pipe" was experienced. The manager searches for "Black Gold stuck pipe", but he must go through all the found documents, identifying the wells and so on. If the company data was stored or internally published as Linked Data using an ontology describing the oil drilling domain, the result could be gained using a single SPARQL query. There are also tools for indexing RDF files on the Internet (Sindice [47]) that could be used within enterprises as search tools if the enterprise was storing Linked Data as RDF files.

2.4.2. Linked Data Approach

In an enterprise exist at least two distinct areas where search technology needs to be applied. On the one hand, there is corporate internal search, which enables employees to find relevant information required for their work. On the other hand, all large enterprises need at least simple search capabilities on their public web portal(s), since otherwise the huge amounts of in-

formation provided may not be reachable for potential customers. Some dedicated companies (e.g. automotive companies) would actually have a need for more sophisticated query capabilities, since the complexity of offered products is very high. Nevertheless, in reality, search, both internal and external, is often solely based on keyword matching. We argue that by employing the Linked Data paradigm in enterprises the classical keyword based search can be enhanced. Additionally, more sophisticated search mechanisms can be easily realized since more information is available in a uniform and machine-processable format.

In cooperation with Daimler a prototype that employs a taxonomy (which is available as Linked Data) as well as other Linked Data sources to provide a uniform search application was developed. By entering simple keywords users can

- find documents that are attached to terms from the taxonomy that match the given query,
- find specific car models that match the criteria given by the user (e.g. more than 6 seats) or
- view at least a description if a term matches, but no document is attached to it.

In the first case the advantage is, that documents can be found even if the content does not mention the keyword. Here an initial set of links from documents to terms can be obtained by automatically crawling (just like in state-of-the-art search engines) the website/-document. In addition to that, links to terms can be added manually, e.g. based on query logs that bring to light poor performing queries. The second case would not be even possible without taking another datasource into account, namely structured information about possible car configurations. This data was extracted from an existing XML format and transformed into RDF. Thus, when a user queries for a keyword that matches a term that is linked to a car related property and also provides a value restriction (e.g. < 10), the system can obtain a list of matching cars (via SPARQL queries) and return them to the user together with some meta-data about the models and a link to the car configurator. The last case is only a fallback solution, but it could be used to show users at least something. Of course in this case there needs to be a way to mark terms as public, such that only approved definitions of terms will be displayed. For the purpose of the prototype, we directly queried a triple store, which contained all the RDF data. Since users are used to search engines, which deliver results in the range of milliseconds, we

propose to take the information available in RDF as the basis for creating an index.

Figure 5 shows the prototype with a query that asks for a consumption value with less than 5 liters per 100 kilometers. This results in a list of three cars that Daimler offers. A click on an entry in the list leads the user to the car configurator pre-populated with the chosen values.

2.4.3. Challenges

In order to implement search systems that are based on a Linked Data approach and that provide a substantial benefit in comparison with traditional search applications, the challenge of *bootstrapping an initial set of high-quality RDF datasources* needs to be tackled first. For example, as a prerequisite for linking documents to terms a hierarchical taxonomy should be created (Section 2.1). Mechanisms then need to be established to automatically create high-quality links between documents and an initial set of terms (e.g. by crawling), since it is not feasible to manually link the massive amount of available documents. Furthermore, the process of semi-automatic linking of

- terms that occur in documents but are not part of the taxonomy yet (as well as their placement in the taxonomy) and
- terms that do not occur in documents but are related and thus useful in a search

needs to be investigated and suitable tools should be developed to support responsible employees.

To provide results beyond those that can be obtained from text based documents directly, other datasets need to be transformed to RDF and queried. For example in order to embed a search for car models based on a set of properties, those information (that is already structured, as it is used in car configuration tools) should be converted to RDF first. The data can then be queried in conjunction with other RDF data in a unified manner via SPARQL.

Finally, although a search engine that queries RDF data directly works (in fact the prototype described above was implemented using this approach), it results in suboptimal performance. The challenge here is to develop methods for improving performance to match traditional search engines, while keeping the advantages of using SPARQL directly.

