

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Towards Better Data Selection for Self-Service Business Intelligence Outputs

Lega, Mathieu

Published in:

Proceedings of the Doctoral Consortium Papers Presented at the 34th International Conference on Advanced Information Systems Engineering (CAiSE 2022), June 06–10, 2022, Leuven, Belgium

Publication date:

2022

Document Version

Peer reviewed version

[Link to publication](#)

Citation for published version (HARVARD):

Lega, M 2022, Towards Better Data Selection for Self-Service Business Intelligence Outputs: a Local Authorities Case Study. in Proceedings of the Doctoral Consortium Papers Presented at the 34th International Conference on Advanced Information Systems Engineering (CAiSE 2022), June 06–10, 2022, Leuven, Belgium.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Towards Better Data Selection for Self-Service Business Intelligence Outputs: a Local Authorities Case Study

Mathieu Lega^{1,2,3}

¹University of Namur, Rue de Bruxelles 61, 5000 Namur, Belgium

²Namur Digital Institute (NaDI)

³PRECISE Research Center

Abstract

Uncertainty belongs to the daily life of decision makers. Be it in the public or in the private sector, most decisions come with a risk of uncertainty. To mitigate this risk and help decision makers, several techniques and tools have been developed, among which data analysis techniques and systems. Business Intelligence (BI), and more specifically Self-Service Business Intelligence (SSBI) on which we focus in this project, are examples of such techniques. SSBI aims to use the data available within an organization to support people when they make decisions, and works on the promise that decision-makers will produce their analyses by themselves, in a “do it yourself” mood. However, the range of usable data readily accessible to decision makers is enormous and constantly increasing. This profusion of data makes it really difficult for SSBI users to know what data to analyze and what data to ignore. The need is real therefore to help these users to manage this data profusion phenomenon. In this project, we thus want to find a way to help SSBI users to select the most important data for their own reporting. Our work is illustrated with an application in the public sector. This paper presents the context, the research questions, the methodology and the contributions that are targeted as part of this project.

Keywords

Decision making, Self-Service Business Intelligence, Value, Data

1. Context

Decision making is a central and critical process in any modern organization. Organizations make numerous decisions – strategic or operational – on a daily basis. Deciding properly – at the right time and with the proper information – has been long and is still recognized as a key differentiation factor for companies [1, 2].

Making the right decision, however, is far from being trivial. The world in which decision makers evolve is characterised by volatility, uncertainty, complexity and ambiguity [3]. Volatility (V) because the organisation situation and environment are unstable. Uncertainty (U) because issues and events are most of the time impossible to predict. Complexity (C) due to the volume


Proceedings of the Doctoral Consortium Papers Presented at the 34th International Conference on Advanced Information Systems Engineering (CAiSE 2022), June 06–10, 2022, Leuven, Belgium

✉ mathieu.lega@unamur.be (M. Lega)

ORCID 0000-0003-1682-4920 (M. Lega)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

of issues that confound. And ambiguity (A) due to the haziness of reality. We thus speak of a VUCA world.

A natural answer to this VUCA world comes from the world of data, which advances the promise of reducing uncertainty as a way to support decision making [4].

Business Intelligence (BI) is one way of realizing that promise. Indeed, BI systems are designed to help decision makers in their strategic and operational decision making processes by providing user-friendly and business-oriented access to integrated information [5]. BI systems use analytical tools to derive complex information from operational data to increase the timeliness and quality of the decision process inputs [5]. In order to achieve this, BI relies on the articulation of a number of technologies, including tools to extract data from various business data sources and integrate them in one centralized data repository [6]. This data repository, ultimately, will be used to feed BI outputs as dashboards and reports (we will refer to these as dashboards in the rest of this document). Dashboards are interactive interfaces supporting the visualization and the analysis of performance metrics [7]. In order to represent the organization of information in a consistent and flexible way, dashboards are most often composed of indicators, graphs, tables and interactive features [8]. Most of the time, dashboards are built manually (from scratch or by customizing off-the-shelf products) based on human knowledge and experience [8].

