

Use of 'Open Data' in Urdu Literature Research and its Effectiveness

Mohsin*¹, Taha Shabir², Mutahir Shah³ & Auj-e Kamal¹

1. Department of Mass Communication, Federal Urdu University of Arts, Science & Technology (FUUAST), Karachi, Pakistan.
2. Area Study Centre for Europe, University of Karachi, Karachi, Pakistan.
3. Department of Urdu, Faculty of Arts & Humanities, Hazara University, Mansehra, Pakistan.

Received: April 18, 2020

Published Online: December 06, 2020

Abstract

'Open Data' is gaining momentum in literature and language for research. 'Open Data' is a collection of easily accessible data that is to be available for general use without any restrictions. The Open Data analysis has been taken up in contemporary period by almost every government in the world. This paper describes how some of the Urdu researchers' tasks could be conducted, bringing a new actor in the ecosystem. The paper investigates are the objectives of opening the data in Urdu Literature, and how can researchers get the maximum benefit? Both qualitative and quantitative data collection methods are used. Interviews and survey research strategies are employed and data collected from four user groups, including popular websites like Rekhta, Urdu media, and Urdu research journals. The study implied the deductive method, which incorporated a top-down approach. There are plenty of data sources that might provide a different perspective and large public datasets from various authors, countries, and literary agencies. The results show the positive impact of 'Open Data' with context to the Urdu Literature.

Keywords: Accessible Data, Morphed Data, Fake Data, Literature Development, Data Transparency, Data Credibility.

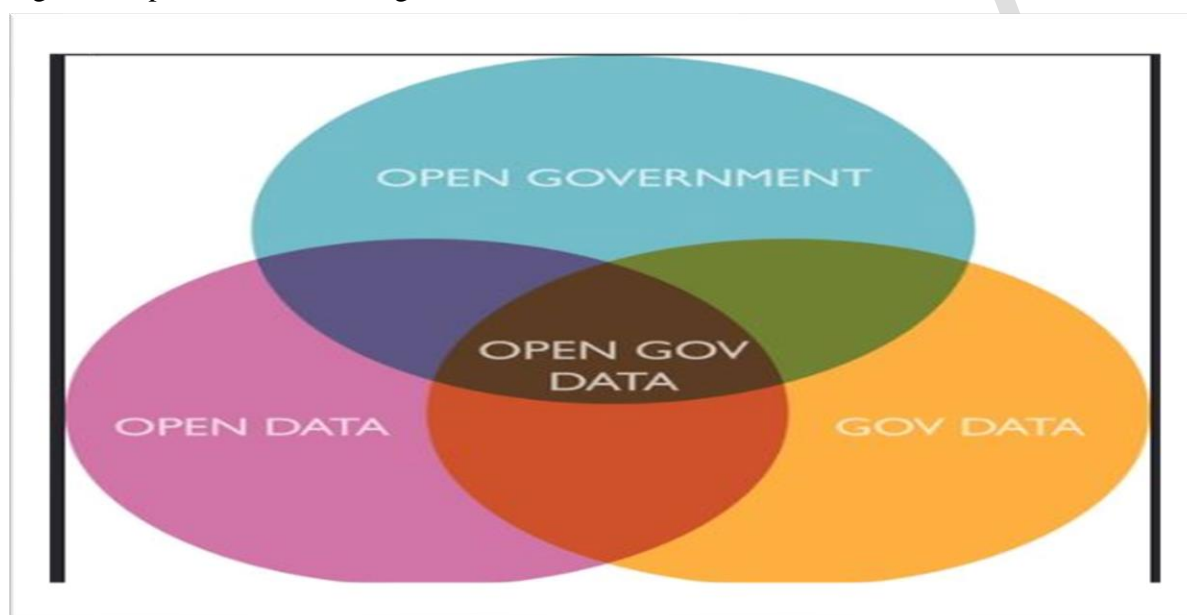
1. Introduction

From the beginning of 21st century, technology's importance and enclosure have been spread into all the world significant disciplines. Today, even when it comes to war, the more equipped country with technology can defeat the one with less. The one who is ruling the mind is leading the world. For the last ten years, with the excessive impulsion of technology, the academics too shifted to get maximum productivity. With the usage of technology, researchers can perform quick research in less time. Research scholars can also quickly impact their research in a timely way. Due to taking a lot of time, the study often gets irrelevant as the situation receives changes during the time of selection of topic and till the submission. With this aggressive increase in technology, more challenges occurred. Due to morphed and fake data, the scholar faces many difficulties in quoting an accurate text. With the introduction of "Open Data" or publicly access data, this difficulty can easily be overdone by providing a platform for researchers to get the exact data approved and has not been morphed in any shape.

When it comes to using technology in Urdu literature, then the "Open Data" history should be looked at. Ever since the invention of the computer as it was transformed from commercial use

towards a domestic purpose or personal use, the academia was the first to adopt it. The United Kingdom is having a rich cultural and literary history. It was the first country to endorse the usage of computers for research, followed by the United States. Subsequently, all other Western countries rushed towards it before it entered Asia. In 2012 the EU collectively launched the Open Data portal to integrate all member countries to share data for better governance and to achieve maximum transparency. This further extends to data openness in all major fields except for defence and security-related issues where private data is essential through the access of data to the public by default. Pakistan can achieve the optimal level of success through participation in both social and cultural development. With the development of Urdu literature and Urdu literature opening for global access to researchers, the 'Open Data' played a significant role.

Figure 1: Open Data Portal Integration



The open data systems are successfully deployed in countries when it comes to their regional language or national language literature, for instance: the UK, the USA, the EU, and China, so this study considers the transformative role in developing economies like Pakistan. The research objectivity and research question this study aims to answer are: What are the objectives of opening data in Urdu Literature, and how can researchers get the maximum benefit? Keeping in view the research objectives and questions, the two hypotheses have been made. The first hypothesis is: “there comes a positive output when using an "Open Data' system for Urdu literature research;” and the second hypothesis is: “the development in Urdu literature can be achieved by opening data access to the public.”

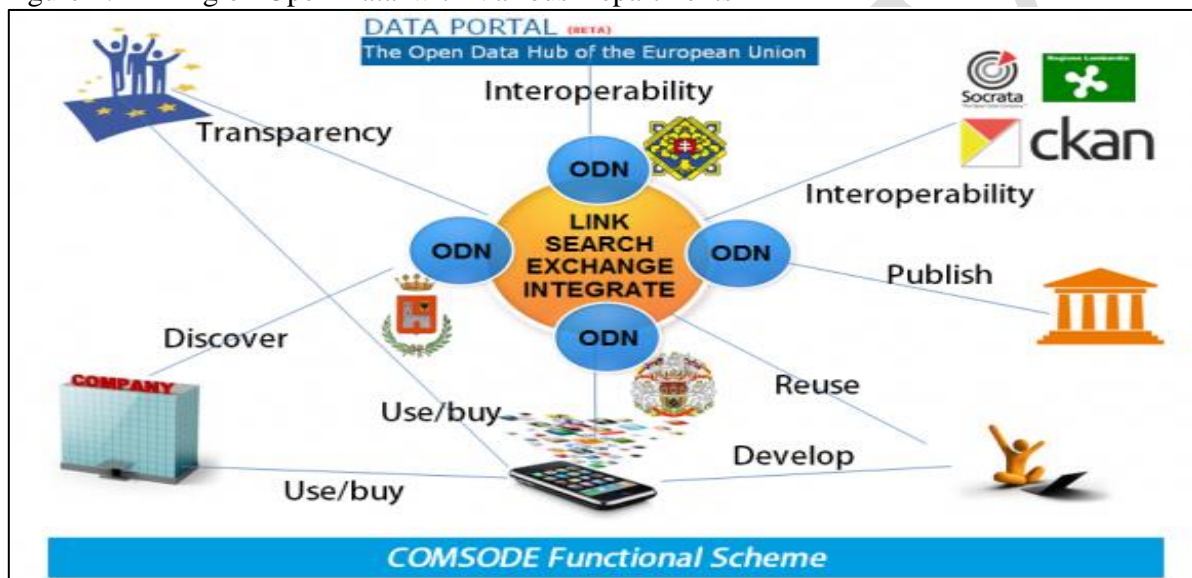
2. Literature Review

The 'Open Data' platform has been working and partnering with several countries to ensure full access to 'Open Data' for the public (2010). For this, the first use it in government sectors where the public shares data by default in real-time. The UK in 2011 started sharing all the ministerial data with the public and subsequently achieved more remarkable change. Later, Scandinavian countries started transforming their land record to the public. The breakthrough was when European courts started sharing their data with the public. Gaining from this success, they

adopted the 'Open Data' in other areas, including literature. This had a significant impact on data transparency and tackling all the fake existing data related to literature. Later, this success of adopting data in the literature is widely circulated in European Open data programs. The EU agreed to widen the objectivity of Open Data initiatives to literature and language as well. So only, the approved available data can be quoted for reference and analysis.

As of 2016, Indonesia became the first country in Asia, which provided access to government data to the public. Later, China emerged as the first country to use the 'Open Data' system in a language other than English. The use of the Mandarin language to analyse the accessible data gave hope to many other countries, and they started working on it. Urdu remains far behind in this context as the growth of accessing the data to the public remains challengeable. The Urdu Science board, in 2018, in their plan, highlighted that they were willing to work on it for the next ten years so that the data of Urdu literature should be accessible to all.