2.5. Database Integration

Relational Database Management Systems (RDBMS) are the predominant mode of data storage in the

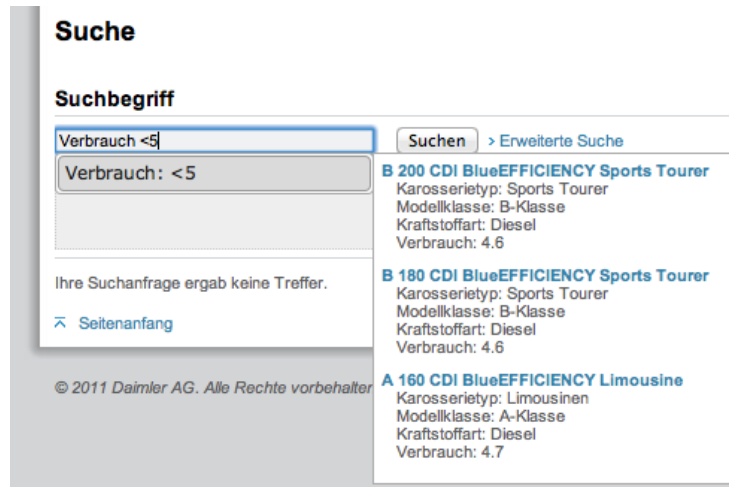


Fig. 5. The screenshot shows a prototype of a search application employing Linked Data. If a user types certain keywords that match properties available for car models, a filtered list matching those restrictions is shown to the user.

enterprise context. RDBMS are used practically everywhere in the enterprise, serving, for example, in computer aided manufacturing, enterprise resource planning, supply chain management and content management systems. We therefore deem the integration of relation data into Linked Data a crucial Enterprise Data Integration technique.

2.5.1. State of the Art

For providing a unified view over different databases multiple methods like data warehousing, schema mediation and query federation have been devised and successfully used. However, problems arise with more heterogeneous data landscapes, where strict schema adherence can not be guaranteed and external data is utilized. The integration of heterogeneous sources requires a costly transformation of the data into the relational model. This has the effect, that only key data sources and thus only a small fraction of the RDBMSes in a typical enterprise are integrated.

2.5.2. Linked Data Approach

The mapping of relational data to the RDF data model adopts relational database integration techniques and augments them. By employing a mapping from relational data to RDF, data can be integrated into an internal or external data cloud. By using URIs for identifying resources, integration with non-relational and external data is facilitated. In terms of the Linked Data life cycle database integration concerns the *extraction* and *storage/query* stages. The general approach of mapping a relational database into RDF us-

ing the RDB to RDF Mapping Language (R2RML) by the W3C RDB2RDF Working Group⁵ is depicted in Figure 6. In essence, the R2RML standard describes how a relational database can be transformed into RDF by means of term maps and triple maps (1). The resulting RDF knowledge base can be materialized in a triple store and subsequently queried using SPARQL (2). In order to avoid a costly materialization step, R2RML implementations can dynamically map an input SPARQL query into a corresponding SQL query (3), which renders exactly the same results as the SPARQL query being executed against the materialized RDF dump. R2RML defines how triples are created from relations, but for facilitating an efficient Enterprise Data Integration it is of paramount importance to provide RDF and SPARQL interfaces to RDBMS. By avoiding a costly materialization of the relational data into a dedicated triple store, a light-weight integration into existing architectures is possible. Consequently, semantic wikis, query federation tools and interlinking tools can work with the data of relation databases. The usage of SPARQL 1.1 query federation [50] allows relational databases to be integrated into query federation systems with queries spanning over multiple databases. This federation allows for example portals in combination with an EKB an integrated view on enterprise data.

