Postprint on Mining the Relationship Between Government Open Data Status and Stakeholder Behavioral Status Under the Data-Driven Paradigm

Authors: Chen Ling, Yaoqing Duan

Date: 2023-04-01T16:15:47+00:00

Abstract

[Purpose/Significance] Under the data-driven paradigm, this study reveals the internal relationships between the open data status of government portal websites and the behavioral status of relevant subjects, thereby advancing the effectiveness and progress of government data openness. [Method/Process] Employing a web crawler method to collect open datasets from the Shanghai Municipal Government Data Portal, this study conducts correlation analysis on various dataset indicators, utilizes Stepwise regression to explore regression relationships and screen variables with higher correlation degrees, and further performs PLS regression testing on variables with significant relationships to derive the internal relationships between government open data status and subject behavioral status. [Results/Conclusions] In the process of government data openness, the subject behavior of government agencies exerts a greater influence on public subject behavior than the objective characteristics of the data itself. Among factors affecting public ratings, the government openness confidentiality level demonstrates the largest impact factor with a significant negative effect, while government update frequency, government initial opening time, and machine readability of data formats exhibit significant positive effects on public ratings.

Full Text

Preamble

Relation Mining Between Government Open Data State and Subject Behavior State in the Data-Driven Paradigm

Chen Ling[1], Duan Yaoqing[1],[2]

[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]Hubei Data Governance and Intelligent Decision Research Center, Wuhan 430079

Abstract:
[Purpose/Significance] In the context of the data-driven paradigm, this study reveals the internal relationship between the open data state of government portals and the behavior state of their main subjects, thereby promoting the effectiveness and progress of government data opening. [Method/Process] Using web crawler methods to capture various open datasets from the Shanghai municipal government data portal, we first conducted correlation analysis on dataset indicators, then employed Stepwise regression to explore their relationships and screen out highly correlated variables. Further PLS regression testing on significantly related variables yielded the internal associations between government open data state and subject behavior state. [Result/Conclusion] In the process of government data opening, government department subject behavior has greater influence on public subject behavior than the objective characteristics of the data itself. Among factors affecting public ratings, government openness and secrecy level shows the largest influence factor with a significant negative impact, while government update frequency, first opening time, and data format machine readability demonstrate significant positive impacts on public ratings.

Keywords: data-driven paradigm; government open data; government portal; data openness state
Classification Number: G203
DOI: 10.13266/j.issn.0252-3116.2020.02.002

Introduction

As the data world increasingly reflects the real world, many traditional management and decision-making processes are becoming data analysis-driven management and decision-making. For a long time, management research has maintained the model-driven paradigm as the mainstream in the field. However, under the big data background, new challenges are emerging that continuously highlight the advantages of the data-driven paradigm. The data-driven paradigm refers to the fourth paradigm of scientific research—a scientific method transformation from traditional hypothesis-driven to exploration based on scientific data, specifically for data-intensive science [1]. This includes directly discovering relationship patterns among specific variables to form problem solutions, and supplementing and expanding the model-driven paradigm to form an integrated paradigm [2]. One important class of relationship patterns discovered by the data-driven paradigm is associations and their extended forms, such as quantitative associations, temporal associations, and pattern associations, which have been widely applied to many fields. Many management

decision-making contexts require not only associations but also causality, which has promoted the data-driven paradigm and its applications to a certain extent.

In the data-driven paradigm context, government departments, as holders of 80% of society's data resources [3], receive attention from all sectors regarding their data opening effectiveness. Researching the relationship between government open data state and subject behavior state helps promote the construction of open government and enables the public to better utilize government data resources, while also helping enterprises create more economic and social value. Current research hotspots in the government open data field can be divided into three categories: Data management-focused research, including data security, data sharing, linked data, and data mining; Data opening-focused research, including open government, open portals, open models, and open systems; Data service and application-focused research, including smart cities, public services, climate change, and public health. Among these, most literature focuses on the current status of government open data and open platform construction issues. In status research, some literature focuses on countries with high government data openness abroad, analyzing their models, legal policies, and promotion mechanisms in countries like Canada [3], the United States [4], and the United Kingdom [5], combining these with China's specific national conditions to learn from successful experiences. Other literature focuses on case analyses of local government data opening platforms in China, comparing and analyzing local government portals in Beijing [6], Shanghai [7], and Wuhan [8] that have achieved better opening results, identifying problems and solutions to promote the establishment of a unified government data opening platform in China. In platform construction research, methods such as DEA data envelopment analysis [9] and BP neural networks [10] have been used to evaluate and analyze the opening effectiveness and construction level of government data websites from perspectives including data volume, data content, connectivity rate, download volume, access volume, data availability, data timeliness, and data comprehensiveness [11-12].

