



University of Tennessee, Knoxville
**TRACE: Tennessee Research and Creative
Exchange**

DataONE Sociocultural and Usability &
Assessment Working Groups

Communication and Information

1-2012

Exploring the motive for data publication in open data initiative: Linking intention to action.

D. S. Sayogo

T. Pardo

Follow this and additional works at: https://trace.tennessee.edu/utk_dataone



Part of the [Library and Information Science Commons](#)

Recommended Citation

Sayogo, D. S., & T. Pardo, T. (2012, January). Exploring the motive for data publication in open data initiative: Linking intention to action. Presentation at the 45th Hawaii International Conference on System Sciences, IEEE Computer Society, Maui, HI.

This Poster is brought to you for free and open access by the Communication and Information at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in DataONE Sociocultural and Usability & Assessment Working Groups by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

Exploring the Motive for Data Publication in Open Data Initiative: Linking Intention to Action

Djoko Sigit Sayogo
Center for Technology in Government
University at Albany, SUNY
dsayogo@ctg.albany.edu

Theresa A. Pardo
Center for Technology in Government
University at Albany, SUNY
tpardo@ctg.albany.edu

Abstract

This research study was designed to broaden understanding of the publishing of research datasets by distinguishing between the intention to share and the action of sharing. The data was generated from preliminary survey results conducted by DataONE work groups. The final data used in this paper is based on 587 observations. The analysis results show support for all of the path coefficients of the theoretical model except for the path of perceived self-efficacy, and legal context and policy variables. The intention to share a dataset was found to be a significant determinant in the action of sharing data. Acknowledging the key determinants of intention to publish datasets arguably entails significant policy implications on data sharing.

1. Introduction

Recent developments in the open data initiative have promoted investigation into the motives and intentions of researchers who publish datasets. The effort to use Information and Communication Technologies (ICTs) for building open data initiatives with the purpose of advancing scientific progress has received increased attention. The development of Web 2.0, which encourages peer production, interactivity, and user-generated innovation [25], has further stimulated the development of open data initiatives for the sharing and distribution of information between several participants to solve social issues. In December 2009, when the Obama Administration issued the Open Government Directive which asserts the three principles of transparency, participation, and collaboration, one focus of the directive was to promote the culture of open government, and more specifically, to encourage collaborative works involving researchers, private sectors, and civilian society [18]. Greater openness and developments in technology have enabled enthusiasts such as Lynn Henning, a family farmer, activist, and the 2010

recipient of the Goldman Prize, to share data with state regulators to encourage stronger enforcement of regulations—and violation of those regulations—regarding water pollution¹.

Promotion of the creative use of ICTs for increased access to data and to encourage further scientific progress through collaboration is not only a function related to government. The effort to develop a global scientific knowledge base through data integration and dissemination across domains and boundaries also manifests in the earth sciences. DataONE is a new collaboration initiative using ICTs to ensure preservation and access to multi-scale, multi-discipline, and multi-national science data; as such, DataONE will transcend domain boundaries. DataONE aims to connect multiple data repositories, managed and organized by different entities, public and private, regardless of size and location². In support of the open access initiative, various policies have been enacted, such as the Berlin Declaration on Access to Knowledge in the Sciences and Humanities, the OECD Communiqué on Science, Technology and Innovation for the 21st Century [20], and the 1996 Bermuda Agreement, followed by the Fort Lauderdale Agreement in 2003 [19].

The relatively recent developments in data sharing policy confront researchers with new uncertainties and raise concerns regarding methods of sharing datasets with the public. Scientists face several challenges regarding the open publication of their datasets. These challenges make necessary the investigation of researchers' motivation for publication of datasets, and their intentions in doing so, in a collaborative network environment. Existing literature has discussed at length the challenges of data publication in the open access initiative [20, 27, 28, 29]. And furthermore, a number of studies have focused on the role of the researcher's motivation and intentions for data publication [20, 27]. However, the matter of how challenges affect the

¹ <http://www.goldmanprize.org/2010/northamerica..>

² DataONE (2011) Retrieved from <http://www.dataone.org>

researcher's motivation to publish their datasets, and whether the researcher's intention to publish correlates with the action of publishing datasets, has received little systematic attention. This research was designed to contribute greater understanding of the nature and function of behavior in publishing research datasets by 1) distinguishing between two principals of sharing behavior – intention to share and action of sharing; and 2) identifying factors which influence the intention to share and action of sharing research datasets.

The rest of the paper is organized as follows. Section 2 will outline the theoretical background that leads to the proposed model linking intention to the action of sharing datasets. Section 3 briefly explains the research design and methodology used in this study. Section 4 presents findings and discussion on those findings and finally, section 5 provides discussion and concluding remarks.

2. The Theoretical Foundation

2.1. Perceived Behavioral Control, Intention and Action

The predominant benefits of research dataset sharing are the ability to enrich and promote the progress of scientific research, and to generate knowledge [19]. By innovating from archival datasets, researchers are encouraged to produce new knowledge, promote advanced discoveries from old datasets, and rethink the meaning of archival datasets to be expanded upon through modern thought [10]. Arguably, sharing datasets through the reconstruction and/or a combination of various existing datasets represents the essence and basis for the generation of knowledge. It is safe then, to assume that the behavior surrounding the sharing of research datasets is similar to knowledge sharing behavior. Gagne (2009) points out the similarities between knowledge sharing behavior and voluntary sharing behavior, and argues for the use of motivation theory to study knowledge sharing behavior [16]. Interestingly, information sharing has been defined as “the voluntary act of making information available to others” [14, p.87].

