E-Commerce and the Web of Data

Mariana Curado Malta
Instituto Politécnico do Porto, Portugal & University of Minho, Portugal

Ana Alice Baptista
University of Minho, Portugal

INTRODUCTION: THE WEB OF DATA

The "Web of Data" - opposed to the "Web of Documents" - is a term used to present a "giant global graph" (Berners-Lee (2007) cited by Heath & Bizer (2011)) on the Web. The Web of Documents is made of documents read by human beings that navigate between documents located in servers through hyperlinks; on the other hand the Web of Data is made of structured data read by agents that relate structured data located in servers through the existence of links – this data is also called Linked Data since it is data connected through links. These two Webs or representations of the Web live side by side and one does not replace the other; instead, they co-exist for different purposes and they are, at least partly, complementary. The Web of Data or Linked Data or even the Semantic Web, three ways of expressing similar concepts, have technologies that "enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies" that started to emerge in 1999 with the issuance of the first Resource Description Framework (RDF) specification by the World Wide Web Consortium (W3C)[1] - c.f. W3C (1999). It is about common formats for integration and combination of data from different sources (W3C, 2012). This data is mostly what is being called metadata, in the way that it is "data about data" (DCMI, 2011) and follows well defined rules of metadata schemas.

In order for meaning to be shared across applications, structured data must be syntactically and semantically interoperable: organisations can share their data with their business partners, provide data to third-parties, or open parts of it (Linked Open Data – LOD) in order to build a more intelligent Web, opening new ways of empowerment of e-commerce businesses. Moreover, from this global Web of Data organisations can create or infer relationships and extract meaning from those relationships or from others that meanwhile may be created or inferred. For all these to happen organisations, business partners or groups of interest need to have a common ground of understanding. In other words structured interoperable data available *per se* or in annotated Webpages can be used by organisations to enhance the information about their products in terms of semantic Search Engine Optimization (SEO).

There are several Linked Data related rules and recommendations that are issued by several organisations from which the Dublin Core Metadata Initiative (DCMI)[2] and the W3C are very good examples. DCMI, in particular, defines four levels of interoperability for metadata – c.f. Nilsson, Baker, & Johnston (2009). The fourth level, the one that assures more interoperability, requires the definition of Metadata Application Profiles (MAP), that are generic constructs "for designing metadata records" (Baker & Coyle, 2009). MAP allow the definition of constraints on metadata and some of those constraints are established by the usage of Vocabulary Encoding Schemes (VES) or Syntax Encoding Schemes (SES). MAP are sets of terms that are defined to serve a specific community of interest. The terms belong to different metadata schemas (also called vocabularies or even ontologies) describing a reality or context. Examples of these metadata schemas are the Good Relations Ontology[3] and the ESSGlobal vocabulary[4],

DOI: 10.4018/978-1-4666-9787-4.ch075

which were conceived having as goal to be able to describe the context and the interactions that exist in commerce and trade. The Schema.org metadata schema has also some e-commerce dimensions even if it was developed with a broader objective.

This chapter will guide the reader through the basic concepts of the Web of Data, such as metadata and metadata schemas- presenting in detail the ones that are central to e-commerce -, and metadata application profiles. It will present the RDF data model, the standard way to represent Linked Data, and some of the most used RDF serialisations, including the use of micro-formats.

HOW TO STRUCTURE DATA

The data on the Web of Data is called metadata. Metadata is structured data that uses different metadata schemas (or vocabularies) to describe resources. Next sub-sections will guide the reader through the concepts of metadata and metadata schemas, and will present three different metadata schemas that can be used to describe particular contexts of e-commerce: the Good Relations ontology, the ESSGlobal vocabulary and the schema.org ontology.

What is Metadata?

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information" (NISO, 2004, pp.1). A resource can be anything, physical or non-physical. It can be a book or any kind of item that you want to sell. But it can also be non-physical like the transaction information about a selling, or information about a train trip, i.e., a service. So Metadata "denotes machine-readable descriptions of things" (Rühle, Baker, & Johnston, 2011). The "things" are often called "resources", any "thing" is a resource, any "thing" may be subject to description, provided that it is identifiable.

There are several types of metadata and, as in any professional or scientific context, there are several approaches to the metadata typification. The National Information Standards Organization (NISO) considers three types of metadata (NISO, 2004):

- **Descriptive Metadata:** "Describes a resource for purposes such as discovery and identification". They can include a name, a price, a colour, a manufacturer, etc.
- **Structural Metadata:** "Indicates how compound objects are put together", for example, how the pages of a book are ordered to form chapters;
- **Administrative Metadata:** "Provides information to help manage a resource". For example on how and when it was created, by whom, type of file, and who can access the resource.

The description of a resource using metadata is "stored" in a metadata record. A metadata record is made of pairs of attribute - value. Metadata elements can also be referred as properties, a term used in the LOD community. An example of a resource can be the XPTO Cell-phone that is sold in an online store where its the manager wants this information to be crawled and further processed by automata in the Web of Data. The XPTO Cell-phone can be described using its name, description, colour, price and validity of the sale. Follows an example on how this description could be done: