DISCOVERING AND OPTIMISING PROCESSES FROM EVENT LOGS OF DOCUMENT MANAGEMENT SYSTEMS

Marek MĘDREK[1*], and Łukasz WIECHETEK[1]

[1] Management Information Systems Department, Faculty of Economics, Maria Curie-Sklodowska University, Lublin; marek.medrek@umcs.pl, lukasz.wiechetek@umcs.pl
* Correspondence: marek.medrek@umcs.pl

Abstract: Modern information management systems are often driven by workflow engines, which require an accurate, detailed, and structured process description. Instead of theoretical modelling of such processes we propose a technique which can be a connection between traditional management (organisational structure) and business process management approach. In our work we present a case study where the automatic process mining techniques are applied to administrative processes in an environment of event logs recorded by DMS. We use event log data to discover and describe the process map of the organisation and to model structured process descriptions. Our approach allows for the extraction of models based on facts and it is a direct connection between the process model and raw data generated by the organisation. Such scenario increases consistency between the mapped model and the reality of the organisation.

Keywords: workflow, process mapping, process discovering, BPD, BPMN.

1. Differences between traditional and process driven management

Process driven management can be defined as management approach that is concentrated not on organisational structure (like the traditional approach) but on the course of processes in the organisation, seen as a collection of interrelated activities or tasks that lead to archiving a specified goal or solving a specific issue. Process-based management treats the organisation as a collection of processes performed to provide value (right service or product) to the customer at the right time and place.

There are many types of business processes in the organisation: management processes, operational processes, and supporting processes. However all the types of processes should share common features such as: definability (defined boundaries, input and), client (the recipient of the results of the process), order (the sequence of tasks), value adding (must give new value for the recipient), embeddedness (strictly defined place in the organisational structure), and multi-functionality (should include many functions). Each process should also

have the owner, so the person responsible for process course and its improvement. Therefore process driven management is a complex task and should include in a holistic manner all processes in the organisation both discovering, documenting, performance analysing, and improvements implementing. The main differences between traditional management and business process management is shown in Table 1.

Table 1.
Main differences between traditional management and business process management

No	Area	Traditional management	Business process management
1.	Accountability	Defined by hierarchical organisation structure.	Defined by the enterprise owner, and process owner and business system owner.
2.	Boundaries	Organisation boundaries typically undefined.	Clearly defined, plays crucial role in BPM. Source of many improvements.
3.	Compliance	Mostly based on difficult to create and even more difficult to maintain SOPs.	Based on process maps that are easier to create and maintain than SOPs.
4.	Customer Experience	Focused mostly only on the external customer.	Focused on internal customer needs that deliver value for external customers.
5.	Improvement	Seen as series of projects created to solve isolated problems.	Managed as a series of business and process improvement projects.
6.	Knowledge	Centralised, the employees are reluctant to share it, trying to protect their own jobs.	Transparent, documented, shared, distributed, and stored at the point of use.
7.	Measurement	Collected and managed at the department level using KPIs.	Business process level capturing and monitoring.
8.	Strategy	Difficult to implement. The cause-and-effect relationship between strategies and execution is often not established.	Organisational leadership better understands the business-processes cause-and-effect relationships.
9.	Structure	Collection of departments, functions.	Built around business processes, tasks.
10.	Workflow	Mostly undefined, sometimes partially defined by the IT system.	Defined by business processes and business systems.

Note: Adapted from: "Ten Key Differences Between Traditional Management And Business Process Management" by J. Bockerstette. Retrieved from https://www.linkedin.com/pulse/ten-key-differences-between-traditional-management-joe-bockerstette.

Process driven management requires a change in approach to the organisation, needs additional efforts for communication, training, IT deployment (Wu et al., 2000), and changes in education programs for both employees and managers (Borthnick et al., 2012), however it gives many advantages e.g.: organisation performance, customer satisfaction, product quality improvement; reducing the number of bad decisions; better risk prediction; better control over resources; better accountability; faster adoption to the internal and external consumer (environment) needs. BPM is crucial especially when the organisation grows rapidly and needs to restructure (Jadhav, 2011).

2. Techniques and tools for process auto discovering

2.1. Business process discovering

Appropriate process discovery is a foundation for successful business process management and critical to the success of the organisation (Jadhav, 2011). One of its main drivers is competitive pressure (Kemsley, 2015). Business process discovery (BPD) is a collection of manual or automated techniques for creation, analysing, optimising, and visualising of the business process flow (Figure 1). It is all about finding hidden processes, exposing them in order to have the opportunity for improvement (Kemsley, 2015).

BPD can create a nearly exact (AS-IS) model of real process but also identify process hierarchy and attributes like: owners, entities, tasks, and statuses. The BPD techniques explore the data gathered by the IT systems used in the organisation or collected via interviews.

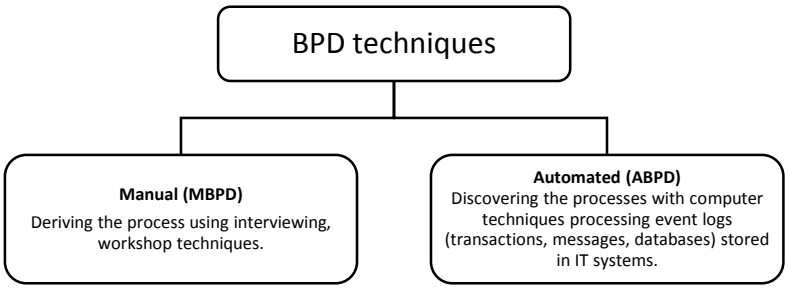

Figure 1. Business process discovering techniques. Own work.

Manual business process discovering (MBPD) uses interview techniques. Analysts have to combine pieces of information (often stored in different format) acquired from the processes stakeholders. The main benefits of manual BPD are: having a bright view on practical problems; releasing stakeholders knowledge. Sometimes manual techniques are the only way for processes discovery in the organisation (Jadhav, 2011).

Automated business process discovery (ABPD) is used when organisations become more complex, competitive and try to automate the way of operation. ABPD builds processes starting from event logs and seeks the patterns. ABPD tools merge a great amount, different formats of data, which are hardly (economically inefficient) or even impossible to process manually. It brings a lot of benefits: examination of process variations (Gartner, 2016); acceleration in process discovery; exact AS-IS visualisation; data collection with less impact on day-to-day organisational functioning (Jadhav, 2011). ABPD also gives an opportunity for wide simulation and can be used for processes optimisation.

However ABPD techniques use only IT stored data, so in order to improve automatically discovered patterns some manual techniques for capturing informal links and data are still needed. Manual and automated techniques may be used complementary, especially in medium sized organisations equipped with IT systems e.g. ERP that cover many areas of the organisation.

Due to an increase in data volume the importance of business process discovery increases rapidly. Future development of BPD will concentrate on (Jadhav, 2011):

- preparing better analysis techniques not only for AS-IS process discovering but also with the possibility of building TO-BE maps,
- intense development work on applications for automated business process discovery, to prepare "one click", almost real time processes visualisations,
- wider integration ABPD with end-to-end business modelling specifications in order to automatically visualise the processes in an appropriate, widespread standard.

2.2. Business modelling specifications

In many cases process driven management can be supported by the IT solutions. For the past few decades information technology changed not only the course of business processes but also transformed the way of business processes management. Business process modelling was introduced for acceleration of business processes redesign (Borthnick et al., 2012; Wu et al., 2000). Implementation of a new management approach and BPM supporting tools needs some agreements and standards for describing business systems and processes in an understandable language for all stakeholders.

An example of an international consortium established for standards integration is The Object Management Group – OMG ("About the Decision Model", 2015). One of OMG task forces established for developing requests for proposals of integrated management models is The Business Modeling & Integration Domain Task Force (BMI DTF). The BMI DTF main areas of interest in business are: planning, motivation modelling, BPM, performance monitoring, integration, B2B and B2C collaboration, security issues, and language and vocabulary standardisation ("The Business Modeling", 2016). The specifications published by OMG in the area of business modelling contains ("About the Decision Model", 2015): Business Motivation Model (BMM), Business Process Definition Metamodel (BPDM), Business Process Maturity Model (BPMM), Business Process Model and Notation (BPMN™), Case Management Model and Notation (CMMN™), Date-Time Vocabulary (DTV™), Decision Model and Notation (DMN™), Production Rule Representation (PRR), Semantics of Business Vocabulary and Business Rules (SBVR™),Value Delivery Modeling Language (VDML™), and Workflow Management Facility (WfMF).