Self-Service BI (SSBI) Systems propose an approach in which the end users – the decision makers – choose directly the data and the visuals to use [9]. Three different levels of self-service have been identified by [10]: (i) usage of information where end-users have access to the already created information such as existing reports; (ii) creation of information where end-users have access to disaggregated data and create the information themselves and; (iii) creation of information resources where end-users even have the opportunity to discover new data sources and to combine it with already existing corporate data. These three levels are presented from the lowest level of self-reliance and system support to the highest level. By offering all these possibilities to the end-users, SSBI reduces the time-to-delivery by removing all discussion between the business and the IT – and therefore all subsequent requirements analysis effort and validation – and improves the alignment of the SSBI outputs with the business requirements [10].

The remainder of this paper is structured as follows. Section 2 presents the problem that we address in this paper and the research questions of this project which aims to help SSBI users in the selection of the most important data for reporting. In section 3, we detail our research methodology. Section 4 elaborates on the five expected contributions of this project. Finally, we conclude the paper in section 5.

2. Problem Statement and Research Questions

The adoption of SSBI faces some significant challenges, classified by [10] into two main categories: (i) the challenges about the access and use of data and (ii) the challenges about the self-reliance of users. The first category is itself composed of various sub-problems [10]: (i) making data sources easy to access and use; (ii) identifying data selection criteria; (iii) using correct data queries; (iv) controlling the integrity, security and distribution of data; (v) defining

policies for data governance and management; and (vi) preparing data for visual analytics. The second category contains four challenges [10]: (i) making SSBI tools easy to use; (ii) making SSBI tools easy to consume and enhance; (iii) giving the right tools to the right user; and (iv) educating the users for the selection, interpretation and analysis of data in a context of decision making.

With the development of data over the last decades, organizations produce and manage increasing amounts of data, so that data repositories in SSBI become more and more rich and complex. While this phenomenon of data profusion creates several opportunities (such as allowing decision makers to use this data to better understand the needs of their customers, to improve the quality of their services, and to better predict and prevent risks [11]), it also impacts directly the two categories of SSBI challenges presented above. First, the access and the use of data becomes more difficult due to the fact that a bigger quantity of data must be managed and accessed. Then, it also makes it more difficult for the users to be self-reliant because they may drown in the volume of data.

One way to reduce the challenge of data profusion for decision makers in SSBI rely in offering some guidance and assisting end-users in the selection of pieces of data that really matter to them. In this project, we want to elaborate on the problem of data profusion for BI outputs with a focus on SSBI outputs. More precisely, we plan to develop a SSBI framework designed to help the end-users in the selection of the most important data for reporting.

We plan to address the problem above in the particular domain of e-government (or simply egov) and more particularly smart governance. Egov is the field studying the way information systems may be used in the public sector. This field is quite recent (late 1990s) and is rather multi-disciplinary, combining fields such as public administration, information systems and political sciences [12]. Egov covers notably smart governance, defined in [13] as "the intelligent use of information and communication technologies (ICTs) to improve decision-making", what BI and SSBI are specifically designed for.

This egov domain appears then as well suited to apply our work because an essential aspect of smart governance is the access to timely and actionable information which can be eased by ICTs, among which SSBI [14].

In order to treat these problems, we develop several research questions.

- RQ1 How do cognitive loads influence the chance of adoption of a BI or SSBI output?
- RQ2 Which selection criterion may theoretically be used to select the data to use in BI and SSBI outputs to keep control on the cognitive loads?
- RQ3 How can this selection criterion be operationalized practically to select the data that will be used for reporting?
- RQ4 What is the situation of the public sector in terms of current practices and needs for decision making support, and more specifically BI and SSBI?
- RQ5 How can we integrate the previous selection criterion into a SSBI process for public decision making?

3. Research Methodology

The different research questions presented in the previous section are designed to feed each other. RQ1 feeds RQ2 and RQ4 by supplying knowledge about cognitive loads. RQ2 feeds RQ3 with a theoretical data selection criterion for SSBI outputs. Finally, RQ3 and RQ4 both feed RQ5 by respectively supplying an operationalized data selection criterion and knowledge about the particular SSBI-related needs of the public sector. The articulation of the different research questions is illustrated in figure 1.

