The Web has been evolving from the initial "Web of documents" to a new "Web of Data" being the driving force behind this change the Semantic Web initiative lead by the World Wide Web Consortium (W3C). In the Semantic Web, the information available on the Web is readable not only by humans but also by machines. This change is supported by a data model, the Resource Description Framework (RDF), that is used to formally describe Web resources (which can be, literally, anything) identified by Uniform Resource Identifiers (URIs). In the last ten years, we have witnessed a constant growth in the amount of RDF repositories available on the Web for several domains such as bibliographic data, biological data, and governmental data. The Semantic Web initiative plus its associated technologies and standards, provided a first step towards the big goal of having a Web that can be queried and analyzed as a huge single database. But the drastic change came in the form of a new initiative developed on top of the Semantic Web standards: The Web of Open Linked Data.

The Web of Open Linked Data is a shift towards a data model in which every piece of RDF information describes itself and its relations with other pieces of data in a highly decentralized way. Linked Data is based on the a set of principles to publish data on the Web. These principles state that each piece of data should be published as a Web-accessible URI such that when this URI is accessed by some application/people/machine, an RDF specification describing that URI should be provided. But the most important requirement is that every URI should be described in terms of RDF statements that point to other accessible URIs, thus forming a huge data space of interconnected statements about resources on the Web.

There are several issues posed by the above mentioned principles that are not satisfactorily covered from a classical database point of view. Among the most important we can observe that data is highly distributed and at a fine grain and that any URI can make statements about (provide links to) any other URI. These issues together with the scalability issues faced when querying data at Web scale, the graph structure of the data at a global level, plus the particularities of the RDF data model (incomplete information, special semantics, etc.), make Linked Data a very challenging scenario for database research.

Despite the importance and the potential applicability of the Web of Linked Data, only little research has been pursued to study the fundamental new problems towards efficiently and effectively querying this emerging data space. In particular, there is no query language specifically designed for this new data model. The design, formal study, implementation and evaluation of such a language is the final goal of this project.

One of our main hypothesis is that classical database technologies are not enough to provide a comprehensive solution to the problem. Thus, we need to adapt the previous technology, but
more importantly, leverage the power of emerging technologies to make a substantial contribution for this topic. Besides basing our proposal in the classical literature of declarative query languages, distributed databases, and data integration, we plan to incorporate new technologies such as column oriented databases, the map reduce paradigm, and cloud computing that have shown to be very useful when applied to Semantic Web data management.

We do believe that efficiently and effectively querying the Web of Linked Data is a fundamental topic in the materialization of the Semantic Web vision. Thus our proposal for designing and formally studying a query language for Linked Data can have a considerable practical impact and the potential of laying the foundations for the next-generation Semantic Web tools.