CrossMark

EDITORIAL

Linked Data in Business

Witold Abramowicz · Sören Auer · Tom Heath

1 The Potential of Linked Data in Business

The linked data principles[1] were proposed 10 years ago and since then have received ever increasing attention from researchers, developers, companies, and governments as a means of data distribution and integration that is consistent with the architecture of the World Wide Web.[2] This last decade has seen an explosion in the availability of data, driven by a range of factors such as open data initiatives worldwide, the increasing use of sensors to create a so-called internet of things, and by continued interest in the concept of big data.

In parallel to these trends, the broader Semantic Web vision has also evolved. While the Semantic Web stack was originally seen as rather monolithic and, at times, inaccessible to developers or different technology ecosystems, the RDF data model now appears a more truly lingua franca of data integration, bridging different knowledge representation formalisms, data serializations, conceptualizations, and technology ecosystems. For example, there are now, in addition to XML, bindings of RDF to:

- tabular and relational data with the W3C R2RML[3] and the CSV on the Web[4] standards,
- JSON with the W3C JSON-LD standard[5] providing a minimally invasive way of equipping standard JSON documents with an RDF mapping preamble,
- HTML with RDFa[6] as a mechanism for embedding RDF data into HTML documents.

Figure 1 illustrates this evolution of the semantic technologies stack by juxtaposing the original version from 2001 and a new one created now 15 years later.

Linked data is beneficial where heterogeneous data needs to be exchanged between a variety of distributed systems or stakeholders. The lightweight knowledge representation formalisms propagated by linked data have been successfully applied in a number of domains, including:

- Life sciences, e.g., the OpenPhacts[7] initiative that integrates commonly used data resources in pharmacology, to support drug discovery.
- Web Commerce, Web Search, and Semantic SEO, e.g., with the schema.org[8] initiative supported by major search engines, to help Web publishers embed

Prof. Dr. W. Abramowicz
Department of Information Systems, Faculty of Informatics and Electronic Economy, Poznan University of Economics, Al. Niepodleglosci 10, 60-875 Poznan, Poland
e-mail: w.abramowicz@kie.ue.poznan.pl

Prof. Dr. S. Auer (✉)
Computer Science Department/Enterprise Information Systems, University of Bonn, Römerstr. 164, 53117 Bonn, Germany
e-mail: auer@cs.uni-bonn.de

Dr. T. Heath
Arup, 13 Fitzroy St, London W1T 4BQ, UK
e-mail: tom.heath@arup.com

1 http://www.w3.org/DesignIssues/LinkedData.html (Accessed 10 July 2016).
2 https://www.w3.org/TR/webarch/ (Accessed 10 July 2016).
3 https://www.w3.org/TR/r2rml/ (Accessed 10 July 2016).
4 https://www.w3.org/2013/csvw/wiki/Main_Page (Accessed 10 July 2016).
5 https://www.w3.org/TR/json-ld/ (Accessed 10 July 2016).
6 https://www.w3.org/TR/xhtml-rdfa-primer/ (Accessed 10 July 2016).
7 https://www.openphacts.org/ (Accessed 10 July 2016).
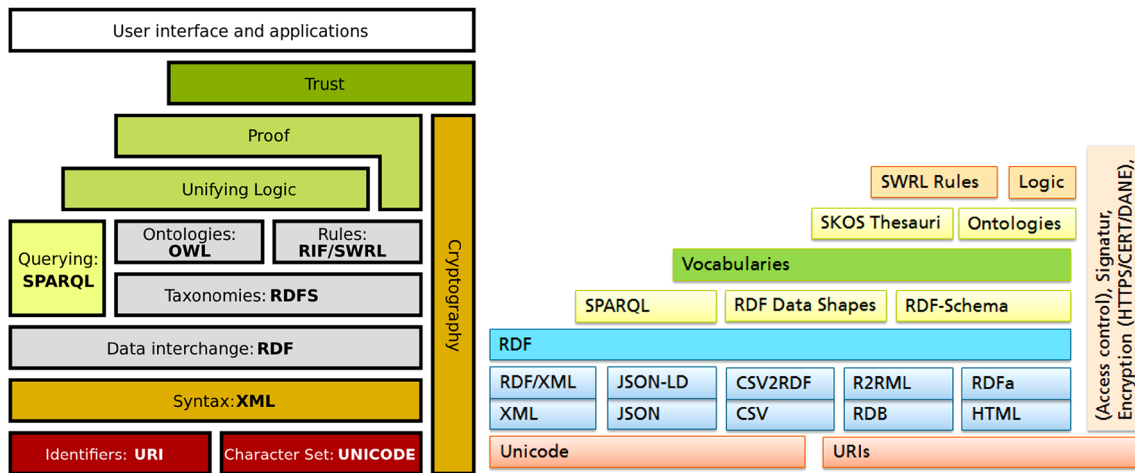8 http://schema.org (Accessed 10 July 2016).

Fig. 1 The semantic web layer cake in 2001 (*left*) and 2016 (*right*)

structured data in their pages and benefit from improved presentation in search engine results pages.

- Digital Libraries and aggregators such as Europeana and national digital libraries, which aggregate metadata about millions of artifacts using the Europeana Data Model (EDM).[9]

However, there remain a significant number of domains where applying the Linked Data principles would be beneficial, but the concept is still largely unknown. These include the wider business domain, but also concrete verticals such as:

- Finance, where heterogeneous data about governance, risk management, compliance and other regulatory requirements needs to be exchanged between a variety of stakeholders and organizations.
- Manufacturing and production, which recently gained attention under the various banners of Industry 4.0 (Germany), industrie du futur (France), or Industrial Internet (USA) initiatives.
- Logistics and supply chain management, where the scale and diversity of actors involved, and the inherently distributed nature of related data, requires a robust mechanism for representing connections between those actors and the data they generate.
- Enterprise data integration, where the challenges of monitoring and integrating corporate data assets create barriers to more efficient commercial operation and innovation.

In the following we will explore in more depth three such domains and illustrate the role linked data can play.

2 Linked Data in Automation and Manufacturing

In the engineering and manufacturing domain, there is currently an atmosphere of departure to a new era of digitized production, where traditional industrial engineering methods are synergistically combined with IT and Internet technologies. Industry 4.0 (I4.0) is a term coined in Germany to refer to the "fourth industrial revolution". This is understood as the application of modern IT concepts such as the internet of things (IoS), cyber-physical systems (CPS), the internet of services (IoS), and data-driven architectures in industrial contexts to generate new and innovative products, services, and added value. Although the vision of digitizing production and manufacturing gained much traction lately, it is still not clear how this vision can actually be implemented with concrete standards and technologies. The physical network connection problem is meanwhile largely solved using technologies such as Profibus/Profinet and OPC-UA. However, the more challenging problem is to make smart industrial devices able to communicate and understand each other as a prerequisite for cooperation scenarios.

To address this problem, there is a need for techniques and standards to represent and exchange information, data and knowledge between devices participating in manufacturing and production processes. Such standards must be flexible to accommodate new features and usage scenarios, cover multiple domains and device categories, and to bridge organizational boundaries. Most importantly, they must be able to evolve seamlessly over time to facilitate the swift realization of new features and scenarios as they become apparent.

Within the Industry 4.0 initiative, the concept of an Administration Shell was devised to respond to these requirements. The Administration Shell is planned to provide a digital representation of all information (and

---

[9] http://pro.europeana.eu/page/edm-documentation (Accessed 10 July 2016).

services) being available about and from a physical manufacturing component.

In Grangel-González et al. (2016) a first version of an RDF-based Administration Shell was presented. The work also identifies the six challenges of interoperability, global unique identification, data availability, standardization compliance, integration, multilinguality for industry 4.0 and how the features of RDF can be utilized for solving them. However, the capability of semantic alignment with the Admin Shell and Reference Architecture for I4.0 (RAMI4.0) model hierarchy levels was not included. The hierarchy levels of the RAMI model are based in the IEC 62264 standard. In addition, crucial concepts such as units of measurements and provenance were not covered in the first version of the RDF-based Admin Shell. As a result, linked data can help in the Industry 4.0 context to establish the required semantic models for communication between sensors, devices and machines. By transforming existing IEC, IEEE and ISO standards into linked data vocabularies, a network of independently evolving semantic models can emerge, which enables at the same time the conceptual representation of information as well as its operational linking, integration, and execution.

3 Linked Data in Enterprise Architecture and Data Management

In the business and enterprise domain, there is still a gap between conceptual approaches for modeling architectures, systems, data models as well as their implementation, operationalization and execution. Despite approaches aiming at bridging this gap, such as model-driven architecture, this is due to the fact that systems, domains, organizational structures and applications are too heterogeneous to be tackled by an integrated approach or methodology. Similarly, as loosely coupled architectures (e.g., containerization, micro-services) become more popular in software architectures, similar loosely-coupled approaches will become more important in the enterprise architecture and data management domain. The linked data concept can serve as a basis for realizing loosely-coupled enterprise systems and data architectures, while at the same time improving automation, standardization and collaboration.

A promising avenue for implementing loosely-coupled, linked data-based enterprise systems and data architectures can be following a vocabulary-based integration methodology. The core pillars of such a methodology are

- A vocabulary development and collaboration infrastructure, which empowers domain experts, knowledge engineers and enterprise architects to develop human and machine-readable information models.
- Creation and maintenance of enterprise knowledge graphs (Galkin et al. 2016) that bind together disparate internal data sources and, in doing so, embody an organization-wide consensus regarding those information models.
- Provide mapping and transformation layers/interfaces between input data sources and knowledge graphs, such that sources are not arbitrarily omitted or excluded.

