

ExConQuer: Lowering barriers to RDF and Linked Data re-use

Editor(s): Name Surname, University, Country

Solicited review(s): Name Surname, University, Country

Open review(s): Name Surname, University, Country

Judie Attard *, Fabrizio Orlandi and Sören Auer

Enterprise Information Systems, University of Bonn, Regina-Pacis-Weg 3, 53113 Bonn, Germany

E-mail: attard@iai.uni-bonn.de, orlandi@iai.uni-bonn.de, auer@cs.uni-bonn.de

Abstract.

A major obstacle to the wider use of semantic technology is the perceived complexity of RDF data by stakeholders who are not familiar with the Linked Data paradigm, or are otherwise unaware of a dataset's underlying schema. In order to help overcome this barrier, we propose the ExConQuer Framework (Explore, Convert, and Query Framework) as a tool that preserves the semantic richness of the data model while catering for simplified and workable views of the data. With the aim of encouraging and enabling further re-use of Linked Data by people who would otherwise shy away from this task, this framework facilitates the publication and consumption of RDF in a variety of generic formats. In this manner, any stakeholder can export and work with RDF data in the formats they are most accustomed with, radically lowering the entry barrier to the use of semantic technologies, and possibly enabling the exploitation of Linked Data to its full potential. Through the ExConQuer Framework we provide a comprehensive set of tools that enable users to easily query linked datasets, download the results in a number of formats, and re-use previously-executed queries and transformations. With this framework we hence attempt to target the evident niche in existing tools that are intended to be used by non-experts to consume Linked Data.

Keywords: Linked Data, Consumption Framework, Publishing

1. Introduction

The radical advances in technology, particularly through the advancement of the World Wide Web, have created new means to share knowledge. However, although barriers to information access have been lowered through various means (e.g. hypertext links, web search engines, REST APIs), accessibility to raw data was only afforded the same importance [2] in recent years. One of the catalysts for this change is the increasing adoption of Linked Data practices, as indicated by the extraordinary growth in the Linked Open Data Cloud's¹ volume over the past eight years.

Whereas raw data used to be published in barely-interpretable formats such as CSV, the implementation of Linked Data practices has achieved a more meaningful representation of the same data on the Web. Yet, this does not mean such data is easier for the average stakeholder to locate, access, or most importantly, re-use. Individuals facing these hurdles are typically more acquainted with file formats such as generic JSON, XML, basic CSV or other legacy formats such as XML-based Keyhole Markup Language (KML) or GPS Exchange Format (GPX); finding the sophisticated nature of the RDF format overwhelming. Unfortunately, the emergence of a wide number of tools supporting people to publish their data as Linked Open

*Corresponding author. E-mail: attard@iai.uni-bonn.de.

¹<http://lod-cloud.net/>

Data², has not been complemented by approaches supporting them to consume existing Linked Data in formats other than RDF [2]. While such publishing tools are useful in order to ensure the best quality data is published, it is of no use if the consumers do not have the tools or the expertise to exploit it.

We here propose the *ExConQuer Framework*³ (Explore, Convert, and Query Framework); a set of tools whose aim is (i) to facilitate the publication and consumption of RDF data in a wide variety of generic, legacy or domain-specific formats⁴, as well as (ii) to enable stakeholders to easily re-use persisted transformations. The framework is based on the concept of *RDF softening*. In contrast to the semantic *lifting* of data into RDF, which addresses the enrichment, mapping, and transformation of semantically shallow formats, the softening process is then:

The generation of domain-specific RDF data views in semantically-shallow representation formalisms.

This will enable stakeholders to more easily obtain, interpret and re-use existing Linked Data in conventional formats. Moreover, any transformations executed on the data are persisted to enable their re-use. Initiatives such as the one undertaken by the W3C CSV on the Web working group⁵, which aims to standardise JSON-LD serialisation, promise to lower the entry barrier to Linked Data re-use. Yet to the best of our knowledge, very few approaches address the need for the provision of semantically-rich RDF data in shallower formats. Although this might appear to be counter-productive, it is favourable to offer the reduction of a degree of semantics in favour of an increase in the *degree of (re)usability* by stakeholders who would otherwise shy away from using the data. Through retaining provenance information we also ensure that the softening process does not result in the loss of the richness of RDF representation, and users are also given the option to lift back the results to RDF.

Based on the motivation of providing stakeholders with a tool that enables them to consume Linked Open Data easily without requiring previous knowledge of

RDF, SPARQL, or the datasets' underlying schema, we provide the following contributions as part of the ExConQuer Framework:

- The **Query Builder Tool**⁶: enables users to explore, query, and convert datasets (or subsets) through endpoints;
- **RDF2Any API**: provides the functionality to query and convert RDF datasets into a number of different formats through RDF softening;
- The **ConQuer Ontology**⁷: used to represent transformations carried out in the Query Builder;
- **The PAM Tool**⁸: a provenance-aware management system that enables users to explore and re-use Linked Data Publications;
- **Usability Evaluation**: an evaluation of 27 users on the tools within the ExConQuer Framework to analyse its usability.

