

The Linked Data Value Chain: A Lightweight Model for Business Engineers

Atif Latif, Anwar Us Saeed

(Graz University of Technology, Austria)

{atif.latif,anwar.ussaeed}@student.tugraz.at)

Patrick Hoefler

(Know-Center Graz, Austria)

phoefler@know-center.at)

Alexander Stocker, Claudia Wagner

(Joanneum Research, Graz, Austria)

{alexander.stocker,claudia.wagner}@joanneum.at)

Abstract: *Linked Data* is as essential for the Semantic Web as hypertext has been for the Web. For this reason, the W3C community project *Linking Open Data* has been facilitating the transformation of publicly available, open data into Linked Data since 2007. As of 2009, the vast majority of Linked Data is still generated by research communities and institutions. For a successful corporate uptake, we deem it important to have a strong conceptual groundwork, providing the foundation for the development of business cases revolving around the adoption of Linked Data. We therefore present the *Linked Data Value Chain*, a model that conceptualizes the current Linked Data sphere. The Linked Data Value Chain helps to identify and categorize potential pitfalls which have to be considered by business engineers. We demonstrate this process within a concrete case study involving the BBC.

Key Words: Linked Data, Linking Open Data, Value Chain, Business Case, Business Models

Category: H.m, L.1.4, M.0, M.4

1 Introduction

For several years now, the *Semantic Web* [Berners-Lee et al. 2001] has been of great interest to the international research community. As a subtopic, the concept of *Linked Data* has gained much attention in the recent months.

Linked Data is based on four simple rules [Berners-Lee 2006]:

1. Use URIs as names for things
2. Use HTTP URIs so that people (and machines) can look up those names (see also [Sauermann et al. 2008])
3. When someone looks up a URI, provide useful information
4. Include links to other URIs so that they can discover more things

To seed the Semantic Web with Linked Data and to promote its adoption, the W3C community project *Linking Open Data*¹ was founded in 2007 [Bizer et al. 2007]. The project helps to solve the causality dilemma (chicken-egg problem) between Semantic Web content and Semantic Web applications by providing RDF² data sets from existing open data repositories. To enable intelligent applications that generate a valuable output for the end user, a critical amount of high-quality interlinked datasets across different domains is a crucial precondition, as shown by [Jaffri et al. 2008] and [Raimond et al. 2008]. The vision of the scientific Linked Data community can therefore be described as follows: First, facilitate the generation of semantically enriched Linked Data, and as a result, semantic applications will be built on top of this data.

Linked Data incorporates a lot of potential for enterprises [Servant 2008]. However, there is a significant difference between the aims of a scientific community and the demands and requirements of enterprises, such as revenue flow and generated value. Furthermore, every successful commercial adoption requires the discussion of inherent technical, social and business risks connected to the Semantic Web and Linked Data.

We propose that limited commercial Semantic Web adoption is, among other reasons, caused by the lack of conceptual work supporting the development of business cases and the identification of associated risks. Our publication is motivated by these factors and intends to start a discussion which moves the Semantic Web and Linked Data closer to businesses.

In section 2 we present the Linked Data Value Chain, a model of the Linked Data life cycle along with participating entities and involved roles and types of data. In section 3, we apply the Linked Data Value Chain to an existing business case from the BBC and use the aforementioned model to highlight potential pitfalls. We conclude our results and present an outlook to potential future research in section 4.

2 The Linked Data Value Chain

As a prerequisite for the development of successful business cases in the emerging context of Linked Data, three concepts have to be introduced first: *Participating Entities*, their assigned *Linked Data Roles* and processed *Types of Data*, as depicted in Fig. 1.

Our contribution tries to support business engineers with the process of assigning Linked Data Roles to Entities, modelling interactions and responsibilities of Linked Data Roles, and transforming data from Raw Data to Linked Data and Human-Readable Data, thereby increasing its value along the way.