The BMM, BPDM, BPMM, BPMN™, PRR concern the area of business process design and provide notation that should be understandable for all business users (Borthnick et al., 2012). The specifications are used for description of the organisation, expression of business process models, providing the maturity levels for the organisation (Van Looy et al., 2011), and graphical notation of business process for better understanding of their performance. CMMN is used for graphical expressing of cases and exchange models among cooperating IT tools. DTV specification offers a business vocabulary for the concept of date and time, models; continuous, discrete time and relationship of situations and events (Linehan et al.,

2012). DMN is a common notation understandable by all business users. It is used for describing repeatable decisions in a form that is interchangeable across organisations ("About the Decision Model", 2015). SBVR introduces rules and vocabulary for documenting the semantics of business vocabularies and rules ("About the Semantics", 2015). VDML describes a standard modelling language for designing and analysing an enterprise's operation. It focuses on value creation and exchange ("About the Value Delivery", 2015). WfMF specifies rules for workflow management processes as a kind of automatic processes involving humans, machines, and IT tools and defines interfaces used to manipulate and execute workflow objects ("About the Workflow", 2015).

The set of business modelling specifications is quite extensive, however, it allows for a comprehensive and detailed description of the business processes, and gives the ability for better process understanding, and management. It also brings a possibility of automation and stronger processes linking among cooperating organisations. Modern business process discovery tools often offer mechanisms for automatic process visualisation according to an appropriate standard.

2.3. Tools for business processes discovering

The first IT applications due to their constraints were used only for simple repetitive calculations especially in accounting and finance departments. Development in the area of data storage and the increase in computing power allowed for storing and processing larger sets of data. The stored data were then analysed and aggregated. Due to technical limitations the business processes often had to adapt to the limitations of IT systems. In some ways the organisational structure and processes were dependent on IT tools. The next step in the IT system development was the transition from the data-driven approach to process oriented management. Management information systems (MIS) equipped with workflow management modules allowed for greater flexibility and better IT tools customisation to business needs. At present, IT tools have a very wide range of application and they allow for information flow management, supervising individual activities, mapping and visualisation of processes, building electronic forms, and designing and supervising task paths (workflow). IT solutions can also be used for automatic business process discovery. Some characteristics of IT tools used for business processes discovery are shown in Table 2.

Table 2.
IT tools used for business processes discovery

Tool	Description
BMC Atrium Discovery	The system for process discovering and mapping. It is composed of discovery engine, data model, reasoning engine, pattern language, administration and delivery tools. An advanced platform offering tools for process discovering, business mapping, reports management, data import and export, and many others. *More information: https://docs.bmc.com/docs/display/DISCO90/Documentation*
Fujitsu Process Analytics Software	A tool that integrates Fujitsu process discovery technologies with business monitoring and analytics. It allows for automatic visualisation of business process flows, real-time monitoring and alerting, the use of event collection sensors, easy process management using the Rich Dashboard and Management Console, and continuous process improvement. *More information: www.fujitsu.com/global/products/software/middleware/application-infrastructure/interstage/solutions/bpmgt/bpma/features/*
Fluxicon Disco	A tool built by experienced academics. It contains fast process mining algorithms with user-friendly interfaces. It has a built in mechanisms for: automated process discovery, process map animation, presenting detailed process statistics, cases view, inspecting history of the relevant cases, event log filtering, and data import and export. *More information: www.fluxicon.com/disco*
Lexmark Perceptive Process Mining	An IT tool for automated collection and quick analyses of the organisation's activity. It can be used for mining and refining process insight, identifying and assessing threats to compliance. It contains alerting, equipped with a filtering mechanism controlling the amount of analysed data and shown details. *More information: www.lexmark.com/en_us/products/software/workflow-and-case-management/process-mining.html*
Metastorm ProVision	A professional and easy-to-use tool for process modelling and visualisation. It enables discovering, monitoring, and improving all business events in order to improve the organisation's effectiveness. It supports: advanced capabilities in enterprise modelling, analysis and simulation of processes, group work, and integration with Microsoft Office. *More information: www.softwareag.com/corporate/products*
PRoM (Process Mining framework)	An open-source framework for implementing process mining tools. It plays a role of an easy to use and easy to extend platform for process mining algorithms for both users and developers. PRoM tends to be a standard process mining platform in the academic world with strong community of contributors and users. *More information: www.processmining.org/prom/start*
QPR Process-Analyzer	Process mining software for automated generation of processes flowcharts. It generates process visualisation and process metrics (KPIs). It has a built-in data integration mechanism. *More information: www.qpr.com/products/qpr-processanalyzer*
Software AG ARIS Architect & Designer	A tool for designing and analysing a digital enterprise. It can be used for process, enterprise architecture, governance, risk analysis; process designing; harmonising business activities with IT systems. Its main capabilities are: support for business notations, central administration, cross analytics mechanism, configurable meta model, and extension packages. The free version is the ARIS Express. *More information: www.dsa.com.pl/produkty/metastorm-enterprise-suite/opentext-provision,metastorm-provision.html*
StereoLOGIC Enterprise Discovery Suite	A tool for visualising and measuring business processes in real time. It contains mechanism for process discovery, process measuring, deviations analysing, finding root causes and inefficiencies, and preparing processes documentation. Is equipped with a fact-based process improvement mechanism for accelerating business operations. *More information: www.stereologic.com/products/enterprise-discovery-suite/*
Worksoft Analyze	A cloud based solution for discovering, visualising, and analysing end-to-end business processes. It provides to both analysts and technical staff information about business process execution, it can analyse process variants and impact on the organisation. *More information: www.worksoft.com/products/worksoft-analyze*

