Data in Brief 57 (2024) 110906

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data of dwelling time process at container terminal: Multi-perspective based dataset

Hanung Nindito Prasetyo[a,b], Riyanarto Sarno[a,*], Dedy Rahman Wijaya[b], Raden Budiraharjo[c], Indra Waspada[d], Kelly Rossa Sungkono[a]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[b] Department of Information System Diploma, Telkom University, Bandung, Indonesia
[c] Department of Information Systems, Institut Teknologi Nasional, Bandung, Indonesia
[d] Department of Informatics, Universitas Diponegoro, Semarang, Indonesia

ARTICLE INFO

Article history:
Received 10 November 2023
Revised 27 August 2024
Accepted 28 August 2024
Available online 4 September 2024

Dataset link: Event Log Dwelling Time Dataset (Original data)

Keywords:
Dwelling time dataset
Containers
Event logs
Multi-perspective

ABSTRACT

We present the Event log Dataset on the Container Dwelling Time process at the Container Terminal. The dwelling time process is the process and time calculated from the time the container is unloaded and lifted from the ship until the container leaves the port terminal through the main door. Dwelling time processes at ports are currently managed by systems in almost all major ports in the world. The mechanism for withdrawing the Container dataset is carried out for 6 months from the web server in the Container Terminal information system. The event log recording data that we took was data for 3 months of the dwelling time process that occurred at the port. We anonymized several parts of the data to prevent data leaks due to ethical practices that have been agreed with the institution that owns the dataset. The dataset obtained can be used to develop research related to process science such as process mining which includes process discovery, conformance checking, enhancement or research related to process anomaly detection. The resulting dataset is not only the basis for variables in the mining process such as Case_ID, Activities, and Timestamp, but we also complete other indicators such as workers, waiting time and costs as a multi-perspective point of view. The dataset we

* Corresponding author.
E-mail address: riyanarto@if.its.ac.id (R. Sarno).

https://doi.org/10.1016/j.dib.2024.110906

collect will be very helpful and will be liked by other researchers who want to develop knowledge in the field of Process Science, especially process mining.

Specifications Table

Subject	Dwelling Time Process at the Container Terminal
Specific subject area	Traffic network inside the Container Terminal.
Type of Data	Event log
How data were acquired	Data is obtained during normal operations within the Container Terminal information system network. Collection of traffic network and port information system web server logs in a three months period, namely July, August, and September 2021 where the collection dates are November 4 and 9 2021
Format Data	It has been normalized in the form of an Event log which can be used in process mining research
Description of data collection	Determine the variables needed in research by setting indicators. After understanding the required indicators, Network traffic is captured by the server operator. Activity traffic records are captured as trace packets by setting a recording capture period limit of three months. The raw data is then processed with the Pandas application to obtain the event logs needed for research.
Data source location	Container Terminal
Data accessibility	The event log data used to support the findings of this study are deposited in the Mendeley Data repository (DOI: 10.17632/yvp2b4rtp3.1)
Direct link to the dataset:	https://data.mendeley.com/datasets/yvp2b4rtp3/1

1. Value of The Data

- In general, event log datasets that record process activities are quite limited and not widely available in open repositories accessible to the public, especially regarding dwelling time in ports. This event log dataset for process dwelling time is not only limited to the main attributes needed for process mining research, such as transaction ID, activity, and timestamp, but also includes additional attributes related to the recorded activities, thereby enriching the dataset source.
- This event log dataset can be used in process mining research aimed at improving process modelling algorithms, measuring efficiency and optimizing processes, detecting bottlenecks in processes, understanding process behavior related to dwelling time, comparing workloads, and modelling and predicting aspects related to dwelling time at the Container Terminal in the port.
- Researchers and students can utilize the event log dataset on process dwelling time for broader research areas such as process science using machine learning or deep learning approaches, compliance, and process auditing. It can also be used to monitor the performance of IT systems at port terminals, such as server response times, CPU usage, and user interactions related to dwelling time processes at the Container Terminal.