¹<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

²<http://www.w3.org/RDF/>

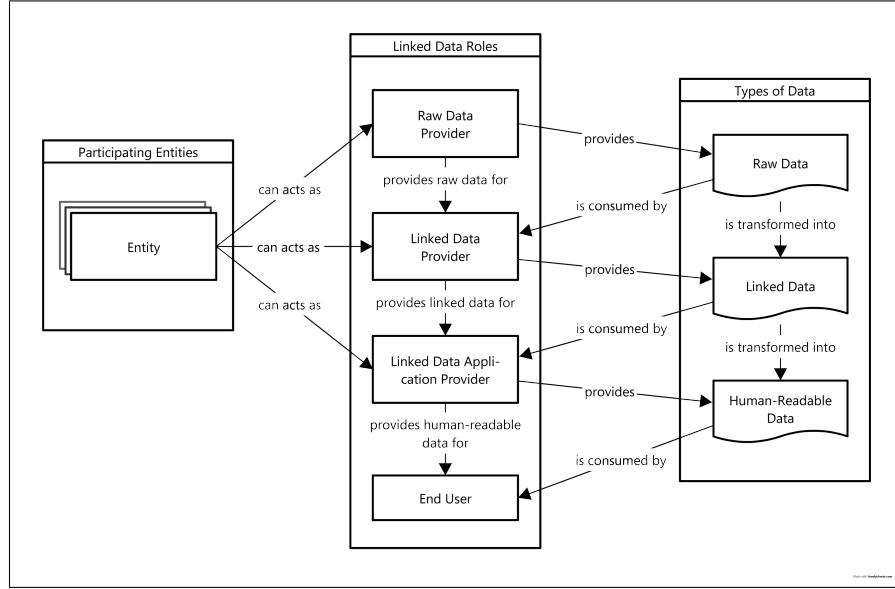


Figure 1: The Linked Data Value Chain

We propose the Linked Data Value Chain as a lightweight model, which builds upon the concepts of Linked Data and Value Chains and makes the interdependencies of entities, roles and different types of data – as the output of the value creation process – explicit.

The value chain as introduced by [Porter 1985] is a concept from the business domain: In a nutshell, a value chain is a chain of activities producing outputs, each activity increasing the value of its particular output, finally shaping a highly valuable end product. In the case of Linked Data with respect to business cases, Human-Readable Data is the most valuable output for the targeted End User.

2.1 Participating Entities & Linked Data Roles

In the context of Linked Data, participating entities – both corporate and non-corporate, e.g. persons, enterprises, associations, and research institutes – can occupy one or more of the following roles:

- A *Raw Data Provider* is a role that provides any kind of data in any non-RDF format.
- A *Linked Data Provider* is a role that provides any kind of data in a machine-readable Linked Data format. Such data is currently provided through dereferenceable URIs, a SPARQL endpoint or an RDF dump.

- A *Linked Data Application Provider* is a role that processes Linked Data within an application and generates human-readable output for human end users.
- An *End User* is a human, consuming a human-readable presentation of Linked Data. He or she does not directly get in touch with Linked Data, and typically does not even want to.

The Linked Data Value Chain allows a flexible assignment of roles to entities: In most cases, one entity just occupies one role, but it may – in extreme cases – also occupy all roles at once. For example, one enterprise could own the role of a Data Provider, a Linked Data Provider, and a Linked Data Application Provider all at the same time. The Linked Data Value Chain also supports multiple sources of data: A Linked Data Provider may acquire Raw Data from more than one Raw Data Provider simultaneously, and will usually provide Linked Data to more than one Linked Data Application Provider.

2.2 Types of Data

Three different types of data can be identified within the Linked Data Value Chain:

- *Raw Data* is any kind of data (structured or unstructured) that has not yet been converted into Linked Data. Such data usually has some structure, but generally less structure than Linked Data, and is in most cases also not universally identifiable.
- *Linked Data* is data in a RDF format that uses dereferenceable HTTP URIs to identify resources and is linked with other RDF data. This data can be generated by the Linked Data Provider itself, or data provided by a Raw Data Provider can be "RDFized". Linked Data is intended to be consumed and processed by machines only.
- *Human-Readable Data* is any kind of data which is intended, arranged and formatted for consumption by humans. Consuming this data generates value for the human end user, which is crucial to the success of any Linked Data business case.

2.3 Interaction between Roles and Data

Entities can act in different roles. These roles are closely connected through three types of data, which they provide and/or consume: A *Raw Data Provider* provides *Raw Data* as the input for a *Linked Data Provider*, who turns it into *Linked Data*, increasing its value by semantically enriching it. Linked Data in

turn serves as the basis for a *Linked Data Application Provider*, who generates *Human-Readable Data* as the most valuable output for a human *End User*.