⁵<http://www.w3.org/2001/sw/rdb2rdf/>

2.5.3. Challenges

A primary concern when integrating relational data is *scalability and query performance*. With our R2RML based tool SparqlMap⁶ we show that an efficient query translation is possible, thus avoiding the higher deployment costs associated with the data duplication inherent in ETL approaches. The challenge of closing the gap between triple stores and relational databases is also present in SPARQL-to-SQL mappers and drives research. A second challenge for mapping relational data into RDF is a current lack of best practices and *tool support for mapping creation*. The standardization of the RDB to RDF Mapping Language (R2RML) by the W3C RDB2RDF Working Group establishes a common ground for an interoperable ecosystem of tools. However, there is a lack of mature tools for the creation and application of R2RML mappings. The challenge lies in the creation of user friendly interfaces and establish best practices for creating, integrating and maintaining those mappings. Finally, for a *read-write integration* updates on the mapped data need to be propagated back into the underlying RDBMS. An initial solution is presented in [15]. In the context of Enterprise Data an integration with *granular access control* mechanisms is of vital importance.

2.6. Enterprise Single Sign-On

As a result of the large number of deployed software applications in large enterprises, which are increasingly web-based, single sign-on (SSO) solutions are of crucial importance.

2.6.1. State of the Art

While identity management in the sense of consolidated user credentials is already deployed in most larger companies, users still have to provide their credentials frequently when using different applications. Even in cases where a SSO solution is available, it is based on a centralized approach resulting in a single point of failure. Single sign-on solutions (e.g. Microsoft Active Directory [40]) are often based on Kerberos [43], which after an initial credential request uses a mechanism based on tickets to grant access to compatible applications. Another well-known approach for SSO, although originating in the Web 2.0 community and not commonly deployed within company intranets is OpenID [52]. It follows a more distributed approach than other SSO solutions, since

it allows users to use the OpenID provider of their choice (potentially their own). Nevertheless in an enterprise scenario this would very likely be a centralized provider again, thus eliminating this benefit. Recently in [64] the authors exploited some logical security flaws in popular web SSO systems, including OpenID, which hinders OpenID adoption in large enterprises even more. Furthermore, authentication mechanisms employed in enterprises often require the users to remember passwords, which makes the system vulnerable if users do not secure their credentials sufficiently [58]. On the other hand cryptographic smart cards allow for single sign-on solutions based on certificates where all keys computation is done on-card within integrated circuits. As a result users are prevented from unintentionally contaminating login credentials (passwords as well as certificates), since the design of cryptographic smart cards does not allow for extraction of this sensible information. Nevertheless, card-based identities are very limited in regard to attaching background knowledge to them, since there is currently no chance to link them to other information resources as well as extend them with additional arbitrary metadata. The same applies for one-time password (OTP) token based authentication mechanisms, which are frequently deployed in large enterprises, even though usually for accessing the intranet through the internet.

2.6.2. Linked Data Approach

We think that *WebID* [59], which is also based on Linked Data, as well as *X.509 client certificates* [25], are building blocks for a solution in this context. With WebID authentication usernames and passwords can be removed from the login procedure completely. Instead (possibly self-signed) X.509 client certificates are employed, which include the users unique identifier as an extension (namely by specifying a Subject Alternative Name). As long as the user (e.g. via a certificate installed in a web browser) can prove that it knows the private key for a given certificate, the relying party can use the public key in conjunction with the identifier (a URI) to ensure that the connecting user indeed owns that WebID. The URI is de-referenced via HTTPS and a RDF based profile is retrieved, since the WebID is accessible through the Linked Data principles. The public key which is found in the profile is then compared with that given by the certificate. In case this process is successfully completed the system can assert, whether an access to a certain resource is allowed. While X.509 based certificates can only con-

⁶<http://askw.org/Projects/SparqlMap>

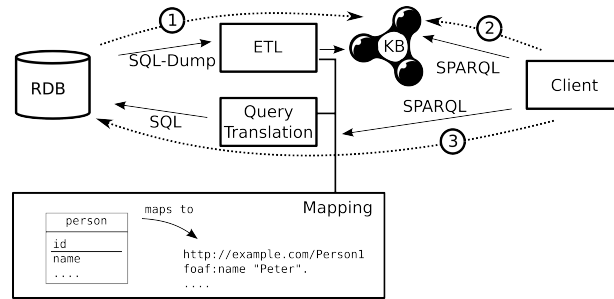


Fig. 6. Mapping based database integration employing query translation and Extract Transform Load (ETL).

