

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/383493160>

BPMN Diagram Dataset: Comprehensive Collection of Functional Requirements

Article in Data in Brief · August 2024

DOI: 10.1016/j.dib.2024.110882

CITATIONS

0

READS

26

2 authors:



Sholiq Sholiq

Institut Teknologi Sepuluh Nopember

45 PUBLICATIONS 287 CITATIONS

SEE PROFILE



Muhammad Ainul Yaqin

Universitas Islam Negeri Maulana Malik Ibrahim Malang

176 PUBLICATIONS 330 CITATIONS

SEE PROFILE



Data Article

BPMN diagram dataset: Comprehensive collection of functional requirements

Sholiq Sholiq^{a,*}, Muhammad Ainul Yaqin^b^a Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia^b Department of Informatics, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia

ARTICLE INFO

Article history:

Received 25 June 2024

Revised 6 August 2024

Accepted 21 August 2024

Available online 28 August 2024

Dataset link: [BPMN Diagrams Dataset of Functional Requirements \(Original data\)](#)*Keywords:*

Requirements

Software engineering

Software requirements

Business process model

ABSTRACT

The dataset in this paper explains the functional requirements in the form of Business Process Modelling Notation (BPMN) diagrams from three software applications: employee cooperative, store, and mini hospital. Data were collected from clients or owners of software in three business software projects developed by development teams (including the authors) through requirements elicitation activities. The functional requirements are presented as a BPMN diagram. There are 67 data objects, each presenting a single functional requirement or process. Apart from being presented in diagram form, the BPMN diagram of the functional requirements is also provided in the form of an XML Process Definition Language (XPDL) file, making it possible for users of this dataset to export it using business process management platform tools. Due to the rarity of functional requirements presented in the form of BPMN diagrams for research and education, this dataset plays an important role in research and education in the fields of requirements engineering, software engineering, software metrics, business process modelling, and other relevant fields.

© 2024 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

* Corresponding author.

E-mail address: sholiq@is.its.ac.id (S. Sholiq).Social media: [@yaqinov](#) (M.A. Yaqin)

Specifications Table

Subject	Computer Science.
Specific subject area	A more specific field could be software engineering, requirements engineering, and requirements in business process modelling diagram (BPMN).
Type of data	Graph, Text Analysed.
Data collection	Data were collected from clients or owners of software in three business software projects, an employee cooperative, store, and mini hospital, developed by development teams (including the authors) through requirements elicitation activities.
Data source location	Institution: Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/9tgnvj9nsg.2 Direct URL to data: https://data.mendeley.com/datasets/9tgnvj9nsg/2
Related research article	S. Sholiq, R. Sarno, and E. S. Astuti, "Generating BPMN diagram from textual requirements," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no 10, pp. 10,079–10,093, 2022, doi: https://doi.org/10.1016/j.jksuci.2022.10.007 .

1. Value of the Data

- This dataset fills a gap by containing real functional requirements from software applications in the form of BPMN diagrams.
- This dataset is valuable because it represents real functional requirements, enables cross-sector analysis, and supports solution development. Thus, it is an important learning resource in BPMN education, training, and business process management.
- These datasets are useful for software development firms, consultants, researchers, students, end users, and academics for solution design, requirement analysis, research, and education. Students and end users can also use it as a case study and reference to better understand information system requirements.
- For researchers in the fields of software requirements engineering, software metrics, and software engineering, these datasets are important for validation, comparative analysis, new model development, algorithm testing, software requirements analysis, software size estimation, functional complexity estimation, and other research topics.
- Researchers can validate, replicate, and develop functional requirements in BPMN diagrams. They can also perform comparative analysis between functional requirements presentation models with various diagrams, test algorithms, and datasets for research, education, and training.

2. Background

The motivation and initial context behind this dataset collection was driven by several important factors. First, there is an urgent need to improve the understanding and use of BPMN diagrams that were initially more widely used to visually model business processes are used for other things, namely, to illustrate the functional requirements of the software. This is intended to make it easier for software developers, researchers, and practitioners to identify and define the functional requirements of software in a clear and structured manner. The authors of Refs. [1,2] stated that BPMN diagrams could be used for software requirements modelling. Likewise, the authors in Ref. [3] noted that the use of BPMN and Use Case models did not have a significant difference. In fact, the BPMN model was better for requirements that contain sequential activities [3].