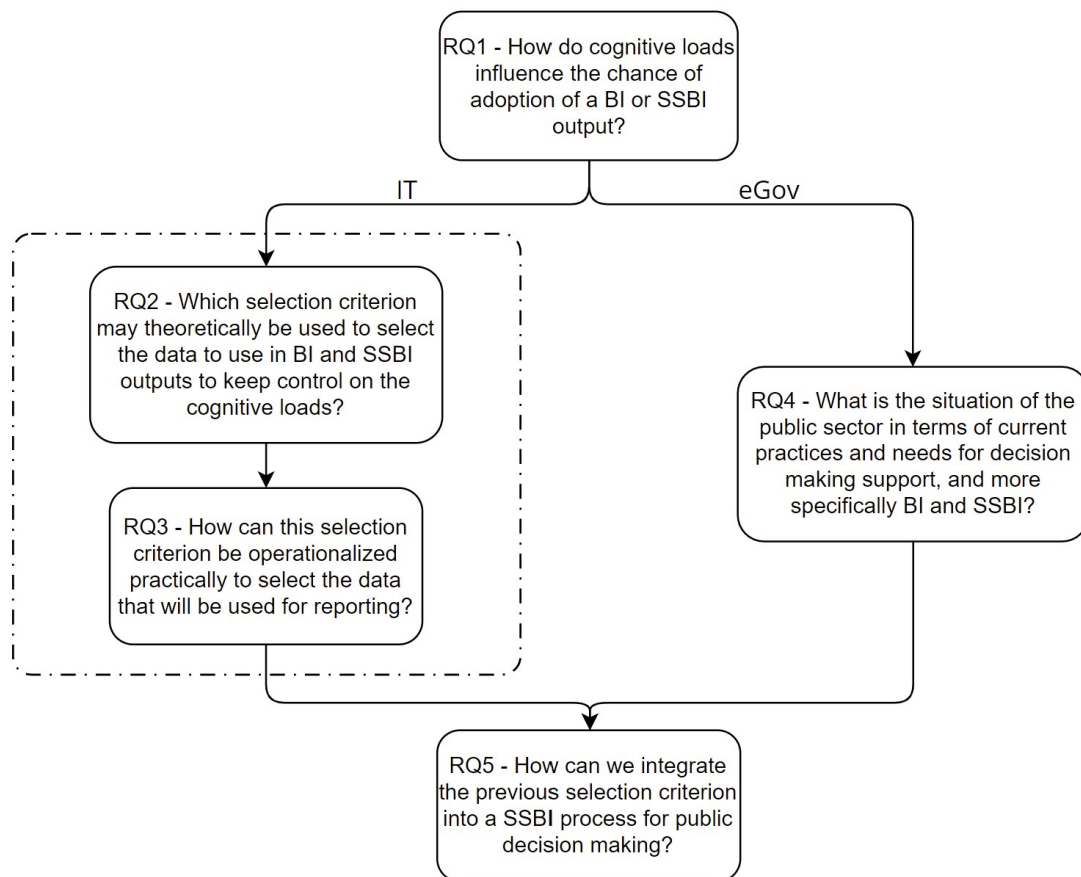


Figure 1: Articulation of the research questions

In order to answer these research questions, a Design Science Research (DSR) approach will regularly be used. This latter is a scientific problem-solving methodology that focuses on the design of new and innovative artifacts in order to increase human and organizational capabilities [15].

The goal of this approach is the creation of the "what is effective" based on a good understanding of the problem domain. To do so, a three-cycle view of DSR has been proposed in [16].

The relevance cycle initiates the project with an application context including the requirements and the acceptance criteria to evaluate the results. The rigor cycle allows to incorporate past knowledge to the project. Finally, the design cycle represents the heart of the DSR project and consists of several iterations made of the construction and evaluation of an artifact and of the production of feedback. Used iteratively and in an interrelated way, these three cycles allow to refine the design of the artifact.

