Drivers and Barriers to Data Sharing Practices in the New Digital Age – An Exploratory Study

ISBN 978-1-943295-24-1

Manjunath B Swarna
Anjula Gurtoo
Indian Institute of Science, Bengaluru
(manjunathbswarna@gmail.com)
(anjula@iisc.ac.in)

Data plays a crucial role in most fields in the new digital age. To harness the potential of data, it should be Findable, Accessible, Interoperable, and Reusable (FAIR principle). The big question is how to make data FAIR. The private sector is the forerunner in acquiring and utilizing data, but government sectors are much neglected; education is one of them. The authors explored what are the drivers and barriers of bringing data to data exchanges from department silos. The authors conducted a Pan-India survey of 1300+ respondents to gather insights on one of the data platforms of the central government ministry.

Keywords: Data Exchange, Data Regulation, Social Good, Data Sharing Perception, Data Sharing Involvement, Digital Value, Institutional Factors, HQ Datasets

1. Introduction

Data plays a crucial role in most fields in the new digital age. To unlock the actual data value, data should be Findable, Accessible, Interoperable, and Reusable -FAIR principle. The FAIR principle, introduced in the last decade((https://www.go-fair.org/fair-principles/), aspires to improve data's smooth and efficient usage in data-driven processes. The big question is how to make data FAIR. Implementing FAIR principles ensures that educational sector data is managed equitably, allowing diverse stakeholders to access and utilize the data effectively (Bowers & Choi, 2023). Implementing a successful data exchange based on FAIR standards can improve usability and contribute to informed decision-making and policy development. Data exchange platforms built on FAIR standards and driven by new-age data exchange technologies like blockchain, new-gen encryption, and privacy-preserving algorithms can stimulate the data ecosystem development and unlock its value.(Fabrice, Tocco., Laurent, Lafaye. (2022).)

The government is one of the critical collectors and sharers of data. Many government departments have their portals that share data. Many departments publish data in various open data portals to encourage users to create new and innovative services (Abella et al., 2015). However, producing more data does not imply creating value or appreciated services for citizens. Creating a new service or product by sourcing and combining various data sources urgently needs a holistic framework to attract long-term investments. According to authors Jetzek, Avital, and Bjorn-Andersen (2014a, 2014b), Innovating new products and services is one of the primary value-creation processes. Innovation is nurtured when data is shared and linked between different organizations. However, due to a lack of proper governance and framework, it remains poorly accessible or not accessible at all.(Fabrice, Tocco., Laurent, Lafaye. (2022).

Creating a data marketplace is a multistakeholder approach where various platforms connect data producers and data users with value-added services by technology providers (Eisenmann et al., 2006). This data exchange platform can create value for producers and users through better data discovery and smooth transactions (Bakos, 1991; Soh et al., 2006).

2. Literature Review

Data sharing is a process of transferring data between two or more individuals or departments or organisations (Harvey & Tulloch, 2006). Many researchers have demonstrated the importance of data sharing (Dallmeier-Tiessen et al., 2014; Williamson et al., 2016; Verhulst & Young, 2016). Data sharing can pave the way for creating and smoothing access to large, high-quality data sets. Efficient data-sharing practices provide ways for improving business processes efficiency and creating value through new business models and services.

- **Data Exchange Platforms**

The data economy is gaining importance, and innovative business models and data-sharing uses are continuously developing. Data-sharing platforms as trusted intermediaries can facilitate data exchange between different market players and reduce the entry barriers in data-driven markets. (Richter, H., & Slowinski, P. R. 2019)Trust between different stakeholders plays an important role in a data sharing ecosystem. Therefore, creating and maintaining trust is an important function of data-sharing platforms. (European Commission and Everis (2018); IDC and Lisbon Council 2017; Hofheinz and Osimo 2017) Despite data-sharing platforms playing the role of intermediaries, many issues like transparency, high market power, and governance issues can act as barriers to the development of a data-sharing culture. A general set of frameworks for data-sharing ecosystems can reduce these barriers to data harnessing. (Richter, H., & Slowinski, P. R. 2019).