Second, accurately defining functional requirements is essential in software development. This dataset provides a real-world example of how functional requirements can be expressed through BPMN diagrams, making it easier for developers to accommodate user requirements and business processes in software design.

For related papers [4], Some of these datasets have been used as test data and validation against the proposed model. This dataset provides real-life examples of how the functional requirements of the software can be expressed through BPMN diagrams.

3. Data Description

This dataset is a collection of functional requirements from three business software applications, namely employee cooperative, store, and mini hospital in the form of BPMN diagrams. Each BPMN diagram represents a functional requirement or a functional process (FP). Software functionality comprises of functional processes [5,6]. In ref [7], functional requirements are expressed in user stories consisting of functional processes.

The BPMN diagram of a cooperative application illustrates the functional processes associated with cooperative application operations, such as user management, adding new members, making cash transactions, making loan repayments, and others (see Table 1). The BPMN diagram of a store application illustrates the functional processes that occur in retail operations, such as sales processes, making customer payments, purchase orders, making new suppliers, and others (see Table 1). BPMN diagrams from hospital applications include functional processes related to patient management, registration of patients, general inpatient registration, emergency room triage, and others (see Table 1). The BPMN components used are limited to the basic control structure, which can then be used to calculate the number of each component in the process model [8].

One example of FP i.e. Sale items from the application store (FP number 38 in Table 1) is presented in Fig. 1. The process can be described as follows:

1. An officer asks an employee to sell an item, and then the system displays a form indicating the sale item.
2. The officer enters an item code, and then the system queries the Goods database based on the item code. Then, the system displays the item name and price on the list.
3. If the added item is not completed, proceed to step 2
4. The officer enters buyer data, including customer type, payment type, and customer ID.
5. If the payment method is credit, the system calculates instalment.
6. The system queries the Customer database based on the customer ID, then the system displays the customer information.
7. If the save menu appears, the system saves the sales information.

Each FP in the dataset was presented in two forms: (1) BPMN diagrams in graphic form, as presented in Figs. 1, and 2 BPMN diagrams in XPDl file format. An XPDl file is a format standardized by the Workflow Management Coalition (WFMC) to swap business process definitions between different workflow products. XPDl defines an XML schema to define declarative parts of a workflow/business process. XPDl was developed to describe and exchange process models between various applications and their proprietary models [9,10].

The descriptive statistics of this dataset are given in Table 2. In the table, BPMN elements in the dataset include Node (activity and event), activity, arc, branch, cyclic, and data store. From the data provided, the number of nodes representing various points in the diagram varied between 5 and 24, with an average of 13.50. The number of activities, which included tasks or activities in process, ranged from 3 to 16, with an average of 8.24. Arc, which describes the relationship between nodes, ranges from 4 to 26, with an average of 14.34. Branches that appear branches in the diagram include XOR, OR, and AND have a minimum number of 0 and a maximum number of 3, with an average of 1.34. Loops, represented by cyclic elements, have a maximum of 1, and some diagrams do not include loops. In addition, the number of data stores representing data storage ranges from 1 to 7, with an average of 2.19.

This dataset can be used by various parties, including software development companies, consultants, researchers, academics, educators, students, and end users. These datasets can be leveraged for validation, comparative analysis, new model development, algorithm testing,

Table 1
Profile of dataset for BPMN diagram representing functional requirements.