2. Background

The geographical conditions of a country, especially countries with waters that are much wider than their land, certainly pay great attention to sea transportation. Ports, as part of maritime transportation and economic gateways, absolutely must be able to contribute, among other things, to reducing distribution costs, which have an impact on purchasing power, competitive-

ness, and multiplier effects on the growth and national income of a country that is directly connected to the sea. For this reason, the length of the loading and unloading process at the port is one of the causes of inefficiency in a country's economy. The container loading and unloading process, or what is commonly referred to as the dwelling time process, plays a very big role in supporting the smooth running of sea transportation. Dwelling time is the time required from the time the goods leave the ship or the goods are stockpiled until the goods leave the port [1]. Basically, all ports in the world want the dwelling time process to be carried out effectively and efficiently. To make this happen, a lot of research is needed, which is supported by data that supports this condition.

3. Data Description

The dwelling time process is the process and time that is calculated from the time the container is unloaded and lifted from the ship until the container leaves the port terminal through the main door [1]. All activities that occur in the dwelling time process are recorded in the information system used by the container terminal agency. All interactions that occur during the dwelling time process are recorded as events and stored on the web server. Activities before the container disembarks from the ship until the container leaves the port terminal are recorded in a system such as Importer ID, Different activities carried out by containers via the red route, green route or quarantine, and custom are recorded via timestamp recording. Attributes such as officers performing activities and costs can be captured and pulled into raw data. The raw data obtained is then converted and anonymized to obtain an event log dataset according to related research needs.

The description of the Event Log dataset which is the final result is generally shown in Table 1.

The total number of attributes is 12. Generally, the event log attribute consists of Case_ID, Activity, Start Timestamp, End Timestamp, and duration. However, in addition to the main attributes in process mining, the dataset also comes with other attributs such as CTR_SIZE, CTR _TYPE, Gross, Yard_Block, yard_Slot, Document Type, and Role. CTR_SIZE is the size of the container that enters the Port. CTR_TYPE consists of two types of RFR and DRY. RFR (refrigerated container) is a type of container designed to transport goods that require temperature control or cooling such as food, medicines, and temperature-sensitive chemicals. while DRY (Dry Container) is a container used to load and ship various types of cargo in dry conditions. Then the gross weight attribute is the total weight of the container and the goods in it. For the attribute Yard_Block is a part of a yard container that is limited/divided to facilitate the setting of the container layout. Whereas Yard _Slot is part of the yard block to set the line of container placement. There is a JOB_DEL_DOCTYPE attribute that contains the container document type. Then the Role

Table 1
Dwelling Time Process Event Log Description.

No	Description Item	Total
Overview		
1	Events	952,069
2	Cases	95,648
3	Activities	20
4	Attribute	12
Activities		
5	Minimal Frequency	18
6	Median Frequency	26,940
7	Mean Frequency	47,603.45
8	Maximal Frequency	95,648
9	Frequency Std. Deviation	45,089.45

attribute is an Attribute that indicates the role or staff involved in an activity. An example of an attribute explanation can be seen in Table 2.

From Table 2, it can be seen that Case_ID with container number ACCU2049701-2021-07-13 has a size of 20 feet, the type of container is DRY which has a gross of 30.08 tons, placed in Yard Block J in slot 111, the accompanying document is SPPB, recorded activities such as BALIE, VESSEL_ATB etc. which are equipped with start Timestamp and end Timestamp for each activity. Then the duration attribute shows the duration of each activity, and who the officers are who are involved in the activity. The number of activities in the process detected was 20 activities with the proportions as shown in Table 3.

The dataset provided is the basis for ongoing research development such as research in detecting anomalies in processes [2,3], risk measurement in the dwelling time process [4], and business process collaboration optimization [5].

3.1. Data format

The data format that is generally used in datasets is in the form of an Event Log. The basic required attributes are user identity, activity and timestamp. Then we added multi-perspective attributes recorded in the process, namely the role of officers and costs. Specifically for user identity and officer roles, the data has been anonymized.