- **Drivers**

Drivers can be broadly classified into social and organizational benefits. Several scholars have pointed out that adequately designed data exchange platforms can create value both for the public and organizations (Attard et al., 2015; Lourenço, 2015; Dawes & Helbig, 2010). Many open data initiatives aim to improve accountability, promote transparency, improve public service delivery, and enhance e- (Berrone et al., 2016; Davies & Perini, 2016; Smith et al., 2015). Interagency government data-sharing initiatives play a significant role in coordination and improved public service delivery (Wang, F.,2018). Linking and combining data from various departments can positively impact developing economies in the Global South. For example, linking Aadhar data with other government data to identify real benefits for social benefits(Pati et al.,2015)

- **Barriers**

Poor data quality and lack of standardization can hinder effective sharing and utilization(Drummond & Christie, 2022). Organizations often need more skills and resources to manage data sharing effectively, leading to a preference for limited sharing practices(Houtkoop et al., 2018). No universal definition of data quality can be used across all data domains. Data quality is subjective in nature; data can be suitable for one purpose but not for another purpose, and the user expectations of what they define as helpful data. Infrastructure barriers can be a considerable concern when implementing data-sharing initiatives in developing economies. In the past, many researchers have contributed valuable insights into implementing data regulations and governance protocols (Alhassan et al., 2016; Brous et al., 2016; Weber et al., 2009). Many studies have demonstrated that designing data governance is a challenge; what is even more challenging is implementing it at the ground level (Benfeldt Nielsen, 2017).

Research Questions

1. What are the significant drivers and barriers for data sharing with data exchange platforms
2. Are there significant differences in drivers and barriers between government and private users?

Firstly, the authors explored what are the drivers and barriers to bringing data to data exchanges from the department silos. In the second step, authors use various statistical tests to find the significant difference between different types of users. Finally, we draw a conclusion based on the data analysis.

3. Research Methodology

The study employs an exploratory approach to identify the various drivers and barriers. The authors selected survey and expert interviews for the pilot study and a large-scale survey for the main study. The authors used the Mann-Whitney two-sample test to compare the distribution of drivers and barriers across government vs private users.

- **Pilot Study**

The pilot study was conducted to confirm the validity and reliability of the scale. Primary data was collected via field visits across 10+ taluks of south Karnataka. Users of one of the state government portals were the target population. Fifty responses and two expert interviews were collected to understand the holistic picture.
The validity and reliability of barriers and drivers were established. The scale adopted for this study was validated with empirical data. Additional measurement items are added based on expert feedback

- **Main Study**

The main study was a full-scale online survey across the country. The target population was users of one of the central government portals. A total of 1300+ responses were received. The questionnaire was circulated online through various channels like email ID, Linkedin, etc.

4. Data Analysis

The authors used the SPSS software package to conduct exploratory factor analysis, Normality test, and Independent K sample test. Smart PLS software for scale validation

- **Constructs Measurement**

Overall, 37 measurement items were selected after the pilot study: 16 for drivers and 21 for Barriers. The driver items converged into two factors, namely social drivers and organizational drivers, and the barrier items converged into three factors: infrastructure barriers, funding barriers, and privacy barriers

Exploratory Factor Analysis

EFA, or Exploratory Factor Analysis, is a multivariate statistical method used to identify the underlying relationships between variables by identifying latent factors that explain observed data. EFA is particularly valuable in fields such as social sciences, health sciences, and economics, where it helps researchers reduce data complexity and categorize items into meaningful factors("Exploratory factor analysis," 2023) (Sürücü et al., 2024).

Table 1 *Total Variance Explained*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.491	32.729	32.729	9.491	32.729	32.729	8.076
2	5.484	18.910	51.639	5.484	18.910	51.639	7.177
3	1.784	6.150	57.789	1.784	6.150	57.789	5.479
4	1.124	3.878	61.667	1.124	3.878	61.667	4.700
5	1.110	3.828	65.494	1.110	3.828	65.494	5.106
6	0.839	2.894	68.388				
7	0.694	2.394	70.782				
8	0.661	2.278	73.059				
9	0.633	2.184	75.243				
10	0.557	1.919	77.162				
Extraction Method: Principal Component Analysis.							

- **Construct Reliability**

Reliability was assessed by the criteria of Cronbach's alpha being more significant than 0.7 (Chin, 1998). The reliability values of all variables are more than 0.7. All the driver and barrier constructs satisfy the conditions prescribed by (Hair et al., 2009). Thus, the construct validity condition is satisfied.
According to (Hair et al., 2009), construct validity can be assessed by the below three parameters

1. item loadings (lambda) greater than 0.50 at minimum and ideally greater than 0.70
2. average variance extracted from each variable is more significant than 0.5; and
3. construct reliability exceeding 0.7.