Application name	Module name	No	Functional process	Σ node	Σ activity	Σ arc	Σ branch	Σ cyclic	Σ data store
Employee cooperative	Main	1	User management	15	9	16	1	1	1
	Main	2	Add new member	17	11	18	1	1	2
	Main	3	Make a cash transaction	14	8	15	2	0	1
	Main	4	Display member list	14	10	18	1	0	1
	Main	5	Change work unit	14	8	15	1	1	3
	Main	6	Make a report of the mandatory savings bill	12	8	13	1	0	2
	Main	7	Make a report of the mandatory savings bill	16	12	18	1	0	3
	Main	8	Make a report on the mandatory savings account book	14	10	14	1	0	2
	Main	9	Displays the history of moving work units	11	9	8	0	0	3
	Main	10	Login	12	6	13	1	1	1
	Main	11	Make cash inflow	10	6	10	0	1	1
	Main	12	Display mandatory savings	16	10	17	2	0	2
	Main	13	Make cash outflow	10	6	10	0	1	1
	Main	14	Resign member	14	8	15	1	1	2
	Main	15	Make member recap	11	7	12	1	0	1
	Main	16	Pay principal deposit	12	8	13	1	0	2
	Main	17	Update member status	11	7	11	1	0	1
	Main	18	Validate member mandatory deposit	20	12	22	3	0	2
	Main	19	Display cash inflow	10	6	10	1	0	1
	Main	20	Withdraw mandatory deposits	14	10	14	1	0	2
	Main	21	Display cash outflow	10	6	10	1	0	1
	SIPA	22	Make a bill report	13	7	14	2	0	2
	SIPA	23	Make a loan repayment	15	9	16	2	0	4
	SIPA	24	Make new member loan	14	10	14	1	0	3

(continued on next page)

Table 1 (continued)

Application name	Module name	No	Functional process	Σ node	Σ activity	Σ arc	Σ branch	Σ cyclic	Σ data store
Store	SIPA	25	Make a member loan payment	18	12	19	2	0	3
	SIPA	26	Realize a member loan	13	9	13	1	0	3
	SIPA	27	Make a list member loans	17	11	18	2	0	3
	Electronic	28	Make a new member loan	15	11	15	1	0	3
	Electronic	29	Make a new sales transaction	15	9	16	2	0	4
	Electronic	30	Make a bill report	14	8	15	2	0	2
	Cashier	31	Login	12	6	13	1	1	1
	Cashier	32	Makes goods return	14	8	15	1	1	2
	Cashier	33	Make recapitulation	13	9	14	1	0	1
	Cashier	34	Make customer payments	19	13	22	2	0	2
	Cashier	35	Make monthly reports	14	8	15	2	0	3
	Cashier	36	View items in store	7	5	6	0	0	1
	Main	37	Add a new item	11	7	11	1	0	1
	Main	38	Sale item	18	10	20	2	1	3
	Main	39	Update item	11	7	11	1	0	1
	Main	40	Purchase order	13	7	14	1	1	4
	Main	41	View items	16	8	18	3	0	1
	Main	42	Make a new item category	8	4	8	1	0	1
	Main	43	Make new rack	8	4	8	1	0	1
	Main	44	Make new unit	8	4	8	1	0	1
Main	45	Make new supplier	8	4	8	1	0	1	
Warehouse	46	Make new item entrance	15	7	17	2	1	3	
Warehouse	47	Setup initial stock	9	5	9	1	0	2	
Warehouse	48	Make new purchase	17	9	19	2	1	3	
Warehouse	49	View currently stock	9	5	9	1	0	2	
Warehouse	50	View items sold	13	7	14	2	0	2	
Mini Hospital	Front Office (FO)	51	Registration of old patient	12	8	12	1	0	7
	FO	52	General new patient registration	12	8	12	1	0	6
	FO	53	Corporate new patient registration	13	9	13	1	0	6

(continued on next page)

Table 1 (continued)