Through review, we find that existing literature either adopts a qualitative analysis perspective to classify the logical attributes of opening indicators and establish standards, or adopts a quantitative analysis perspective to study single-attribute relationships among opening indicators. The innovation of this paper lies in quantitatively analyzing different attribute states of government dataset indicators on the basis of logical attribute classification, studying the correlation relationships among indicators, and adopting a combination of qualitative and quantitative methods.

Government open data state refers to the indicator content of each dataset published on open data portals, which can be divided into data state and subject behavior state according to logical attributes [13-14]. Data state refers to the attribute characteristic values of the data itself. Subject behavior state can be further divided into government behavior state and public behavior state according to different subject attributes [15-16]; government behavior state refers

to the behavioral attribute values of government subjects, while public behavior state refers to the behavioral attribute values of public subjects.

This study's relationship mining refers to measuring the internal relationships among dataset indicator contents under different states based on logical attribute classification for datasets published on open data portals [17-19].

The research first uses association mining methods under the data-driven paradigm to analyze correlation relationships among different logical attribute states of government data opening, then explores their regression relationships to reduce variable space and combination scale. Based on relationship detection results, we construct regression paths between government data opening states, test regression relationships among variables, and reveal internal associations among data state, government subject behavior state, and public subject behavior state.

---

2. Research Sample and Indicator Selection

2.1 Research Sample

Considering that China currently has no unified national government data opening portal, and local government data opening websites often adopt specialized data website forms [20] with relatively uniform webpage content, this study limits the research scope to local government open data websites. Through public reports and search engines, we identified regions with online open data platforms, comprehensively considering local open data maturity levels, administrative levels, and geographical distribution [21], while also considering whether they provide indicators such as page views, downloads, and ratings for published datasets, as well as the number of open datasets [22]. Based on these conditions, this paper selected the Shanghai municipal government open data website as the research sample to study and analyze relationships among government portal data opening states.

2.2 Research Variables and Indicator Selection

Reasonable indicator selection is an important prerequisite for mining and analyzing relationships among government portal data opening states. The open state categories and research indicators constructed in this paper are based on the actual operation of the Shanghai municipal government open data website, as shown in Table 1 . Among them, data state includes data format diversity and data format machine readability; government behavior state includes government openness and secrecy level, government update frequency, and government first opening time; public behavior state includes public browsing, public downloads, and public ratings.

Table 1. Measurement Indicators of Government Open Data State

State Category	Indicator	Measurement Standard	References
Data State	Data Format Diversity	Number of format types	Literature [4][7][21]
	Data Format Machine Readability	Berners-Lee's 5-star evaluation level	Literature [7][11][20]
Government Behavior State	Government Openness and Secrecy Level	Data openness attributes: having data opening authorization agreement, free access, free utilization	Literature [5][17-20]
	Government Update Frequency	Data update frequency: yearly, monthly, weekly, daily, etc.	Literature [6-8][10][20]
	Government First Opening Time	Government data's first release date, initial publication time	Literature [7][22][23]
Public Behavior State	Public Browsing	Statistics of page views for government open data	Literature [9][12]
	Public Downloads	Statistics of download volume for government open data	Literature [9][12]
	Public Rating	Public rating of government open data	Literature [10]

2.3 Data Collection

For different classification themes on the website, this study primarily used crawler software combined with manual observation to capture the above indicator data. Data collection was conducted until 10:00 on December 31, 2018, collecting a total of 1,233 datasets, all from the Shanghai municipal government open data portal, ensuring authenticity and reliability. The number of datasets for each theme category is shown in Table 2 .

Table 2. Statistics of Datasets by Theme

Theme Category	Number of Datasets
Urban Construction	
Road Traffic	
Public Safety	

Theme Category	Number of Datasets
Institutional Groups	
Education & Technology	
Economic Construction	
Livelihood Services	
Social Development	
Health & Wellness	
Culture & Leisure	
Credit Services	
Resource Environment	
Total	**1,233**

3. Data Processing and Testing

3.1 Data Cleaning and Numericalization

This study examines relationships among government portal data opening states. During data cleaning, we removed 605 datasets lacking data format, openness attributes, update frequency, first release date, page views, download volume, or ratings, resulting in 628 valid datasets.