3.2. Repository data structure

The Data Repository contains data files, namely **Anonymized Dwelling Time Dataset** - the dataset has been anonymized for several attributes. The dataset consists of 1 file, the file has the extension .xlsx

4. Experimental Design, Materials and Methods

The methods we use in capturing, processing and publishing datasets are carried out for the purposes of our research. The methods we use are data pre-processing techniques especially in the data integration and data transformation section. The first step is to provide a dataset that contains the information needed for the research we are conducting and potential future research. The event log dataset is generated and available through the process of converting raw data into an event log dataset with basic indicators [6] and additional indicators in multi-perspective form [7]. The second important step is to anonymize the process of converting the dataset. This mechanism can be seen in Fig. 1.

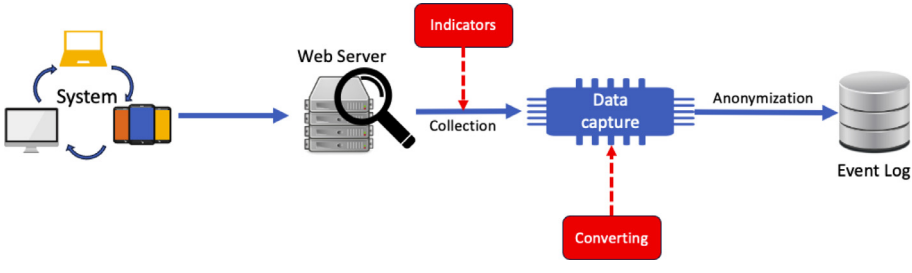

Fig. 1. Data processing scheme - capture, collection, conversion, and anonymization.

Table 2
Event Log Dataset View of dwelling time process.

Case_id	CTR SIZE	CTR TYPE	GROSS	YARD BLOCK	YARD SLOT	JOB DEL DOCTYPE	activity	Start Timestamp	End Timestamp	Duration	Role
ACCU2049701-2021-07-13	20	DRY					BAPLIE	13/07/21 15.22	13/07/21 16.51	01.28.16	EDI BAPLIE
ACCU2049701-2021-07-13							VESSEL_ATB	13/07/21 19.25	13/07/21 20.20	00.54.03	Dock spv
ACCU2049701-2021-07-13			30.08	J	111		DISCHARGE	13/07/21 21.54	13/07/21 23.25	01.30.15	tally, CC OP
ACCU2049701-2021-07-13							STACK	13/07/21 23.43	14/07/21 00.16	00.32.13	RTG OP
ACCU2049701-2021-07-13							CUSTOMS_DEL	18/07/21 22.55	19/07/21 00.00	01.04.54	Custom Staff
ACCU2049701-2021-07-13						SPPB	JOB_DEL	21/07/21 08.44	21/07/21 09.53	01.08.03	CT Staff

Table 3
Proportion of activity frequency in the Dwelling Time process.

Activity	Frequency	Relative Frequency
Baplie	95,648	10.050%
Vessel_ATB	95,648	10.050%
Discharge	95,648	10.050%
Customs_Del	95,648	10.050%
Truck_In	95,647	10.050%
Truck_Out	95,647	10.050%
Stack	95,624	10.040%
Unstack_To_Truck	95,540	10.030%
Job_Del	95,147	9.990%
Has_Quarantine_Flag	26,940	2.830%
Quarantine_Release	26,940	2.830%
First_Stack_Quarantine_Blk	13,639	1.430%
Sp_Red_Flag_Pib	5,155	0.540%
First_Stack_Behandle	4,032	0.420%
First_Job_Customs_Behandle	3,924	0.410%
Last_Complete_Behandle	3,924	0.410%
Stack_Main_Yard	3,844	0.400%
Stack_Trk_Bf_Main_Yard	2,951	0.310%
Job_Move_Yard_Pile_Up	505	0.050%
Behandle_Quarantine_Job	18	0.000%

4.1. Data collection

Collection data is taken from the Port Container Terminal network server. The entire dwelling time process goes through a national scale port information system and is connected to the National Single Windows (NSW) system. A diagram of the infrastructure is shown in Fig. 1. Data collection is monitored and stored directly on the individual web server in the local log. Data withdrawal time is carried out at the time when processing transactions are lowest.