tain a very limited set of additional metadata, a WebID profile can contain arbitrary information. Hence the decision, whether access to a resource is granted can be based on very sophisticated assessments (e.g. the assistant of a trusted subject with access may also be granted access, by assessing that the trusted subject has a *assistant* relation to the subject in question). The use of passwords with this approach is minimized, since they are only required to secure the certificate store or by putting WebID certificates on smart cards to grant initial access to the certificate.

2.6.3. Challenges

In order to deploy a WebID based SSO solution in large enterprises, a first challenge is to *transfer user identities* to the Enterprise Data Web. Those Linked Data identities need to be enriched and interlinked with further background knowledge, while maintaining privacy. Thus mechanisms need to be developed to assure that only such information is publicly (i.e. public inside the corporation) available, that is required for the authentication protocol. Another challenge that arises is related to *user management*. With WebID a distributed management of identities is feasible (e.g. on department level), while those identities could still be used throughout the company. Though this reduces the likeliness of a single point of failure, it would require the introduction of mechanisms to ensure that company-wide policies are enforced. Distributed group management and authorization is already a research topic (e.g. dgFOAF [56]) in the area of social networks. However, requirements which are gathered from distributed social network use-cases differ from those captured from enterprise use-cases. Thus, social network solutions need a critical inspection in the enterprise context.

3. Related Work

The enterprise integration survey [12] clearly identifies Web-based technologies (Web Services, SOA) as reasonable technical interoperability solutions. However, at the same time, they are not sufficient to be considered a whole interoperability architecture. A need for a simple, bottom-up, best practice-based approach is expressed. We argue that the Linked Data paradigm is such an approach. In [20], the authors describe the problem of data integration throughout the history. They clearly identify the challenges in the field and survey some approaches to easing the problems. As well as in [51, p. 3], the authors of [20] emphasize the role of XML as the fuel of data integration in the last decade. As future challenges, the authors identify:

- *pay-as-you-go data management* meaning that some level of integration should be provided out-of-the-box and the integrator can later gain further precision and value by investing in better tailored configurations.
- *uncertainty and lineage*, which is a problem of dealing with inconsistent data and tracking the origins of data.
- *reusing human attention* so that we can exploit human interaction with the data to improve its semantic description.

All of these problems can be addressed to some extent by the Linked Data paradigm. The *pay-as-you-go data management* is addressed by usage of appropriate vocabularies. Such data sources provide an initial level of integration immediately thanks to available Linked Data tools. The *uncertainty* is inherent in Linked Data and therefore poses no additional challenge. The *lineage* of data can be managed and tracked by using tools such as OntoWiki [8].

Other approaches to the implementation of semantic web based technologies in enterprise follow. In [26],

a Semantic Web-based Enterprise Application Integration architecture using ontologies and reasoning is proposed. In [16], the authors describe a tool for T-box ontology management using Microsoft Sharepoint, showing that Linked Data projects in the enterprise environment are starting to show up. In [1] OpenLink Virtuoso is presented as an information hub that allows the integration of heterogeneous databases. With its manifold features, Virtuoso provides the services required for an deployment as an EKB. A more loosely coupled integration is presented in [33], where a system federating queries over mapped databases is described. A thorough overview over the modes of mapping databases and other data sources into RDF is given in [23].