The numericalization process for each dataset's indicators is shown in Table 3 . Each dataset includes eight indicators: format diversity, format machine readability, openness attributes, update frequency, first release date, page views, download volume, and rating level. We numericalized them according to their attributes and characteristic values [23-25].

Openness attributes include general openness and specific openness, numericalized as 1 and 2 respectively. Update frequency includes one-time, every five/ten years, yearly, semi-annually, quarterly, monthly, biweekly, weekly, on-demand, and real-time/immediate, numericalized as 1-10 respectively. First release date by year includes 2012-2018, numericalized as the year difference from 2018 (1-7 respectively). Data formats include PDF, RAR, ZIP, XLS/XLSX, DOC/DOCX, XML, CSV, etc. Format machine readability was numericalized as 1-3 according to Berners-Lee's 5-star evaluation level. Format diversity was numericalized according to the number of format types. Page views and download volume are numeric data, representing cumulative values. Rating levels include to , numericalized as 1-5 respectively.

Table 3. Numericalization of Government Open Data Indicators

Indicator Content	Numericalization
Openness Attributes	General Openness → 1, Specific Openness → 2

Indicator Content	Numericalization
Update Frequency	One-time → 1, Every 5/10 years → 2, Yearly → 3, Semi-annually → 4, Quarterly → 5, Monthly → 6, Biweekly → 7, Weekly → 8, On-demand → 9, Real-time/Immediate → 10
First Release Date	2018 → 1, 2017 → 2, 2016 → 3, 2015 → 4, 2014 → 5, 2013 → 6, 2012 → 7
Format Machine Readability	PDF, RAR, ZIP → 1; XLS/XLSX, DOC/DOCX → 2; XML, CSV → 3
Format Diversity	Number of format types
Page Views	Numeric value
Download Volume	Numeric value
Rating Level	→ 1, → 2, → 3, → 4, → 5

3.2 Descriptive Statistics and Standardization

This study conducted structural variable statistics on the sample data, obtaining minimum values, maximum values, means, standard deviations, and other statistics, as shown in Table 4 . The standard deviations of rating level, openness attributes, update frequency, first release date, format machine readability, and format diversity are all less than 2, indicating relatively small differences and dispersion among variables. However, the standard deviations of page views and download volume are 3,436.144 and 2,751.472 respectively, because the first six variables are scale-level data while page views and download volume are numeric data.

Accordingly, we performed Z-score standardization on the sample data, subtracting each variable's mean from each value and dividing by the standard deviation. This standardization eliminates dimensional and magnitude effects, removing data unit restrictions and converting them into dimensionless pure values. The conversion function is:

$$X^* = (x - \mu)/\sigma$$

where X^* is the standardized variable value, x is the actual observed value, μ is the mean of all sample data, and σ is the standard deviation of all sample data. Partial standardization results are shown in Table 5 .

Table 4. Descriptive Statistics of Government Open Data

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Public Rating	628	1	5	3.324	1.265
Openness Attributes	628	1	2	1.509	0.500
Update Frequency	628	1	10	5.454	2.175
First Release Date	628	1	7	2.287	1.355
Format Machine Readability	628	1	3	1.659	0.659
Format Diversity	628	1	5	1.175	0.168
Page Views	628	0	50,911	2,659.09	3,436.144
Download Volume	628	0	40,287	700.21	2,751.472

Table 5. Z-Standardization Results of Government Open Data (Partial)

Z_{Public Rating}	Z_{Public Browsing}	Z_{Public Download}	Z_{Government Openness}	Z_{Government Update}	Z_{Government First Open}	Z_{Data Format Readability}	Z_{Data Format Diversity}
-3.00917	-0.52038	-0.17344	-0.48063	-1.68386	-1.94299	-2.07729	0.26013
-3.00917	-0.56025	-0.19597	-0.48063	-1.68386	-1.94299	-2.07729	0.26013
-3.00917	-0.50292	-0.20978	2.07729	-1.68386	-1.15271	-2.07729	0.26013

3.3 CMV Test

To address common method bias issues, this study conducted analysis and control after database entry and data organization. Using Harman's single-factor test, the results are shown in Table 6 . The factor analysis yielded 8 factors with eigenvalues greater than 1, with the first factor totaling 2.613, accounting for 32.668% of variance, below the critical value of 40%. Therefore, common method bias has minimal impact on this study.

