Chapter 15
Open Data Repositories in Knowledge Society

Nadim Akhtar Khan
University of Kashmir, India

Sara Sohrabzadeh
Tehran University of Medical Science, Iran

Garvita Jhamb
University of Delhi, India

ABSTRACT

Open data is data that can be freely used, re-used, and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike. Open data is an important resource in present knowledge-based society used for supporting decision-making activities in government agencies, research Institutions, higher educational institutions, etc. The research activities as well as policy-making decisions are highly benefited by the availability of open data. There should be no restrictions to its reuse like legal, financial, or technological. The chapter discusses the concept of open data and its importance in present times for supporting advanced research activities in different domains of human knowledge. It also provides a glimpse of open data repositories at global level and attempts to showcase select open data repositories set by different organizations and institutions for making different datasets available in varied areas for facilitating specialized research and development activities.

INTRODUCTION

"Open means anyone can freely access, use, modify, and share for any purpose subject, at most, to requirements that preserve provenance and openness. (Open Definition, 2015)." Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike(Open Data Handbook, n.d). The term "Open Data" thus refers to data and information beyond just governmental institutions and includes those from other relevant stakeholder groups such as business/industry, citizens, Non Profit Organizations(NPOs) and Non Government Organizations(NGOs),

science or education including the World Bank, the United Nations, The Guardian and the Open Knowledge Foundation (Bauer & Kaltenbock, 2012).Open data is regarded as an essential component of modern day research as it facilitates carrying advance research activities in different subject domains at much broader levels. Open Data is an important resource for supporting decision making activities in Government agencies; research Institutions, Higher Educational Institutions etc. The research activities as well as policy making decisions are highly benefited by the availability of open data. Many countries have come forward to embrace the concept of Openness especially facilitating Open access to government data as such data is deemed to be public data by law. The open data facilitates transparency and democratic control, participation, self-empowerment, improved or new products and services, innovation, improved efficiency and effectiveness of government services, impact measurement of policies, new knowledge from combined data sources and patterns in large data volumes. Open data as such is of much economic importance and is being reused world over in different setups. It improves government efficiency in addressing different issues concerning common people. It is paving ways for modern social fabrication and adding economic values to the existing data sets for efficient utilization resulting in new knowledge creation and more specializations. There should be no restrictions to its reuse like legal, financial or technological etc (Open Data Handbook, n.d).

The key requirements for Open Data include free availability of data and access and also format for redistribution and reuse of that data (SPARC, 2013). Good open data can be linked to so that it can be easily shared and talked about; is available in a standard, structured format, so that it can be easily processed; has guaranteed availability and consistency over time, so that others can rely on it; is traceable, through any processing, right back to where it originates, so others can work out whether to trust it (Open Data Institute, 2015). Harnad (2005) discusses that open access has two complementary routes: the gold road and the green road. Publishing in open-access journals is the golden road. It calls on publishers to adopt open access policies whereas green road calls on researchers to self-archive their works in institutional repositories and publishes it on the author's institutional Web site. The more data is made openly available in a useful manner, the greater the level of transparency and reproducibility and hence the more efficient the scientific process becomes, to the benefit of society. This viewpoint is becoming mainstream among many funders, publishers, scientists, and other stakeholders in research (Molloy, 2011).

Heery and Anderson (2005) define digital repositories as digital collections in which the content is deposited and organized in a database, either by the content creator or some other means. Such a repository is based on an architecture that manages content as well as metadata and offers a minimum set of basic services while the purpose of a digital repository is to store and manage learning objects (McGreal, 2007). These repositories provide services to faculty, researchers, and administrators who want to collect research, historic, and creative materials (Drake, 2004). One characteristic common to all open repositories is the fact that they have been created by using software packages with open source licensing and developed by working groups linked to the open archives initiative community. This is based on the implementation of a shared protocol: Open Archives Initiative-Protocol for Meta data Harvesting (OAI-PMH). The essential feature of OAI-PMH is that it provides great ease of implementation because components of the OAI model basically involve a metadata harvester, a search and retrieval interface and a shared repository made up of a collection of individual repositories. These are combined with a set of requests and responses made through the hypertext transfer protocol (Bravo & Diez, 2007).

Data Repository is a logical partitioning of data where multiple databases which apply to specific applications or sets of applications reside (Online Learning, 2014).It is a somewhat general term used

Related Content

Information Technology/Systems Offshore Outsourcing: Key Risks and Success Factors
Mahesh S. Raisinghani, Brandi Starr, Blake Hickerson, Marshelle Morrisonand Michael Howard (2008). *Journal of Information Technology Research (pp. 72-92).*
www.irma-international.org/article/information-technology-systems-offshore-outsourcing/3693

Establishing Preconditions for Spanning the Boundaries in Public Private IT Megaprojects
Roman Beck, Oliver Marschollekand Robert Wayne Gregory (2010). *International Journal of Information Technology Project Management (pp. 20-37).*
www.irma-international.org/article/establishing-preconditions-spanning-boundaries-public/47184

Integrating Fact-Oriented Modeling with Object-Oriented Modeling
Terry Halpin (2001). *Information Modeling in the New Millennium (pp. 167-188).*
www.irma-international.org/chapter/integrating-fact-oriented-modeling-object/22987

Comparison of Tied-Mixture and State-Clustered HMMs with Respect to Recognition Performance and Training Method
Hiroyuki Segi, Kazuo Onoe, Shoei Sato, Akio Kobayashiand Akio Ando (2014). *Journal of Information Technology Research (pp. 1-17).*
www.irma-international.org/article/comparison-of-tied-mixture-and-state-clustered-hmms-with-respect-to-recognition-performance-and-training-method/116635

ERP Implementation for Production Planning at EA Cakes Ltd.
Victor Portougal (2006). *Cases on Information Technology Planning, Design and Implementation (pp. 137-149).*
www.irma-international.org/chapter/erp-implementation-production-planning-cakes/6366