We continue this paper by discussing related work in the literature in Section 2. We provide our approach in Section 3. Then we discuss the led evaluation in Section 4, and we finally give our concluding remarks in Section 5.

2. Related Work

Our approach is varied in nature, comprising data exploration, query generation, data views, and a provenance-aware management system. To the best of our knowledge, there is no Linked Data consumption framework with all the functions as the one we propose. Yet, there are a number of tools that tackle the different approaches separately.

2.1. Linked Data Exploration Systems

In the ExConQuer Framework we enable users to explore datasets in order to identify if and how the data they require is represented in existing open datasets. Therefore we here explore various data exploration systems. The author of [8] distinguishes between *lookup* and *exploratory search* activities. Lookup activities are done to satisfy specific information needs, such as searching for a known item, where the user has defined keywords to use. On the other hand, *exploratory search* refers to cognitive consuming search

²<http://www.w3.org/wiki/LinkedData>

³More information on the framework, including source code and evaluation results, can be found here: <http://eis.iai.uni-bonn.de/Projects/ExConQuer>

⁴While hundreds are in existence: http://en.wikipedia.org/wiki/List_of_file_formats, we here focus on the most popular ones such as JSON, CSV and RDB

⁵http://www.w3.org/2013/csvw/wiki/Main_Page

⁶<http://purl.org/net/exconquer/builder>

⁷<http://purl.org/eis/vocab/cqo>

⁸<http://purl.org/net/exconquer>

tasks, such as learning or investigation. Here, the information need is less well-defined than in a lookup activity and the keywords are not known in advance, therefore also evolving during the activity. In our approach we cater for both activities, where users are given both results that exactly match the specified keyword, and also results that are related to that keyword.

The authors of [16] attempt to facilitate exploratory search by extending their own base browser through the implementation of three search paradigms; keyword-based, view-based, and content-based. The browser also enables dataset exploration through adaptive result overviews and incremental graph-based resource exploration. A drawback for using this approach is the possibility of information overload, since a huge dataset might result in an enormous amount of facets or nodes.

The authors of [6] use Facet Graphs in their approach to build semantically unique queries. Users are given the option to choose the result set they need, as well as the facets to filter it. Both are represented as nodes in a graph visualisation and enable them to produce a personalised interface to build search queries. Compared to the previous approach in [16], by enabling users to enter keywords the authors reduce the risk of information overload.

In [13], the authors present Explorator, a tool for exploring RDF data through direct manipulation. Users are enabled to explore a semi-structured RDF database through browsing and searching. While the led experiments and studies indicated that users with a basic knowledge of RDF were able to use the tool, the authors also point out that the Explorator is better suited to advanced users who have solid knowledge about RDF, further motivating our approach.

In [10], the authors propose Visor, a multi-pivot approach that allows users to explore datasets from multiple points in the graph. Visor consists of a generic data explorer tool that can be configured on any SPARQL endpoint. Here, a user is able to explore existing classes in the dataset at hand, related the properties and classes, and individual instances. A graph is then rendered in order to show the user selection and the relations between them (if any). Visor enables users to query a user's selection by creating custom spreadsheets, and then convert them to CSV or JSON.

While numerous tools that enable users to explore Linked Data exist, most of them are targeted for more experienced users who have some knowledge of either RDF or the data's underlying schema. Therefore, such

tools are unsuitable to fit our aim of lowering the entry barrier towards re-using Linked Open Data.

2.2. SPARQL Query Builders

The first process towards achieving reusability is data access. Linked Open Data is usually accessible on data portals or catalogues through SPARQL endpoints or data dumps. The latter method for accessing data has the disadvantage of generally resulting in a large bulk of data, with the user having no control to get specific data (such as a subset) from the data the provider made available as a dump. Moreover, data might also be outdated. While SPARQL endpoints allow thorough control over what data to access, then there is the disadvantage of having to use SPARQL, and using SPARQL to search through data stores is a tedious process and limits data access to Semantic Web practitioners [3,4]. This is mainly due to two reasons; (i) because of the syntax barrier, and (ii) due to the heterogeneity of the data and its schema. As yet, there are few tools that help inexperienced users with respect to the creation and editing of SPARQL queries.

The authors of [12] present NITELIGHT, a tool that enables users to create SPARQL queries using a set of graphical notations and GUI-based editing actions. NITELIGHT uses a visual query language, vSPARQL, to provide graphical formalisms for SPARQL query specification. Users can construct a query through dragging and dropping ontology elements. This approach, while suitable for users with at least a minimal understanding of the SPARQL query language, is not suitable for users who do not know SPARQL or the underlying schema of the dataset to be queried.

Similar to NITELIGHT, the authors of [5] also implement a visual approach. The authors define it to be a novel approach for visual SPARQL querying based on the filter/flow model. Thus, no structured text input is required, rather, queries can be generated entirely through the use of graphical elements, and filter restrictions are shown, rather than a representation of the complete query. While this approach does not require knowledge of the SPARQL query language, users are expected to be familiar with the Semantic Web and the filter/flow concepts. Moreover, while this approach allows users to query a dataset, they need to know if and how the information they need is available in the dataset in question.