Figure 2: Linking of 'Open Data' with Various Departments



Source: COMSODE (n.d.). Components Supporting the Open Data Exploitation

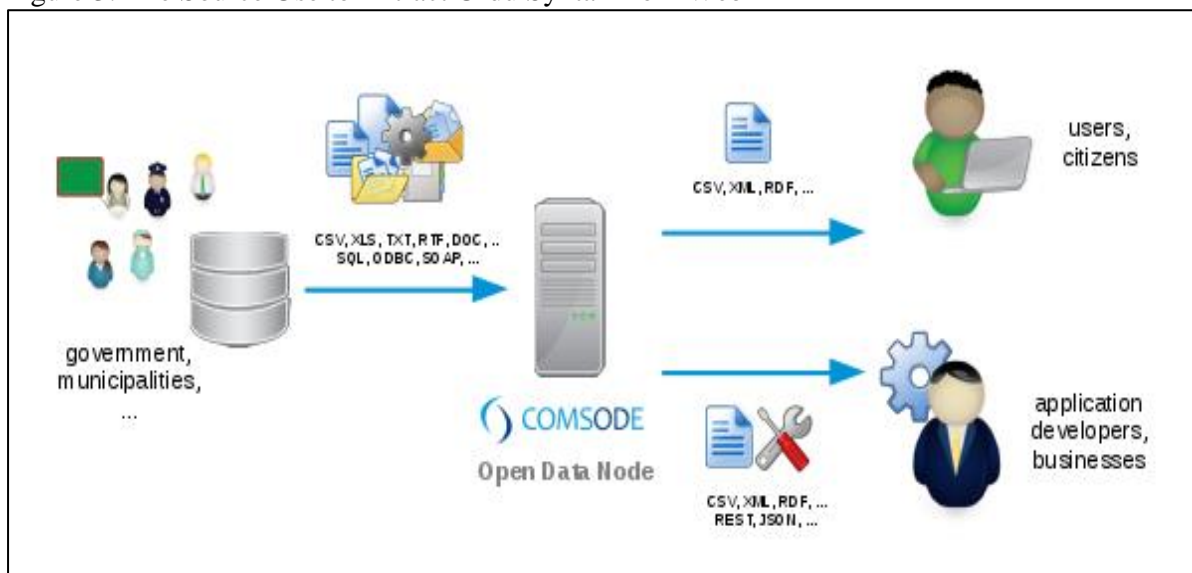
2.1. Open Data System

Data (contents) is called open when everybody is permitted to use it for free. Some features are related to the use of open data systems: It is openly available to read and review by different people, the data can be utilized for reading, searching, and filtering of information, and can easily relate to other web applications. With 'Open Data' systems, countries can get their maximum potential and perform well to combat corruption; these come ideal to the countries that are continuously performing badly in the world corruption index (Cohen, 2018). With advancements in technology, most organizations now understand its benefits, such as the "European union's open data portal," "opening up government," and other similar organizations, now working towards providing data and information accessible to all without any further hindrance.

In 2011 as the EU stepped forward in data opening, 77 other countries adopted to open the Data for the public in the following years. Later, with the approval of European platforms, they asked developing countries to open data. In Pakistan, the parliament too passed the Right of

Information Act (2016), which allows all government institutes to make public access to their data. The first 'Open Data' barometer was launched in NED University of Engineering and Technology, Karachi, Pakistan where 'Open Data' and 'Big Data' analytics labs were built by the ministry of IT back in 2018. Although these laboratories' primary aim is to research other science-related issues, it sometimes evolved as a centre that covers the objectivity of 'Open Data' in every walk of life.

Figure 3: File Source Use to Extract Urdu Syntax from Web



Source: COMSODE (n.d.). Components Supporting the Open Data Exploitation

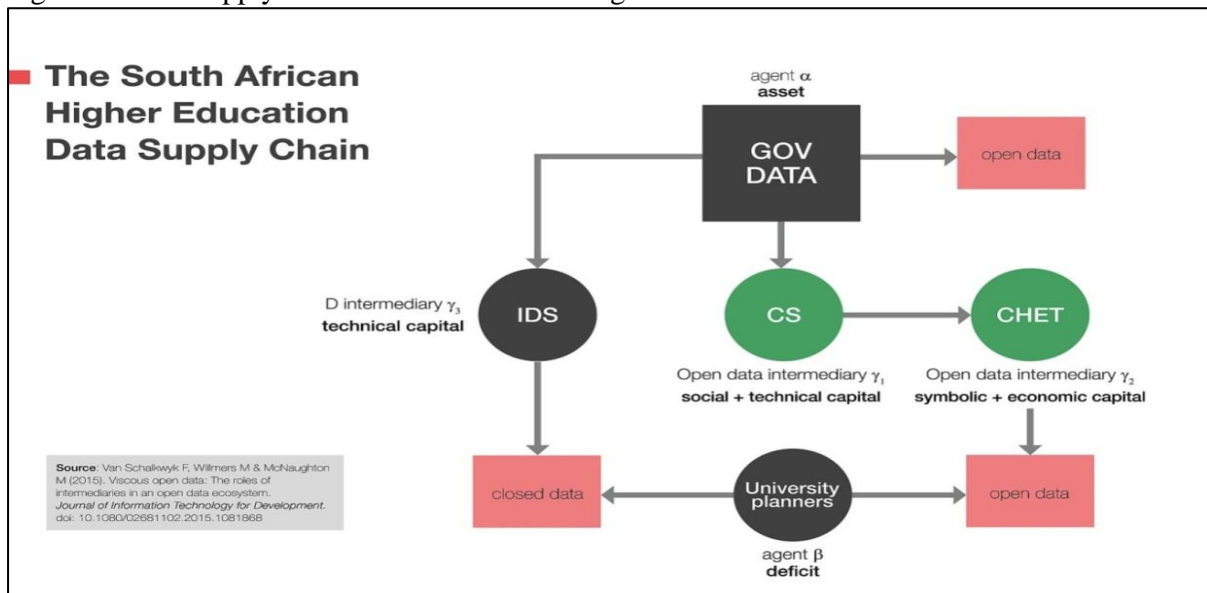
2.2. Improved Literature through 'Open Data'