Note: own work based on: https://docs.bmc.com; www.fujitsu.com; https://fluxicon.com; www.lexmark.com; www.softwareag.com; www.processmining.org; www.qpr.com; http://dsa.com.pl; www.stereologic.com; www.worksoft.com.

The next step of BPM tools development will be (or already is) driven by the development of e-business – using ICT to support as many business activities, as it is possible. E-business means automation, interoperation across organisations, integration, great amount and variety of data and less time to deliver service, product, or information. Fast development of communication techniques, the increasing popularity of mobile technologies, social media, introduction appliances that can exchange data without human intervention, Internet of Things – IoT (Xia et al., 2012) or even Internet of Everything – IoE (Macaulay et al., 2015), development of cloud technology, requires preparing new standards, law regulations and tools for automatic data processing and visualisation.

Using modern, powerful, widespread, based on Internet communication IT tools allows for automatic gathering of data (logs) concerning not only single people, companies, organisations, but also social groups or whole societies. Logs collected by modern IT systems sometimes underestimated, ignored and simply closed in the archives begin to play a very important role. They could be used for modelling business, local-government, social, and demographical processes, especially with the use of business processes discovery tools.

3. Case study: discovering and optimising processes

In our work we describe an attempt of the process mapping which was performed in a medium-sized, self-government district institution in the east of Poland. In this institution employment is about 70 office workers who are involved in business processes of this organisation. The organisational structure of the analysed district office is presented in Figure 2.

The organisational structure is vertical and aligned through departments, functions, and jobs at the desk level. Accountability is traditionally defined through a top-down, hierarchical structure, with responsibilities assigned from the district council, through the district board, managers, and directors to front line officers (see Figure 2). The theoretical workflow is well described by an official, government document – office instruction which defines document categorisation, numbering, transmission, etc. In practice the workflow is reflected by some IT systems, especially the document management system (DMS) which uses role definition in close connection with the organisational structure. Each official is identified in the DMS by a unique account and for document classification the consistent itemised file list (CIFL) is used, where each document is assigned to a specified part of hierarchical, five level CIFL.

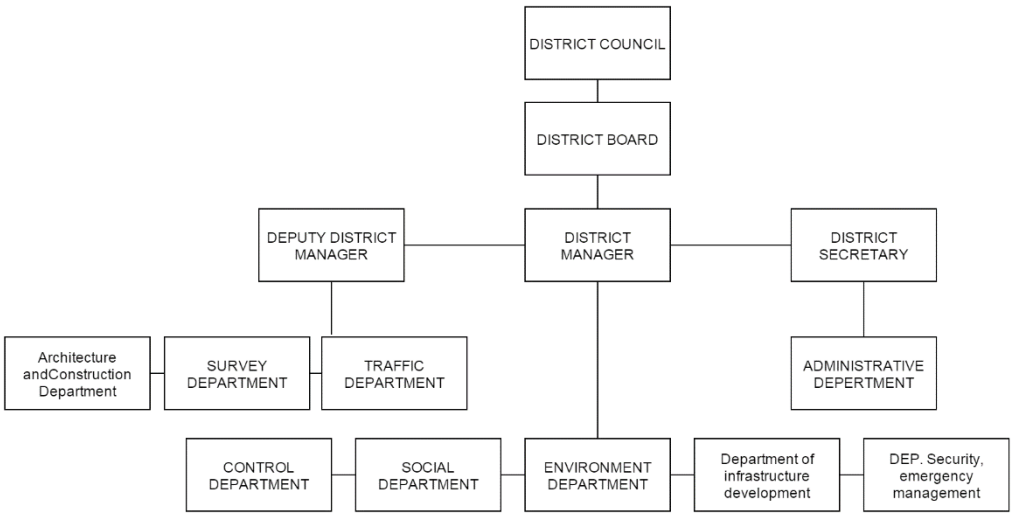

Figure 2. Organisational structure of the analysed district office. Own work.