Building on the above-mentioned study, this study assumes that data sharing behavior will also share similarities with voluntary sharing behavior, and therefore, certain arguments from motivation theory have been incorporated here. This research study bases the theoretical framework on the theory of planned behavior, particularly focusing on the importance of perceived behavioral control (PBC) to explain intention to and action of sharing data in a collaborative network. Azjen (1991) argues that

individual intention is significant to the successful carrying out of an action or behavior to share [2, 21]. Intention represents an indication of the level of effort individuals are willing to exert. The theory of planned behavior posits that behavior is a direct function of attitude, subjective norms, and perceived behavioral control [2]. Each of the three factors in the theory of planned behavior is mutually exclusive of the others. Taylor and Todd (1995) argue for the differing roles of each factor as determinants of IT usage. The role of the subjective norm as a determinant of IT usage is considered less definite [24, 31]. On the other hand, extant literature demonstrates that PBC is an important determinant of IT usage [22, 24, 31]. PBC has been found to be a significant predictor of behavioral intention of IT usage, both directly [24] and indirectly [4]. Ajzen (1991) defines perceived behavioral control (PBC) as the perception of individuals of the ease or difficulty in carrying out a certain behavior [2]. Azjen (1991) posits that the constructs of PBC consist of two elements: internal individual notion and external resource constraints [2, 31].

Internal individual notion is compatible and shares similarities with the concept of self-efficacy by Bandura [2]. Bandura (1977) claims that people undertake activities based on assessment of their own capabilities to manage the task [3]. Higher self-efficacy is argued to significantly affect behavioral performance [2, 3]. Higher levels of self-efficacy were found to be a significant determinant of intention and usage of IT [13] and performance and effort expectancy in using collaborative technology [8]. Likewise, earth observation research is very specific in terms of its purpose, events/phenomenon, methodology, and duration [6], and it is very heterogeneous [28] and highly dependent on local context [35, 36]. Consequently, analyzing environmental data involves a great deal of human judgment [17]. Preparing data for publication is a labor intensive process [6] and researchers must invest great deals of effort before data is available for sharing [10]. Tucker's (2009) investigation into cancer research data sharing reveals that one of the subjective barriers to the sharing of data is a researcher's guilt and embarrassment regarding unorganized data [32]. These difficulties in sharing earth observation datasets arguably necessitate greater self-efficacy from researchers in order for their datasets to be shared publicly.

The second element to PBC consists of external resource constraints or facilitating conditions. Information technology usage literature classifies facilitating conditions into two categories, namely: economic resources and technology related factors, such as compatibility [31]. The literature argues that existence of these facilitating conditions may or may

not support behavioral intention and action, while the absence of facilitating conditions represents a barrier for intention and action [2, 31]. Azjen (1991) hypothesizes that facilitating conditions are in the form of time and money [2], and Taylor & Todd (1995) add technological compatibility, specifically to predict intention behavior and usage of information technology [31]. This paper argues that behavioral intention, usage, and behavioral control are also affected by institutional contexts and barriers for sharing data in collaborative networks in terms of organizational, technological, and policy aspects [29]. These barriers affect the likelihood that researchers will share their data [30]. Tucker (2009) found that privacy policy was a major barrier for sharing data from the research center for cancer [32]. In this regard, this study will incorporate technology, organization, and policy as determinants that affect the relationship between behavioral control, intention, and usage/action to predict the data sharing behavior in collaborative networks such as DataONE.

Technology is found to be a significant predictor which affects the likelihood of researchers sharing their data [30]. Ecological and earth observation datasets are dispersed, heterogeneous, and context-dependent [28]. Data was collected from various locations, habitats, and ecosystems. Zimmerman (2007, 2008) argues that secondary use of ecological data will always be confronted with the problem of data complexity [35, 36]. Researchers need to have adequate data management skills to prepare their data for publication. In addition, there are various metadata concepts that researchers could use to describe their data in the earth sciences and ecology, such as Dublin Core, Directory Interchange Format, Ecological Metadata Language, etc. Metadata is generally defined as information which describes data, comprising information necessary to understand the data [26]. Different levels of metadata are necessary as support for various functions of data sharing [26]. The existence of various metadata concepts could create enormous logistical challenges in encouraging data publication in the open data initiative [26]. An increase in the user-friendliness of metadata management will improve compatibility and reduce researcher opposition to publishing research datasets [26].

Organizational support for managing data for open publication is found to be a significant determinant of the researchers' likelihood to share data [30]. A supportive organizational climate is found to be a positive predictor of knowledge sharing [5]. Bock et al (2005) found organizational support to be a significant predictor of the knowledge sharing behavior [5]. However, organizational support is found to be stronger in its indirect rather than direct influence [5].

Similarly, Lin (2006) found that the influence of organizational support on researchers' intentions of sharing knowledge is mediated by individually subjective predictors such as perceptions and trust [23]. Additionally, reward and attribution were also regarded as significant predictors of sharing behavior. Perceived intrinsic and/or extrinsic rewards show a significant correlation to the action and intention of sharing knowledge [5, 11]. A number of studies have argued for the importance of attribution and acknowledgement of sharing datasets [1, 19, 32, 34]. Attribution and acknowledgement of the researcher responsible for datasets will likely result in higher instances of openly sharing research datasets with the public [32]. In contrast, lack of sufficient reward and recognition could inhibit the intention and action of sharing [5, 12].