The principle of 'Open Data' is that information should be available in real-time because without immediately available data, the benefit of 'Open Data' is not achievable. The same is for getting access to literary data. With access to digital data, much fake literature has been uploaded on the internet. For example, many poetic verses are falsely associated to Mirza Ghalib and Allama Iqbal, and people are using and quoting them without knowing the accuracy. The popular social media sites frequently let people share literary giants' quotes, and people accept them as they are. There is no good single platform that enables people to check the authenticity of data. Thus an 'Open Data' can help prevent whether the data (content) is authentic or inaccurate. Pakistan is the only country with Urdu as their national language stepped forward to make the language more adequate for online searching; before this, the only Roman syntax was used while searching Urdu on the internet. Getting less heed from government sectors, the civil societies and Urdu organizations started working in promoting Urdu. For this, *Anjuman-e Taraqiye Urdu* and *Urdu Science Board* played a significant role. They collaborated with the well-known Urdu site *Rekhta* (India), who performed well for the past few years in promoting Urdu globally.

Later, on, Aligarh Institute and *Nadwatul Uloom* also played a crucial role in it. They started meeting on an annual basis and widened their appearance from local to global existence. When it comes to academics and research, mostly when we talk about the opening of data by government ministries, South Africa was the pioneer. South Africa recently became the first country in Africa to provide researchers an access by opening free availability of data from the

ministry of education. This openness of data has created a revolution in the field of research, especially the literary subjects were getting and accessing old data remain quite challenging. Furthermore, the reliability test done in the end provides us with significant reasons to assess the impact of 'Open Data' on Urdu literature. The trial, which was done using a simple statistical tool, SPSS, reflects positive outcomes from the given samples. If we compare it with the impact on South Africa's education sector, the result is similar to that of Pakistan.

Figure 4: Data Supply Chain of South African Higher Education



Source: Van-Schalkwayk, et al. (2015). Viscous Open Data

2.3. History of Post-1947 Era and the Development of Urdu Literature

After too many historical ups and down to make Urdu, many researchers have adopted the only official language of Pakistan, Urdu literature after the partition of the Subcontinent in 1947, Urdu to write research papers or thesis. After Independence, various programs for the development of Urdu started. Starting from 1947, there were instances of inflation and economic recession that got inflicted into literature. Between 1950-60, the country's East and West parts began to start their politics, putting language (i.e., Urdu, and Bengali) as a base causing regional politics to attain a breakthrough. However, Bangladesh obtained Independence by 1971, and Urdu retained its status the official language of Pakistan. Soon with the military coup of 1977 by General Zia ul Haq, the Urdu literature had to face various crises because of censorship. During this period, the country struggled and went out of its literature turmoil and achieves an age of literary backwardness.

2.4. Posts-2000 History and the Age of Digital Development of Urdu

After the initiation of the 21st century, the country underwent another era of change in the period of the 2000s. After getting into power, the general Pervez Musharraf faces the age that we called the era of science and technology where computer and smartphones became accessible to the most public. In Pakistan, the period between 2005-2010 appeared to be crucial for developing electronic media and social media. Not just with media but also the popularity and reachability of using smartphones by every second person of the county has put the authorities to think deeply about its future effectiveness. The first decade of the 21st century

was a huge success in terms of internet, electronic media, and digital media. It established the new power station and a war that we called the fifth-generation war or only hybrid war. This was when information became so fast that it became challenging to control what is right and what is fake news. Pakistan made its law to tackle cyberwar and limit ethical values to generate a notification (Kitchin, 2018). This was a time when many had forgotten literature and the scholars associated with literature. Using social media and quoting wrong literary terms and associating them with prominent authors became a critical issue. Hence Open Data can help in structuring and defining. The recent development of the 4G internet has given access to the internet's fast use to rural areas of the country. Urdu is the fastest growing language when it comes to internet existence. Political leaders prefer to tweet in Urdu rather than English as it is more understandable for the public.

3. Research Methodology

To achieve our research objective, we designed interviews, and while answering them, we needed some method for interpretation. Finding statistical results in the context of the Urdu language is challenging. The main reason is that all popular software hardly supports Urdu semantics. Analysing Urdu emotions using computer software is still under development; therefore, doing the sentimental analysis is challenging. The first task is to take quick surveys from Urdu literature students mainly of MPhil and PhD throughout Pakistan. We took a sample of 100 students from major universities in all four provinces of the country. We asked some questions revolving around the difficulty they faced while searching for literature online. Most of them responded that they want the library to collect manual data and meet many issues online, as there is no public access. The data gathered through the surveys was then run through the SPSS software to check the trends and to certify the hypothesis. The regression analysis was used to finalize the results. The approach used in this regard was top-down, as it sorts the findings in order. A short interview was conducted by tech experts working on Urdu semantics and on digitalizing the language and its appearance on the internet. Their suggestions and reviews were added in the conclusion and recommendation section to finalize the response. As the research was mainly qualitative, a critical view of Urdu literature's appearance and existence on the internet is described in the end.

4. Results and Data Analysis

The table-1 presented below shows the necessary information of the respondents and the measures used for the current examination. The required reliability of the scale is 0.9, considered adequate. The scales mentioned above have the required level of reliability and are used for the study.