A typical document flow is presented in Figure 3 – a document is registered in the office and next it is categorized and routed to a specific person who performs administrative work on the case. The document may be transferred to different persons several times, until the decision or final document is made. Each step in the document flow is reported in the log event of the DMS.

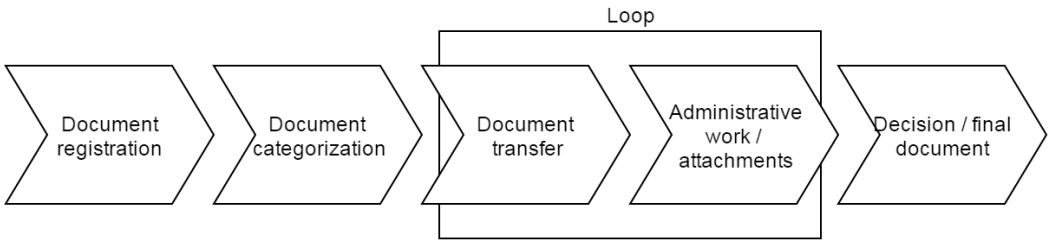

Figure 3. Typical work (document) flow scenario in the analysed organisation. Own work.

The main purpose of our work is to create a framework based on event logs of DMS which discover process maps and identify non-optimised paths and bottlenecks in the overall structure of the workflows. In the next two sections we present the structure and transformations of the real DMS log event and finally, in section 3.3., we present some results of our work.

3.1. Log event formats

The log event of the analysed DMS is stored in the database (PostgreSQL) in a single table, which consists of the following attributes (see Figure 4):

- event id (id) – record number in the log,
- event type (type) – relation to dictionary table which contains a set of seven different events types,
- event description (description) – description containing details of the event,
- user (user) – relation to user table, which contains a full description of the DMS users,
- time (time) – time of the event.

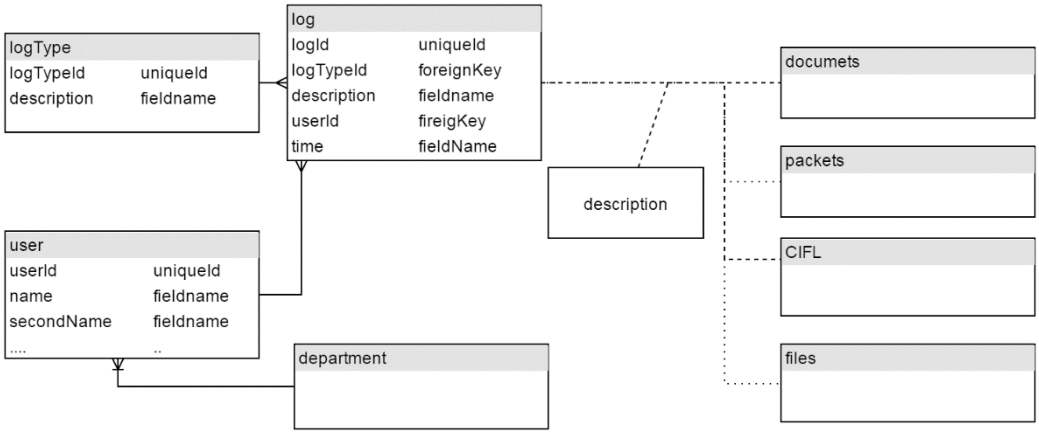

Figure 4. Database diagram of a log event and the directly connected and associated tables. Own work.

As a result of the non-normalised log architecture and the format of the event description it is impossible to analyse the data without recourse to other system tables – see Figure 4. The table log references to other tables: logType and user and the relations are realised by atomic values in the attributes: logTypeId and userId respectively. The above references are normalised, but there are also non-normalised several associations which result from the description of the log event – these associations are denoted by the dotted lines in Figure 4. The attribute description consists of the human readable sentence containing hidden references to other DMS database tables – detailed analysis of this example description is presented in Figure 5.

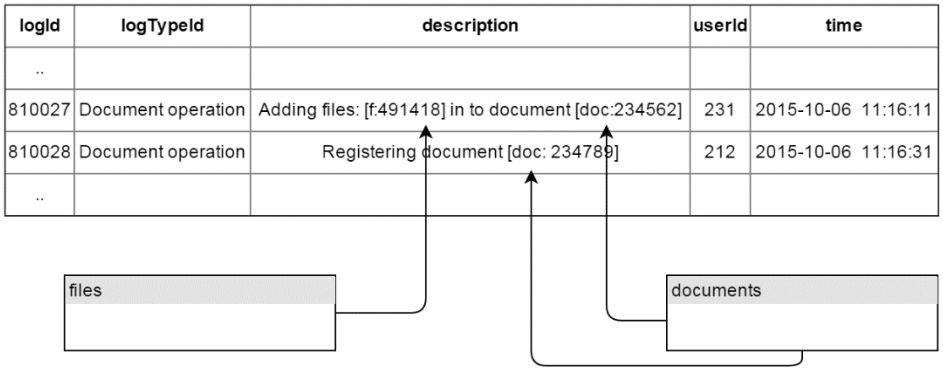

Figure 5. The structure of the sample event description and association with other DMS tables. Own work.