The result of the data withdrawal stage is Raw Data. The mechanism for withdrawing data from the data network is carried out on a limited basis from the discussion stage, determining indicators, the period for which the dataset is withdrawn, until the raw data withdrawal process occurs for approximately 6 months. This mechanism is carried out from 4 June 2021 to 9 November 2021. Raw data obtained from the network data system is stored in 3 files. The 3 files produced are monthly data withdrawals, namely data for July, August, and September 2021. The captured raw data contains all the information needed to form the desired dataset. The raw data obtained in 3 files was then converted into a combined dataset, then using the Pandas application.

4.2. Data conversion

When carrying out data conversion, there are several provisions that need to be considered to change raw data so that it meets XES or CSV standards as a form of event log dataset which is commonly used in process mining analysis. The thing to note is that each record is event data (from a case), each event is collected sequentially in one case, each case. contains only active events (activities), each case must have a unique ID, it is necessary to distinguish between data with case coverage and data that is only generated by certain events. In this section, monthly data is aggregated and converted into a Pandas Dataframe. The steps taken are:

(1) Combine monthly data and convert it into a Pandas Dataframe.
Technical steps as shown in Fig. 2.

```
df_07 = pd.read_excel('./data/tps_Juli2021.xls', delimiter=";")
df_08 = pd.read_excel('./data/tps_Agustus2021.xls', delimiter=";")
df_09 = pd.read_excel('./data/tps_September2021.xls', delimiter=";")
```

```
# vertical concat
df_all = pd.concat([df_07, df_08, df_09], axis=0)
```

Fig. 2. Data Integration via Pandas Dataframe.

```
df_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 95648 entries, 0 to 32338
Data columns (total 28 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   CONTAINER_KEY                   95648 non-null  object
 1   CTR_SIZE                        95648 non-null  int64
 2   CTR_TYPE                        95648 non-null  object
 3   GROSS                           95648 non-null  float64
 4   VESSEL_ATB                      95648 non-null  datetime64[ns]
 5   BAPLIE_TS                       95648 non-null  datetime64[ns]
 6   DISC_DATE                       95648 non-null  datetime64[ns]
 7   YARD_BLOCK                      95624 non-null  object
 8   YARD_SLOT                       95406 non-null  float64
 9   STACK_DATE                      95624 non-null  datetime64[ns]
 10  CUSTOMS_DEL_DATE                95648 non-null  datetime64[ns]
 11  HAS_QUARANTINE_FLAG             26940 non-null  datetime64[ns]
 12  BEHANDLE_QUARANTINE_JOB_TS      18 non-null     datetime64[ns]
 13  FIRST_STACK_QUARANTINE_BLK_TS   13639 non-null  datetime64[ns]
 14  QUARANTINE_RELEASE              26940 non-null  datetime64[ns]
 15  JOB_PLP_TS                      505 non-null    datetime64[ns]
 16  JOB_DEL_DATE                    95147 non-null  datetime64[ns]
 17  JOB_DEL_DOCTYPE                 95647 non-null  object
 18  TRUCK_IN_DATE                   95647 non-null  datetime64[ns]
 19  UNSTACK_TO_TRUCK                95540 non-null  datetime64[ns]
 20  TRUCK_OUT_DATE                  95647 non-null  datetime64[ns]
 21  SP_JALUR_MERAH_PIB_DATE         5155 non-null   datetime64[ns]
 22  CUSTOMS_BEHANDLE_COUNT          95648 non-null  int64
 23  FIRST_JOB_CUSTOMS_BEHANDLE_TS   3924 non-null   datetime64[ns]
 24  FIRST_STACK_BEHANDLE_TS         4032 non-null   datetime64[ns]
 25  LAST_COMPLETE_BEHANDLE_TS       3924 non-null   datetime64[ns]
 26  STACK_TRK_BF_MAIN_YARD_TS       2951 non-null   datetime64[ns]
 27  STACK_MAIN_YARD_TS              3844 non-null   datetime64[ns]
dtypes: datetime64[ns](20), float64(2), int64(2), object(4)
memory usage: 19.7+ MB
```

Fig. 3. Information on merged data in Pandas Dataframe.