Table 6. Total Variance Explained for Government Open Data

Component	Initial Eigenvalues	Extraction Sums of Squared Loadings
	Total	% of Variance
1	2.613	32.668
2	1.958	24.480
3	1.349	16.861
4	0.779	9.729
5	0.622	7.780
6	0.393	4.910

Component	Initial Eigenvalues	Extraction Sums of Squared Loadings
7	0.271	3.383
8	0.015	0.188

4. Exploration of Relationships Between Government Data Openness States

4.1 Correlation Analysis Between Government Data Openness States

After data cleaning, numericalization, standardization, and CMV testing, we conducted correlation analysis on research variables and measurement indicators, with results shown in Table 7 . Public rating shows high correlation with data state and government behavior state variables, but not significant correlation with public browsing and public downloads. Additionally, public browsing and public downloads have significant correlation with each other, but not with other research variables.

Table 7. Correlation Analysis Between Government Data Openness States

	Government Openness & Secrecy	Government Update Frequency	Government First Open Time	Data Format Readability	Data Format Diversity	Public Browsing	Public Download	Public Rating
Government Openness & Secrecy	1							
Government Update Frequency	-0.525**	1						
Government First Open Time	0.509**	-0.063	1					
Data Format Readability	0.343**	0.108**	-0.005	1				

	Government Openness & Secrecy	Government Update Frequency	Government First Open Time	Data Format Readability	Data Format Diversity	Public Browsing	Public Download	Public Rating
Data Format Diversity	0.454**	-0.052	-0.287**	0.175**	1			
Public Browsing	-0.355**	-0.228**	-0.287**	-0.012	0.190**	1		
Public Download	-0.228**	-0.355**	-0.228**	0.190**	0.410**	0.659**	1	
Public Rating	0.175**	-0.005	0.178**	0.168**	0.410**	0.659**	0.659**	1

Note: ** Correlation is significant at the 0.01 level

4.2 Regression Analysis Between Government Data Openness States

We used Stepwise regression to further explore regression relationships between public behavior state and data state/government behavior state variables. To avoid potential multicollinearity issues in regression models, we centered all variables. First, we established univariate regression models between the dependent variable Y and each of the p regression independent variables X_1, X_2, …, X : $Y = \beta_0 + \beta X$ (1), i = 1, …, p. We calculated the F-test statistic values for regression coefficients of variable X , denoted as $F^{(1)}$, …, $F^{(1)}$, and selected the maximum value $F_1^{(1)} = \max\{F_1^{(1)}, \ldots, F^{(1)}\}$. For a given significance level α with corresponding critical value $F^{(1)}$, if $F_1^{(1)} \geq F^{(1)}$, variable X_1 was introduced into the regression model, with I_1 recorded as the selected variable indicator set. Next, we established binary regression models between dependent variable Y and independent variable subsets $\{X_1, X_1\}$, …, $\{X_1, X_{1\,1}\}$, $\{X_1, X_{1\,1}\}$, …, $\{X_1, X\}$, totaling p-1 models. We calculated F-test statistic values for variable regression coefficients, denoted as $F^{(2)}$ $(k \neq I_1)$, selected the maximum $F_2^{(2)} = \max\{F_1^{(2)}, \ldots, F_{1\,1}^{(2)}, F_{1\,1}^{(2)}, \ldots, F^{(2)}\}$, with corresponding independent variable subscript recorded as i_2. For a given significance level α with corresponding critical value $F^{(2)}$, if $F_2^{(1)} \geq F^{(2)}$, variable X_2 was introduced into the regression model; otherwise, the variable introduction process terminated. Finally, we considered regression of dependent variable on variable subset $\{X_1, X_2, X\}$, repeating the above steps by selecting one variable each time from those not yet introduced into the regression model until no variables could be introduced.

Stepwise regression results show that public rating has significant regression relationships with government openness & secrecy level, government update frequency, data format readability, and government first open time. There are 4 hypothetical models: Model 1 includes only government openness & secrecy level; Model 2 adds government update frequency based on Model 1; Model 3 adds data format readability based on Model 2; Model 4 adds government first open time based on Model 3.

As shown in Table 9 and Table 10 , the Sig coefficients of all 4 hypothetical models are below 0.001, indicating that government openness & secrecy level, government update frequency, data format readability, and government first open time all have significant regression effects on public rating. However, Model 4's R-square and adjusted R-square values are the largest among the four models, indicating the best explanatory effect for public rating.