Extant literature asserts the importance of legal and policy factors in sharing information/data/knowledge [7]. However, the exact role of regulatory factors as determinants of data sharing is somewhat unclear. Tucker (2009) regards legal and policy factors as the most complex and misunderstood area of data sharing [32]. On one hand, regulation and policy could enhance data sharing by ensuring proper and accountable use of data and information [34], and lack of regulations and policies does not guarantee a neutral sharing environment [34] or improved flexibility in sharing datasets [10]. On the other hand, rigidity of policies and regulations could inhibit the development of the data sharing initiative. For instance, complex privacy concerns could lead to discouragement about sharing data [32]. Accordingly, regulations and policy are considered the greatest obstacles in the creation of a knowledge network [15]. Unresolved legal issues can deter or restrain the development of collaborative data sharing, even if scientists are prepared to proceed [7].

2.2. Research Model and Hypotheses

Drawing from the literature review, this study identifies the following factors which influence the behavior of researchers in sharing research datasets openly: perceived self-efficacy, intention to share datasets, facilitating conditions in terms of technology, policy, organizational and local contexts. In light of those factors, this study generates a path model illustrated in Figure 1. This path model is unique compared to typical linear regressions. The nature of causal relationships can be direct or indirect. The combined effects of both direct and indirect causalities are of special interest in this study. For example, organizational support may directly affect the action of sharing datasets, or the action may be mediated by the individual intentions of researchers to share datasets.

A rich, cumulative body of empirical research has established the influence of perceived behavioral control on behavioral intention and action, yet no one has applied the concept to data sharing in collaborative networks. Perceived behavioral control could influence the actions of an individual directly, or it could be mediated through behavioral intention [2]. Perceived behavioral control consists of two elements: the individual's self-efficacy and facilitating conditions. Higher self-efficacy will significantly affect behavioral intention and action. Literature on information technology usage asserts that self-efficacy is a significant determinant of intention and IT usage [31]. Lack of facilitating conditions corresponds to difficulties with self-efficacy and behavioral intention and action [31]. As a result of specificity, heterogeneity, and data collection processes which are highly dependent on local context in earth science and ecology have resulted in barriers which obstruct or complicate the open sharing of datasets. These barriers have been classified into four categories, namely: technological, organizational, policy based, and local context [29]. These four categories of barriers represent facilitating conditions that might affect the behavior of researchers in regard to openly sharing their datasets. This study hypothesizes four causal relationships between key determinants derived from existing literature, corresponding to the research question addressed in the introduction.

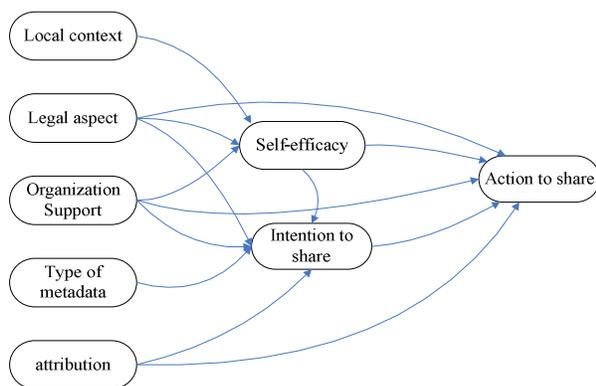


Figure 1. The Structural Model of Motivation to Data Sharing

Hypothesis 1. A researcher's behavioral intention to share their datasets positively influences the action of openly sharing datasets.

Hypothesis 2. A set of determinants categorized as technological, organizational, policy,

and local context significantly influence the researcher's behavior and actions of openly sharing their datasets.

Hypothesis 3. A researcher's self-efficacy positively influences behavioral intention and action to openly share datasets.

Hypothesis 4. Technological, organizational, policy, and local context determinants significantly influence the researcher's self-efficacy, behavioral intentions, and actions to openly share datasets.

3. Research Methodology

3.1. The Respondent

This paper uses the preliminary survey results conducted by the Usability and Assessment working group, and the Socio-Cultural working group. Data were collected using online surveys. The links for this baseline assessment survey were open from October 7, 2009 to July, 2010. The research sample was randomly selected from identified stakeholders. The Usability and Assessment and Socio-Cultural working group identified a set of stakeholders, comprised mainly of scientists, librarians, computer scientists, decision makers, citizen scientists, students, and teachers. 1,329 total respondents participated. After eliminating certain participant responses due to missing values, 587 final observations were used in this paper 587.

Table 1 Sample Characteristics (587 scientists)

| Characteristics | Category | % |
|----------------------------|-------------------------|----|
| Regional Distr. (region) | North America | 73 |
| | Europe | 15 |
| | Other Regions | 12 |
| Age Distr. (age) | 20 to 39 years | 37 |
| | 40 to 50 years | 30 |
| | Over 50 years | 33 |
| Status & Position (status) | Professor or lecturers | 49 |
| | Grad./Post Doc students | 19 |
| | Researchers | 21 |
| | Other occupation | 11 |

The respondents were mostly mature adults (an average of 44 years old) and well-educated (49 percent are employed as professors or lecturers; 19 percent are graduate or post-doctoral students). Respondents' distribution comprised a North American majority with 73% of the sample, 15% of participants from European regions, and the remaining participants from various other regions. For the distribution of respondents, refer to Table 1.

