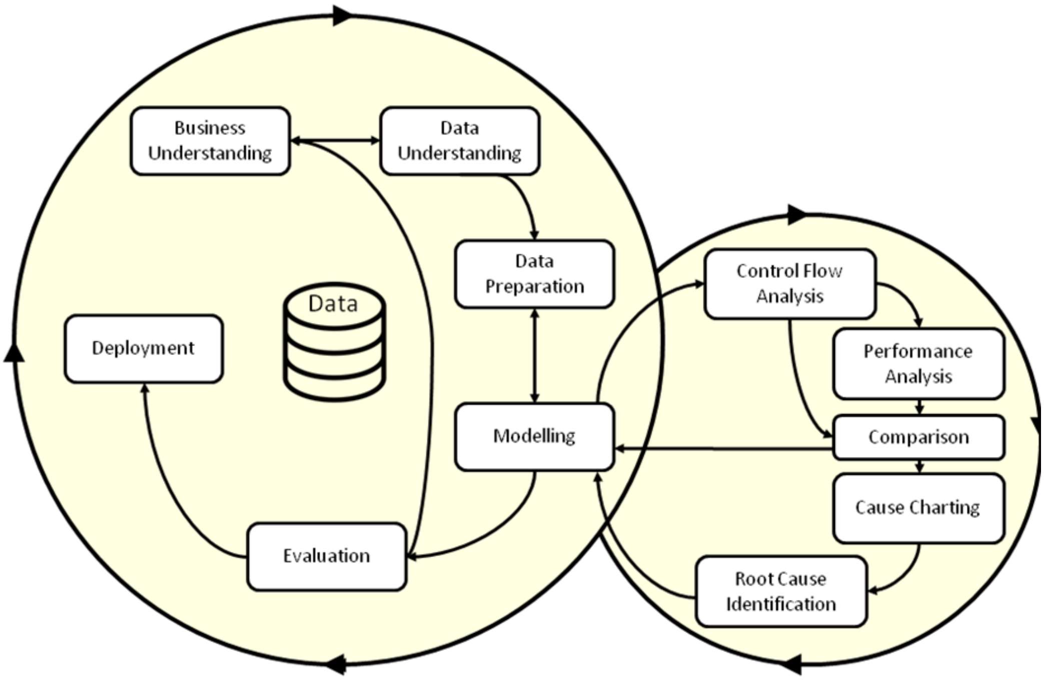
WHAT DO YOU MEAN?

The CIRCA-DIPS method for root cause analysis of data interoperability problems within aviation information systems



Cassandra Renes
Marco Spruit

Technical Report UU-CS-2019-011
December 2019

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

WHAT DO YOU MEAN?

The CIRCA-DIPS method for root cause analysis of data interoperability problems within aviation information systems

Cassandra Renes and Marco Spruit
Utrecht University, Department of Information and Computing Sciences, Utrecht, the Netherlands

Abstract

Data interoperability is not merely a quality and a tool to gain a competitive advantage. It is increasingly becoming necessary for survival and to achieve information superiority. Data interoperability issues typically occur in a networked value constellation. An example of a networked value constellation is the aviation industry where airport, airline and air traffic control together create value for their environment. The aviation industry is characterized by a huge set of different information systems which typically not only operate within organizational boundaries. These interorganizational systems enhance the probability of data interoperability problems. While collaboration is very important within the aviation industry, data interoperability problems make this collaboration difficult.

Using a systematic literature review approach, the CRISP-DM extension for Root Cause Analysis of Data Interoperability Problems (CIRCA-DIPS) conceptual method is created. Then, an extensive case study is executed to validate the method at an international airport to examine the interoperability between the airport, air traffic control and airlines. CIRCA-DIPS identifies the root causes of data interoperability problems using business process mining within aviation information systems.

Keywords: Data interoperability, Aviation information system, Knowledge discovery, Root cause analysis, Business process mining.