In contrast to the above, the authors of [11] present an approach where users can enter a natural language query that is then translated into a formal graph query

through the use of query patterns. The aim behind this approach is to hide the complexity of formulating a query expressed in graph query languages such as SPARQL, thus enabling end users to use natural language queries to query ontology-based knowledge bases. The approach described here still has some usability issues. For instance, only English and French can be used as natural languages for the input query. Besides, users who might know the data they need, but not exactly how it is represented in the dataset, will find difficulty in expressing the correct query even if a natural language is used.

QueryMed [14] is the tool that is most similar to our approach for query generation. Focused on the medical domain, this tool enables users with no knowledge of SPARQL to run queries across SPARQL endpoints. The tool requires users to input specific search terms. Users are then given the possibility to filter the results and restrict the query further. A key difference between our approach and QueryMed is that the authors base their search on properties. Thus, when a user selects one or more data stores, the tool displays all the properties within these stores. Apart from resulting in an information overload, this approach is not particularly useful when there many domains involved (e.g. DBpedia), specifically due to the heterogeneity of the data.

2.3. Data Transformations and Management Systems

There are a myriad of tools available for converting between data formats, such as Any23⁹, Datalift¹⁰, Db2triples¹¹, and METAmorphoses¹². However, there are very few tools that ‘soften’ RDF to other formats (such as [15]). Considering RDF is much more expressive than most other formats, it is understandable that efforts and interest are focused in that direction, however we need to cater for users who require the conversion of Linked Open Data (which is generally available in RDF) to a format they understand which is compatible to their native systems. Albeit this might result in some loss of information, the advantages outweigh this shortcoming since it will encourage users to exploit such data, rather than shying away from it due to unfamiliarity with Linked Data or RDF.

The PAM Tool, a provenance-aware management system, is a core contribution within this paper. The

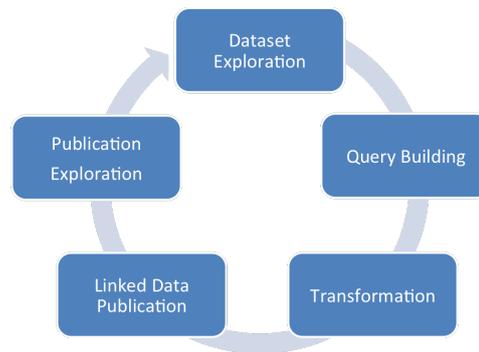


Fig. 1. Abstraction of the ExConQuer Framework Processes

aim behind this tool is to provide a means for users to explore and re-use what we call *Linked Data Publications*. A Linked Data Publication consists of all the information generated in the transformation of data, including the SPARQL query used, its description, the dataset(s) queried, the initial and target data formats, and the user generating the Linked Data Publication instance. The authors of [9] survey existing Linked Data based exploration systems, however all the systems they review are based on exploring data, rather than Linked Data Publications which represent the data, as well as the transformations made on it. SPARQLpedia¹³ is more similar to what we propose, in that it is a service that allows users to submit SPARQL queries in a searchable repository. The PAM Tool follows the same concept, however through retaining provenance information we enable users to not only browse existing queries, but also re-execute them to get updated results or even edit them to refine their query.

3. Approach

The ExConQuer Framework assists data publishers and consumers in exploiting and re-using Linked Data by providing tools that enable them to easily and simply explore, query, transform, and publish Linked Data. Figure 1, shows an abstract overview of the processes within the framework.

Consider a user who requires to use data on actors from the UK. Through the first stage (*Dataset Exploration*), the user can explore DBpedia. The user discovers that actors are represented by the class ‘Actor’.

⁹<https://any23.apache.org/download.html>

¹⁰<http://datalift.org/>

¹¹<http://www.w3.org/2001/sw/wiki/Db2triples>

¹²<http://metamorphoses.sourceforge.net/>

¹³<http://composing-the-semantic-web.blogspot.nl/2009/01/sparqlpedia-sharing-semantic-web.html>

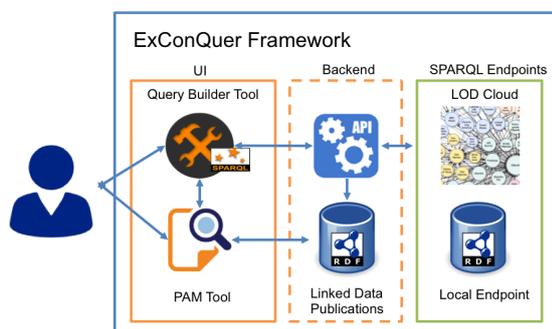


Fig. 2. ExConQuer Framework Architecture

The user then generates a SPARQL query in the *Query Building* step, adding a filter in order to obtain data only about actors having UK as their nationality. The user then has the option to *Transform* the query results into various formats. Since the user wants to explore the data in Microsoft Excel, he converts the results to CSV. The querying and transformation processes are then represented as a *Linked Data Publication*. Through the PAM Tool, the user can explore Linked Data Publications and proceed to re-use, share, or edit them by executing further transformations. Deciding he wants actors over 30 years of age, the user finds the Linked Data Publication and edits his query by adding a filter, and re-downloads the new results in CSV.