3.2. Variable Measurement

- a. **Action to share datasets.** This variable represents the actual behavior of sharing datasets by putting research datasets online or in research network databases. There are six items designed to measure the action of sharing datasets. These items reflect respondents' report on past actions of putting their datasets on the PI website, Organizational website, National research network database, Regional research network database, and Global research network database. These items were measured on a 4-point ordinal scale (1: has never published data, to 4: published all data). The factor analysis indicates only one factor has acceptable properties (eigenvalue greater than 1.0), and therefore we extract only one factor. Reliability estimates (Cronbach's alphas) for the action to share datasets is .77 (table 3). A conventional rule of thumb regards any value above .70 as acceptable. For the result of Cronbach's alphas please refer to Table 3.
- b. **The intention to share datasets** variable measures the intention of researchers to share their datasets with others. This variable is measured by an indication of willingness to share datasets, based on agreement with a statement such as "I would be willing to place all of my data into a central data repository with no restrictions" or "I would be willing to share data across a broad group of researchers who use data in different ways." In addition, this variable is measured by agreement with statements representing attitudes of openness to sharing such as, "I share my data with others" and "Others can access my data easily." There are 5 statements designed to construct this variable, measured by using a 5-points Likert scale (1: strongly agree, to 5: strongly disagree). Factor analysis indicates only one factor with acceptable properties (eigenvalue greater than 1), and reliability estimates (Cronbach's alphas) for this variable is .75.
- c. **Perceived self-efficacy in managing data** measures the confidence and efficacy of individual researchers regarding their data management skills in preparing their datasets for publication. A new measure to assess self-efficacy in sharing data is developed for use in this study. The measure consists of 8 items assessing the respondent's satisfaction with their skill in managing data for publication on a scale ranging from 1 (strongly agree) to 5 (strongly disagree). Only one factor with acceptable properties is derived from the factor analysis. This variable has reliability estimates (Cronbach's alphas) of .85. A conventional rule of thumb regards any value above .70 as satisfactory.
- d. **Local context and specificity.** This variable represents a new measure developed for this study to depict the heterogeneity and context-dependent nature of earth observation datasets. This measure consists of 3 items assessing respondents' views on the complexity of earth observation datasets and possible negative impacts for data re-use, using the 5-point Likert scale (1: strongly agree, to 5: strongly disagree). Only one factor is acceptable based on the factor analysis result and the Cronbach's alpha for this variable is .78.
- e. **The organizational support** variable measures the level of support, in terms of storing, funding, training, managing, and technical support that the organization provides to researchers to encourage the sharing of research datasets. This variable is measured by agreement with statements that represent organizational involvement in data sharing on a scale ranging from 1 (strongly agree) to 5 (strongly disagree). This variable has reliability estimates (Cronbach's alphas) of .92.
- f. **Legal context and policy.** This variable is measured by 6 items which assess the researcher's view on the importance of having regulations and policy support to encourage data sharing, on a scale ranging from 1 (strongly agree) to 5 (strongly disagree). This variable has reliability estimates (Cronbach's alphas) of .83.
- g. **Incentive to share.** This variable represents the respondent's view on the importance of proper attribution, acknowledgement, and citation to encourage data sharing. The variable is measured by 4 items on a dichotomous scale. The reliability of this estimate (Cronbach's alphas) is .82.
- h. **Type of Metadata.** This variable measures the types of metadata that respondents are currently using to describe their data on 9-point nominal scales which pertain to 9 different metadata.

4. Result and Discussion

4.1. Assessment of Structural Model

Table 3 presents the correlations and descriptive results among all study variables. Table 4 reports results from the model analysis and the structural path coefficients. Table 5 presents the model fit statistics. The data analysis of this research consisted of two stages. The first stage is creating the construct using Confirmatory Factor Analysis. The reliability of the constructs/variables was examined using Cronbach's alpha values (Table 3). All Cronbach's alpha values were above 0.70, representing acceptable levels for confirmatory research (Table 3).

Table 3 Cronbach's alpha and Descriptive Statistics

| Variables | Means | s | α |
|--------------------------------------|---------|--------|----------|
| 1. Action to share datasets | 0.0302 | 1.6220 | 0.77 |
| 2. Intention to share datasets | 0.0590 | 0.8785 | 0.75 |
| 3. Self-efficacy in sharing datasets | 0.1073 | 2.0044 | 0.85 |
| 4. Local context and specificity | -0.0237 | 1.3917 | 0.78 |
| 5. legal context and policy | -0.0876 | 1.0030 | 0.83 |
| 6. Organizational support | 0.0859 | 2.1724 | 0.92 |
| 7. Type of metadata | 7.8109 | 1.9056 | - |
| 8. Incentive to share datasets | -0.0237 | 1.5976 | 0.82 |

The structural model was tested with the structural equation modeling approach using LISREL 8.80. The assessment of the overall model fit was based on multiple criteria.

