Literature review

Accompanying with the rapid development of Internet, information explosion brings both challenges and opportunities for traditional web structure. Although individuals and organizations publish numerous amount of data on web via protocols such as TCP/IP and HTML, very few of them could be understood by computers, since that the main content of traditional web technologies is designed for human to read rather than machines. On the current web the scale of content for computer to read and understand is very limited (A Semantic Web Primer). As a consequence, it is difficult to integrate and reuse data on the web if they are difficult to understand by computers (Semantic Web Technology Systems). Besides, information explosion not only brings great amount of data, but also highly increases the complexity of data. With the requirement for more agile approaches to handle links and mash-up data in complex processes, conventional approaches of data integration would collapse in most situations (Challenges and Opportunities). In order to deal with these problems, the concept of Semantic Web is proposed and added to the current web layer to make the machine to understand web document (Berners-Lee, Hendler, & Lassila, 2001)

In order to support Semantic Web technology, the term “Linked Data” is introduced, referring to the method by which data could be described via Uniform Resource Identifier and Resource Description Framework, so that these data could be exchanged, displayed, linked and published (From Bibliographic Records to Data). According to the rules raised by Berners-Lee in 2006, from the bottom the raw data is named after HTTP (Hypertext Transfer Protocol) URI strings so that people can locate this data on Internet, and standards such as SPARQL and RDF to describe the relationships between different data to provide useful information for people to look up these data. During this process, data is expressed using standards (RDF) which could help machine to understand and utilize these data. Once data is machine-readable, which means it is explicitly defined, the link between local data set and external data set could be established (Linked Data-The Story So Far)

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), 28–37. doi:10.1038/scientificamerican0501-34 PMID:11341160