The abstract overview in Figure 1 is implemented through the tools provided within the ExConQuer Framework; namely the *RDF2Any API*, the *ConQuer Ontology*, the *Query Builder Tool*, and the *PAM Tool*. Figure 2 shows an overview of the architecture within the framework. The user can create a SPARQL query through the Query Builder Tool, then query a datastore (through a SPARQL endpoint) through API calls. Information pertinent to the executed processes is then persisted in a triple store as Linked Data Publications. The latter are represented with the ConQuer Ontology which we propose for recording provenance data of the transformation. The represented data includes the queried dataset, the SPARQL query, the format conversion, etc. A user can access all this relevant information through the PAM Tool, which allows a user to re-use existing resultsets, modify them through the Query Builder, or otherwise re-use them in semantic-based systems.

3.1. Query Builder Tool

In the ExConQuer Framework we enable users to explore existing open datasets. We target users who either do not know the content of the dataset in question, or otherwise do not know how specific data is represented in this dataset. Our approach is intended to be particularly user friendly and simple, to allow non-experts to easily use the tool to achieve the goal of re-using open data. Through the *RDF2Any RESTful API*, the Query Builder Tool (shown in Figure 3, available online <http://purl.org/net/exconquer/builder>), enables users to navigate through classes, subclasses, instances, and properties in a somewhat similar manner to a faceted browser, without requiring to know the structure of RDF data. The API calls concerned with this exploration task are made up of a number of actions which essentially hide the RDF data model and help in the exploration of RDF data and the underlying structure (e.g. to get class labels). Since the functionality of this tool is provided through an API, this tool can be attached to other frameworks and re-used or extended easily.

3.1.1. Dataset Exploration

Figure 3 shows different parts of the UI of the Query Builder Tool. The following exploration functions are particularly useful for users who do not know exactly what data from the available linked datasets is useful for their purpose, or for those who do not know the underlying schema behind the dataset in question. In Step 1, the user can select any dataset from the auto-complete drop down list or otherwise adding a new endpoint. In Step 2 the user can then proceed to explore the classes contained in the selected dataset. Here the user can either view all classes, or view the classes which match a given keyword. This search can be executed either only on the labels of classes, or otherwise extended to the classes' URIs. The user also has the option (by clicking on the plus button) to expand the view and show the subclasses of the selected class. In this step, along with the classes and subclasses, a number of instances are displayed. Once a class is selected, the user can proceed to the Step 3; the property view, where all the properties of the selected class are shown. The view is divided into *Object Properties* and *Data Type Properties*. The former is when a triple has a URI as its range, while the latter is when a triple has a literal as its range.

Consider as an example the API call required in Step 2. This call abstracts the complexity required to get all

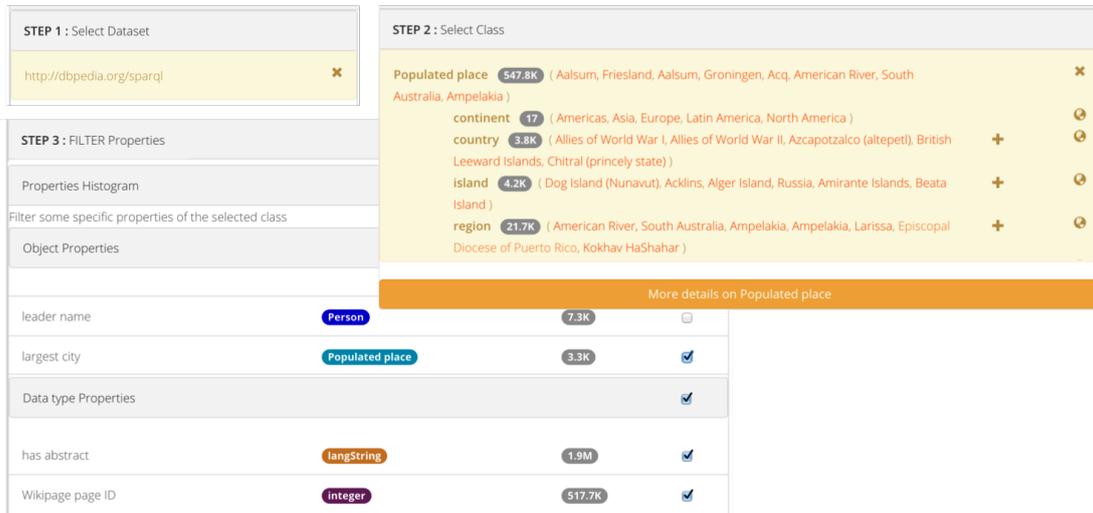


Fig. 3. Query Builder Tool: Different parts of the UI

classes matching a given keyword. After the user has selected the dataset to explore (Step 1), the user enters a keyword and the API call containing the following SPARQL query is executed:

```
SELECT distinct ?class ?label
WHERE {
  {?class rdf:type owl:Class} UNION {?class rdf:
  type rdfs:Class}.
  OPTIONAL {?class rdfs:label ?label}
  FILTER((bound(?label) && (REGEX(?label, <
  search_string>) || REGEX(str(?class), <
  search_string>))))
} ORDER BY ?class
```