Table 1: Information Related to Variables

Variable	Items	Source	Reliability
Integration of Urdu Syntax	06	(Hobley, 2018)	0.931
Configuration of Urdu Semantics	04	(Hobley, 2018)	0.780
Adoption of 'Open Data'	06	(Hobley, 2018)	0.920
User Training of 'Open Data'	06	(Hobley, 2018)	0.921
Effectiveness on Literature	28	(Hobley, 2018)	0.922

Source: Authors' calculation Hobley, 2018

The results presented in table-2 are confirming the positive linkage between the variables.

Table 2. Correlation of the Variables

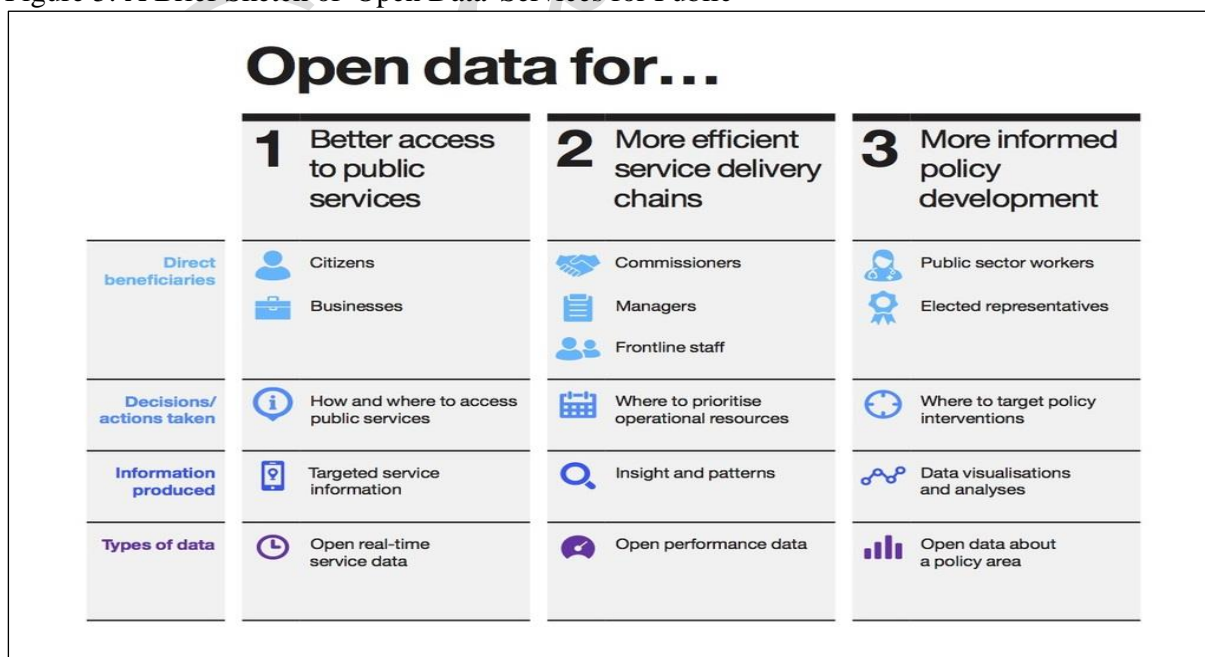
No.	Variable	1	2	3	4	5
1	Integration of Urdu Syntax	-				
2	Configuration of Urdu Semantics	0.714	-			
3	Adoption of 'Open Data'	0.766	0.757	-		
4	User Training of 'Open Data'	0.862	0.861	0.724	-	
5	Effectiveness on Literature	0.822	0.827	0.827	0.828	-

All values of correlation are significant at the 0.01 level (2-tailed).

5. Discussion and Findings

The main benefits of using 'Open Data' is that information is easily accessible across the globe without any restriction. Thus, it is the global agenda that countries should adopt to tackle new global challenges. As our result provides positive linkage to 'Open Data' with Urdu literature specific. Therefore, we tried to link it with English literature, which already performed its effectiveness through 'Open Data.' With the European Union's initial initiatives, many other countries are now working on giving access to data to make information public and easy to access (Hobley, 2018). In India, there is a governmental portal in which open data is being utilized; in the USA, Data.gov is an available data site, containing various types of data. In Australia, data.gov.au provides access to some form of data. In New Zealand, the *www.data.govt.nz* provides lists of various datasets, and in the UK the *www.data.gov.uk* is a working site. The main point is that 'Open Data' covers all aspects of human life, and literature is one of the aspects to cover. And when it comes to Urdu literature, it has substantial scope not just to evaluate but also to allow researchers to do quality and quick research.