Fig. 2 shows the process carried out to unify previously separated datasets in the months of July, August, and September. The information obtained from the combined data includes attribute names and data types as in Fig. 3.

(2) Based on the merge results, CONTAINER_KEY cannot be used as case_ID because the same container can arrive at different times so it is not unique. To make it unique it is necessary to combine CONTAINER_KEY with the arrival date.
These steps are as shown in Fig. 4.

(3) The next step is to perform a mechanism to select an activity that has a timestamp and separate an activity without a time stamp in the sense that this condition is not activated. Here's the mechanisms in the phyton with the pandas:
   a. Initialize an empty list called Event Log.
   b. Iterate through each case in df activities.
   c. Initialize an empty list called Case Fix to store the data to be processed.

```
df_seleksi.head()
```

	ID	CONTAINER_KEY	CTR_SIZE	CTR_TYPE	GROSS	YARD_BLOCK	YARD_SLOT	JOB_DEL_DOCTYPE	VESSEL_ATB	BAPLIE	...	JOB_DEL	T
0	AAAU9001220-2021-06-28	AAAU9001220	40	DRY	12.20	N	98.0	SPPB	2021-06-28 20:00:00	2021-06-27 11:09:00	...	2021-07-05 17:13:00	
1	AAMU4003031-2021-06-20	AAMU4003031	40	RFR	31.30	Q	8.0	SPPB	2021-06-20 12:45:00	2021-06-19 14:50:00	...	2021-06-29 15:56:00	
2	AAMU4004969-2021-07-07	AAMU4004969	40	RFR	33.70	B	8.0	SPPB	2021-07-07 15:50:00	2021-07-06 19:09:00	...	2021-07-08 21:34:00	
3	AAMU4008291-2021-07-07	AAMU4008291	40	RFR	33.70	B	18.0	SPPB	2021-07-07 15:50:00	2021-07-06 19:09:00	...	2021-07-08 21:34:00	
4	ACCU2049701-2021-07-13	ACCU2049701	20	DRY	30.08	J	111.0	SPPB	2021-07-13 20:20:00	2021-07-13 16:51:00	...	2021-07-21 09:53:00	

Fig. 4. Determines the unique Case_ID of the Process Dwelling Time case.

```
# MEnyeleksi aktifitas yang memiliki timestampn (aktifitas tanpa timestamp artinya tidak diaktivasi)
col = column # id and activities
eventLog = []
for i in range(len(df_activities)):
    case = df_activities.iloc[i] # telusuri case i (berisi timestamp)
    case_fix = []
    for j in range(len(case)):
        ts = case[j] # timestamp tiap activity
        if isinstance(ts, pd.Timestamp): # hanya ambil yang ada timestamp nya
            case_fix.append({'activity': col[j+8], 'timestamp': ts}) # +8 karena skip kolom id dan data
    eventLog.append(case_fix)
# eventLog
```

```
eventLog[3]
```

```
[{'activity': 'VESSEL_ATB', 'timestamp': Timestamp('2021-07-07 15:50:00')},
 {'activity': 'BAPLIE', 'timestamp': Timestamp('2021-07-06 19:09:00')},
 {'activity': 'DISCHARGE', 'timestamp': Timestamp('2021-07-08 11:46:00')},
 {'activity': 'STACK', 'timestamp': Timestamp('2021-07-08 12:31:00')},
 {'activity': 'CUSTOMS_DEL', 'timestamp': Timestamp('2021-07-08 00:00:00')},
 {'activity': 'HAS_QUARANTINE_FLAG',
  'timestamp': Timestamp('2021-07-07 10:50:00')},
 {'activity': 'QUARANTINE_RELEASE',
  'timestamp': Timestamp('2021-07-08 21:22:00')},
 {'activity': 'JOB_DEL', 'timestamp': Timestamp('2021-07-08 21:34:00')},
 {'activity': 'TRUCK_IN', 'timestamp': Timestamp('2021-07-08 23:13:00')},
 {'activity': 'UNSTACK_TO_TRUCK',
  'timestamp': Timestamp('2021-07-08 23:22:00')},
 {'activity': 'TRUCK_OUT', 'timestamp': Timestamp('2021-07-08 23:28:00')}]
```