As a result, in order to make an analysis, the event log must be combined with other tables of DMS, and the transformation is presented in section 3.2.

3.2. Log event transformations

In order to receive information which would allow us for a full analysis of the business processes in the organisation we made some extraction and transformation of the original log data. Our purpose is to identify the cases of business processes with participants and its positions in the organisational structure and all data (files, documents, packet of documents) which is transmitted during the workflow. To this end we use ETL tool Pentahoo data integration, which allows us to make all necessary transformation in real time and pass completed and standardised log event to the process mining tool.

During the extraction we use transformations offered by Pentahoo: String operations, Value mapper, Splitting fields and Regular expression evaluation. After transformations we write all data in a single, flat table, which is the source for Fluxicon DISCO software, which we use as the main process mining tool. This step is the most important part of the analysis, because we have to correctly include all data needed by our future exploration. In our case these data come from multiple tables and we have to use whole database structure for proper data preparation: in the log, an event should correspond to an activity that is executed in the process and multiple events should be linked in a process instance or case ordered by the time attribute. Figure 6 shows the component diagram of the system used for log event preparation for data mining.

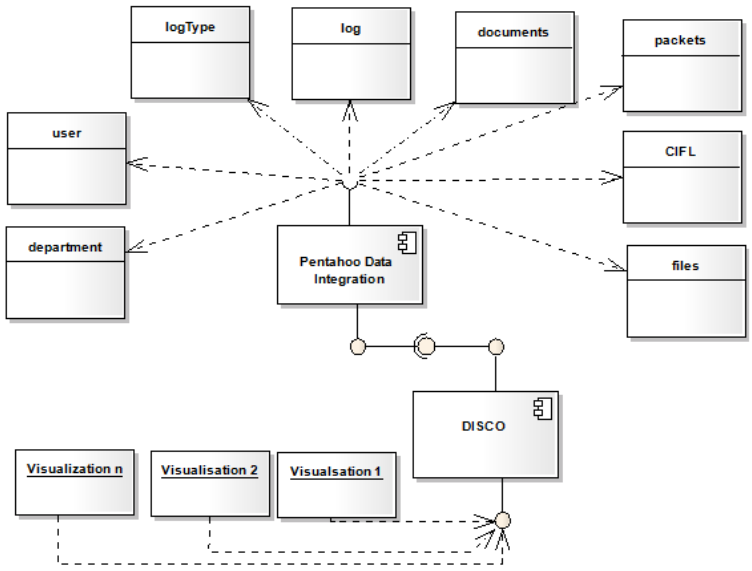

Figure 6. The structure of the sample event description and association with other DMS tables. Own work. For our research we choose generally a part of the event log, which time frame is described by the following equation: $time\ frame = expected\ time\ of\ a\ single\ case\ \times 4 \times 5$, where expected time of a single case is process completion time for a typical case (Rozinat, 2014). In the case of the analysed organisation, the typical time of process completion is 30 days, so our time frame covers about 1.5 years.

3.3. Process map generation

The final step in our analysis is the preparation of the process maps. The log event format required for visualisation should contain in each record the information about one event or executed activity, at least one case ID column, an activity, and one or more timestamps (Rozinat, 2014). The selection of case ID columns depends on the target analysis model, e.g. when we want to illustrate a document flow map with nodes in the form of organisational units the case IDs will be the organisation unit and the document ID, when we draw a map of the document flow between users we use as case ID columns the user and document ID.

We prepare several process maps based on the organisational structure, on document categorisation and on log event type. Figure 7 shows the sample process map – panel a) shows the numbers of different activities recorded in the log and panel b) describes the document flow between the organisational units (Dep1-Dep6) and the thickness of the line is related to the number of cases correlated with the analysed connection. Panel c) of Figure 7 displays one of the analytical views of the global statistics available in the DISCO software – the number of cases as a function of case duration. The number of analysed cases grows with time and achieves its local maximum in about 13-16 days and in about 28-30 days (the second extremum is related to the official deadline of the case proceeding).