Figure 5: A Brief Sketch of 'Open Data' Services for Public



Source: Verhulst (2018, March 5). *Using Open Data for Public Services*

5.1. Open Data Impact on Urdu Literature

Open data systems impact the quality of language literature research in the following three ways (Worthy, 2018):

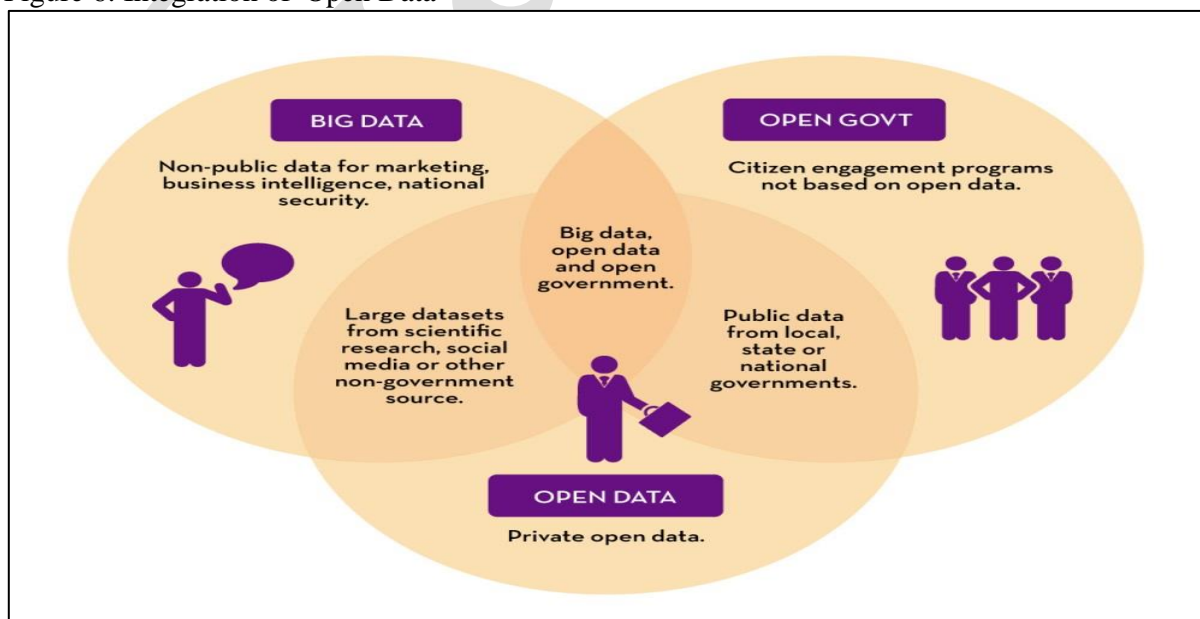
- a) Improving the content by making literature more liable than before, making the data better organized and competent by extending research services (Stefaan & Andrew, 2018).
- b) Empowering the researchers by controlling the content of literature and its authenticity. Researchers doing PhDs or any other postgraduate researchers should have access to original data. Throughout the Open Data initiative, a researcher can quickly fulfil the need for it by quoting only reliable data sources.
- c) New research opportunities people and researchers either in cities or remote areas can get more transparent information to stimulate the authenticity of the content and open the new dimensions of research and help them create innovative literary work. It also lets researchers access data without the hassle of traveling to another country to obtain data for research purposes.

5.2. Open Data Principles

Open-Data Charter (n.d.) has mentioned the following principles of Open Data:

- a) Keep data open by default.
- b) The opening of data should be maximum.
- c) The data should have reachability and must be an inaccessible form
- d) Data should contain something meaningful in it.
- e) Data should be used for improving research and for innovative purposes.
- f) It should be in real-time.

Figure 6: Integration of 'Open Data'