1 Introduction

The ability for an organization to interoperate within and with other organizations is not only a quality and a tool to gain a competitive advantage anymore, it is necessary for survival (Chen & Daclin 2006). It has become critical and it is the keystone to achieve information superiority (Morris et al. 2004). Data interoperability issues exists in all types of organizations (Dell'Erbaa et al. 2003; Qamar & Rector 2007; Shen et al. 2016; Meulendijk et al. 2017). These issues especially occur in interorganizational systems. The aviation industry is characterized by a huge set of different information systems which typically not only operate within organizational boundaries (Dell'Erbaa et al. 2003). This is because the aviation industry is an example of a networked value constellation. Networked value constellations are sets of organizations who together create value for their environment (Tapscott et al. 2000).

Aviation requires collaboration of airports, airlines and air traffic control. These main three stakeholders are responsible for the whole commercial flight process. The flight process cannot exist with one of these three stakeholders missing. Collaboration between them is therefore of key importance. IATA (2011) states that failures in the aviation industry structure can be overcome when these stakeholders work together. This stakeholder collaboration could lead to:

- Better customer service and satisfaction
- Simpler pricing for consumers
- Reduced costs of fragmentation
- Reduced congestion, delay, environmental impact
- Fewer layoffs and more predictable wages for workers

However, within the aviation networked value constellation, the different stakeholders use different definitions for the same concept. For example, the concept *flight* has a different meaning for each stakeholder. Whereas an airport sees a flight as an aircraft which either departs from or arrives at that specific airport, the airlines define a flight as every flight of the aircraft owned by the airline, no matter the origin or destination. The air traffic control define a flight as an aircraft located within their radar. The destination, origin or owner of that aircraft is not interesting. This simple example illustrates the scale of the data interoperability issues existing in the aviation industry. Even a simple misconception about a definition can lead to inefficient collaboration. While each stakeholder benefits from good collaboration within the networked value constellation. Therefore, it is necessary to identify these data interoperability problems and eliminate them.

2 Research approach

The following research question will be central throughout this research: *How can the root causes of data interoperability problems be discovered using business process mining in an aviation information system?*

To answer this main research question, we investigate through several sub-questions the fields of data interoperability and business process mining to identify how data interoperability problems can be discovered within an aviation information system, and how its root causes and underlying problem types can be uncovered and visualized.

We start from the existing body of knowledge, identified and analyzed according to the Systematic Literature Approach proposed by Moher, Liberati, Tetzlaff and Altman (2009). For each sub question, several keywords are selected. To ensure these keywords accurately represented the concept, a validation check has been executed.

Figure 1 is used to visualize the SLR activities. Based on these sub questions, our conceptual model is created. This conceptual model is validated by the case study. The improvements generated by the case study will be used to adjust the conceptual method. After changing the method accordingly, it is validated again. This process repeats itself until the conceptual method is found correct. Then, the method is finalized.