Where the `<search_string>` variable will be replaced by the keyword entered by the user. This query uses the selected dataset to look for resources of type `owl:Class` or `rdfs:Class`. Here the user also has a choice (through the UI) to search for the latter resources either through just the resources' labels, or otherwise extend the search to also include resources' URIs. In the case of the former, the query would be a little different, in that both the `OPTIONAL` clause and the `REGEX` component for the class would be removed. The `REGEX` component of the query will enable the classes to be searched and returned to the user on the fly in an auto-complete manner.

3.1.2. Query Generation

Apart from enabling users to explore datasets, the main task of the Query Builder tool is to aid users to generate a SELECT SPARQL query, without requiring prior knowledge of SPARQL or the dataset's underlying schema. This tool enables users to gener-

ate a SPARQL query through a user-friendly interface equipped with auto-complete features. Similar to the exploration function of the Query Builder, the query building function is also enabled through the consumption of the RDF2Any API, where the user selection for dataset, classes, and properties selection is converted into a SPARQL query.

In order to generate a SPARQL query, the user can follow exactly the same procedure as explained in Section 3.1.1. After selecting the dataset and class to query, the user can proceed to select the properties wanted in the resultset. Properties can be freely selected to be included or excluded from the results, and a click on a property allows the user to define a filter. Throughout the query building process, the query, which is generated on the fly, is displayed. Thus, once the user has made the preferred selections, he/she can view and edit the generated query if this is required. Finally, the user has the option to first preview a subset of the results, and then proceed to export the full result set.

3.1.3. Data Transformation

Also provided within the Query Builder Tool, the Transformation function is aimed towards users who need the resultset in a different format than RDF. This might be because their native system is in another format, or simply because they find results in another format more easily readable and interpretable. This *softening* does indeed result in a certain degree of loss in semantics. Yet, this is compensated through retaining links with the original RDF data and other relevant information through the ConQuer Ontology, as

discussed in Section 3.2.1. This means that it is always possible to obtain the original data in RDF through exploiting the provenance information recorded for every transformation and resultset.

The transformation process consists in converting the results in RDF to a number of different formats through the consumption of the RDF2Any API. Currently, the conversions provided are from RDF to CSV, JSON, and RDB, as well as a more advanced configurable conversion. The latter allows a user to convert RDF into potentially any output format, such as XML, KML, TSV (tab separated values), etc. The exception are formats which require memory storage, such as RDBMS serialisation, which requires the storing of foreign key values. The use of the Generic Conversion requires some knowledge about the dataset to be converted, and the user is required to pass required parameters through a template. Apart from being easily extendible with further converters, this tool provides the additional advantage that a user can directly convert the required subset of the dataset in question, rather than converting a bulky data dump. We manually validated the correctness of the various conversions for various queries on different datasets. While we confirm there is a loss from the rich representation of RDF, the essence of the data is retained and the provenance information allows us to retain the link to the original data and the transformations for the reproducibility.

3.2. PAM Tool

All the processes executed through the ExConQuer Framework generate what we call a Linked Data Publication, which is basically what users can share, re-use, explore, and edit. Thus, a Linked Data Publication consists of all the generated information, including the SPARQL query used, its description, the dataset(s) queried, the initial and target data formats, and the user generating the Linked Data Publication instance. We represent all this data using the *ConQuer Ontology*, shown in Figure 4. All generated Linked Data Publications can then be explored using the *PAM Tool* (available online: <http://purl.org/net/exconquer>), which furthermore enables users to re-execute or edit existing queries.

The main aim of the PAM Tool is to provide all stakeholders with the potential to explore all existing queries and transformations executed on different datasets. In this way, a user is given the opportunity to find any results that match his or her requirements. Moreover, if the results are not exactly as the user re-

quires, for example if they are in a different format, or the resulting data is too generic/specific, the user can proceed to edit or update the results with minimal effort, through re-loading the Linked Data Publication on the Query Builder Tool. The Linked Data Publications can also be loaded into semantic-based systems.

3.2.1. ConQuer Ontology

The ConQuer ontology, through the represented information, not only allows us to represent all possible transformations on an entity through querying and converting, but it also allows us to replicate the resulting Linked Data Publications and edit them to achieve different results. Figure 5 shows how, starting from a transformation on a specific dataset (Original Transformation), a user can re-use the query but execute a different conversion on the resultset, or otherwise edit the original SPARQL query in order to obtain different (more generic, more specific, or otherwise) results. Thus, using the ConQuer ontology to represent our transformations allows us to *soften* RDF into semantically shallower formats without actually compromising on the richness of RDF representation, as any resultsets in formats other than RDF are linked back to the original data in RDF. Additionally, through the provenance information, we can track the changes to each entity, and potentially also assign a reputation for the different agents generating the Linked Data Publications.