This approach is well adapted to our project (and to most projects with an information system context) because it allows to focus on the creation and evaluation of creative IT artifacts that will help organizations facing information-related tasks [15].

4. Research Contributions

Our contributions to the research questions mentioned above will be developed within several studies that we present in this section.

4.1. Empirical Study on Cognitive Loads in BI Systems

Objective. In this study, we try to better understand the problem of overload that may rise when a user is confronted to SSBI systems and more precisely dashboards. Indeed, while the aim of SSBI and dashboards is to help decision makers to gain insights and to make better decisions, some SSBI systems/dashboards are adopted while others are rejected. The aim of this study is thus to create a Dashboard Adoption Model extending the well-known Technology Adoption Model (TAM) in the context of dashboards.

Method. To achieve our goal, we plan to conduct a survey experiment where the respondents are confronted to dashboards and are asked to adopt them or not. The presented dashboards should vary in terms of informational and non-informational load. The responses will then be analysed in order to develop knowledge that will be used to develop a Dashboard Adoption Model based on the existing TAM model.

Related Work. Two main articles will be used as basis for our work. The first is the original TAM model studying the factors of technology adoption [17]. The second is the work of [18], studying the factors of adoption of hedonic information systems.

Current State of the Research. This study is finished and will be soon submitted in an article.

4.2. Decision-Making Data Value Taxonomy

Objective. Once we better understand how the cognitive loads impact the chance of adoption of a dashboard, we want to investigate a way to better control these loads. As the problem of data selection for reporting becomes more difficult with the growing amount of available data and the content (i.e. the data used for) of a SSBI output impacts the cognitive loads, we want to investigate a way to simplify this process of data selection. The aim of this study is to introduce the concept of “decision-making value” as a selection criterion to determine which data should

be kept for reporting in a database. The underlying idea is that all columns within a database are not equally valuable to a decision maker, and we want to identify those columns which have the higher values. To do so, we define the concept of decision-making data value (DMDV) and build a DMDV taxonomy.

Method. To fulfill our objectives, we realize a literature review on data value and its dimensions. Then, we present a methodology to select the dimensions of data value to take into account for decision making based on what we retrieved. Finally, we apply this methodology to build our DMDV taxonomy.

Related Work. Different articles tackle the necessity to define the value of data. The authors of [19] analyze four context-independent challenges of value-driven data governance retrieved using the literature and their experience. In [20], a value assessment framework is presented based on a decomposition of data into several characteristics. Our work differentiates itself in the way we define data value.

Current State of the Research. The literature review has been performed and the methodology for the selection of the dimensions has been defined. Our taxonomy is currently under finalization.

4.3. Decision-Making Data Value Assessment Framework

Objective. The aim of this study is to create a DMDV assessment framework based on the taxonomy built in the previous study and to validate this framework with end-users. Indeed, while the previous study stays theoretical, the goal of this study is to allow the identification of the value-optimal set of columns for reporting and to validate it practically.

Method. In order to attain our goal, we develop metrics for the different dimensions of our DMDV taxonomy and study aggregation techniques in order to combine these metrics into a single indicator for DMDV. We consider different levels of granularity for our DMDV indicator but we focus on the set of columns level. Optimization algorithms are also considered to find a way to select the value-optimal set of columns for reporting. Then, we design a survey experiment to compare the results of our algorithm with human choices. Based on the results, we adapt our indicator.

Related Work. Several recent works exist on the assessment of the value of data. In [21], the authors develop a data value assessment technique based on the survey of data professionals and academics. In [22], an automatic and metric-based data value assessment approach is presented and tested in a use case. Finally, the authors of [23] combine both the human input and the data processing in their data value assessment tool. Our work differentiates itself in the way we define and use data value. We do not use it to maximise a monetary aspect but to optimize decision making.

Current State of the Research. Different aggregation techniques have been retrieved and

analyzed. We also identified several challenges to address for a robust metric for the value of a set of columns. Finally, the survey experiment design is currently in progress.