Table 4 Structural Parameter Estimates

| Path Coefficients | Restricted Model | Theoretical Model |
|--------------------------------|------------------|-------------------|
| Intention → action | -0.294 * | -0.294 * |
| Self efficacy → action | -0.015 | -0.015 |
| Self efficacy → intention | 0.035 | 0.050 ** |
| Local context → self efficacy | -0.078 * | -0.074 * |
| Legal & policy → action | | 0.000 |
| Legal & policy → self efficacy | | 0.057 ** |
| Legal & policy → Intention | | -0.225 * |
| Org. support → self efficacy | 0.186 * | 0.181 * |
| Org. support → action | -0.152 * | -0.152 * |
| Org. support → intention | 0.096 * | 0.112 * |
| Type of Metadata → intention | 0.199 * | 0.188 * |
| Appropriation → intention | 0.112 * | 0.124 * |
| Appropriation → action | 0.051 ** | 0.051 ** |

*significant at 0.05 **significant at 0.10

The results of goodness of fit indexes are presented in Table 5. Table 5 indicates that overall, the fitness test signifies an adequate fit model. Results from the structural analysis indicate that the theoretical model provides an adequate explanation for the structural relationships among variables. The theoretical model comparative fit index of 1.00 indicates a perfect fit with the structural portion among the variables.

Table 5. Analysis of Overall Model Goodness-of-fit

| Model goodness-of-fit indexes | Cut-off value | Result from this study | |
|-------------------------------|--------------------|------------------------|--------------|
| | | Rest. model | Theory model |
| χ^2 | $< \chi^2_{table}$ | 4.402 | 3.701 |
| $\chi^2_{significance} (p)$ | $p > .05$ | 0.493 | 0.539 |
| Goodness-of-fit index (GFI) | ≥ 0.90 | 0.998 | 0.998 |
| Normed fit index (NFI) | ≥ 0.90 | 0.975 | 0.983 |
| Comparative Fit Index | ≥ 0.90 | 1.000 | 1.000 |
| Incremental Fit Index (IFI) | ≥ 0.90 | 1.004 | 1.000 |
| RMR | ≤ 0.10 | 0.026 | 0.021 |
| RMSEA | ≤ 0.10 | 0.000 | 0.000 |
| p-value of RMSEA | $p > .05$ | 0.929 | 0.953 |

Comparing the model fit among the two different models indicates that the theoretical model has the best fit among the models. The χ^2 for the theoretical model is 3.701 larger than the χ^2 cut-off value of 11.07 (df=5, $\alpha=0.05$). The exclusion of regulations and policy context from the model diminished the significance of perceived self-efficacy in researchers' intentions to share datasets, and decreased the fit index norm below the cut-off value, from 0.9 to 0.89.

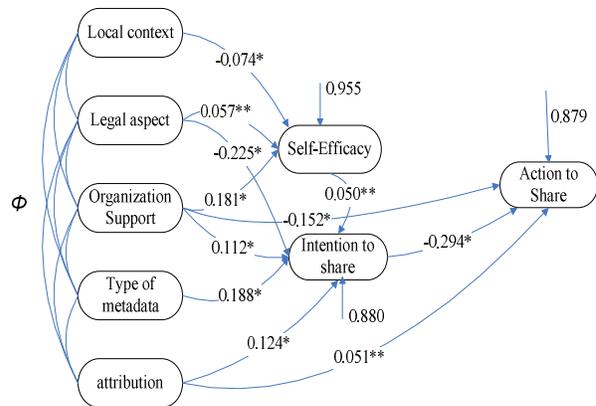


Figure 2. Derived Path Coefficients of the Theoretical Model³

Figure 2 presents significant standardized path coefficients for the theoretical model. The error terms in standardized form represent the proportion of variance in each equation not accounted for in the structural model. The error terms indicate that the theoretical model may be best viewed as addressing

³ Upon request, we will provide larger diagrams of path visualization

the antecedents and consequences of intention to share and action to share, as opposed to the antecedents of perceived self-efficacy. The proportion of variance explained in the theoretical model for action to share and intention to share is 12.1% and 12% respectively, while the proportion of variance for self-efficacy is only 4.5%.

4.2. Behavioral Intention, Action to Share Datasets, and the Antecedents

Strong support was found for the determinants of action to share datasets with varying degrees of significance. Intention to share was found to significantly influence the action of researchers to share their datasets online (t -value = -7.217). This study also found both direct and indirect influences of intention as a strong predictor of data sharing behavior. The findings also indicate strong support for the impact of exogenous variables (organizational support, legal context and policy, type of metadata, and attribution and reward) on the researchers' intention and action of sharing their datasets. All of these variables are statistically significant at 0.05 levels (Table 4).

Organizational support was found to significantly impact the action of sharing, both directly (t -value = -4.658) and indirectly, through the intention to share (t -value = -2.657). On the other hand, this study found a weak support for the direct impact of incentive to share on the action of sharing datasets in terms of coefficient estimates and significant levels. The path coefficient of incentive and attribution to the action of sharing is 0.051, and is significant at 0.10. This finding also shows a strong but negative indirect impact of incentive and attribution on the action of sharing that is mediated by intention to share (t -value = -2.657). Additionally, the data shows the significant indirect influence of the exogenous variables through the intention to share. Types of metadata (t -value = -3.990) and legal context and policy (t -value = 4.317) were found to have significant indirect impact on the action or sharing mediated through the intention to share datasets.