The main concepts in the ontology are the following:

- **Transformation:** A Transformation represents all the information required to achieve a Linked Data Publication, as described above.
- **Query:** A Query represents a set of statements forming a SPARQL query.
- **Data Resource:** A Data Resource is used to represent a data store. This can be anything from a Linked Open Dataset with a SPARQL endpoint such as DBpedia, to a database or a CSV document.
- **Agent:** An Agent is any entity, whether machine or human, that has some sort of responsibility or control over the generation of a Transformation instance.

To describe the ontology in an informal manner, a *Transformation* has a *Query* that is executed on one or more instances of a *DataResource* (enabling the representation of federated queries). The latter must be a *QueryableResource*, or, in other words, it should be

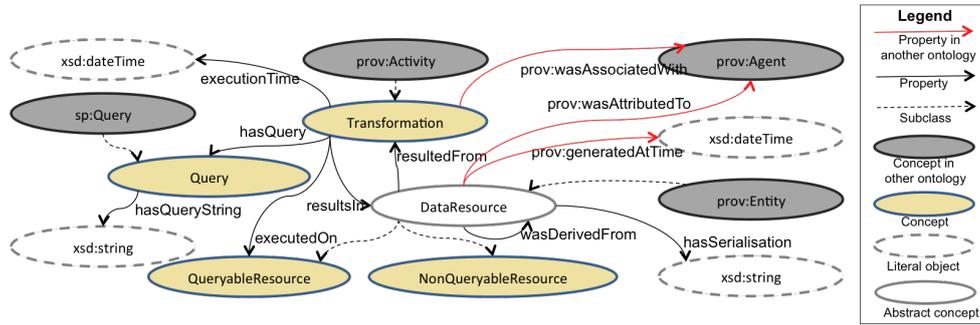


Fig. 4. ConQuer Ontology for Modelling Linked Data Publications

expressed in one of the serialisations of the RDF data format (RDF/XML, NTriple, Turtle, etc.). The resulting *DataResource*, on the other hand, can be either a *QueryableResource* or a *NonQueryableResource* (formats such as CSV, PDF, etc). Finally, each *Transformation* and *DataResource* are linked through the relevant properties.

Since the ConQuer Ontology is representative of Transformations, thus making the latter class the main concept within the ontology, we define a *Transformation* T as follows:

Definition 1 $T = \{q, d, f_d, r, f_r, a, t\}$

where q is a *Query*, d and r are *DataResource* instances (original resource(s) and resultset), f_d and f_r are the serialisation formats of d and r respectively, a is an *Agent*, and t is the time the transformation

was executed. Hence, a generates T , which represents a Linked Data Publication instance. The latter results from applying q to d and then obtaining the final Linked Data Publication by converting f_d to f_r . This means that $r \subseteq d$, as the user can query to get all, or part of resource d .

In the ConQuer Ontology we re-use concepts from the SPIN vocabulary¹⁴, which is used to represent reusable SPARQL queries as templates, and the PROV-O ontology [1], used to represent provenance information. The use of SPIN to represent SPARQL queries not only enables the direct querying of the queries themselves, but also allows the represented knowledge to be re-used in any frameworks or tools using the SPIN vocabulary. The re-used concepts are:

- **sp:Query**: A SPIN concept which represents a SPARQL query. This concept enables us to search within the persisted *Query* instances.
- **prov:Activity**: A PROV-O concept representing something that occurs over a period of time and either interacts with or acts upon *prov:Entity* instances. *prov:Activity* instances can include transforming, consuming, using, or generating entities.
- **prov:Entity**: An *Entity* can be physical, digital, conceptual, or any other thing with a fixed set of aspects.
- **prov:Agent**: This concept represents something or someone who bears some sort of responsibility for an *Activity* taking place or for the existence of an *Entity*.

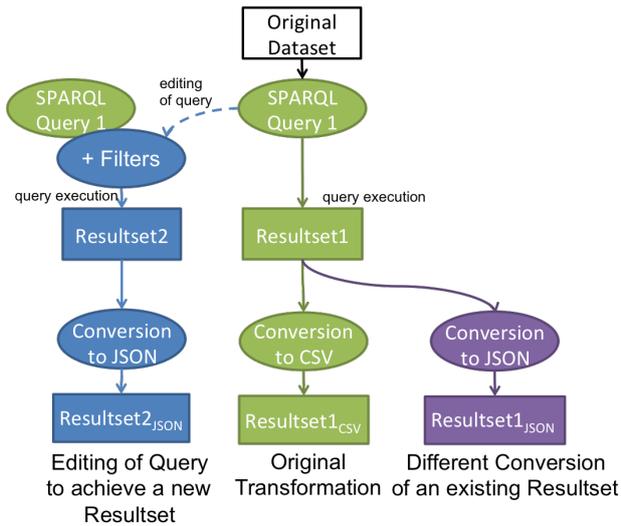


Fig. 5. Example of possible Linked Data re-use scenarios enabled by the ExConQuer Framework and the underlying provenance-aware ConQuer Ontology