Fig. 5. Activity selection results.

d. Iterate through each ts in the case.
e. If ts contains timestamp data, create a dictionary with keys 'activity' (the value from the corresponding column) and 'timestamp' (the value of ts), then append it to Case Fix.
f. Append Case Fix to Event Log.
g. Iterate through each case in Event Log.
h. Sort the case based on the 'timestamp' value.
i. Initialize an empty list called Case Fix to store the data to be processed.
j. Iterate through each event in the case.
k. Append a list containing 'activity' and 'timestamp' of the event to Case Fix.
l. Append Case Fix to Event Log fix.
Advanced mechanisms and result data sets as shown in Figs. 5 and 6.
Then the Event log results in each case must be sorted based on timestamp as shown in the steps in Fig. 6.

```
# sort berdasarkan timestamp

def myFunc(e):
    return e['timestamp']
eventLog_fix = []
for case in eventLog:
    case.sort(key=myFunc) # urutkan aktifitas2 dalam case
    case_fix = []
    for event in case:
        case_fix.append([event['activity'], event['timestamp']])
    eventLog_fix.append(case_fix)
```

```
eventLog_fix[2] # case ke-x berisi event-event utk case ke-x
```

```
[['BAPLIE', Timestamp('2021-07-06 19:09:00')],
 ['HAS_QUARANTINE_FLAG', Timestamp('2021-07-07 10:50:00')],
 ['VESSEL_ATB', Timestamp('2021-07-07 15:50:00')],
 ['DISCHARGE', Timestamp('2021-07-07 17:32:00')],
 ['STACK', Timestamp('2021-07-07 18:21:00')],
 ['CUSTOMS_DEL', Timestamp('2021-07-08 00:00:00')],
 ['QUARANTINE_RELEASE', Timestamp('2021-07-08 21:22:00')],
 ['JOB_DEL', Timestamp('2021-07-08 21:34:00')],
 ['TRUCK_IN', Timestamp('2021-07-08 23:21:00')],
 ['UNSTACK_TO_TRUCK', Timestamp('2021-07-08 23:39:00')],
 ['TRUCK_OUT', Timestamp('2021-07-08 23:45:00')]]
```