Each combination of roles and entities as well as every transformation step of data holds inherent risks, some of which will be presented in the next section along with a concrete case study. Knowledge about such risks is crucial for the development of successful business cases. If not considered properly, these risks may become pitfalls to business success.

We identified two main areas where pitfalls may arise, grouping them into Role-Related Pitfalls and Data-Related Pitfalls. In a nutshell, Role-Related Pitfalls are either related to individual roles or to the interaction of different roles. Data-Related Pitfalls are either related to the data itself or the data transformation process. We will explain selected pitfalls emerging in the following BBC case study.

3 Applying the Linked Data Value Chain

3.1 BBC Case Study

As the idea of Linked Data is still young, there are not many appealing Web interfaces for human end users yet. One enterprise that is on the cutting edge, both in regard to deployed Semantic Web technologies and the end-user interface, is the BBC³. Furthermore, the BBC is a pioneer when it comes to adopting Linked Data within a business case. Their system utilizes Linked Data technologies to interconnect distributed micro-sites within the BBC network, e.g. BBC News⁴ and BBC Music⁵, and reuses external data from DBpedia and MusicBrainz [Kobilarov et al. 2009]. By doing so, the BBC generates additional value for the human end users, while allowing them to immediately consume

Linked Data Roles	Participating Entities
Raw Data Provider	BBC Wikipedia
Linked Data Provider	BBC MusicBrainz DBpedia
Linked Data Application Provider	BBC

Table 1: Linked Data Roles and Participating Entities in BBC case study

³<http://www.bbc.co.uk/>

⁴<http://news.bbc.co.uk/>

⁵<http://www.bbc.co.uk/music/>

contextually and semantically relevant content from third party sites as well as interconnected BBC sites.

We apply the Linked Data Value Chain to examine role assignments along with their interactions as well as the data transformation processes within the BBC business case. As summarized in Table 1, BBC acts as Raw Data Provider and Linked Data Provider for their own data as well as Linked Data Application Provider for all data, including external data from Wikipedia (Raw Data Provider) via DBpedia (Linked Data Provider) and from MusicBrainz (Linked Data Provider).

3.2 Discussion of Potential Pitfalls

In the BBC case study, BBC micro-sites utilize data from DBpedia as an important input for the business case. Unfortunately, *transforming Raw Data* (from Wikipedia) to *Linked Data* (via DBpedia) is a very *time-consuming* and, at best, *semi-automated* effort, which is currently undertaken by a team of researchers [Auer et al. 2007]. Linked Data generated this way is therefore hardly ever *complete, correct and up-to-date* [Jaffri et al. 2008]. There are no service level agreements or similar contracts between BBC and DBpedia or DBpedia and Wikipedia securing all these issues. If not considered well, such performance and data quality risks may become pitfalls.

Second, BBC end users may want to edit content on BBC Websites which is provided by DBpedia / Wikipedia. Unfortunately, BBC does not provide an *automated feedback loop leading back to the Linked or Raw Data Providers*, in this case DBpedia and Wikipedia. Such feedback loops are neither implemented nor, most of the time, conceptualized yet. Currently, users are requested by BBC to directly edit the respective articles in Wikipedia. Unfortunately, a synchronization of data between Wikipedia and DBpedia will take a very long time, depending on Wikipedia's *data dumping* and DBpedia's *transformation intervals*, possibly annoying a human end user, who certainly is not interested in such technological issues, but wants to see his or her changes promptly, if not in real time.

Third, BBC provides related *links to third party sites*. Such a procedure is a pitfall to successful commercialization, because users may leave the site of the Linked Application Provider (BBC). Reusing data should be based on *wIDGETS and embedded content*, thereby making users stay on the site longer, but still having the benefits from consuming third party contents.

Fourth, users need information about the *provenance of Linked Data* in order to be able to assess whether the displayed data comes from a *trustworthy provider*. Therefore, Linked Data Application Providers should state clearly from which Linked Data Providers and Raw Data Providers they present data. Completeness, correctness and actuality of data strongly depends on the involved

Raw Data Provider, the underlying algorithms and techniques which convert Raw Data into Linked Data as well as on the quality of service provided by Linked Data Provider.

Fifth, the BBC is currently integrating content from only a few *selected Linked Data Providers* (DBpedia, MusicBrainz), thereby not utilizing the full potential of the Linked Data sphere. On the other hand, the BBC retains *firmer control* over the displayed content and also avoids problems that may occur when displaying large amounts of semantic data from different data sources in a human-readable way [Heath 2008].