3.2.2. Provenance Exploration and Management

We implemented the *PAM Tool* as a provenance-aware publishing and consumption management tool

¹⁴<http://www.w3.org/Submission/spin-overview/>

that enables the exploration of Linked Data Publications with the aim of encouraging their re-use. The motivation behind providing such a tool is that queries are re-usable, and a single query might be the answer to many users' requirements. Besides, the PAM Tool also enables users to persist and re-use complex SPARQL queries. We query the persisted instances of the Linked Data Publications and publish them through a faceted browser (Exhibit¹⁵). Through the use of the ConQuer ontology, the Linked Data Publications have queryable metadata that enables users to search for specific instances using various criteria, such as by the datasets used and the classes queried for. Moreover, a user would be able to search by Agent if the user is required to log in before using the Query Builder Tool¹⁶. Through the persistence of such provenance information, users could query Linked Data Publications according to Agents who have the reputation of providing the best data for the intended use. This tool thus allows users to share, explore, and directly edit (through the Query Builder or otherwise) and re-use Linked Data Publications, whilst keeping data lineage intact.

4. Evaluation

The purpose of this section is to discuss the usability evaluation led on the ExConQuer Framework. Our framework is intended for the use of stakeholders who are not familiar with RDF or the Linked Data Paradigm. This does not exclude stakeholders who already use Linked Data in one way or another, simply because users who are not familiar with the underlying complexity are unable to exploit Linked Data to its fullest potential. For example, a user downloading a data dump of a linked dataset is hardly exploiting the potential of the data in question. For the above reasons, the framework requires to be very user-friendly and provide simple access to the required functionality, whilst also abstracting the underlying complexity. We constructed a survey consisting of 25 questions and split it into three sections, namely questions on the current means and methods of accessing and using Linked Open Data (if any), questions on the Query Builder Tool, and finally questions on the PAM Tool. Where relevant, we used the Likert scale [7] to assess the evaluators' perception of the tools.

4.1. Evaluators' Profile

Since the aim of this evaluation is to identify whether the ExConQuer framework helps or encourages people in the re-use of Linked Data, we shared this evaluation with relevant partners or colleagues who, to some extent or another, had contact with Linked Data. In total we had 27 evaluators. Their domains differ in nature (such as education, healthcare, research, consulting, industry and marketing). 18 of them use Linked Data in their work (mostly for analysis, visualisation and integration), however 9 of them do not. All the evaluators specified more or less the same processes while interacting with Linked Data, namely searching for existing data, accessing and gathering it, cleaning it, integrating it, leading out analyses, and consuming it by visualising it or in other ways such as data mashups. Apart from other issues, nearly all evaluators pointed out that the format of the data hindered them from reusing it, and very commonly data is also incomplete or invalid. Moreover, data might not be accessible at all. 11 out of the 27 evaluators are not familiar with the SPARQL query language, so we were able to interpret the results considering the two different target users.

4.2. Query Builder Tool

In order to have a better insight, the evaluators were asked to describe their current process of querying Linked Data. 7 of the evaluators directly specified they use SPARQL to query Linked Data. The rest either do not use or query Linked Data (13), export the datasets to Microsoft Excel (2), or use other methods for querying such as running test queries or using query designers (5). The evaluators were then asked to access the Query Builder Tool, explore a dataset, formulate a SPARQL query including filters, and download and convert the results in the preferred format.

Questions 1 and 2 in Table 1 show the results for the evaluators' impression of the Query Builder tool. The evaluator who replied with 'disagree' in both question 1 and 2 was of the opinion that tutorials or demo videos would have been helpful with executing the given task. When asked if the Query Builder is a better approach than their current way of consuming Linked Data, only 5 replied 'not sure' while the others all agreed that it would be better. For question 2 in Table 1, almost all the evaluators (22) agreed that they would find this tool useful (to some degree or another) in their SME/Company/Academic Entity. From this part of the evaluation

¹⁵<http://www.simile-widgets.org/exhibit/>

¹⁶This is not currently implemented in our online demo

		Strongly Agree		Agree		Neither Agree nor Disagree		Disagree		Strongly Disagree	
		LD Users	Non Users	LD Users	Non Users	LD Users	Non Users	LD Users	Non Users	LD Users	Non Users
Query Builder											
1.	Do you agree that it was easy to execute this task?	6	1	9	6	2	2	1	0	0	0
2.	Do you agree that this tool would be useful in your SME/Company to access, explore and query open datasets?	4	2	11	5	2	2	1	0	0	0
PAM Tool											
3.	Do you agree that it was easy to execute this task?	3	1	9	5	2	2	2	1	2	0
4.	Do you agree that this tool would be useful in your SME/Company to re-use saved data access queries?	6	1	6	6	3	2	2	0	1	0

Table 1

Results for Tool Evaluation (Complete results: <http://eis.iai.uni-bonn.de/Projects/ExConQuer.html>)

we can conclude that while this tool still requires some improvements with regards to usability, it is however generally deemed to be useful by the target stakeholders (both experts/non-experts, and users/non-users of Linked Data) and is an improvement on their current methods of exploiting Linked Data (if any).