From the collinearity statistics in Table 11 , all variables have tolerance values above 0.1 and VIF values far below 10, indicating no multicollinearity issues in the regression. Further examining the standardized and unstandardized coefficients in Table 11 shows that government openness & secrecy level has a significant negative impact on public rating, while government update frequency, data format readability, and government first open time have significant positive impacts.

Table 8. Variables Entered/Removed

Model	Variables Entered	Variables Removed	Adjusted R Square
1	Government Openness & Secrecy Level		.525
2	Government Update Frequency		.644
3	Data Format Readability		.681
4	Government First Open Time		.691

. Dependent variable: Public Rating
. Predictors: (Constant), Government Openness & Secrecy Level
. Predictors: (Constant), Government Openness & Secrecy Level, Government Update Frequency
. Predictors: (Constant), Government Openness & Secrecy Level, Government Update Frequency, Data Format Readability
. Predictors: (Constant), Government Openness & Secrecy Level, Government Update Frequency, Data Format Readability, Government First Open Time

Table 9. Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics	Durbin-Watson
					R Square Change	F Change
1	.725	.525	.525	238.044	.525	172.739
2	.803	.644	.644	149.459	.119	130.202
3	.826	.681	.681	56.732	.037	96.986
4	.832	.691	.691	16.379	.010	74.891

Table 10. ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	172.739	1	172.739	627.000
	Residual	454.261	627	.724	
2	Regression	260.404	2	130.202	627.000
	Residual	366.596	626	.586	
3	Regression	290.957	3	96.986	627.000
	Residual	336.043	625	.538	
4	Regression	299.565	4	74.891	627.000
	Residual	327.435	624	.525	

Table 11. Coefficients

Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	Collinearity Statistics
	B	Std. Error	Beta		Tolerance
1	(Constant)	1.023E-013	.042		.000
	Government Openness & Secrecy Level	-.343	.022	-.525	-15.429
2	(Constant)	1.024E-013	.037		.000
	Government Openness & Secrecy Level	-.384	.031	-.588	-12.441
	Government Update Frequency	.289	.023	.509	12.225
3	(Constant)	1.023E-013	.036		.000
	Government Openness & Secrecy Level	-.413	.039	-.632	-10.712
	Government Update Frequency	.235	.022	.413	10.083

Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	Collinearity Statistics
	Data Format	.126	.013	.343	9.073
	Readability				
4	(Constant)	1.027E-013	.036		.000
	Government	-.413	.032	-	-12.935
	Openness &			.632	
	Secrecy Level				
	Government	.236	.018	.416	12.884
	Update Frequency				
	Data Format	.127	.010	.345	12.225
	Readability				
	Government First	.234	.018	.290	13.350
	Open Time				

. Dependent variable: Public Rating

5. Testing Relationships Between Government Data Openness States

5.1 PLS Regression Path Construction

We conducted PLS regression testing on regression paths between government portal data opening states. In PLS testing, relationships between research variables and observed indicators can typically be expressed through three matrix equations:

$= \mathrm{B} + \Gamma +$
$\mathrm{X} = \Lambda + \delta$
$\mathrm{Y} = \Lambda +$

where Λ represents the relationship between exogenous observed variables and exogenous latent variables (the factor loading matrix of exogenous observed variables on exogenous latent variables); Λ represents the relationship between endogenous observed variables and endogenous latent variables (the factor loading matrix of endogenous observed variables on endogenous latent variables); B represents path coefficients indicating relationships between endogenous latent variables; Γ represents path coefficients indicating effects of exogenous latent variables on endogenous latent variables; and represents residual terms of structural equations reflecting unexplained portions.

PLS results show that the explanatory power of government behavior state and data state on public browsing is only 0.015, on public downloads only 0.007, and on public rating 0.478. The overall explanatory power of government behavior state and data state on public behavior state is 0.470, less than 0.478. Therefore,

this study only presents further results for public rating PLS testing. During testing, data format diversity's significance T-value for public rating was only 1.455, so it was removed. The final path model is shown in Figure 1 [Figure 1: see original paper].

We conducted reliability and validity tests on the regression path model, with overall results shown in Table 12 . As sample data are objective data crawled from government websites, all variables' Cronbach's Alpha coefficients and C.R. coefficients are above 0.7, indicating good internal consistency and stability. AVE values are all above 0.5, indicating good convergent validity and accuracy. Additionally, public rating's R-square is 0.478, indicating high explanatory power.