Application name	Module name	No	Functional process	Σ node	Σ activity	Σ arc	Σ branch	Σ cyclic	Σ data store
	FO	54	General inpatient registration	24	16	26	3	0	3
	FO	55	Corporate inpatient registration	24	16	26	3	0	3
	FO	56	Search for patient data	10	6	10	1	0	1
	FO	57	Looking for room data	8	6	7	0	0	1
	Cashier	58	General patient cashier	22	16	23	1	1	3
	Cashier	59	Down payment	15	7	17	2	1	1
	Cashier	60	Reprint hospitalization receipts	16	8	18	2	1	1
	Cashier	61	Outpatient claims	11	7	12	1	0	1
	Cashier	62	Daily revenue report	15	9	17	2	0	1
	Emergency Department (ED)	63	Updated Patient Information	12	8	12	1	0	6
	ED	64	View emergency room action reports	15	9	17	2	0	1
	ED	65	Emergency room Triage	21	9	25	3	0	3
	ED	66	View today's triage results	5	3	4	0	0	3
	ED	67	Register referral patients	17	9	20	2	1	3

Table 2

Descriptive statistics of the dataset of the BPMN diagram of functional process.

BPMN elements	Min	Max	Range	Average
Node	5	24	5–24	13,50
Activity	3	16	3–16	8,24
Arc	4	26	4–26	14,34
Branch	0	3	0–3	1,34
Cyclic	0	1	0–1	0,26
Datastore	1	7	1–7	2,19

training, and education. Thus, this BPMN diagram dataset has significant value in understanding, analyzing, and improving software business processes in cooperatives, retail, hospitals, and other sectors.

4. Experimental Design, Materials and Methods

The dataset in the form of BPMN diagrams was obtained through a series of activities, namely requirements analysis, BPMN diagram construction, and product owners' review and validation, as presented in Fig. 2. The resulting artefacts included user stories, a list of requirements, BPMN diagrams, and validated BPMN diagrams. The dataset in this paper is a validated BPMN diagram that was reviewed together between the development team (authors) and the product owner or client.

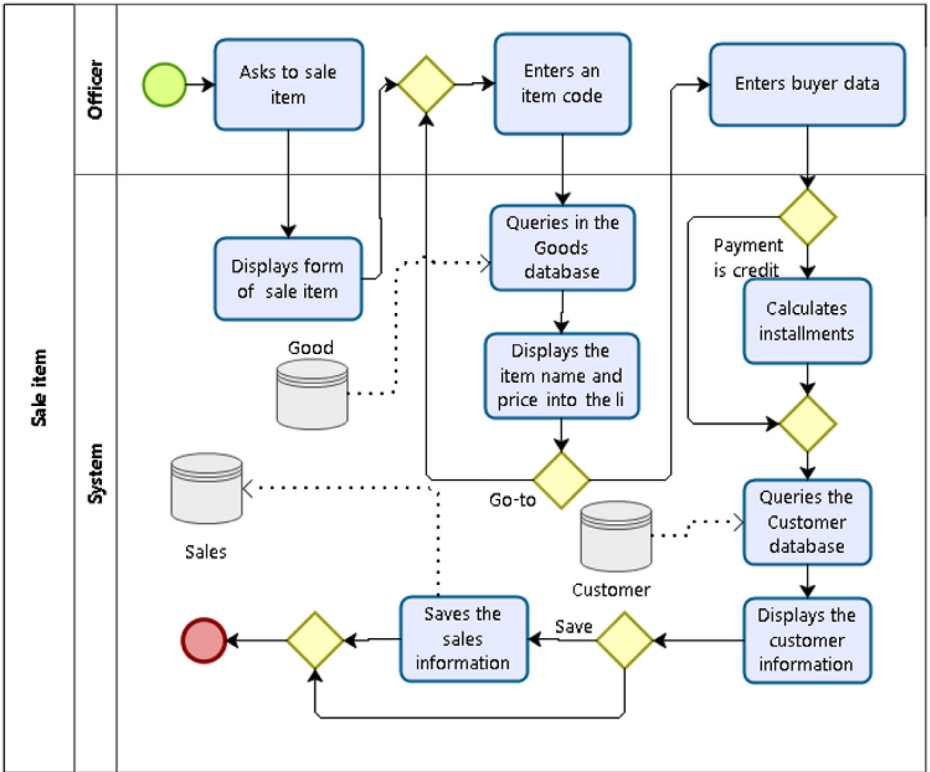


Fig. 1. Functional process for sale items in the store application.