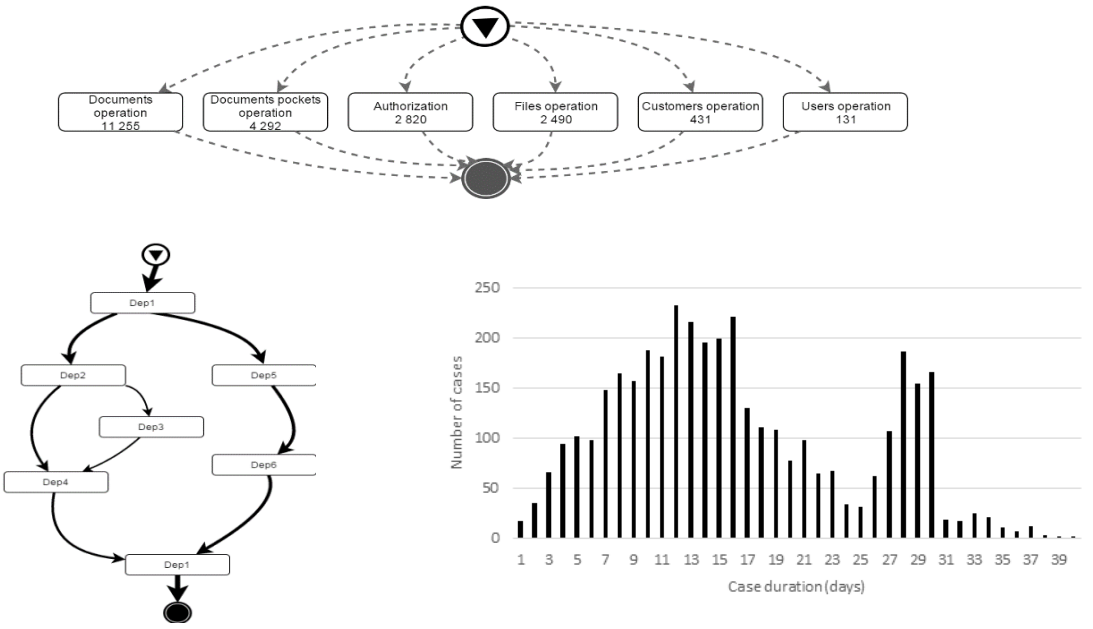

Figure 7. Sample visualisations of the identified processes. Own work.

Apart from the presented analysis, properly prepared log data allow to prepare other quantitative statistics, for example variant analysis. The variant is a specific sequence of activities and sometimes it can be useful to decompose the process map where the activities are shown for all cases together into a process map for identified variants. Figure 8 shows an example of two variants of the example process: usually (over 68%) the process consists of activities shown in Figure 8 panel b), where all activities take place in Department of Architecture and Construction (see Figure 2), while Figure 8 panel a) shows a more complicated scenario, where the documents are passed also to the Environment Department (see Figure 2).

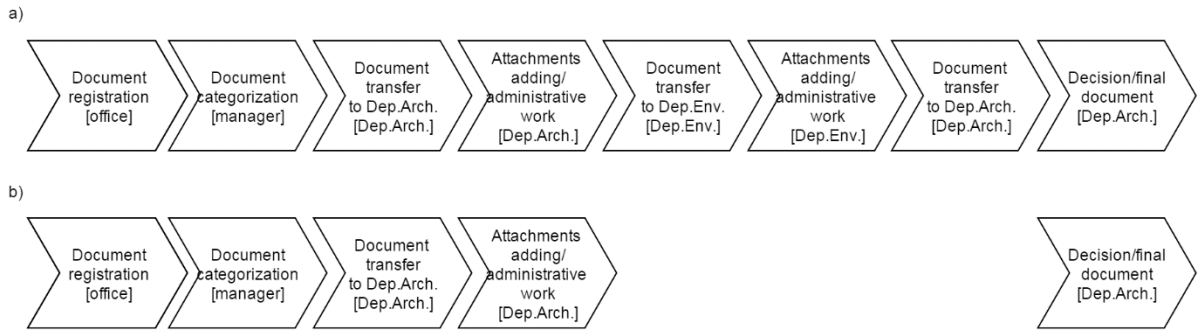

Figure 8. Two variants (activity sequence) of one case of administrative work in the analysed organisation. Own work.

The case of administrative work presented in Figure 8 shows that the workflow does not reflect the organisational structure and the metrics should be built around business processes and tasks rather than on departments and functions.

4. Conclusions

In our paper we showed a model which was developed for an automated business process discovering in one of the public institutions in Poland. We tried to build maps of the processes starting from event logs and other databases resources of the DMS used in the organisation. In our framework we merge a great amount and different formats of data to discover the relationships between activities recorded in the log.