Table 12. Overall Test Results

	Composite Reliability	Cronbach's Alpha	Communality	Redun
Government Openness & Secrecy Level	1.000000	1.000000	1.000000	0.2425
Government Update Frequency	1.000000	1.000000	1.000000	0.4777
Government First Open Time	1.000000	1.000000	1.000000	0.0000
Data Format Readability	1.000000	1.000000	1.000000	0.0000
Public Rating	1.000000	1.000000	1.000000	0.0000

Figure 1. Regression Path Model

5.2 PLS Regression Path Significance Testing

Regression path significance testing results are shown in Table 13 . The T-values for the four regression paths from government first open time, government openness & secrecy level, government update frequency, and data format readability to public rating are all greater than 1.96, indicating significant effects on public rating. This is consistent with previous stepwise regression results.

Table 13. Regression Path Test Results

Path	Original Sample	Sample Mean	Standard Deviation (STDEV)	Standard Error (STERR)	T Statistics (
Government Update Frequency → Public Rating	0.289697	0.289624	0.028723	0.028723	10.083308
Government First Open Time → Public Rating	0.234257	0.235662	0.028253	0.028253	8.341236
Government Openness & Secrecy Level → Public Rating	-0.343027	-0.343342	0.035436	0.035436	9.689101
Data Format Readability → Public Rating	0.126839	0.126215	0.024756	0.024756	5.098408

5.3 PLS Regression Path Coefficient Testing

After significance testing, we further examined path coefficients in the model, with results shown in Figure 2 [Figure 2: see original paper] and Table 14 . Public rating's R-square is 0.478, indicating good model fit. Path coefficients for Government First Open Time → Public Rating, Government Openness & Secrecy Level → Public Rating, Government Update Frequency → Public Rating, and Data Format Readability → Public Rating are 0.126, -0.343, 0.290, and 0.236 respectively. In summary: Government openness & secrecy level has the greatest impact on public rating with significant negative effect; government first open time, government update frequency, and data format readability have significant positive effects on public rating.

Figure 2. Regression Path Coefficients

Table 14. Correlation Coefficient Matrix

	Government Openness & Secrecy Level	Government Update Frequency	Government First Open Time	Data Format Readability
Government Openness & Secrecy Level	1.000000			
Government Update Frequency	-0.524881	1.000000		
Government First Open Time	0.508731	-0.286754	1.000000	
Data Format Readability	0.342839	0.190371	0.168021	1.000000

6. Conclusions and Recommendations

6.1 Summary

The innovations of this paper are: On the basis of logical attribute classification of government dataset indicators, it quantitatively analyzes different attribute states and examines correlation relationships among research indicators using a combination of qualitative and quantitative methods. Under the data-driven paradigm, it uses association mining methods to analyze correlation relationships among different logical attribute states of government data opening. On the basis of reducing variable space and combination scale, it tests regression

relationships among variables to reveal internal associations among data state, government subject behavior state, and public behavior state.

Through correlation analysis, stepwise regression exploration, and PLS regression testing of government data opening states, we found: Public browsing and public downloads have no significant regression relationships with other openness states and are poorly explained by government behavior state and data state. In contrast, public rating has significant regression relationships with other openness states and is well explained, hence further presentation and explanation.

Comparing and listing exploration and verification results of relationships among government data opening states, as shown in Table 15 :

Table 15. Exploration and Verification Results of Relationships Among Government Data Opening States

Government Behavior State Variables	Data State Variables	Public Behavior State Variables
Government Openness & Secrecy Level	Stepwise Regression:	PLS Test:
Government Update Frequency	Stepwise Regression:	PLS Test:
Government First Open Time	Stepwise Regression:	PLS Test:
Data Format Readability	Stepwise Regression:	PLS Test:
Data Format Diversity	Stepwise Regression:	PLS Test:

From Table 15 and above analysis:
(1) Public browsing and public downloads have significant correlation with each other but low association with other openness states, and are poorly explained by government behavior state and data state.
(2) Public rating has high association with government behavior state and data state variables. Specifically: In government data opening processes, government department subject behavior has greater influence on public rating behavior than data's objective characteristics. Among factors affecting public rating, the influence degree order is: government openness & secrecy level > government update frequency > data format readability > government first open time.