Table 2 *Construct Validity*

Factor	Cronbach's alpha	Number of items	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)	Factor Loadings
B1	0.891	7	0.907	0.913	0.601	0.742 - 0.810
B2	0.804	3	0.88	0.88	0.711	0.744 - 0.904
B3	0.929	8	0.866	0.926	0.612	0.672 - 0.904
D1	0.826	5	0.829	0.878	0.591	0.698 - 0.817
D2	0.902	7	0.902	0.922	0.629	0.754 - 0.819

- **Discriminant Validity**

Discriminant validity was calculated by the method that the correlation of one construct with other constructs must be less than the square root of average variance explained as demonstrated by authors (Fornell & Larcker, 1981).

Table 3 *Discriminate Validity*

	B1	B2	B3	D1	D2
B1	0.775				
B2	0.604	0.843			
B3	0.683	0.588	0.782		
D1	0.128	0.119	0.16	0.769	
D2	0.113	0.123	0.126	0.702	0.793

- **Multicollinearity**

The smallest possible value of VIF (1) indicates the absence of multicollinearity, and VIF values less than three indicate that multicollinearity is not severe (Diamantopoulos & Siguaw, 2006). As a thumb rule (Vörösmarty et al., I. (2020), a VIF value that exceeds 5 or 10 indicates a multicollinearity issue. The values obtained are less than five, which rules out the multicollinearity issue in this study.

Table 4 *Multicollinearity*

	VIF				
D1_1	1.702	**B1_1**	2.425	B16_1	2.262
D2_1	1.982	**B2_1**	2.845	B17_1	2.881
D3_1	1.816	**B3_1**	2.357	B18_1	3.187
D4_1	1.453	**B4_1**	1.859	B19_1	3.047
D5_1	1.582	**B5_1**	2.012	B20_1	2.688
D9_1	2.207	**B6_1**	2.155	B21_1	2.874
D10_1	2.496	**B7_1**	1.96		
D11_1	1.898	**B8_1**	1.881		
D13_1	2.331	**B10_1**	1.576		

D14_1	2.539	**B11_1**	1.871		
D15_1	2.409	B12_1	1.861		
D16_1	2.203	B14_1	2.333		

- **Test for Normality**

The normality assumption is crucial to identify whether to use a parametric or non-parametric test. There are many methods to test the normality of the continuous data. The two well-known normality tests, the Kolmogorov–Smirnov test and the Shapiro–Wilk test, are the most used methods to test the normality of the data.

The Shapiro–Wilk test is an appropriate method for small sample sizes of less than 50 samples, although it can also handle larger sample sizes(Shapiro & Wilk, 1965), while the Kolmogorov–Smirnov test is used for n greater than or equal to 50.(Wenjun et al., 2021).

The null hypothesis: Data is taken from the normally distributed population

Alternative hypothesis: Data is not taken from the normally distributed population

A significant p-value indicates that the data set is normally distributed, and a low p-value means it is not. (Ghasemi, A., & Zahediasl, S., 2012). As per both tests, P values are <0.05. A low p-value indicates that data is not normally distributed. Thus, the nonparametric test is used for further analysis.

Table 5 *Test of Normality*

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
1 Privacy Barriers	0.118	689	0.000	0.931	689	0.000
2 Infrastructure Barriers	0.120	689	0.000	0.937	689	0.000
3 Organisational Drivers	0.080	689	0.000	0.927	689	0.000
4 Social Drivers	0.069	689	0.000	0.937	689	0.000
5 Fund Barriers	0.066	689	0.000	0.978	689	0.000

- **Independent K Sample Test**

The Mann-Whitney two-sample test is a nonparametric method used to compare two independent samples. It is handy when the assumptions of normality are not met, making it a robust alternative to parametric tests. This test evaluates whether one of the two samples tends to have larger values than the other based on the ranks of the data rather than the raw values. (Mann., & Whitney., (1947), Nachar, N. (2008))

The test does not assume any specific distribution, making it suitable for ordinal data or non-normally distributed continuous data(Meléndez et al., 2020).