Sixth, the BBC has *chosen the data* which they retrieve from Linked Data Providers *in respect to its provenance*. BBC Music takes their introductory texts for bands from Wikipedia, whose content is inherently unstructured. Discographies and related links are taken from MusicBrainz, which is a well-structured music database by design. By not yet using e.g. content extracted from Wikipedia info boxes, BBC avoids the problems related to the transformation process of less-structured data into Linked Data.

4 Conclusion and Outlook

Our contribution is dedicated to facilitate the commercial uptake of the Semantic Web vision. We have presented the Linked Data Value Chain as a lightweight model for business engineers to support the conceptualization of successful business cases. Thereby, we identified three main concepts: Different *Entities* acting in different *Roles*, both consuming and providing different *Types of Data*.

We demonstrated that the assignment of roles to entities, the combination and involvement of roles, the data selected as well as the data transformation process hold inherent business risks. We exemplarily applied the Linked Data Value Chain within a concrete case study from BBC to showcase selected business risks.

Our future research will deal with a more detailed classification scheme for pitfalls arising when businesses strive to adopt the Semantic Web, and we will elaborate our concept of the Linked Data Value Chain.

5 Acknowledgement

This contribution is partly funded by the Know-Center and the Higher Education Commission of Pakistan.

The Know-Center is funded within the Austrian COMET program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

- [Auer et al. 2007] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: "DBpedia: A Nucleus for a Web of Open Data"; 6th International Semantic Web Conference, Busan, Korea (2007), <http://richard.cyganiak.de/2008/papers/dbpedia-iswc2007.pdf>
- [Berners-Lee et al. 2001] Berners-Lee, T., Hendler, J., Lassila, O.: "The Semantic Web"; Scientific American Magazine (May 2001), <http://www.sciam.com/article.cfm?id=the-semantic-web>
- [Berners-Lee 2006] Berners-Lee, T.: "Linked Data – Design Issues"; (July 2006), <http://www.w3.org/DesignIssues/LinkedData.html>
- [Bizer et al. 2007] Bizer, C., Heath, T., Ayers, D., Raimond, Y.: "Interlinking Open Data on the Web"; Demonstrations Track at the 4th European Semantic Web Conference, Innsbruck, Austria (May 2007), <http://www.eswc2007.org/pdf/demo-pdf/LinkingOpenData.pdf>
- [Heath 2008] Heath, T.: "How Will We Interact with the Web of Data?"; IEEE Internet Computing, vol. 12, no. 5, pp. 88-91, (2008)., <http://tinyurl.com/ct3gx2>
- [Jaffri et al. 2008] Jaffri, A., Glaser, H., Millard, I.: "URI Disambiguation in the Context of Linked Data", Linked Data on the Web Workshop at the 17th International World Wide Web Conference, Beijing, China (2008), <http://events.linkeddata.org/ldow2008/papers/19-jaffri-glaser-uri-disambiguation.pdf>
- [Kobilarov et al. 2009] Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., Lee, R.: "Media meets Semantic Web - How the BBC uses DBpedia and Linked Data to make Connections"; To appear In ESWC 2009 Conference: The 6th Annual European Semantic Web Conference Heraklion, Greece (2009), <http://www.georgikobilarov.com/publications/2009/eswc2009-bbc-dbpedia.pdf>
- [Porter 1985] Porter, M. E.: "Competitive advantage: Creating and sustaining superior performance"; Free Press (1985)
- [Raimond et al. 2008] Raimond, Y., Sutton, C., Sandler, M.: "Automatic Interlinking of Music Datasets on the Semantic Web"; In WWW 2008 Workshop: Linked Data on the Web (LDOW2008), Beijing, China (2008), <http://citeseervx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.9753&rep=rep1&type=pdf>
- [Sauermann et al. 2008] Sauermann, L., Cyganiak, R., Ayers, D., Vlkel, M.: "Cool URIs for the Semantic Web"; W3C Interest Group Note (2008), <http://www.w3.org/TR/2008/NOTE-cooluris-20081203/>
- [Servant 2008] Servant, F.-P.: "Linking Enterprise Data"; Linked Data on the Web Workshop at the 17th International World Wide Web Conference, Beijing, China (2008), <http://events.linkeddata.org/ldow2008/papers/21-servant-linking-enterprise-data.pdf>