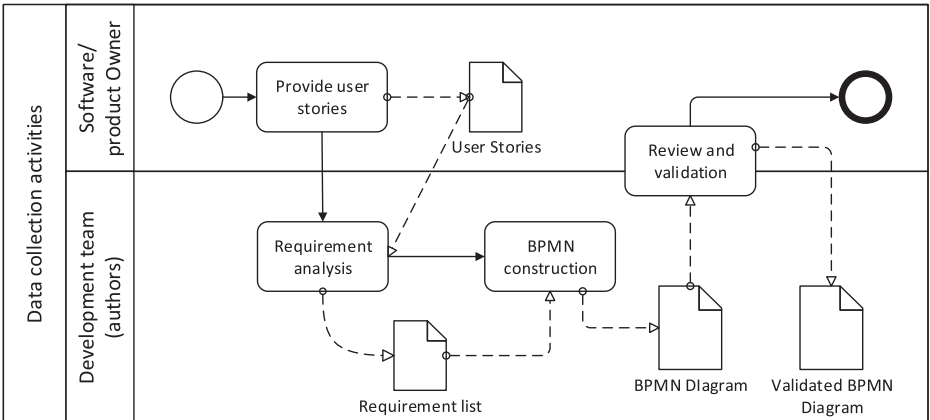


Fig. 2. BPMN diagram of the data collection process.

A brief description of the activities and artefacts in 2 is given below.

1. User stories.
2. The stage where user stories are obtained from product owners or clients focuses on understanding the requirements and expectations of the end user for the software application to be created. User stories are short semi-structured sentences that describe requirements from

a user's perspective and can be used to provide a concise explanation of user expectations [11]. User stories can be expressed in one of the following simple formats [12]:

- As a [aspect of who], I want [aspect of what] so that [aspect of why].
- As a [aspect of who], I need [aspect of what], so that [aspect of why].
- As a [aspect of who], I can [aspect of what], so that [aspect of why], or
- In order to [aspect of why] as a [aspect of who], I can [aspect of why].

It helps identify tasks and expected results of the system. Collaboration with stakeholders, such as product owners and operational users, is critical for uncovering the details of the process, including the applicable steps, actors, and rules. This information is the basis for creating accurate BPMN diagrams following the reality of business operations.

3. Requirement Analysis

The requirement analysis stage involves identifying, understanding, and documenting the requirements of user stories provided by the product owner. The development team works closely with the business owner or client to gather information through in-depth interviews, business process observations, and document reviews. The collected information is organized and translated into a list of requirements that must be implemented in the software system. For the author, the next list of requirements is detailed in the form of a BPMN diagram.

4. BPMN Construction

The BPMN construction stage is a concretization process in which information and analysis are translated into clear and structured BPMN diagrams. The development team uses the BPMN standard to identify important elements from the list of requirements to be detailed using BPMN elements, such as activities, gateways, events, and workflows. These elements are arranged in a diagram using BPMN software to illustrate the sequence and relationships between activities. Each element is labelled with additional details for clarity. Gateways and events mark important decisions and events. This process involves continuous iteration and collaboration with stakeholders to ensure accurate and comprehensive diagrams. This activity produces a BPMN diagram, which is a functional process of the functional requirements of the software to be built [4].

5. Review and Validation

The review and validation stage in the preparation of the BPMN diagram dataset is a crucial step where key stakeholders, including end users and product owners, together with the development team examine and evaluate the BPMN diagram that has been produced. This is intended to ensure its suitability to operational realities and business requirements. Product owners and users provide feedback on the accuracy, completeness, and usability of the diagram, assessing whether all activities, decisions, and workflows are depicted correctly and are easy to understand. Interactive discussions are frequent and require adjustments based on feedback. The primary goal is to ensure that BPMN diagrams reflect existing business processes and can be used for software system development. The result of the activity at this stage is a validated BPMN diagram presented in this paper.

Limitations

No limitations.

Ethics Statement

We have read and followed the ethical requirements for publication in the Data in Brief. We state that this study did not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Sholih: conceptualized, wrote draft, methodology, collected data, analysed, and created BPMN diagrams. **Muhammad Ainul Yaqin:** Investigation, review and validation of BPMN diagrams, and review and editing.