The impacts of the exogenous variables on the intention to share datasets are also strong and significant both directly and indirectly. Legal context and policy were found to be negative and significant predictors of the intention to share. Types of metadata and attribution and incentive are found to be stronger predictors of intention to share datasets directly. The direct impact of organizational support on the intention to share datasets is stronger than if it is mediated by perceived self-efficacy. The findings support hypotheses 1 and 2 of this study.

4.3. Perceived Self-efficacy, Intention and Action to Share

Perceived self-efficacy was found to significantly influence the action of sharing datasets when mediated by the intention to share datasets. On the other hand, findings also show a weaker direct impact of perceived self-efficacy on the intention to share datasets. This study found that the significance of perceived self-efficacy as a predictor of intention to share depends on the inclusion or exclusion of legal context and policy variables. Perceived self-efficacy becomes insignificant when legal context and policy variables are excluded. Analysis results show strong support for the impact of local context and organizational support on perceived self-efficacy to share datasets. Local context and specificity significantly influence perceived self-efficacy to openly share datasets (t -value 1.807). Organizational support was found to be a significant predictor of perceived self-efficacy. The path coefficient of organizational support to perceived self-efficacy is significant with a coefficient estimate of 0.181 and t -value of 4.456. On the other hand, the data shows weaker impact of legal context and policy on perceived self-efficacy. The path coefficient of legal context and policy to perceived self-efficacy is 0.081 with t -value of 1.401. Except for the impact of legal context and policy, the findings support hypotheses 3 and 4 of this study.

5. Discussion and Concluding Remarks

5.1. Organizational support, reward, and legal policy to encourage intention and action

With the objectives of exploring the nature and function of behavior in publishing research datasets, by distinguishing two principals of sharing behavior – intention to share and action of sharing – and by identifying the predictors which influence the development of intention and action of sharing data, this research study's findings are supportive and substantial. The data shows that the intention of researchers to openly publish their datasets to public correlates significantly with the carrying out of the action of sharing datasets. This finding thus supports Azjen's argument (1991) that individual intention was a critical predictor of sharing behavior [2]. Intention in this study reflects the level of a researcher's effort and willingness to publish their research datasets in an open data initiative. Considering that the preparation of scientific datasets for publication requires plenty of effort from researchers [32], the willingness to work hard in an effort toward publishing their data is a

significant necessity of researchers. Greater willingness to exert more effort to publish data will result in greater likelihood of actual data publication action. Building on the argument that publishing research datasets is a personal decision propelled by social influence [32], this study provides support for the significance of behavioral intention to share as the key and critical predictor of the actual sharing of research datasets. Thus, encouraging collaborative works to support scientific progress through data integration necessitates nurturing the intention to share in addition to the predictors that are likely to influence behavioral intention. Along with the testing of hypothetical causal effects, the analysis highlighted the significance of organizational support and rewards for sharing datasets while also revealing the somewhat intriguing impact of legal and policy factors.

This study found that organizational support is significant to encouraging the action of publishing research datasets. The data show the significance of both direct and indirect impact of organizational support on the action of sharing research datasets. This finding contrasts with the argument asserted by a number of other knowledge sharing studies which have found the indirect impact of organizational support to be stronger in its influence on the action of knowledge sharing [5, 23]. One plausible explanation for this is the differing nature between data sharing and knowledge sharing. Knowledge sharing can be classified into tacit and explicit [9]. Arguably, sharing knowledge requires more personal judgment and social relations, particularly in the sharing of tacit knowledge. As result, the impact of organizational support on knowledge sharing is mediated according to personal decision. On the other hand, earth science research datasets are heterogeneous [28], very specific [6], and highly dependent on local context [35, 36]. Therefore, proper preparation and presentation of the dataset for sharing becomes more significant [32], due to the specificity of the data, and support to improve skills in the production and preparation of data become the imperative factor for sharing. Nonetheless, this assertion warrants further research. Future research could ascertain the impact of organizational support, distinguishing between data and knowledge sharing, and explore causes of possible differences.

This study also found a strong indirect impact of reward, in the form of attribution and acknowledgement, than the direct impact of reward to encourage the action of sharing datasets. Rewards will significantly impact the action of sharing datasets if mediated through the behavioral intention of the researcher. This finding contradicts the argument that rewards will have a direct and significant impact on the likelihood of sharing research datasets [1, 30, 32, 33].

On the other hand, the contradictory finding could also be the result of the differing constructs in measuring the attribution variable in this study. This study measures the attribution mostly according to extrinsic rewards for the dataset's owner, such as acknowledgment or the opportunity for collaboration. In this way, this study does not take into account the influence of intrinsic rewards. Arguably, intrinsic rewards in term of self-satisfaction or altruistic behavior in the sharing of datasets could influence behavioral intention and action of sharing datasets. In this regard, future research could re-examine the model by incorporating intrinsic rewards and distinguishing the impact of intrinsic and extrinsic rewards.