Government openness & secrecy level has the most significant negative impact on public rating. This variable's observation indicator—openness attributes—mainly includes general openness and specific openness. Due to government data confidentiality, specifically-opened data with higher confidentiality receives relatively lower public favorability. Government update frequency has significant positive impact on government open data rating. Collected data update

frequencies range from one-time to real-time/immediate; higher frequency data has better timeliness and gains more public favorability. Data format readability has significant positive impact. Structured data like XML and plain text data like CSV gain more public favorability; document and table data like XLS/XLSX and DOC/DOCX also receive good ratings; compressed formats like RAR and ZIP receive lower ratings due to inconvenience. Government first open time has significant positive impact. As this study's crawled data represent a static time point, earlier-released government data has accumulated higher public favorability.

6.2 Recommendations

Based on the above analysis, we propose the following recommendations to promote government open data effectiveness and progress:

(1) Although public browsing and downloading behaviors have low association with government behavior state and data state, they have significant strong correlation with each other. Therefore, government data opening processes should adhere to a "public-centered" approach and constantly focus on public experience [26], taking providing needed data to the public as the central goal of building government open data portals. Government open data's social impact should be expanded to make it more relevant to public life and solve problems closely related to public concerns.

(2) Public rating behavior is greatly influenced by government behavior state and data state variables. Government departments should open more high-value data to improve economic and social value. This includes: Under the premise of following confidentiality principles and respecting personal privacy, improve government data openness attributes and reduce secrecy levels to achieve true openness; Increase data update frequency, striving for real-time and timely updates to enhance data stability and timeliness; Improve open data machine readability [27] to enhance usability, making it easier for computer automatic reading and processing and facilitating user access and utilization; Open and publish various thematic data types as early as possible to expand the scope of open data.

(3) Regarding problems discovered in this study, we recommend further promoting the establishment of a national-level government data opening portal while unifying local government and departmental platform construction standards [28]. China's government data opening and utilization has not yet developed into a nationwide action, and local government departments lack relevant policy guidance in the opening process. Therefore, the state should promptly formulate and improve laws and regulations related to government data opening, establish open data standards and norms to ensure orderly government data opening work. Local governments should actively participate in open platform construction to lay foundations for a unified, integrated national data platform.

6.3 Limitations and Future Directions

China currently has no national-level government data opening portal, and local open platform construction levels and standards are not unified. Local portals vary in dataset numbers, openness attributes, and metadata indicators, making full-sample analysis and comparative analysis difficult. Considering local platform status and theoretical research value, this study only selected Shanghai's government open data website as the research sample, selecting variables and indicators based on the portal's actual operation and theoretical foundations. Future research should address full-sample analysis and comparative analysis across local levels, and further deepen and explore variable and indicator selection.

References

[1] Liu Yunong, Shi Qin. Research on Innovation of Humanities and Social Sciences Knowledge Services Under the Data-Driven Paradigm [J]. Library and Information Service, 2019(1): 24-30.
[2] Deng Zhongsheng, Li Zhifang. Evolution of Scientific Research Paradigms—The Fourth Paradigm of Scientific Research in the Big Data Era [J]. Information and Documentation Services, 2013, 34(4): 19-23.
[3] Yang Feifei. Research on Foreign Government Data Opening Portal Construction [D]. Hebei: Yanshan University, 2016.
[4] Hou Renhua, Xu Shaotong. Analysis of Management and Utilization of U.S. Government Open Data—Taking www.data.gov as an Example [J]. Library and Information Service, 2011, 55(4): 119-122.
[5] Li Chongzhao, Huang Huang. Policy and Governance Structure of UK Government Data Governance [J]. E-Government, 2019(1): 25-36.
[6] Huang Sijin, Zhang Yanhua. Problems and Countermeasures in Current Chinese Government Data Opening Platform Construction—Taking Beijing and Shanghai Government Data Opening Websites as Examples [J]. China Management Informationization, 2015(14): 175-177.
[7] Gu Tiejun, Xia Yuan, Xu Kewei. Research on Sustainable Development of Shanghai Government's Transition from Information Disclosure to Data Opening—Based on Practical Investigation of 49 Government Department Websites and Shanghai Government Data Service Network [J]. E-Government, 2015(9): 14-21.
[8] Chen Tao, Li Mingyang. Research on Data Opening Platform Construction Strategy—Taking Wuhan Government Data Opening Platform Construction as an Example [J]. E-Government, 2015(7): 46-52.
[9] Ma Haiqun, Wang Jin. Efficiency Evaluation of Government Open Data Websites Based on DEA [J]. Digital Library Forum, 2016(6): 2-7.
[10] Zou Chunlong, Ma Haiqun. Research on Government Open Data Website Evaluation Based on Neural Networks—Taking 20 U.S. Government Open Data Websites as Examples [J]. Modern Information, 2016, 36(9): 16-21.