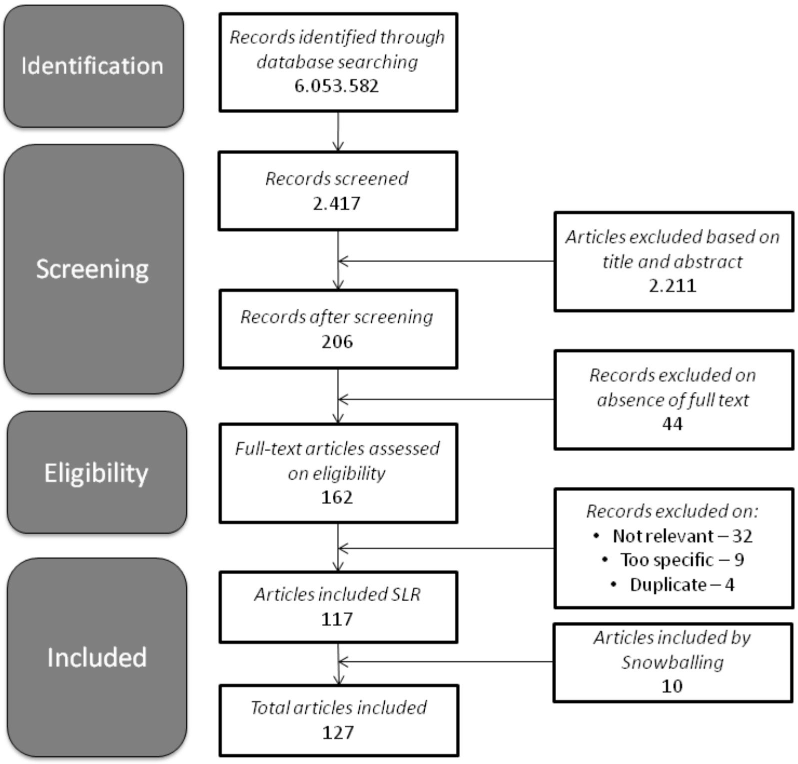


Figure 1. *SLR process of this research, based on* (Moher et al. 2009)

3 Related literature

In this section the selected articles through the SLR process are described. The section is divided into six sub sections: Data interoperability, process mining, discovering data interoperability problems, identifying root causes, visualization techniques and, lastly, the conceptual model is proposed.

3.1 Data interoperability

Literature has many definitions for data interoperability. Chen and Daclin combined the definitions of IEEE, Vernadat and IDEAS into a less restricted definition: "The ability to (1) communicate and exchange information; (2) use the information exchanged; (3) access to functionality of a third system" (Chen & Daclin 2006).

Different levels of data interoperability are defined throughout literature (Yang & Zhang 2006; Ouksel & Sheth 1999; Sheth 1999; Rh 2001; Gottschalk 2009). Although they do not agree about the specific levels, they do agree however, that semantic interoperability is the most challenging type of interoperability. To achieve semantic interoperability, systems must be able to exchange data in such a way that the precise meaning of the data is readily accessible and the data itself can be translated by any system into a form that it understands (Heflin & Hendler 2000; Heiler 1995; Pokraev et al. 2005; Sheth 1999). Making the precise meaning of the data explicit has proven extraordinarily difficult for several reasons (Heiler 1995).

A more broad definition of semantic interoperability is given by Veltman (2001): "The ability of information systems to exchange information on the basis of shared, pre-established and negotiated meanings of terms and expressions, and is needed in order to make other types of interoperability work".

3.2 Process mining

The goal of process mining is to extract information about processes from event logs. It is assumed that (1) each event refers to an *activity* (i.e., a well-defined step in the process), (2) each event refers to a *case* (i.e., a process instance), (3) each event can have a *performer*, also referred to as originator (the person executing or initiating the activity) and (4) events have a *timestamp* and are totally ordered (A. K. Alves De Medeiros 2005; van der Aalst 2012; van der Aalst et al. 2007; van der Aalst et al. 2012; van Dongen et al. 2005). Process mining enables organizations to discover and analyze business processes based on raw event data (van der Aalst & Dustdar Schahram 2012). Process mining can achieve many advantages for the organization. The challenge is to exploit event data in a meaningful way. For example, to provide insights, identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures and streamline processes (van der Aalst et al. 2012). Another challenge is to identify all possible behaviours (van der Aalst 2009).
To make process mining a repeatable process, the process diagnostics method is created (Bozkaya et al. 2009). This method is visualized in Figure 2.