Data Availability

[BPMN Diagrams Dataset of Functional Requirements \(Original data\)](#) (Mendeley Data).

Acknowledgements

We are happy to thank the Directorate of Research and Community Service of the Institut Teknologi Sepuluh Nopember in Surabaya, Indonesia. This research was supported by the ITS 2024 Scientific Research scheme with contract no 1181/PKS/ITS/2024.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] C. Monsalve, A. April, A. Abran, On the expressiveness of business process modeling notations for software requirements elicitation, in: IECON Proceedings of the (Industrial Electronics Conference, 2012, pp. 3132–3137, doi:[10.1109/IECON.2012.6389398](https://doi.org/10.1109/IECON.2012.6389398).
- [2] A. Przybyłek, A business-oriented approach to requirements elicitation, in: *ENASE 2014 - Proceedings of the 9th international conference on Software Engineering*, 2014, pp. 152–163, doi:[10.5220/0004887701520163](https://doi.org/10.5220/0004887701520163).
- [3] O.L. Vega-Márquez, J. Chavarriaga, M. Linares-Vásquez, M. Sánchez, Requirements Comprehension Using BPMN: An Empirical Study, in: D. Lübke, C. Pautasso (Eds.), *Empirical Studies on the Development of Executable Business Processes*, Springer, Cham, 2019, doi:[10.1007/978-3-030-17666-2_5](https://doi.org/10.1007/978-3-030-17666-2_5).
- [4] S. Sholih, R. Sarno, E.S. Astuti, Generating BPMN diagram from textual requirements, *J. King Saud Univ. - Comput. Inf. Sci.* 34 (10) (2022) 10079–10093, doi:[10.1016/j.jksuci.2022.10.007](https://doi.org/10.1016/j.jksuci.2022.10.007).
- [5] C. Committe, Introduction to the COSMIC method for measuring software, The Common Software Measurement International Consortium (COSMIC), 2019 [Online]. Available:www.cosmic-sizing.org.
- [6] K. Khattab, H. Elsayed, H. Soubra, Functional size measurement of quantum computers software, in: *CEUR Workshop Proceedings*, 3272, 2022.
- [7] T. Hacıoğlu, O. Demirors, An exploratory case study using events as a software size measure, *Inf. Technol. Manag.* 24 (3) (2023) 293–312, doi:[10.1007/s10799-023-00394-y](https://doi.org/10.1007/s10799-023-00394-y).
- [8] M.A. Yaqin, R. Sarno, S. Rochimah, Measuring scalable business process model complexity based on basic control structure, *Int. J. Intell. Eng. Syst.* 13 (6) (2020) 52–65, doi:[10.22266/ijies2020.1231.06](https://doi.org/10.22266/ijies2020.1231.06).
- [9] A. Haller, M. Marmolowski, E. Oren, W. Gaaloul, A process ontology for business intelligence, *Business* (2008) 19 [Online]. Available:<http://sib.deri.ie/fileadmin/documents/DERI-TR-2008-04-01.pdf>.
- [10] K.P. Kim, An XPDL-based workflow control-structure and data-sequence analyzer, *KSII Trans. Internet Inf. Syst.* 13 (3) (2019) 1702–1721, doi:[10.3837/tiis.2019.03.034](https://doi.org/10.3837/tiis.2019.03.034).
- [11] I.K. Raharjana, F. Harris, A. Justitia, Tool for generating behavior-driven development test-cases, *J. Inf. Syst. Eng. Bus. Intell.* 6 (1) (2020) 27, doi:[10.20473/jisebi.6.1.27-36](https://doi.org/10.20473/jisebi.6.1.27-36).
- [12] I.K. Raharjana, D. Siahaan, C. Fatichah, User stories and natural language processing: a systematic literature review, *IEEE Access* 9 (2021) 53811–53826, doi:[10.1109/ACCESS.2021.3070606](https://doi.org/10.1109/ACCESS.2021.3070606).