Fig. 6. Ordering Each Case_ID.

```
# Langkah akhir menyusun Event Log mengacu format XES
col = column
case_id_fix = [item for sublist in df_case_id.values.tolist() for item in sublist] # case id
# data_fix = [sublist for sublist in df_data.values.tolist()] # tiap data dalam list
data_baplie = [sublist for sublist in df_data_baplie.values.tolist()] # tiap data dalam list
data_discharge = [sublist for sublist in df_data_discharge.values.tolist()] # tiap data dalam list
data_job_del = [sublist for sublist in df_job_del.values.tolist()] # tiap data dalam list
data_other = [sublist for sublist in df_data_other.values.tolist()] # tiap data dalam list

eventLog_fix = eventLog_fix

eventLog_final = []
for i in range(len(eventLog_fix)): # jumlah case, untuk tiap case
    case_final = []
    for events in eventLog_fix[i]: # ['BAPLIE_TS', Timestamp('2021-04-27 15:30:00')] --> case ke-i
#         event = []
        event = [case_id_fix[i]] # +case id
#         event.extend(data_fix[i]) # +kolom2 data level case
        if events[0] == 'BAPLIE':
            event.extend(data_baplie[i]) # tambahkan data pada baplie
        elif events[0] == 'DISCHARGE':
            event.extend(data_discharge[i]) # tambahkan data pada discharge
        elif events[0] == 'JOB_DEL':
            event.extend(data_job_del[i]) # tambahkan data pada job_del
        else:
            event.extend(data_other[i]) # aktivitas tanpa data kosong
        for e in events: # +activity dan timestamp ['BAPLIE_TS', Timestamp('2021-04-27 15:30:00')] --> tiap event dalam case ke-i
            event.append(e)
        case_final.append(event) # tambahkan event ke case ['CAIU3407429-2021-04-27 07:10:00','CAIU3407429',20,'DRY',3.03,'N',61.0
    eventLog_final.append(case_final)
```

Fig. 7. Embed data into appropriate activities.

The final step is to embed the data into the appropriate activity and add column names so that a complete final result is obtained as shown in Figs. 7 and 8. Then export the event log data into CSV format.

(4) The next step is to anonymize the dataset to become public data.

```
data = [activity for event in eventLog_final for activity in event]
df_eventlog_final = pd.DataFrame(data)
df_eventlog_final.columns = ['case_id', 'container key','CTR_SIZE','CTR_TYPE','GROSS','YARD_BLOCK','YARD_SLOT','JOB_DEL_DOCTYPE','
```

```
df_eventlog_final[:20]
```

	case_id	container key	CTR_SIZE	CTR_TYPE	GROSS	YARD_BLOCK	YARD_SLOT	JOB_DEL_DOCTYPE	activity	timestam
0	AAAU9001220-2021-06-28	AAAU9001220	40.0	DRY	NaN	None	NaN	None	BAPLIE	2021-0 11:09:(
1	AAAU9001220-2021-06-28	AAAU9001220	NaN	None	NaN	None	NaN	None	VESSEL_ATB	2021-0 20:00:(
2	AAAU9001220-2021-06-28	AAAU9001220	NaN	None	12.2	N	98.0	None	DISCHARGE	2021-0 22:22:(
3	AAAU9001220-2021-06-28	AAAU9001220	NaN	None	NaN	None	NaN	None	STACK	2021-0 22:40:(
4	AAAU9001220-2021-06-28	AAAU9001220	NaN	None	NaN	None	NaN	None	HAS_QUARANTINE_FLAG	2021-0 09:06:(
5	AAAU9001220-2021-06-28	AAAU9001220	NaN	None	NaN	None	NaN	None	FIRST_STACK_QUARANTINE_BLK	2021-0 01:43:(
6	AAAU9001220-2021-06-28	AAAU9001220	NaN	None	NaN	None	NaN	None	CUSTOMS_DEL	2021-0 00:00:(

Fig. 8. Complete data with column names ready to be exported in CSV format.

4.3. Data anonymization

After obtaining the Event log dataset that suits your needs. We see that data anonymization steps are needed to avoid internal data leaks and an ethical agreement that is our agreement with the institution that owns the process. Based on the ethical agreement that is the agreement, it is not permitted to mention the name of the agency that owns the process. Then, if there is a personal name in the process recorded in the event log, a mechanism for changing the personal name to a role assignment is needed. The anonymization mechanism is our part in respecting Privacy. Privacy is a fundamental human right, as stipulated in the United Nations Declaration of Human Rights (UDHR) 1948 [8].

4.4. Multi-perspective data

Multi-perspective data refers to a collection of data that is analyzed or viewed from various different points of view or perspectives. In this context, data is not only viewed from one single point of view, but is also analysed from a variety of different perspectives to gain a more comprehensive and in-depth understanding of the subject being observed. In this context, the resulting dataset can be seen as not just an activity data flow but can also be explored through various approaches. Event log datasets on the dwelling time process can be explored, such as the number of activities per Case_ID, process duration per Case_ID, the role of the officers involved, Documents and Identity attributes of the Container. The multi-perspective point of view really depends on what the research needs.