Interestingly, this study found that the legal context and policy factors significantly impact the intention to share, but negatively. The impact of legal context and policy on the action of sharing datasets is mediated by the researcher's intention to share datasets. The negative connotation of legal context and policy is intriguing since it indicates that an increase in regulations and policy will result in lower intentions of sharing research datasets. Some studies have found stringent legal and policy requirements to negatively impact researchers' intentions of sharing datasets, particularly if such policy and regulations are related to data with privacy rights. For instance, a study by Tucker (2009) on medical research data sharing in a cancer center found that onerous privacy policy requirements to protect the privacy of patient information will negatively impact researchers' intentions of publishing research datasets [32]. However, this study is in no way arguing that regulations and policy inhibit the sharing of research data. Arguably, this study's finding signifies that the impact of legal context and policy depends on the nature of the data being shared. Thus, researchers follow and pay attention to the existing policies to guide their intentions of sharing data. Knowing that a legal framework exists to support and guide data sharing may provide assurance for dataset owners that their datasets will not be misused or poached [30]. This finding warrants further investigation on the impact of legal context and policy on researchers' intentions to share data by specifying different regulatory levels.

5.2. Lesson Learned and Implication for Open Data

Open government has emerged as an initiative to encourage knowledge development and resolve public issues through collaboration, data integration, and data dissemination. One objective of open government in the United States is to promote a culture of open government that encourages the collaborative networking of researchers, private sectors, and citizens

[18]. With this openness and technology development, environmentalists like Lynn Henning, the recipient of the 2010 Goldman Prize, can work collaboratively with state regulators and others by sharing her data to encourage stronger enforcement of environmental protection, or to tackle any other social issue. Arguably, sharing data in a collaborative network is influenced by personal decisions and social relations [32]. Thus, it is important to understand the determinants which affect an individual's decision to share data. Building on the investigation of motives for sharing earth observational research datasets in an initiative for open data such as DataONE, this research study argues that acknowledging the key determinants of motivation to publish data entails significant policy implications that will extend to the public sector. Although the object is different, the personal decision process leading to sharing is most likely the same in open government as in research network collaboration. Thus, by understanding the determinants of sharing data in collaborative network, government could encourage more enthusiasts such as Lynn Henning. This paper argues four plausible implications:

1. This study's finding is somewhat unclear in regard to legal context and policy issues. Thus, it calls for consideration to specify different levels of policy according to individual users to encourage sharing. Arguably, understanding the psychological processes of the act of sharing data will be beneficial in increasing the adherence to and effectiveness of proposed policy in data sharing. This process could be crucial to encouraging collaborative sharing among stakeholders with various backgrounds and interests in open government such as researchers, citizens, and government officials.
2. This study found that organizational support plays a significant role in encouraging data sharing. Organizational support could reduce the existing complexity of the data sharing process. Considering the heterogeneity of collaborators in open government, more attention to the existence and non-existence of organizational support is necessary. In this regard, the proposed policy should emphasize organizational support to increase researchers' intentions of sharing datasets, or develop strategies to accommodate collaborators without an organizational background.
3. Reward was found to have a significant indirect impact on data sharing, which leads to the issue of considering the balance between intrinsic and extrinsic rewards to encourage sharing behavior in collaboration. For instance, environmentalist such as Lynn Henning might be motivated by self-satisfaction or altruistic behavior, but private sector

collaborators might require a different type of reward.

5.3. Limitations of the Study

This study attempts to provide preliminary assessment on the determinants of a researcher's action of sharing datasets from the perspective of a personal decision. These findings need to be interpreted with consideration of the study's limitations. First, almost 75% of survey respondents were from the North American region. Thus, the findings of this research study are best interpreted as evidence of researchers' motivation to share datasets in the North American region. Further research is needed to establish the generalizability of these findings to other regions of the world. Second, this study does not account for the possible non-recursive causality between variables. Finally, the confirmatory assessment of the construct only considers a relaxed parameter analysis and generates one factor from the measurement model. Future research could compare the restricted and non-restricted parameters in the measurement model.

6. Acknowledgments

The authors would like to thank the socio cultural and usability and assessment working group of the DataONE project for their permission to use the project's data and information.

7. References

- [1] Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhler, P., & Wouters, P. "An international framework to promote access to data", *Science*, vol. 303, issue 5665, pp. 1777-1778, 2004.
- [2] Azjen, I. "The theory of planned behavior", *Organizational Behavior and Human Decision Processes*, vol.50, p.179, 1991.
- [3] Bandura, A., "Self-efficacy: Toward a unifying theory of behavioral change", *Psychological Review*, vol. 84, issue 2, pp.191-215, 1977.
- [4] Barker, E.W., Al-Ghatani, S.S., & Hubona, G.S. "The Effects of gender and age on new technology implementation in a developing country: Testing the theory of planned behavior (TPB)". *Information Technology & People*, vol.20, pp. 352-375, 2007.
- [5] Bock, G., Zmud, R.W., Kim, Y., & Lee, J. "Behavioral intention formation in knowledge sharing: examining the roles of extrinsic motivators, social-psychological forces, and organizational climate", *MIS Quarterly - Special Issue on Information Technologies and Knowledge Management*, vol. 29, pp. 87-111, 2005.