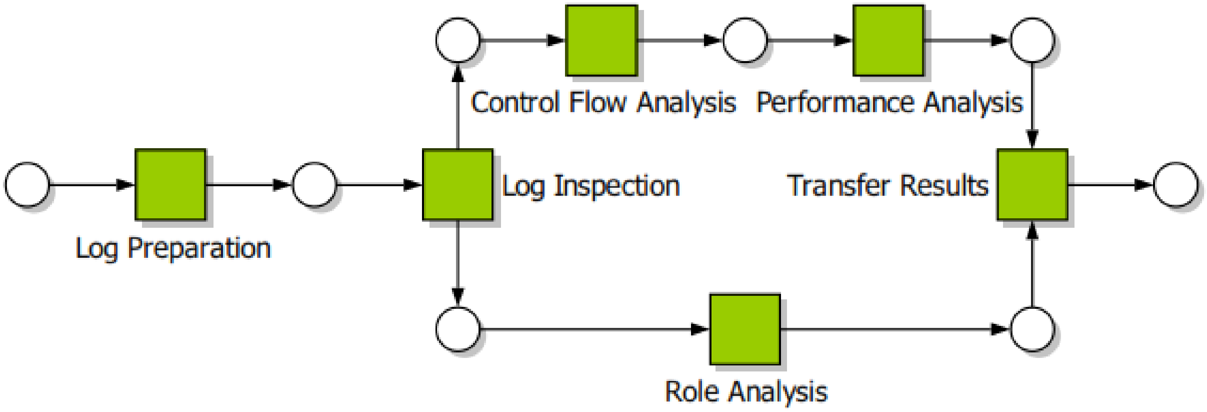


Figure 2. Process diagnostic method based on (Bozkaya et al. 2009)

3.3 Discovering data interoperability problems

An interorganizational system can be defined as "an automated information system shared by two or more companies" (Johnston & Vitale 1988). Several aviation information systems can be considered as interorganizational systems. This aviation interorganizational system operates within a networked value constellation of airport, airlines and air traffic control. A networked value constellation requires (1) alignment of the participating organizations and (2) alignment of four separate perspectives described by Derzsi and Gordijn (2006). To achieves these two goals, the e^3 alignment framework was created (Pijpers et al. 2013).

Another related framework is the CRISP-DM framework, created for data mining projects. It aims to make large data mining projects less costly, more reliable, more repeatable, more manageable and faster (Wirth & Hipp 2000) by standardizing on best-practice process steps. Next to the e^3 alignment framework and the CRISP-DM, related methods like ETL, data linkage and process analytics are considered. From each method, useful method fragments are selected and summarized. These method fragments, presented in Figure 3, form the basis for the conceptual method proposed in this paper.