Table 6 *Mann- Whitney Two-Sample Test*

	Social Drivers	**Organisational Drivers**	**Infrastructure Barriers**	**Fund Barriers**	**Privacy Barriers**
Mann-Whitney U	31361.000	35910.000	36806.000	38433.000	34482.000
Wilcoxon W	130151.000	134700.000	135596.000	54723.000	50772.000
Z	-4.215	-1.985	-1.546	-0.748	-2.685
Asymp. Sig. (2-tailed)	0.000	0.047	0.122	0.454	0.007
Grouping Variable: Type of users - Government, Private users					

Null Hypothesis

H0 = There is no significant difference in social drivers/Organisational drivers/Infrastructure barriers/Fund Barriers /Privacy barriers between government and private users

Alternative Hypothesis:

H1 = There is a significant difference in social drivers between government and private users

H2 = There is a significant difference in organizational drivers between government and private users

H3 = There is a significant difference in Infrastructure barriers between government and private users

H4 = There is a significant difference in Fund barriers between government and private users

H5 = There is a significant difference in Privacy barriers between government and private users

Based on Mann- Whitney test results

For Social drivers: P value < 0.05 Reject Null Hypothesis

There is a significant difference in social drivers between government and private users

For Organisational drivers: P value < 0.05 Reject Null Hypothesis

There is a significant difference in organizational drivers between government and private users

For Infrastructure barriers: P value > 0.05 Failed to reject Null Hypothesis

For Fund Barriers: P value < 0.05 Failed to reject Null Hypothesis

For Privacy Barriers: P value < 0.05 Reject Null Hypothesis

There is a significant difference in social drivers between government and private users

5. Conclusion

Firstly, the authors identified various drivers and barriers to data sharing in this paper. The results show that all measurement items converge into five factors: Social and Organisational drivers, Infrastructure and Funding barriers, and Privacy barriers. This can be a starting point for policymakers in implementing a data exchange platform where multiple departments can create value for citizens and the organization.

Second, the authors compared the differences in drivers and barriers between the two types of users, which can be an input for creating different strategies for different users. One size fits all does not work anymore in this highly specialized world. Regarding drivers, both government and private users have significant differences, whereas in barriers, only privacy barriers have significant differences, and with Infrastructure and funding barriers, there is no significant difference. This study highlights the differences between how users experience data-sharing drivers and data-sharing barriers differently.

This research can be a guiding path toward implementing successful data exchange in the education sector to harness the potential of data. These drivers and barriers can be a missing puzzle for many struggling government portals and government-backed data exchange platforms.

6. References

1. Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2015a). Information reuse in smart cities' ecosystems. El Profesional de la Information, 24(6), 838–844.
2. Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2015). Information reuse in smart cities' ecosystems. Profesional de la information, 24(6), 838-844.
3. Alex, J., Bowers., Yeon-joo, Choi. (2023). 1. Building School Data Equity, Infrastructure, and Capacity Through FAIR Data Standards: Findable, Accessible, Interoperable, and Reusable. Educational Researcher, doi: 10.3102/0013189x231181103
4. Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. Government Information Quarterly, 32(4), 399–418.
5. Bakos, J. T. (1991), 'A strategic analysis of electronic marketplaces,' MIS Quarterly, 15(3), 295 310.
6. Benfeldt Nielsen, O. (2017). A comprehensive review of data governance literature. In Selected Papers of the IRIS (Vol. 8).
7. Berrone, P., Ricart, J. E., & Carrasco, C. (2016). The open kimono: Toward a general framework for open data initiatives in cities. California Management Review, 59(1), 39–70.
8. Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: A systematic review of data governance principles.
9. CHIN, W. W. 1998. Commentary: Issues and Opinion on Structural Equation Modeling. MIS Quarterly, 22, 7-16
10. Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Matthews, B., Mele, S., & Wilson, M. (2014). Enabling sharing and reuse of scientific data. New Review of Information Networking, 19(1), 16–43.
11. Davies, T., & Perini, F. (2016). Researching the emerging impacts of open data: Revisiting the ODDC conceptual framework. The Journal of Community Informatics, 12(2).
12. Dawes, S., & Helbig, N. (2010). Information strategies for open government: Challenges and prospects for deriving public value from government transparency. Electronic Government, 6228, 50–60.
13. Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. British journal of management, 17(4), 263-282.
14. Eisenmann, T. R., G. Parker and M. W. Van Alstyne (2006), 'Strategies for two-sided markets,' Harvard Business Review, 84(10), 92–101.
15. European Commission, Everis (2018) Study on data sharing between companies in Europe. European Union.
16. Fabrice, Tocco., Laurent, Lafaye. (2022). Data Platform Solutions. 383-393. doi: 10.1007/978-3-030-93975-5_23
17. FORNELL, C. & LARCKER, D. F. 1981. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. Journal of Marketing Research, 18, 382-388.
18. Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486–489.
19. HAIR JR., J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E. & TATHAM, R. L. 2009. Multivariate Data Analysis, Pearson Prentice Hall.
20. Harvey, F., &Tulloch,D.L.(2006).Local government data sharing: Evaluating the foundations of spatial data infrastructures. International Journal of Geographical Information Science, 20(7), 743–768.
21. Hofheinz P, Osimo D (2017) Making Europe a data economy: a new framework for free movement of data in the digital age. Lisbon Council Policy Brief, vol 11, No 1
22. IDC, Lisbon Council (2017) Opening up private data for public interest. http://datalandscape.eu/sites/default/files/report/Story_1_New_format.pdf.
23. Jesiļevska, S. (2017). Data Quality Dimensions to Ensure Optimal Data Quality. Romanian Economic Journal, 20(63).

24. Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014a). Generating sustainable value from open data in a sharing society. In B. Bergvall-Kåreborn, & P. A. Nielsen (Eds.), Creating Value for All Through IT. TDIT 2014, IFIP advances in information and communication technology. 429. (pp. 62–82). Berlin, Heidelberg: Springer http://dx.doi.org/10.1007/ 978-3-662-43459-8_5
25. Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014b). Data-driven innovation through open government data. Journal of Theoretical and Applied Electronic Commerce Research, 9(2), 100–120 http://dx.doi.org/10.4067/s0718-18762014000200008
26. Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. Government Information Quarterly, 32, 323–332.
27. Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically more significant than the other. The Annals of Mathematical Statistics, 18(1), 50-60.
28. Nachar, N. (2008) The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution, Tutorials in Quantitative Methods for Psychology, 4(1), 13-20.
29. Pati, R. K., Kumar, V., & Jain, N. (2015). Analysis of Aadhaar: A Project Management Perspective. IIM Kozhikode Society & Management Review, 4(2), 124-135.
30. Richter, H., & Slowinski, P. R. (2019). The data sharing economy: on the emergence of new intermediaries. IIC-International Review of Intellectual Property and Competition Law, 50(1), 4-29
31. Richter, H., & Slowinski, P. R. (2019). The data sharing economy: on the emergence of new intermediaries. IIC-International Review of Intellectual Property and Competition Law, 50(1), 4-29
32. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3-4), 591-611.
33. Smith, F., Gerry, W., & Truswell, E. (2015). Supporting sustainable development with open data. Retrieved from https://opendatacharter.net/resource/supportingsustainable-development-open-data/.
34. Soh, C., M. L. Markus and K. H. Goh (2006), 'Electronic marketplaces and price transparency: strategy, information technology, and success,' MIS Quarterly, 30(3), 705–723.
35. Verhulst, S. G., & Young, A. (2017). Open data in developing economies: Toward building an Evidence is based on what works and how. (PB - African Minds)
36. Vörösmarty, G., & Dobos, I. (2020). Green purchasing frameworks considering firm size: a multicollinearity analysis using variance inflation factor. Supply Chain Forum: An International Journal, 21(4), 290–301.
37. Wang, F. (2018). Understanding the dynamic mechanism of interagency government data sharing. Government Information Quarterly, 35(4), 536-546.
38. Weber, K., Otto, B., & Osterle, H. (2009). One size does not fit all — A contingency approach to data governance. ACM Journal of Data and Information Quality, 1(1), 4:1–4:27.
39. Wenjun, Zheng., Dejian, Lai., K., Lance, Gould. (2021). 3. A simulation study of a class of nonparametric test statistics: a close look at empirical distribution function-based tests. Communications in Statistics - Simulation and Computation
40. Williamson, K., Kennan, M.A., Johanson, G., & Weckert, J. (2016). Data sharing for the advancement of science: Overcoming barriers for citizen scientists. Journal of the Association for Information Science and Technology, 67(10), 2392–2403.