Limitations

The dataset in this article has certain limitations, especially in the data collection process. The data collection time is only 3 months a year; this time certainly does not describe the behaviour of the data in a 1-year normal period. This is, of course, a limitation in modelling the dwelling time process.

Ethics Statement

As a statement of ethics in the process of collecting data as a research requirement, we held discussions regarding the required data format. The matters that are the substance of the discussion are what data will be taken from the web server data recording, what period data will be taken considering that the data recording is a type of big data, what data indicators will be used in the data retrieval process, limitations on the form of data used can be distributed and ends with the signing of an ethical agreement on the use of data from researchers. We as researchers always consider the ownership of data owned by agencies as well as the impact on the potential leak of users' personal data. When preparing the dataset, we ensure that the data is not specifically identifiable. After collecting the dataset, we manually check potential personal data and perform anonymization features. We changed some personal names to roles to avoid specifying user data in the process.

CRediT Author Statement

Hanung Nindito Prasetyo: Conceptualization, Methodology, Validation, Investigation, Data curation, and draft writer; **Riyanarto Sarno**: Methodology, Investigation, supervision, and Data Curation; **Dedy Rahman Wijaya**: Conceptualization, Methodology, review & editing, and Supervision; **Raden Budiraharjo**: Data Collector- review & editing; **Indra Waspada**: Data Collector, supervision, and Data Curation; **Kelly Rossa Sungkono**: Investigation, supervision, and Data Curation;

Data Availability

Event Log Dwelling Time Dataset (Original data) (Mendeley Data).

Acknowledgements

This research was funded by Asian Development Bank under HETI ADB Project; Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Flagship Program; the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program; and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and IPR Incentive Program (PPHKI). The authors would like to thank their colleagues from the Computer Science Doctoral Program at the FTEIC Intelligent Information Management laboratory, Institut Teknologi Sepuluh Nopember, Surabaya, as pleasant colleagues during the research discussion and PPM unit of Telkom University that has provided assistance and support in this research process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article

References

[1] I. Kourounioti, A. Polydoropoulou, C. Tsiklidis, Development of models predicting dwell time of import containers in port container terminals–an Artificial Neural Networks application, Transp. Res. Procedia 14 (2016) 243–252.
[2] I. Waspada, R. Sarno, E.S. Astuti, H.N. Prasetyo, R. Budiraharjo, Graph-based token replay for online conformance checking, IEEE Access 10 (2022) 102737–102752.

[3] I. Waspada, R. Sarno, E.S. Astuti, H.N. Prasetyo, R. Budiraharjo, Extending graph-based process discovery for hierarchical block detection and incomplete concurrent relationship handling, IEEE Access 11 (2023) 123382–123391.
[4] R. Budiraharjo, R. Sarno, D.R. Wijaya, H.N. Prasetyo, I. Waspada, IMARA: a new approach to multi-attribute risk assessment based on event data weighting (Case Study in a Container Terminal, IEEE Access 11 (2023) 62292–62306.
[5] H.N. Prasetyo, et al., Optimizing decision making on business processes using a combination of process mining, job shop, and multivariate resource clustering, Appl. Comput. Intell. Soft Comput. 2023 (2023) 1–13.
[6] S. Suriadi, R. Andrews, A.H.M. ter Hofstede, M.T. Wynn, Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs, Inf. Syst. 64 (2017) 132–150.
[7] F. Mannhardt, M. De Leoni, H.A. Reijers, The multi-perspective process explorer, in: 13th International Workshops on Business Process Management Workshops (BPM 2015), 2015, pp. 130–134.
[8] J. Salas, J. Domingo-Ferrer, Some basics on privacy techniques, anonymization and their big data challenges, Math. Comput. Sci. 12 (2018) 263–274.