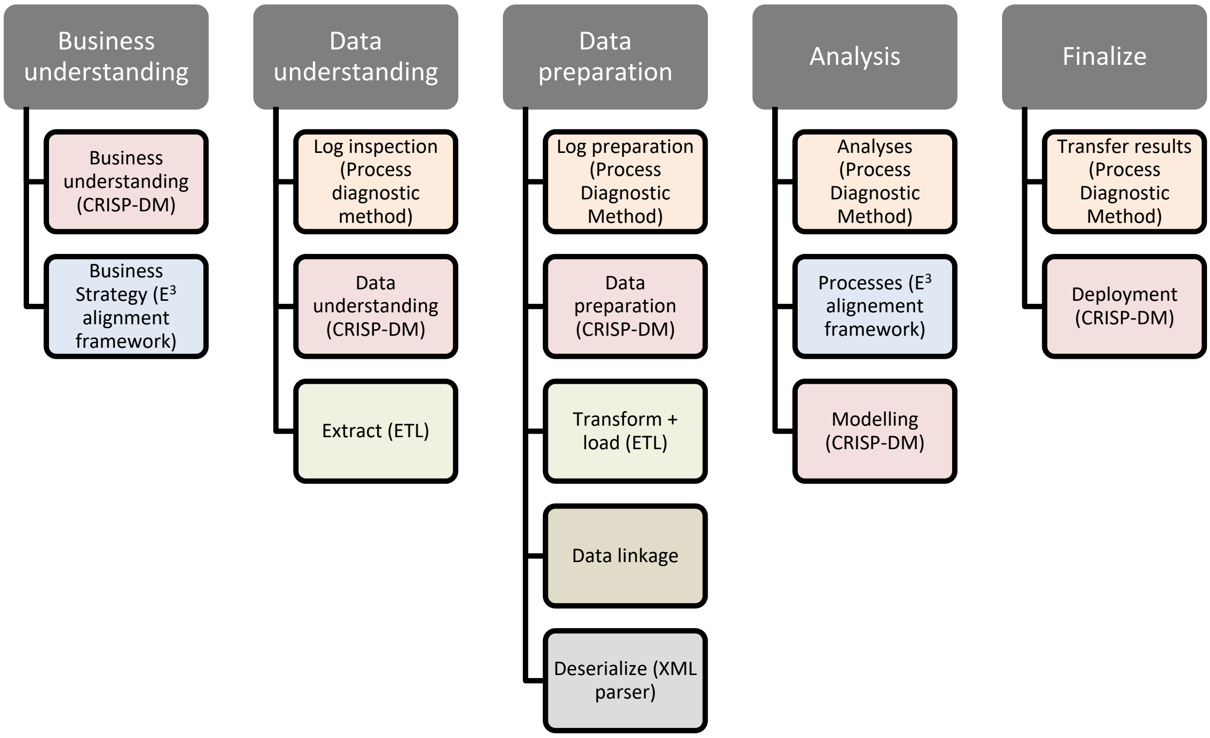


Figure 3. Framework of related method fragments

3.4 Root cause identification

For the discovery of root causes, both exploratory data analysis and root cause analysis methods are considered. Exploratory data analysis aims to identify major features of a dataset of interest and to generate ideas for further investigation (Cox & Jones 1981; Leith et al. 1991). Several exploratory data analysis methods include cluster analysis, classification and association. Root cause analysis (RCA) is an analytical process designed to investigate underlying factors that have contributed to or have directly caused a major event or failure (Staugaitis 2002). The goal of RCA is prevention (Bach et al. 1997; Rooney & Van den Heuvel 2004; Staugaitis 2002).

The relevant method fragments, selected in this section, are added to the already existing framework. This updated framework is visualized in Figure 4. This framework represents all the phases of the conceptual method, from business understanding to a finalize phase.

3.5 Visualization

Several visualization methods are considered including Unified Modelling Language (UML), Process Deliverable Diagram (PDD), Business Process Modelling Notation (BPMN) and Interactive visualizations. UML was created in the mid 1990s and has been adopted as the de-facto standard formalism for software design and analysis by the Object Management Group (Berardi et al. 2005; Dobing & Parsons 2006; Eriksson & Penker 2000). UML can be defined as "a language for specifying, constructing, visualizing and documenting the artefacts of a software-intensive system" (Mellor 2002).