4. Conclusions

In this work we identified several data integration challenges that arise in corporate environments. In particular, we examined six areas, namely enterprise taxonomies, XML schema governance, corporate wikis, (intranet) search, database integration as well as single sign-on. Furthermore, we discussed the use of Linked Data technologies in those contexts and presented some insights gained during the development of corresponding prototypes for Daimler. We conclude from our experiments, that the deployment of Linked Data approaches in enterprise scenarios has huge potential and can result in extensive benefits. However, we are aware that more challenges than the aforementioned need to be tackled when trying to create sophisticated enterprise knowledge intranets. We consider as future work the investigation of database integration as a sub-problem of the data integration challenges in enterprises. A Linked Data layer on top of databases, which involves the Linked Data lifecycle stages extraction, (RDF storage,) querying and linking, may facilitate those challenges. In order to ultimately establish Linked Data as a strategy for enterprise data integration also many organizational challenges have to be tackled. For example, it is relatively easy to determine the return-on-investment for an integration of two information systems, while it is very difficult to precisely assess the cost savings of the Linked Data approach. Also, the added value of the Linked Data approach might only become visible after a critical mass of Linked Data interfaces and resources are already established in the enterprise. This is one of the reasons why Daimler plans to establish a Linked Data compe-

tence center, which will assist many IT-related projects in capitalizing on Linked Data technology wherever possible.

References

- [1] Mapping Relational Data to RDF with Virtuoso's RDF Views. <http://virtuoso.openlinksw.com/whitepapers/relational%20rdf%20views%20mapping.html>.
- [2] Altova. FlowForce Server, 2012. <http://www.altova.com/flowforce.html>.
- [3] Altova. MapForce, 2012. <http://www.altova.com/mapforce.html>.
- [4] Altova. SchemaAgent, 2012. <http://www.altova.com/schemaagent.html>.
- [5] Altova. XMLSpy, 2012. <http://www.altova.com/xmlspy.html>.
- [6] Atlassian. Confluence, 2012. <http://www.atlassian.com/software/confluence/overview>.
- [7] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: A nucleus for a web of open data. *The Semantic Web*, pages 722–735, 2007.
- [8] S. Auer, S. Dietzold, and T. Riechert. OntoWiki - A Tool for Social, Semantic Collaboration. In *The Semantic Web - ISWC 2006*, volume 4273 of *Lecture Notes in Computer Science*, pages 736–749. Springer Berlin / Heidelberg, 2006.
- [9] S. Auer, J. Lehmann, and S. Hellmann. LinkedGeoData: Adding a spatial dimension to the Web of Data. *The Semantic Web-ISWC 2009*, pages 731–746, 2009.
- [10] Autonomy. IDOL Universal Search, 2012. <http://www.autonomy.com/content/Products/idol-universal-search/index.en.html>.
- [11] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, and F. Yergeau. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. W3C, 2008.
- [12] D. Chen, G. Doumeingts, and F. Vernadat. Architectures for enterprise integration and interoperability: Past, present and future. *Computers in Industry*, 59(7):647–659, 2008. Enterprise Integration and Interoperability in Manufacturing Systems.
- [13] J. Clark and M. Makoto. *RELAX NG Specification*. Oasis, December 2001. <http://www.oasis-open.org/committees/relax-ng/spec-20011203.html>.
- [14] Dr. Axel-C. Ngonga Ngomo, Norman Heino. FOX – Federated knowledge extraction framework, 2011. <http://aksw.org/Projects/FOX>.
- [15] V. Eisenberg and Y. Kanza. D2RQ/update: updating relational data via virtual RDF. In *WWW (Companion Volume)*, pages 497–498, 2012.
- [16] C. Fillies and F. Weichhardt. Linked Data Interface, Semantics and a T-Box Triple Store for Microsoft SharePoint. <http://www.semtalk.com/pub/LinkedDataSharepoint.pdf>.
- [17] FOS Wiki. FOS Wiki, 2012. <http://foswiki.org/>.

