

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

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 Metadata 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 to refer to a physical or logical destination

or place designated for data storage for multiple databases, records or files. The term data repository can be used as either a local location that is directly accessible to the user without having to travel across a network or for a network location (Research Data Alliance, 2014). Generally, a data repository serves as the centerpiece of an open data effort. It serves as a central location to find data, a venue for standardizing practices, and a showpiece of the use of that data. Some repository software allows syndication, permitting other organizations to automatically incorporate their own data (How to Open Data, 2015). Open data repository as such is the warehouse of open data, where data warehouse refers to central repository for all or significant parts of the data collected. It includes data that can be freely used, reused and redistributed by anyone (National Informatics Centre, 2013).

Open data repositories as such provide ample opportunities for the scientific communities not only to showcase their research but making their research assets visible and distinct at global level. These repositories thus incorporate much richer features including:

Web Syndication

Web syndication refers to the creation of web feeds available from a particular website in order to provide people with update of recently added content. The two main formats of web syndication include RSS and Atom. This technology is incorporated in open data repositories to make their users updated by retrieving latest information from sites of interest and informing them about the changes in the content.

Support for Varied Data Types

These repositories also have been maintained to host different types of Open Data sets with the help of different software technologies. These include images, graphs, audio-visual data, network-based data, etc. These types of data have been used by



10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/open-data-repositories-in-knowledge-society/184151

Related Content

Literature Review of Augmented Reality Application in the Architecture, Engineering, and Construction Industry With Relation to Building Information

Aydin Tabrizi and Paola Sanguinetti (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 983-993).

www.irma-international.org/chapter/literature-review-of-augmented-reality-application-in-the-architecture-engineering-and-construction-industry-with-relation-to-building-information/183811

What If You Meet Face to Face? A Case Study in Virtual/Material Research Ethics

David Clark (2004). *Readings in Virtual Research Ethics: Issues and Controversies* (pp. 246-261).

www.irma-international.org/chapter/you-meet-face-face-case/28302

Method of Fault Self-Healing in Distribution Network and Deep Learning Under Cloud Edge Architecture

Zhenxing Lin, Liangjun Huang, Boyang Yu, Chenhao Qi, Linbo Pan, Yu Wang, Chengyu Ge and Rongrong Shan (2023). *International Journal of Information Technologies and Systems Approach* (pp. 1-15).

www.irma-international.org/article/method-of-fault-self-healing-in-distribution-network-and-deep-learning-under-cloud-edge-architecture/321753

Internet Addiction and Fear of Missing Out: The Gender Difference of Tertiary Students in Ghana

Joseph Kofi Wireko (2019). *Gender Gaps and the Social Inclusion Movement in ICT* (pp. 253-268).

www.irma-international.org/chapter/internet-addiction-and-fear-of-missing-out/218448

Analysis of the Determinants of Initial Trust on a Virtual Leader

Miguel Guinalíu and Pau Jordán (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 4989-4996).

www.irma-international.org/chapter/analysis-of-the-determinants-of-initial-trust-on-a-virtual-leader/112947