[11] Zheng Lei, Xiong Jiuyang. Research on Chinese Local Government Open Data: Technical and Legal Characteristics [J]. Public Administration Review, 2017, 10(1): 53-73.
[12] Duan Yaoqing, Qiu Xueting, He Siqi. Analysis of Current Utilization Status of Urban Government Open Data in China from Thematic and Regional Perspectives [J]. Library and Information Service, 2018, 62(20): 65-76.
[13] Yang Yongqing, Zhang Jinlong, Man Qingshan, et al. Research on Mobile Internet User Adoption—Based on Perceived Benefits, Costs, and Risks Perspective [J]. Journal of Intelligence, 2012, 31(1): 200-206.
[14] Liu Wenqi. Data Quality Control Model System and Empirical Study for Chinese Public Databases [J]. Scientia Sinica (Informationis), 2014, 44(7): 836-856.
[15] Li Mingshuai, Guan Hua. Analysis of Government Microblog Attention Based on Content Classification—Taking Sichuan Provincial Government Microblogs as Examples [J]. Information Research, 2014(12): 12-15.
[16] Zhao Mianmian, Huang Jiting. Research on Multivariate Nonlinear Regression Model of Chinese Urban Residents' Consumption Expenditure [J]. Mathematics in Practice and Theory, 2011, 41(10): 20-25.
[17] Sola R, Daniels S F, Lopez R, et al. A Model to Guide the Open Government Data Implementation in Public Agencies [J]. Journal of Universal Computer Science, 2014, 20(11): 1564-1582.
[18] Zeleti F A, Ojo A, Curry E. Exploring the Economic Value of Open Government Data [J]. Government Information Quarterly, 2016, 33(3): 535-551.
[19] Charalabidis Y, Alexopoulos C, Loukis E. A Taxonomy of Open Government Data Research Areas and Topics [J]. Journal of Organizational Computing & Electronic Commerce, 2016, 26(1): 41-63.
[20] Cao Yujia. Survival Status of Government Open Data: Investigation Report from 19 Local Governments in China [J]. Library and Information Service, 2016, 60(14): 94-101.
[21] Zhao Rongying, Liang Zhisen, Duan Peipei. Metadata Standards for UK Government Data Opening and Sharing—Investigation and Enlightenment of Data.gov.uk [J]. Library and Information Service, 2016, 60(18): 1-9.
[22] Jing D, Wen L. Study on the Government Strategy of Transformation from Information Publishing to Data Opening [M]. USA: 2017 IEEE 2nd International Conference on Big Data Analysis, 2017.
[23] Thohari A H, Suhardi S. Requirement Engineering for Open Government Information Network Development to Support Digital Startup in Cimahi City Indonesia [C]//Proceedings of International Conference on Information Technology Systems and Innovation. Piscataway: IEEE, 2016.
[24] Huang Ruhua, Wen Fangfang, Huang Wen. Construction of Chinese Government Data Opening and Sharing Policy System [J]. Library and Information Service, 2018, 62(9): 5-13.
[25] Sun Lu, Li Guangjian. Research on Construction of Government Open Data Application Analysis Model [J]. Library and Information Service, 2017, 61(3): 97-108.
[26] Zhou Wenhong. Characteristics and Enlightenment of New Zealand

Government Data Opening [J]. Library and Information Service, 2017, 61(23): 76-82.
[27] Wei Xinling, An Xiaomi, Li Xuemei, et al. Review of Open Government Data Evaluation Systems: Characteristic Analysis [J]. Library and Information Service, 2017, 61(18): 119-127.
[28] Xia Yizheng. On Government Data Opening Risks and Risk Management [J]. Journal of the China Society for Scientific and Technical Information, 2017(1): 22-31.

Author Contributions

Chen Ling: Responsible for outline formulation, data collection, data analysis, and initial draft writing.
Duan Yaoqing: Responsible for topic selection and comprehensive revision of the full paper.

Note: Figure translations are in progress. See original paper for figures.