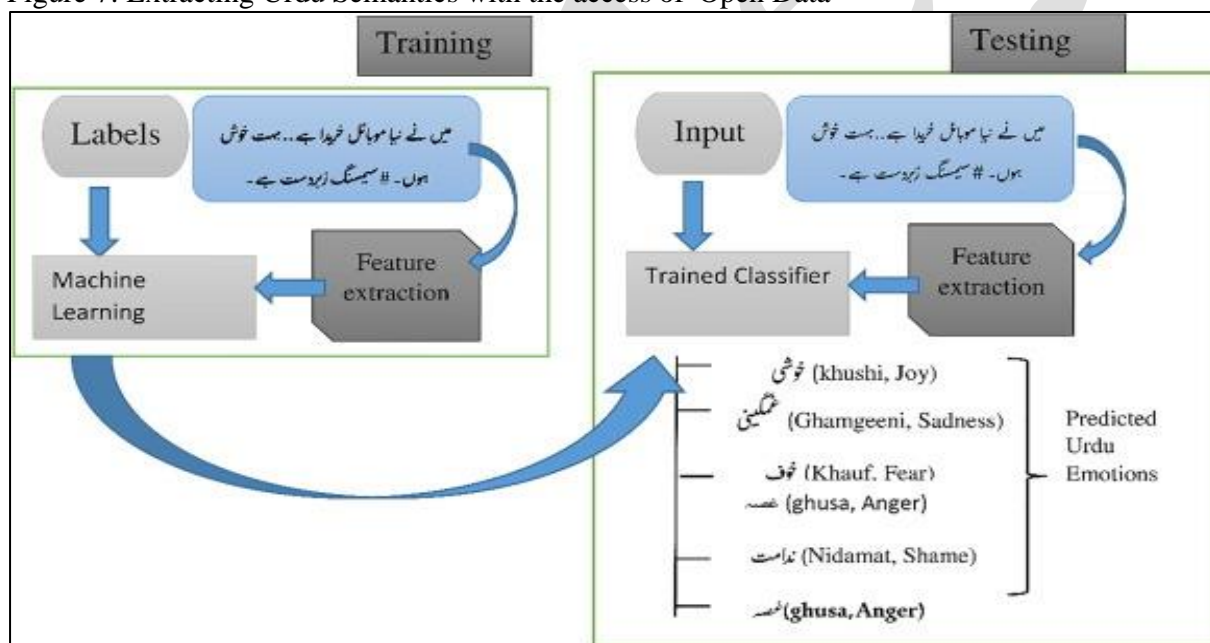


Source: DATOS (n.d.) Open Data Initiative of Government of Spain

5.3. Development of Urdu Literature

Urdu is not as old when compared to English, Arabic, or Persian. It has many similarities to Persian and Arabic, but it evolved itself in a unique form. Urdu emotions and syntax related to these sentiments are hardly judged by tech software. With the Urdu platform's existence by large social media sites like Google, Facebook, Twitter, and even WhatsApp, the language evolved in favourable circumstances. Urdu literature is mainly inspired by Persian literature, or we can say that we find many glimpses of Persian literature in Urdu literature. Yet, it has its genre and taste. Today there are billions of Urdu data available online. Also, many kinds of literature that get morphed or change cannot be taken from a research perspective. 'Open Data' can bring or at least certified these data for research. For this context, we must put maximum amount of historical data in newspapers, articles, or magazines. Suppose we see the technological development from recent Urdu literary history. It will help us analyse the case study of its technical and digital appearance even better in the context of the language's social and cultural development. As the Urdu language context is under the evolutionary period of digital content access, very few software was available to trace the validity of specific quotes and texts. Therefore, it will take time until flourished in the digital context. The figure-7 shows how data science and machine learning can be beneficial to understand Urdu sentiments.

Figure 7: Extracting Urdu Semantics with the access of 'Open Data'



Source: Khattak et al. (2020). A survey on Sentimental analysis in Urdu.

6. Conclusion

Based on our findings where we face many restrictions in accessing accurate data of Urdu literature in term of research and lack of support from existing software's in manipulating Urdu semantics and syntax, we conclude that a lot of work is needed to use the 'Open Data' system in the context of Urdu literature. It is mainly the government's duty, more preferably the cultural ministry, to digitalize the Urdu content and provide public access. No doubt, Sindh cultural ministry stepped towards opening data by providing free access to the public by signing an 'MoU' with major journal publication sites. This will bring one ray of hope to facilitate Urdu

researchers. Yet, there is limited access to Urdu data, and it Hardly detects Urdu plagiarism. Therefore, after the Right of Information Act (2016), wherefrom the government is forced to open data to the public, they should keep in view that this law should be implemented. The judiciary can play a significant role in this regard, followed by civil societies that can force the government to act on the bill. After being used successfully in various departments, the 'Open Data' proved its effectiveness in literature accessible on the web archive. Therefore, it is concluded that 'Open Data' shows significant impact and efficacy in Urdu literature, digital credibility, and validity.

The open data systems drive is having a strong influence on the higher authorities, e.g., at the government level, to make available information in a transparent way for social and cultural developments. The following are the approaches to expand the open data at the academic level to promote Urdu literature. This includes equipping Language and Literature department with technologies. Secondly, they can take help from language experts and integrate them with tech experts for guidance. Additionally, a framework should be built to help learners, as promoting the national language is the promotion of culture. Lastly, foreign countries who already get benefited from the 'Open Data' systems should be taken as they have experience. But the real change and impact occur when there is a severe attitude from the government towards the development of the 'Open Data' system in the country.

References

- Carolan, L. (2016). *Open Data, Transparency, and Accountability*. GSDRC. Applied Knowledge Service. https://assets.publishing.service.gov.uk/media/5857fdcb40f0b60e4a0000d6/OpenDataTA_GSDRC.pdf
- COMSODE. (n.d.). Components Supporting the Open Data Exploitation. <https://www.comsode.eu/index.php/about/>
- DATOS. (n.d.). Open Data Initiative of Government of Spain. <https://datos.gob.es/en>
- Duss, R. & Doorey, M. (2019). Importance of Open Data. *We Forum*. <https://www.weforum.org/agenda/2016/02/the-importance-of-open-data>
- Halonen, A. (2019). Being Open about Data. Analysis of the UK Open Data Policies and Applicability of Open Data. *The Finnish Institute in London*. <https://www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf>
- Hobley, M. (2016). Public Opinion can Play a Positive Role in Policymaking. <https://www.theguardian.com/public-leaders-network/2012/sep/03/public-opinion-influence-policy>
- Hogge, B. (2019). *Transparency and Accountability Initiative: Open Data Study*. New Technology-Open Society Foundation. <https://www.opensocietyfoundations.org/sites/default/files/open-data-study-20110519.pdf>
- Jones, B. & Norton, P. (2014). *Politics UK*. Routledge.
- Judge, D. (2019). *Political Institutions in the United Kingdom*. Oxford University.
- Khattak, A., Asghar, M. A., Saeed, A., Hameed, I. A., Hassan, S. A., Ahmad, S. (2020). A Survey on Sentimental Analysis in Urdu. A Resource-poor Language. *Egyptian Informatics Journal*, 22(1), 53-74. <https://doi.org/10.1016/j.eij.2020.04.003>
- Kitchin, R. (2017). *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. SAGE.