4.3. PAM Tool

For this tool the users were asked to use the PAM Tool to search for the Linked Data Publication they just created in the previous section of the survey, then reload and edit it on the Query Builder Tool. The users were able to use a number of facets to filter the results. For this tool, the responses to question 3 in Table 1 were somewhat varied, however the majority of the evaluators still agreed that the tool is quite easy to use, and that it would be useful to their company. Most of the comments from the negative replies pointed out that the tool took quite long to load, and the user who selected ‘strongly disagree’ commented that we show too many details (such as the SPARQL query). On the other hand, when asked question 4, the evaluators’ replies were mostly positive. Yet again, this indicates that while the tool needs improvement, mostly efficiency-wise, the majority of the evaluators still consider the tool to be useful.

5. Conclusion

It is evident that the use of Linked Data principles is increasing at a fast rate, as indicated through the exponential growth of the Linked Open Data Cloud. This increase is also reflected in tools aiding users in the publishing process, yet, tools aiding users to consume and re-use Linked Data are still not that prevalent. We hence identify a niche with regards to approaches that abstract the complexity beneath exploiting linked

datasets and propose the *ExConQuer Framework*. In order to provide more simple and workable views of the data, in this framework we transform RDF data into a number of different formats, whilst still preserving the semantic richness of the RDF data model. While during this softening we do lose some of the richness of RDF representation, we compromise by preserving the link with the original RDF data through the ConQuer ontology, and still retain the full semantic richness through provenance information. As is evident through the led usability evaluation, the ExConQuer Framework is particularly useful to encourage the re-use of Linked Data by stakeholders who are not familiar with RDF, and are more acquainted with formats such as JSON or CSV. Our framework is also useful for more expert users who are however not able to exploit linked data to its full potential due to not being familiar with RDF, SPARQL or the data’s underlying schema.

Acknowledgements The work presented in this article was performed in the EU FP7 LinDA¹⁷ project.

References

- [1] Khalid Belhajjame, James Cheney, David Corsar, Daniel Garjo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. Provo: The prov ontology. Technical report, 2012.
- [2] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 2009.
- [3] Stephane Campinas. Live sparql auto-completion. In *International Semantic Web Conference (Posters and Demos)*, pages 477–480, 2014.
- [4] Lin Clark. Sparql views: A visual sparql query builder for drupal. In *9th International Semantic Web Conference (ISWC2010)*, November 2010.
- [5] Florian Haag, Steffen Lohmann, Steffen Bold, and Thomas Ertl. Visual SPARQL querying based on extended filter/flow graphs. In *International Working Conference on Advanced Visual Interfaces*, 2014.

¹⁷<http://linda-project.eu/>

- [6] Philipp Heim, Thomas Ertl, and Juergen Ziegler. Facet Graphs: Complex Semantic Querying Made Easy. In *The Semantic Web: Research and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [7] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22:1–55, 1932.
- [8] Gary Marchionini. Exploratory search. *Communications of the ACM*, 49(4):41, April 2006.
- [9] Nicolas Marie and Fabien Gandon. Survey of linked data based exploration systems. In *IESD 2014*, Riva Del Garda, Italy, October 2014.
- [10] Igor O. Popov, M. C. Schraefel, Wendy Hall, and Nigel Shadbolt. Connecting the dots: a multi-pivot approach to data exploration. October 2011.
- [11] Camille Pradel, Ollivier Haemmerlé, and Nathalie Hernandez. Natural Language Query Translation into {SPARQL} using Patterns. In *Proceedings of the Fourth International Workshop on Consuming Linked Data, {COLD} 2013, Sydney, Australia*, 2013.
- [12] Alistair Russell and Paul R Smart. NITELIGHT: A Graphical Editor for SPARQL Queries. In *Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC2008), Karlsruhe, Germany, {CEUR} Workshop Proceedings*, 2008.
- [13] Simone D. J. Barbosa Samur F. C. Araújo, Daniel Schwabe. Experimenting with Explorator: a Direct Manipulation Generic RDF Browser and Querying Tool. 2008.
- [14] Oshani Seneviratne. Querymed: An intuitive sparql query builder for biomedical rdf data abstract.
- [15] Alex Stolz, Bene Rodriguez-Castro, and Martin Hepp. RDF translator: A restful multi-format data converter for the semantic web. *CoRR*, abs/1312.4704, 2013.
- [16] Michal Tvarozek and Mária Bieliková. Generating Exploratory Search Interfaces for the Semantic Web. In *Human-Computer Interaction - Second {IFIP} {TC} 13 Symposium, {HCIS} 2010*. Springer, 2010.