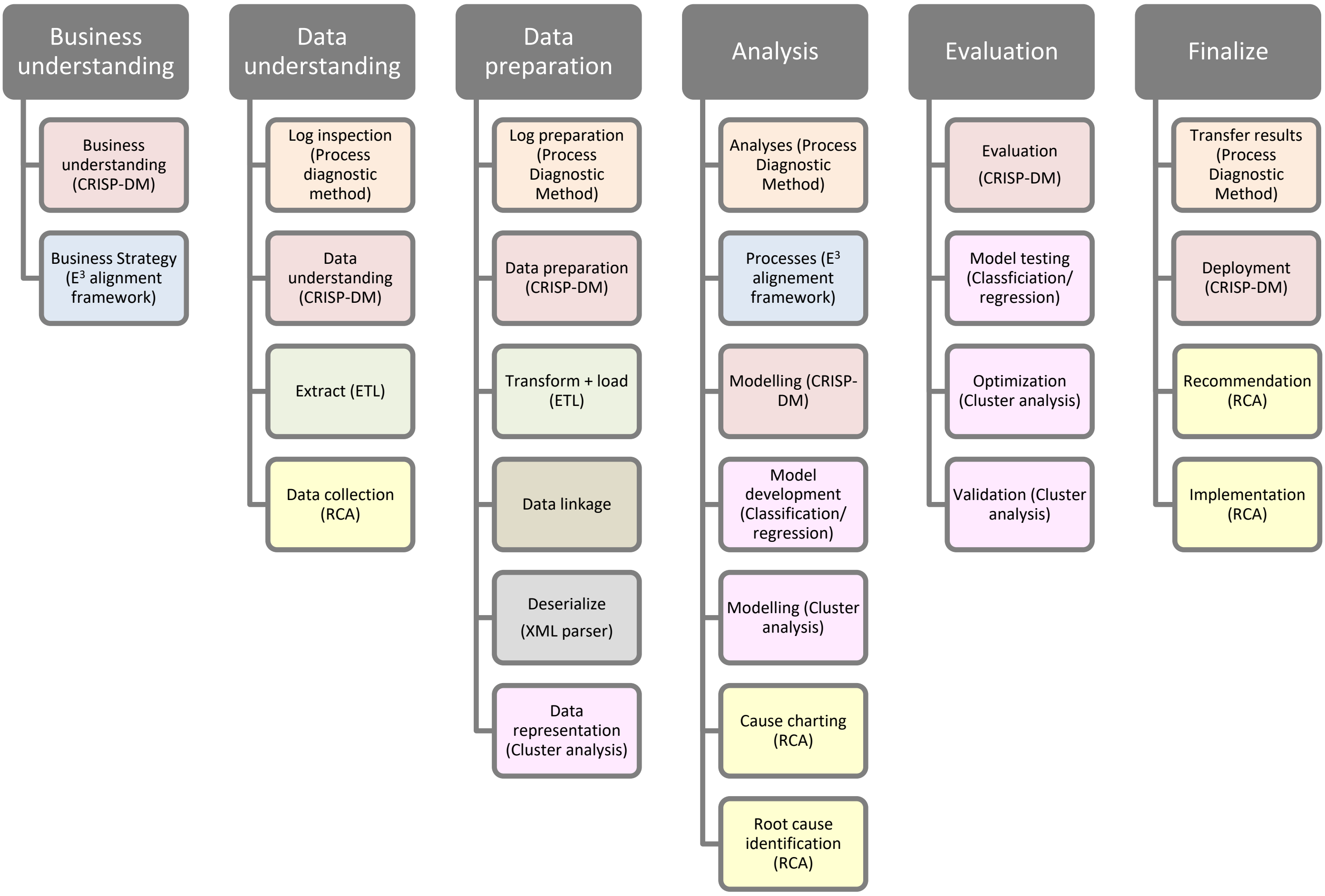
A PDD is a situational method engineering technique. It consists of two integrated diagrams: A process view which is based on a UML activity diagram, and a deliverable view which is based on a UML class diagram (*e.g.* V-anonimized 2010; L-anonimized 2008). The PDD technique is used to tune existing methods to a specific situation (van de Weerd & Brinkkemper 2008).

The goal of BPMN is to provide a notation that is readily understandable by all users. It is based on a flowcharting technique and is tailored for creating graphical models of business process operations (White 2004).

Lastly, interactive visualizations aim to map datasets into visual media for the purposes of assisting users in exploring these datasets or communicating about them to others (Carlis & Konstan 1998). Interactive visualizations can be used to highlight and identify connections between entities in the document (Stasko et al. 2007).

An alternative to building the conceptual method from scratch, is adapting an existing method. As seen in Figure 4, the CRISP-DM method is present in each of the six phases. Using this existing method and extending this with the process mining and root cause analysis phases is preferable to creating an entirely new method since this approach builds upon the familiarity of the CRISP-DM framework.

Figure 4. Framework including exploratory data analytics and RCA methods



3.6 Conceptual method

Including all the related literature, the proposed method is a combination of eight data analytics methods. Since all the steps of the CRISP-DM framework are present in the conceptual method, the CRISP-DM framework is the starting point. Next to the known CRISP-DM steps, another circle is added which extends the modelling phase. These steps elaborate the business process mining and root cause analysis phases. Together the conceptual method is referred to as the CIRCA-DIPS method (**C**R**I**SP-DM extension for **R**oot **C**ause **A**nalysis of **D**ata **I**nteroperability **P**roblem**s**) and is presented in Figure 5.