Collectively, these steps can enable enterprises to progressively adopt the concept of *Semantic Data Lakes* (O'Leary 2014), whereby greater business value can be realized through a more agile and flexible complement to traditional data warehousing approaches.

4 Linked Data in Finance

The financial industry is at a crossroads. Complex legacy infrastructures as well as lax or at least inapt internal governance standards have resulted in in-transparencies and eventually have led to losses that threatened to destabilize the entire financial system. There are three major forces putting pressure on it to change for the better:

- Regulators are forcing the industry to adopt strict governance and compliance standards and thus are effectively enforcing transparency.
- Low interest rates result in a dramatic loss of margins and are putting a high emphasis on cost savings and a better understanding of how and where money is actually made within the bank.
- FinTech start-ups are not threatening the industry as a whole, but they are demonstrating that small, fast, lean entities can produce identical or better services than industry incumbents but at a price point that is an order of magnitude below the industries existing cost structure.

In order to adapt and overcome these challenges financial service institutions have to establish commanding control over their data and thus enabling agility while at the same time dramatically reducing cost of their internal infrastructure and processes. Fragmentation of information across dozens or even hundreds of IT systems, is virtually making it impossible for the industry to achieve the goal of having a commanding control over their business and thus their ability to reduce cost structures and manage risk more effectively. Regarding timeliness, common existing systems do not deliver needed data in time, while supporting only post-transaction, post-end-of-day data that often are only available hours or days after their original creation.

Appropriate reaction or even interruption of hazardous events is hardly possible any more.

A study conducted by the Enterprise Data Management Council in 2015 during which more than 300 experts[10] in the FSI were interviewed and took part in a standardized data management capability assessment (DCAM) underscores the view of the problem at hand: The interviewed CIOs and Chief Data Officers agreed that in fact 93 % of them were not in the position to ensure an unambiguous firm-wide view on data and provide a shared understanding about the meaning of data across their organization. The dramatic flip side to this statistic is that in the case of these 93 % of banks, funds or insurance companies the meaning of data is not clear, but must be established on a case by case basis which of course is effectively preventing the industry to cut cost, increase efficiency, timeliness and effectiveness in managing risks.

Again here in the financial domain, the lightweight, federated linked data approach can help to create the semantic models required to represent and exchange data efficiently. Using RDF vocabularies, the integration of legacy systems can be accomplished (e.g., using RDB2RDF transformation) at the same time as the representation of information for regulatory reporting. A first step in this direction was done by the members of the EDM council with the creation of the FIBO Financial Industry Business Ontology.

5 Looking Forward

In order for linked data to be successful in the long term it needs to continue to demonstrate its value as a technology approach for heterogeneous, distributed data integration in these domains and many more. In this regard, business environments provide an ideal test-bed for application of linked data principles, regardless of whether they operate outside or inside the firewall.

What appears irrefutable is the desire, driven by business requirements, for ever greater integration of data, at ever reducing cost. In this regard, and while the fundamentals of data integration methods may not radically change in themselves, linked data represents the most mature, scalable, coherent and feature rich approach to achieving these goals.

Recently, the newly started industrial data space initiative[11] (Otto et al. 2016) aims at enabling flexible and secure data exchange in industrial value chains based on linked data principles. As a decentralized architecture being based on data space connectors enabling enterprises to publish, share and exchange their semantically represented data, the industrial data space is giving enterprises and businesses back some sovereignty over their data.

In the medium and long term also technologies such as question answering and conversational technologies require the availability of more structured and semantically represented and integrated data and have the ability to be widely as assistive and advisory technologies applied in business settings.

References

Galkin M, Auer S, Kim HL, Scerri S (2016) Integration strategies for enterprise knowledge graphs. Tenth IEEE International Conference on Semantic Computing, ICSC 2016, Laguna Hills, USA

Grangel-González I, Halilaj L, Coskun G, Auer S, Collarana D, Hoffmeister M (2016) Towards a semantic administrative shell for industry 4.0 components. Tenth IEEE International Conference on Semantic Computing, ICSC 2016, Laguna Hills, USA

O'Leary DE (2014) Embedding AI and crowdsourcing in the big data lake. IEEE Intell Syst 29(5):70–73

Otto B, Auer S, Cirullies J, Jürjens J, Menz N, Schon J, Wenzel S (2016) Industrial data space: digital souvereignity over data. Technical Report, Fraunhofer-Gesellschaft

---

[10] Members list of the Enterprise Data Management Council: http://www.edmcouncil.org/councilmembers (Accessed 10 July 2016).

[11] http://industrialdataspace.org/en (Accessed 10 July 2016).