- [18] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. *Commun. ACM*, 30(11):964–971, Nov. 1987.
- [19] J. Grudin. Enterprise Knowledge Management and Emerging Technologies. In *System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on*, volume 3, page 57a, 2006.
- [20] A. Halevy, A. Rajaraman, and J. Ordille. Data integration: the teenage years. In *Proceedings of the 32nd international conference on Very large data bases, VLDB '06*, pages 9–16. VLDB Endowment, 2006.
- [21] D. Hawking. Challenges in enterprise search. In *Proceedings of the 15th Australasian database conference - Volume 27, ADC '04*, pages 15–24, Darlinghurst, Australia, Australia, 2004. Australian Computer Society, Inc.
- [22] N. Heino, S. Dietzold, M. Martin, and S. Auer. *Developing Semantic Web Applications with the OntoWiki Framework*, volume 221 of *Networked Knowledge, Studies in Computational Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [23] S. Hellmann, J. Unbehauen, A. Zaveri, J. Lehmann, S. Auer, S. Tramp, H. Williams, O. Erling, T. T. Jr., K. Idehen, A. Blumauer, and H. Nagy. Report on knowledge extraction from structured sources. Technical Report LOD2 D3.1.1, 2011.
<http://lod2.eu/Deliverable/D3.1.1.html>.
- [24] A. Hogan, A. Harth, and A. Polleres. Scalable authoritative OWL reasoning for the web. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(2):49–90, 2009.
- [25] R. Housley, W. Polk, W. Ford, and D. Solo. Internet X.509 Public Key Infrastructure - Certificate and Certificate Revocation List (CRL) Profile, Apr. 2002.
<http://www.ietf.org/rfc/rfc3280.txt>.
- [26] S. Izza, L. Vincent, and P. Burlat. A Framework for Semantic Enterprise Integration. In D. Konstantas, J.-P. Bourrières, M. Léonard, and N. Boudjlida, editors, *Interoperability of Enterprise Software and Applications*, pages 75–86. Springer London, 2006.
- [27] R. Jelliffe. *The Schematron – An XML Structure Validation Language using Patterns in Trees*. ISO/IEC 19757, 2001.
- [28] JitterBit. JitterBit, 2011.
<http://www.jitterbit.com>.
- [29] Jive. Jive, 2012.
<http://www.jivesoftware.com/social-business/platform>.
- [30] J. Klímek, J. Malý, and M. Nečaský. eXolutio – A Tool for XML Data Evolution, 2011.
<http://exolutio.com>.
- [31] J. Klímek and M. Nečaský. Generating Lowering and Lifting Schema Mappings for Semantic Web Services. In *25th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2010, Biopolis, Singapore, 22-25 March 2011*. IEEE Computer Society, 2011.
- [32] M. Krötzsch, D. Vrandečić, and M. Völkel. Semantic MediaWiki. *The Semantic Web-ISWC 2006*, pages 935–942, 2006.
- [33] A. Langegger, W. Wöß, and M. Blöchl. A semantic web middleware for virtual data integration on the web. In *Proceedings of the 5th European semantic web conference on The semantic web: research and applications, ESWC'08*, pages 493–507, Berlin, Heidelberg, 2008. Springer-Verlag.
- [34] A. Majchrzak, C. Wagner, and D. Yates. Corporate wiki users: results of a survey. In *WikiSym '06: Proceedings of the 2006 international symposium on Wikis*. ACM, Aug. 2006.
- [35] MediaWiki. MediaWiki, 2012.
<http://www.mediawiki.org/wiki/MediaWiki>.
- [36] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems, I-Semantics '11*, pages 1–8, New York, NY, USA, 2011. ACM.
- [37] Microsoft. BizTalk Server, 2011.
<http://www.microsoft.com/biztalk/>.
- [38] Microsoft. FAST Search Server 2010 for SharePoint, 2011.
<http://sharepoint.microsoft.com/en-us/product/capabilities/search/Pages/Fast-Search.aspx>.
- [39] Microsoft. SharePoint Server, 2011.
<http://sharepoint.microsoft.com/>.
- [40] Microsoft. Microsoft Active Directory, 2012.
<http://www.microsoft.com/en-us/server-cloud/windows-server/active-directory-overview.aspx>.
- [41] A. Miles and S. Bechhofer. SKOS Simple Knowledge Organization System Reference. *W3C Recommendation*, 2008.
<http://www.w3.org/TR/skos-reference/>.
- [42] R. Mukherjee and J. Mao. Enterprise search: Tough stuff. *Queue*, 2(2):36, 2004.
- [43] C. Neuman, S. Hartman, and K. Raeburn. The Kerberos Network Authentication Service (V5), July 2005.
<http://www.ietf.org/rfc/rfc4120.txt>.
- [44] M. Nečaský, J. Klímek, J. Malý, and I. Mlýnková. Evolution and Change Management of XML-based Systems. *Journal of Systems and Software*, 85(3):683 – 707, 2012.
- [45] M. Nečaský, I. Mlýnková, J. Klímek, and J. Malý. When conceptual model meets grammar: A dual approach to XML data modeling. *Data & Knowledge Engineering*, 72:1, 2012.
- [46] ontoprise. SMW+ – Semantic Enterprise Wiki, 2012.
<http://www.smwplus.com>.
- [47] E. Oren, R. Delbru, M. Catasta, R. Cyganiak, and G. Tumarello. Sindice.com: A document-oriented lookup index for open linked data. *International Journal of Metadata, Semantics and Ontologies*, pages 37–52, 2008.
- [48] E. S. Poole and J. Grudin. A taxonomy of Wiki genres in enterprise settings. In *WikiSym '10: Proceedings of the 6th International Symposium on Wikis and Open Collaboration*. ACM, July 2010.
- [49] Progress Software Corporation. Stylus Studio, 2012.
<http://www.stylusstudio.com/>.
- [50] E. Prud'hommeaux. SPARQL 1.1 Federation Extensions, November 2011.
<http://www.w3.org/TR/sparql11-federated-query/>.
- [51] M. Rebstock, F. Janina, and H. Paulheim. *Ontologies-Based Business Integration*. Springer Berlin / Heidelberg, 2008.
- [52] D. Recordon and D. Reed. OpenID 2.0: A Platform for User-Centric Identity Management. In *Proceedings of the second ACM workshop on Digital identity management - DIM '06*, pages 11–16, New York, New York, USA, 2006. ACM Press.
- [53] SAP. SAP Netweaver Enterprise Search, 2012.
<http://www.sap.com/platform/netweaver/components/enterprisesearch/index.epx>.