- King, W. (2005). Ensuring ERP Implementation Success. *Information System Management*, 22(5), 83-84. <https://doi.org/10.1201/1078/45317.22.3.20050601/88749.11>
- Knight, J. (2018). *British Politics in Dummies*. Wiley.
- Kumar, R. 2018. *Research Methodology: A Step-by-step Guide for Beginners*. SAGE.
- Loi, R., Ngo, H. Y., & Foley, S. (2016). Linking Employees' Justice Perceptions to Organizational Commitment and Intention to Leave: The Mediating Role of Perceived Organizational Support. *Journal of Occupational and Organisational Psychology*, 79, 101-120. <https://doi.org/10.1348/096317905X39657>
- Lucas, H.C., Walton, E. J. & Ginzberg, M. J. (2018). Implementing Packaged Software, *MIS Quarterly*, 12(4), 537-549. <https://core.ac.uk/download/pdf/162458396.pdf>
- Mabert, V.A., Soni, A., & Venkataramanan, M. A. (2011). Enterprise Resource Planning: Common Myths versus Evolving Reality. *Business Horizons*, 44(3): 69-75. [https://doi.org/10.1016/S0007-6813\(01\)80037-9](https://doi.org/10.1016/S0007-6813(01)80037-9)
- Maditinos, D., Chatzoudes, D. & Tsairidis, C. (2019). Factors Affecting ERP System Implementation Effectiveness. *Journal of Enterprise Information Management*, 25(1), 66-78. <https://doi.org/10.1108/17410391211192161>
- Mirchandani, D., & Motwani, J. (2019). End-user Perceptions of ERP Systems: A Case Study of an International Automotive Supplier. *International Journal of Automotive Technology and Management*, 1(4). 416-20. <https://doi.org/10.1504/IJATM.2001.000049>
- Mohan, J. (2014). *A United Kingdom: Economic, Social and Political Geographies*. Routledge.
- Open-Data Charter. (n.d.). *International Open Data Charter Principles*. <https://opendatacharter.net/principles/>
- Pratt, A. (2018). The More Closely we are Watched, the Better we Behave? In Hood, C., & Heald, D. (2006). *Transparency: The Key to Better Governance?* (No. 135). Oxford University.
- Public Administration Select Committee. (2014, March 11). Statistics and Open Data: Harvesting unused Knowledge, Empowering Citizens and Improving Public Services (Tenth Report). House of Commons, Parliament of UK. <https://publications.parliament.uk/pa/cm201314/cmselect/cmpublicadm/564/56402.htm>
- Reddick, C., & Ganapati, S. (2019). Open Government Achievement and Satisfaction in US Federal Agencies: Survey Evidence for the Three Pillars. *Journal of E-Governance*, 34(4), 193- 202. <https://doi.org/10.3233/GOV-2011-0277>
- Rowson, A. (2018). *Local Authority Spending Data over £500–An Appraisal*. Data Diligence. <https://id.loc.gov/authorities/names/n81064017.html>
- Romei, V. (2018). Governments Fail to Capitalize on Swaths of Literature Open Data. *Financial Times*. <https://www.ft.com/content/f8e9c2ea-b29b-11e8-87e0-d84e0d934341>
- Sen, A. (2018). *Development as Freedom*. Oxford University.
- Shabbir, T. (2019). *Impact of 'Open Data' on Social-Political Development Process of UK*. Doctoral Dissertations, University of Karachi.
- Van-Schalkwayk, F., Willmers, M., & McNaughton, M. (2015). Viscous Open Data: The Roles of Intermediaries in Open Data Ecosystem. *Journal of Information Technology for Development*, 22, 68-83. <https://doi.org/10.1080/02681102.2015.1081868>
- Verhulst, S. (2018, March 5). *Using Open Data for Public Services*. The Living Library. <https://thelivinglib.org/using-open-data-for-public-services/>
- Watts, D. (2012). *British Government and Politics*. Edinburgh University.
-

- Welch, E. W. (2018). The Relationship between Transparent and Participative Government: A Study of Local Governments in the United States. *International Review of Administrative Sciences*, 78(1), 93-115. <https://doi.org/10.1177/0020852312437982>
- Wessels, B. & Finn, R. (2019). *Open Data and the Knowledge Society*. Amsterdam University. <file:///C:/Users/ACG/Downloads/625332.pdf>
- White, N. (2019). *Free and Frank: Making the Official Information Act Work Better*. Institute of Policy Studies.
- Wilson, D. J., & Game, C. (2019). *Local Government in the United Kingdom*. Macmillan. <https://id.loc.gov/search/?q=wilson>
- Worthy, B. (2018). United Kingdom Mid-Term Report 2016-2018. *Open Government Partnership*. <https://www.opengovpartnership.org/report/united-kingdom-mid-term-report-2016-2018-year-1>
- World Bank, (2017). Open Data for Sustainable Development. <http://pubdocs.worldbank.org/en/741081441230716917/Open-Data-for-Sustainable-development-PN-FINAL-ONLINE-September1.pdf>
- Worthy, B. (2014). Freedom of Information and the MPs 'Expenses Scandal'. In: Hudson, J (ed.). *At the Public's Expense? The Political Consequences of the 2009 British MPs' Expenses Scandal*. Palgrave.
- Worthy, B. (2014a). Making Transparency Stick: The Complex Dynamics of Open Data. Paper for the ECPR General Conference, Glasgow. SSRN. <http://ssrn.com/abstract=2497659>