4.4. Analysis of the Situation for Local Authorities

Objective. Local authorities belong to the organizations that have access to a huge amount of data to analyze and that need to optimize their decision making. The aim of this study is thus to analyze the current situation of Belgian local authorities in terms of decision making. Three main questions will guide this study:

1. How do local authorities currently use BI, SSBI and information in general to make their decisions?
2. Which needs do local authorities have in terms of BI, SSBI and information in general?
3. Are there significant differences in the responses based on the characteristics of the studied cities?

Method. In order to perform this analysis, we use semi-structured interviews realised with political and technical staff of Belgian local authorities. The heterogeneity in the questioned people increase the strength of this study because we select people working in cities of different size, technological development and rurality. Once the saturation threshold is reached, i.e. no new observation emerge while performing other interviews, the results are analyzed using an inductive approach.

Related Work. To the best of our knowledge, no existing study really approaches our objectives at the time of writing. The closest studies are the following. In [24], the authors adapt the Technology Acceptance Model (TAM) to study the factors that influence the usage of DSS by Egyptian local authorities. In their study of 1995, the authors of [25] analyze the impact of context and culture on the strong increase in information systems usage among local authorities. Finally, the role and scope of information systems evaluation in the public sector are investigated in [26].

Current State of the Research. So far, we interviewed around ten political or technical participants coming from cities with heterogeneous characteristics and this allowed us to collect a lot of information in order to gain insights about our research questions. More interviews are planned as we did not reach the saturation threshold yet. The analysis of the results using an inductive approach is currently in progress. Preliminary results suggest that there are significant differences among cities and that the size of the city is one of the most important characteristic to explain these differences. The biggest needs seem to be the centralization of the information and a clear vision of the budget.

4.5. Value-Driven Self-Service Business Intelligence Framework for Local Authorities

Objective. The motivation behind this work is that local authorities have a lot of data to manage, often low technical knowledge and rather particular needs. This study aims to develop a value-

based SSBI framework specifically designed to help local authorities make better decisions.

Method. We analyze the different requirements identified in the previous study in order to focus on the most important needs of local authorities. Moreover, we study existing SSBI solutions in order to assess the way these needs and the concept of decision-making data value may be integrated.

Related Work. To the best of our knowledge, there is currently no article trying to include the concept of data value in a SSBI process for local authorities. The following works may however relate to what we plan to do. In [27], the authors propose a hybrid BI solution aiming to enable interoperability for e-Government systems. In [28], data mining techniques are used in a use case of healthcare decisions to demonstrate that using such techniques may increase the quality of decisions. Finally, the authors of [29] present a multidimensional model created to support the financial department of local authorities.

Current State of the Research. As this study is based on the results of all the others, all the progresses in the other studies represent a progress in this one. Moreover, we are currently working on the exact scope of our SSBI solution for local authorities.

5. Conclusion

In this project, we want to tackle the problem of data profusion for decision making in the context of BI and SSBI, with a focus on the domain of local authorities. In order to achieve this global goal, we plan to investigate the concept of decision-making data value as selection criterion, the specific requirements of the local authorities domain for SSBI and a value-based SSBI framework designed specifically for local authorities. The different expected contributions have been presented along with methodological propositions and the related works.

Acknowledgments

I would like to thank the supervisors of this research project, Prof. Corentin Burnay and Prof. Isabelle Linden, for their reviews and support on this paper.

This project was initiated in collaboration with the company Loth-Info and is partially financed by the Fonds Spécial de Recherche (FSR).

References

- [1] A. J. Rowe, J. D. Boulgarides, M. R. McGrath, *Managerial decision making*, Citeseer, 1984.
- [2] P. Rikhardsson, O. Yigitbasioglu, *Business intelligence & analytics in management accounting research: Status and future focus*, *International Journal of Accounting Information Systems* 29 (2018) 37–58.
- [3] R. Raghuramapatruni, S. Kosuri, *The straits of success in a vuca world*, *IOSR Journal of Business and Management* 19 (2017) 16–22.