- [6] Borgman, C.L. "Research Data: Who will share what, with whom, when, and why?" in 5th China-North America Library Conference, pp.1-21, 2010.
- [7] Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E. and Olson, G. "From shared databases to communities of practice: A taxonomy of collaboratories". *Journal of Computer-Mediated Communication*. vol.12, pp.652-672, 2007.
- [8] Brown, S.A., Dennis, A.R., & Venkatesh, V. "Predicting collaboration technology use: Integrating technology adoption and collaboration research", *Journal of Management Information Systems*, vol.27, pp.9-53, 2010.
- [9] Brown, J.S., & Duguid, P. "Knowledge and Organization: A Social-practice Perspective". *Organization Science*, vol.12, pp. 198-213, 2001.
- [10] Bryn, N. "Empty archives". *Nature*. Vol.461, issue.7261, pp.160-163, 2009.
- [11] Cabrera, A., Collins, W.C., & Salgado, J.F. "Determinants of individual engagement in knowledge sharing". *International Journal of Human Resource Management*, vol.17, pp. 245-264, 2006.
- [12] Constant, D., Kiesler, S., & Sproull, L. "What's mine is ours, or is it? A study of attitudes about information sharing", *Information Systems Research*, vol.5, pp.400-421, 1994.
- [13] Compeau, D.R. & Higgins, C.A. "Computer Self-Efficacy: Development of a Measure and Initial Test", *MIS Quarterly*, vol.19, pp.189-211, 1995.
- [14] Davenport, T.H. and Prusak, L. *Information ecology: Mastering the information and knowledge environment*. New York: Oxford University Press, 1997.
- [15] Dawes, S.S., Cresswell, A.M. and Pardo, T.A. "From "Need to Know" to "Need to Share": Tangled Problems, Information Boundaries, and the Building of Public Sector Knowledge Networks", *Public Administration Review*, vol.69, pp.392-402, 2009.
- [16] Gagne, M. "A Model of knowledge-sharing motivation". *Human Resource Management*, vol.48, pp. 571-589, 2009.
- [17] Goldston, D. "Big data: Data wrangling", *Nature*, vol.455, issue 7209, p.15, 2008.
- [18] Harrison, T.M., Guerrero, S., Burke, B.G., Cook, M., Cresswell, A., Helbig, N., Hrdinova, J., & Pardo, T. "Open government and e-government: Democratic challenges from a public value perspective", *Proceeding of the of the 12th International Digital Government Research Conference (Dg.O)*, Maryland, United States, 2011.
- [19] Kaye, J., Heeney, C., Hawkins, N., de Vries, J. and Boddington, P. "Data sharing in genomics—re-shaping scientific practice", *Nature Reviews Genetics*, vol.10, pp.331-335, 2009.
- [20] Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Hock, H., Lautenschlager, M., Schindler, U., Sens, I., & Wachter, J. "Data publication in the open access initiative", *Data Science Journal*, vol.5, pp. 79-83, 2006.
- [21] Kuo, F.Y., & Young, M.L. "A study of the intention – action gap in knowledge sharing practices", *Journal of the American Society for Information Science and Technology*, vol.59, pp. 1224 – 1237, 2008.
- [22] Lee, H. & Chang, E. "Consumer attitudes toward online mass customization: An application of extended technology acceptance model", *Journal of Computer-Mediated Communication*, vol.16, pp.171-200, 2011.
- [23] Lin, H. "Impact of organizational support on organizational intention to facilitate knowledge sharing", *Knowledge Management Research & Practice*, vol. 4, pp.26-35, 2006.
- [24] Mathieson, K. "Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior", *Information System Research*, vol.2, pp.173-191, 1991.
- [25] Mergel, I., Schweik, C., & Fountain, J. "The Transformational Effects of Web 2.0 Technologies on Government", *Social Science Research Network (SSRN)*, 2009.
- [26] Michener, W.K. "Meta-information concepts for ecological data management". *Ecological Informatics*, vol.1, pp.3-7, 2006.
- [27] Piwowar, H.A. & Chapman, W.W., "Public sharing of research datasets: A pilot study of associations", *Journal of Informetrics*, vol.4, p.148-156, 2007.
- [28] Reichman, O.J., Jones, M.B., & Schildhauer, M.P. "Challenges and opportunities of open data in ecology", *Science*, vol.331, pp.703-705, 2011.
- [29] Sayogo, D.S., & Pardo, T.A. "Understanding the capabilities and critical success factors in collaborative data sharing network: The case of DataONE", *Proceeding of the 12th International Digital Government Research Conference (Dg.O)*, Maryland, United States, 2011.
- [30] Sayogo, D.S. & Pardo, T.A. "Exploring the determinants of publication of scientific data in open data initiative". *Proceeding of the 5th International Conference on Theory and Practice of Electronic Governance (ICEGOV2011)*, Tallinn, Estonia, 2011.
- [31] Taylor, S. & Todd, P.A. "Understanding information technology usage: A test of competing models", *Information Systems Research*, vol.6, pp.144-176, 1995.
- [32] Tucker, J. "Motivating subjects: Data sharing in cancer research", *Doctoral Dissertation*, the Virginia Polytechnic Institute and State University, 2009.
- [33] Vogel, L. "The Secret's in: Open data is a foreign concept in Canada", *Canadian Medical Association Journal*, vol.183, pp. 375-376, 2011.
- [34] Zhang, J. and Dawes, S.S. "Expectations and perceptions of benefits, barriers, and success in public sector knowledge networks", *Public Performance & Management Review*, vol.29, pp.433-466, 2006.
- [35] Zimmerman, A.S. "Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse". *International Journal on Digital Libraries*, vol.7, pp. 5-16, 2007.
- [36] Zimmerman, A.S. "New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data", *Science, Technology & Human Values*, vol.33, p.631-652, 2008