The Performed analysis led to several process maps which show the workload and performance of different organisational units and functions. The decomposition of the discovered processes and quantitative statistics offer metrics, which can be used for performance evaluation and processes optimisation. Our work shows that the automated process discovering based on the log events of IT systems can be a useful tool for process driven management in the organisation.

Further work aiming at continuous monitoring of the process management needs to be done. In the future research we plan to introduce into the framework additional data sources like phone calls and other domain IT systems logs to better reflect the real processes in the organisation.

We would like to thank Biala Podlaska District and Devcomm ICT for the DMS data sharing.

Bibliography

1. About OMG®. (2016). Available online http://www.omg.org/gettingstarted/gettingstarted index.htm.
2. About the Decision Model and Notation Specification Version 1.0. (2015, Sep. 1). Available online http://www.omg.org/spec/DMN/1.0/.
3. About the Semantics Of Business Vocabulary And Rules Specification Version 1.3. (2015, May 8). Available online http://www.omg.org/spec/SBVR/1.3/.
4. About the Value Delivery Modeling Language Specification Version 1.0. (2015, Oct. 5). Available online http://www.omg.org/spec/VDML/1.0/.
5. About the Workflow Management Facility Specification Version 1.2. (2015, May 2). Available online http://www.omg.org/spec/WfMF/1.2/.
6. Automated Business Process Discovery (ABPD) (2016). Available online http://www.gartner.com/it-glossary/automated-business-process-discovery-abpd

7. Bockerstette, J. (2015, Nov. 16). Ten Key Differences Between Traditional Management And Business Process Management. Retrieved from https://www.linkedin.com/pulse/ten-key-differences-between-traditional-management-joe-bockerstette.
8. Borthnick, A.F., Schneider, G.P., and Vance, A.O. (2010). Preparing Graphical Representations of Business Processes and Making Inferences from Them. *Issues in Accounting Education*, *25*(3), pp. 569-582. doi: 10.2308/iace.2010.25.3.569.
9. Borthnick, A.F., Schneider, G.P., and Vance, A.O. (2012) Using Graphical Representations of Business Processes in Evaluating Internal Control. *Issues in Accounting Education*, *27*(1), pp. 123-140. doi: 10.2308/iace-50078.
10. Date-Time Vocabulary (DTV). (2013). Available online http://www.omg.org/spec/DTV/1.0/PDF/.
11. Jadhav, S. (2011, Feb. 1). Business Process Discovery. Retrieved from https://www.bptrends.com/business-process-discovery/.
12. Kemsley, S. Business Process Discovery. Retrieved from http://www.tibco.com/multimedia/business-process-discovery-tcm8-2401.pdf.
13. Linehan, M.H., Barkmeyer Jr., E.J., and Hendryx, S. (2012, Oct. 1). The Date-Time Vocabulary. Retrieved from https://lists.oasis-open.org/archives/ws-calendar/201210/msg00002/NIST-Date-Time-Vocab-911074.pdf.
14. Macaulay, J., Buckalew, L., and Chung, G. (2015). Internet of Things in logistics. A collaborative report by DHL and Cisco on implications and use cases for the logistics industry. Retrieved from: http://www.dhl.com/content/dam/Local_Images/g0/Newboutus/innovation/DHLTrendReport_Internet_of_things.pdf.
15. Rozinat, A. Disco User's Guide. Retrieved from https://fluxicon.com/disco/files/Disco-User-Guide.pdf.
16. The Business Modeling & Integration Domain Task Force. (2016, Feb. 2). Available online http://bmi.omg.org.
17. van der Aalst, W., Weijters, A., and Maruster, L. (2003). Workflow mining: a survey of issues of approaches. *Data and Knowledge Engineering*, *47*(2), pp. 237-267. doi: 10.1016/S0169-023X(03)00066-1.
18. van Looy, A., de Backer, M., and Poels, G. (2011). Defining business process maturity. A journey towards excellence. *Total Quality Management & Business Excellence*, *22*(11), pp. 1119-1137.
19. Wu, B., Kay, J.M., Looks, V., and Bennett, M. (2000). The design of business processes within manufacturing systems management. *International Journal of Production Research*, *38*(17), pp. 4097-4111. doi: 10.1080/00207540050204957.
20. Xia, F., Yang, L.T., Wang, L., and Vinel, A. (2012). Internet of Things. *International Journal of Communication Systems*, *25*(9), pp. 1101-1012. doi: 10.1002/dac.2417.