LINKED DATA:
Road to Intelligent Web

Linked data is emerging as a new paradigm for publishing and interlinking data from different data sources and evolving web from "Web of linked documents" into the "Web of linked data". The global database created by 'linked data' provides flexible inter-organizational collaboration and dynamic data integration.

SAVITA BHATNAGAR
Technical Director
savita.bhatnagar@nic.in

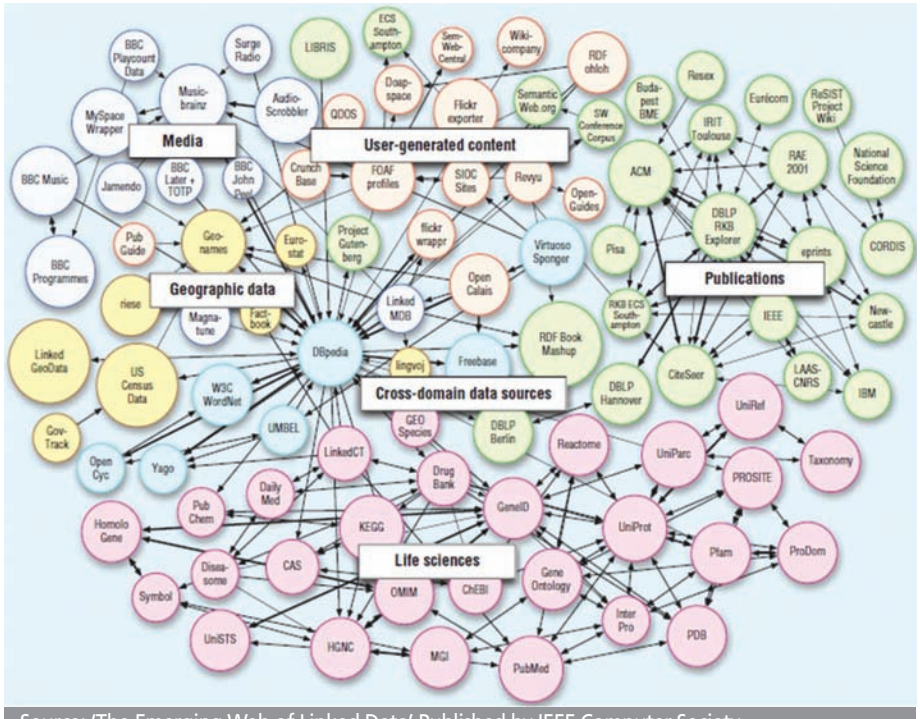
Edited by
ANSHU ROHATGI

SEMANTIC WEB

Rapid advancements in Information Technology have enabled to connect people, places and information in the remotest part of the world over the World Wide Web, in a manner, which was never visualized earlier. These web technologies have benefited the masses, simultaneously creating an enormous volume of data in the public domain. The real challenge now is to make the process intelligent enough so that the data available on the web is linkable, useful, easy to sort and interoperable. Web 3.0 or the Semantic web is emerging as the next phase of technological development which enables people to share content beyond boundaries of application and websites, enabling machines to do work in an intelligent manner. However, this requires the data on the web to be made available in a standard format so that it is easily reachable and manageable by Semantic web tools. Making web of linked data a reality requires a relationship among data on the web as a pre-requisite.

LINKED DATA

The collection of interrelated



Source: 'The Emerging Web of Linked Data' Published by IEEE Computer Society

datasets on the web is called Linked Data. Wikipedia defines it as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using Uniform Resource Identifiers (URIs) and Resource Description Framework (RDF)".

Linked Data can be better understood as using the web for creating typed links between data from different sources. It uses RDF as a standardized data model and provides a mechanism for dynamic data integration originating from multiple sources across the web. It is different from conventional data formatting and publication approaches and may be as diverse as databases of two organisations from geographically different locations, or heterogeneous systems within one organisation, that have not interoperated at the data level. Technically speaking, Linked Data refers to data published on the web in machine-readable format, with meaning explicitly defined and linked to, as well as from, external data sets. Linked data can discover new data sources by following data-level links and deliver a more complete answer in a dynamic environment. It enables users to start with one data source and then move through a potentially endless web of data sources connected by RDF links. Linking of different pieces of information published, with direct reference to a specific piece of information on the web, reduces the data redundancy along with easy verification of the authenticity of published data. Linked data explicitly published under an open license is called linked Open Data.

Web of Linked Data is based on standards for identification, retrieval, and representation of data. Sir Tim Berners-Lee, the inventor of the web has defined four design principles for linked data.

1. Use URIs to uniquely identify things (data entities)
2. Use HTTP URI, so that people can look up those names and information can be retrieved.
3. When someone looks up a URI, provide useful information, using the open standards like RDF, SPARQL.
4. Include links to other URIs, so that people can discover more things.

Linked data require identification of those items of interest, which are called resources, whose properties and relationships need to be described in the data. The information about the resource is represented using RDF. The RDF model encodes data in the form of triples, called subject, predicate, and object. The subject of a triple is the URI identifying described resource. The object can either be a simple literal value, like a string, number, or date; or the URI of another resource that is somehow related to the subject. The predicate indicates what kind of relation exists between subject and object. Triple's subject and object can be defined and described in different datasets. These cross-linked datasets lead to an endless data graph that describes resources in terms of their properties and connections, spanning datasets from countless sources, thereby creating a true web of data.

LINKED OPEN DATA (LOD) PROJECT

Linked Open Data project was initiated in 2007 by World Wide Web Consortium's Semantic Web Education and Outreach (SWEO) Interest Group. The goal of this project is to make common data by converting various open data sources available, as RDF, and setting RDF links between data items from different data sources. The project has evolved from a small team of researchers during the initial stages, to the involvement of large organizations such as Wikipedia, BBC, Thomson Reuters, New York Times etc. The size of this web of data is growing rapidly, having billions of RDFs from datasets in diverse domains such as census information, geographic information, people, online communities, companies, human languages, films, music, scientific publications and books etc.This growth is due to the open nature of the project, where anyone can participate by publishing a data set according to the Linked Data principles, and interlinking it with existing data sets.

FUTURE OPPORTUNITIES

Existence of large amounts of meaningfully interlinked data on Web is a fundamental requirement for the Semantic Web. Linked Data is a promising candidate for addressing challenges in the area of Intelligent Information Management. In coming years, we can expect more and more data to be available as linked data. It can be an enabler to smarter, more efficient services and applications, and fostering creativity & innovation in the area of web technology.

FOR FURTHER INFORMATION:
Savita Bhatnagar
Technical Director
NIC Mumbai
E-mail: savita.bhatnagar@nic.in