- [4] S. Ponde, A. Jain, Bi & bpr: Modern tools for performance management in vuca world, No. 22, Issue 87, APRIL-JUNE 2020 (2020) 202089.
- [5] S. Negash, P. Gray, Business intelligence, in: Handbook on decision support systems 2, Springer, 2008, pp. 175–193.
- [6] C. Elena, et al., Business intelligence, Journal of Knowledge Management, Economics and Information Technology 1 (2011) 1–12.
- [7] H. Chen, R. H. Chiang, V. C. Storey, Business intelligence and analytics: From big data to big impact, MIS quarterly (2012) 1165–1188.
- [8] W. W. Eckerson, Performance dashboards: measuring, monitoring, and managing your business, John Wiley & Sons, 2010.
- [9] P. Alpar, M. Schulz, Self-service business intelligence, Business & Information Systems Engineering 58 (2016) 151–155.
- [10] C. Lennerholt, J. van Laere, E. Söderström, Implementation challenges of self service business intelligence: A literature review, in: 51st Hawaii International Conference on System Sciences, Hilton Waikoloa Village, Hawaii, USA, January 3-6, 2018, volume 51, IEEE Computer Society, 2018, pp. 5055–5063.
- [11] L. Cai, Y. Zhu, The challenges of data quality and data quality assessment in the big data era, Data science journal 14 (2015).
- [12] H. J. Scholl, The egov research community: An update on where we stand, in: International Conference on Electronic Government, Springer, 2014, pp. 1–16.
- [13] G. V. Pereira, P. Parycek, E. Falco, R. Kleinhans, Smart governance in the context of smart cities: A literature review, Information Polity 23 (2018) 143–162.
- [14] H. J. Scholl, M. C. Scholl, Smart governance: A roadmap for research and practice, IConference 2014 Proceedings (2014).
- [15] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS quarterly (2004) 75–105.
- [16] A. R. Hevner, A three cycle view of design science research, Scandinavian journal of information systems 19 (2007) 4.
- [17] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS quarterly (1989) 319–340.
- [18] H. Van der Heijden, User acceptance of hedonic information systems, MIS quarterly (2004) 695–704.
- [19] J. Attard, R. Brennan, Challenges in value-driven data governance, in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, 2018, pp. 546–554.
- [20] K. Kannan, R. Ananthanarayanan, S. Mehta, What is my data worth? from data properties to data value, arXiv preprint arXiv:1811.04665 (2018).
- [21] R. Brennan, J. Attard, P. Petkov, T. Nagle, M. Helfert, Exploring data value assessment: a survey method and investigation of the perceived relative importance of data value dimensions, in: ICEIS 2019-21st International Conference on Enterprise Information Systems, SciTePress, 2019, pp. 200–207.
- [22] M. Bendeche, N. Sudhanshu Limaye, R. Brennan, Towards an automatic data value analysis method for relational databases (2020).
- [23] J. Attard, J. Debattista, R. Brennan, Saffron: a data value assessment tool for quantifying

the value of data assets (2019).

- [24] I. Elbeltagi, N. McBride, G. Hardaker, Evaluating the factors affecting dss usage by senior managers in local authorities in egypt, *Journal of Global Information Management (JGIM)* 13 (2005) 42–65.
- [25] R. A. Hackney, N. K. McBride, The efficacy of information systems in the public sector: issues of context and culture, *International Journal of Public Sector Management* (1995).
- [26] Z. Irani, P. E. Love, T. Elliman, S. Jones, M. Themistocleous, Evaluating e-government: learning from the experiences of two uk local authorities, *Information Systems Journal* 15 (2005) 61–82.
- [27] B. Oumkaltoum, et al., Toward a business intelligence model for challenges of interoperability in egov system: Transparency, scalability and genericity, in: *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, IEEE, 2019, pp. 1–6.
- [28] A. Mourady, A. Elragal, Business intelligence in support of egov healthcare decisions, in: *European, Mediterranean and Middle Eastern Conference on Information Systems: 30/05/2011-31/05/2011*, Information Systems Evaluation and Integration Group, 2011, pp. 285–293.
- [29] A. Costa, M. F. Santos, A. Abelha, A data warehouse schema to support financial process in local egov, in: *World Conference on Information Systems and Technologies*, Springer, 2017, pp. 360–366.