- [54] S. Schaffert, J. Eder, S. Grünwald, T. Kurz, and M. Radulescu. Kiwi—a platform for semantic social software (demonstration). *The Semantic Web: Research and Applications*, pages 888–892, 2009.
- [55] K.-U. Schmidt, D. Oberle, and K. Deissner. Taking Enterprise Search to the Next Level, 2009.
- [56] F. Schwagereit, A. Scherp, and S. Staab. Representing Distributed Groups with dgFOAF. *The Semantic Web: Research and Applications*, pages 181–195, 2010.
- [57] C. Shirky. Ontology is overrated: Categories, links, and tags, 2005.
http://www.shirky.com/writings/ontology_overrated.html.
- [58] G. Skaaf. An alternative to passwords? *Biometric Technology Today*, 15(5):10–11, 2007.
- [59] M. Sporny, T. Inkster, H. Story, B. Harbulot, and R. Bachmann-Gmür. WebID 1.0: Web identification and Discovery. W3C Editors Draft, Dec. 2011.
<http://www.w3.org/2005/Incubator/webid/spec/>.
- [60] Syncro Soft. oXygen XML editor, 2012.
<http://www.oxygenxml.com/>.
- [61] H. S. Thompson, D. Beech, M. Maloney, and N. Mendelsohn. *XML Schema Part 1: Structures (Second Edition)*. W3C, 2004.
- [62] TWiki. TWiki, 2012.
<http://twiki.org/>.
- [63] H. Wache, T. Voegelé, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner. Ontology-based integration of information—a survey of existing approaches. *IJCAI-01 workshop: ontologies and information sharing*, 2001:108–117, 2001.
- [64] R. Wang, S. Chen, and X. Wang. Signing Me onto Your Accounts through Facebook and Google: a Traffic-Guided Security Study of Commercially Deployed Single-Sign-On Web Services.