For the first phase (business understanding) the e^3 alignment framework is used. When creating the e^3 value and e^3 forces model, the business is clearly captured. To create a data understanding, data specialists should be interviewed. These two phases are mainly important whenever the method is used in an unknown environment. For the data preparation the following steps should be executed: Data consolidation, data cleaning, data transformation, data reduction and data loading. After the preparation steps are executed, the modelling phase begins. For this phase, the control flow analysis and performance analysis are executed using a business process mining tool. These steps help identify bottlenecks. When the bottlenecks are clear, the cause charting and root cause identification phases of the RCA are used to discover the root causes of the bottlenecks. To finalize the process, the results are evaluated with the relevant stakeholders. The results and evaluation together create recommendations which can be used to ensure the bottlenecks are tackled. The arrows in the model indicate that after evaluation, the previous phases can always be repeated. Moreover, the sequence of the phases is not strict, the arrows only indicate the most important and frequent dependencies between the phases.

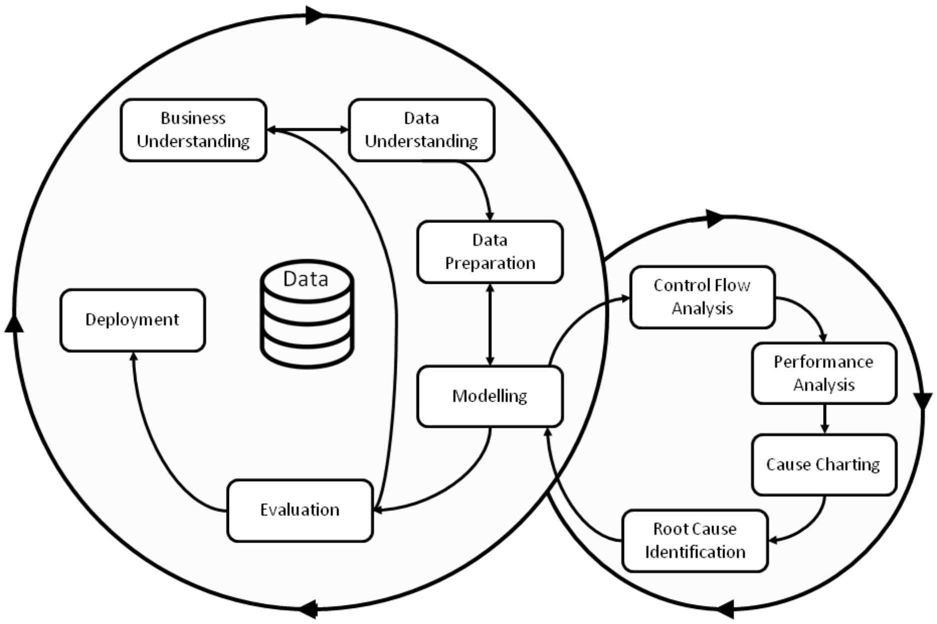


Figure 5. Draft version of the CIRCA-DIPS conceptual method.

4 Validation

In this section the conceptual method is validated through a case study, executed at an international airport in Western Europe (which will remain anonymous). The interoperability between the airport, air traffic control and airlines is examined. For this case study, data from three different systems was used.

4.1 Case study 1

For the business understanding phase, the e^3 value and e^3 forces models are created. These models gave a good understanding of how the networked value constellation was connected. For the next phase—data understanding—the complete view of the data and the connecting systems was visualized. This was done in collaboration with some data experts. A timeslot of 14 days in Spring 2015 was collected with data from three different aviation information systems of around 1.5 million CSV rows and three columns, adding up to around 4.5 million data entries.

After the data collection the data needed to be transformed first because the structure was different for each system. For the transformation, a Python script was used. After the transformation, the relevant data were selected. For the Disco tool used to investigate the data interoperability problems, only the activity, the corresponding timestamp and identifier are needed. Again a Python script was used to select the relevant data. After loading the data in the tool, the tool automatically cleaned. Since there were no missing cases or events, nothing was excluded.

Once the data were loaded into the tool, the analysis started. For this analysis both the control flow analysis and the performance analysis functionalities of the tool were used. Since no bottlenecks were found using the performance analysis, no root causes could be identified. Therefore, the following two steps could not be executed and the next phase was evaluation. Furthermore, the data from one of the systems were extracted too rigorously. Therefore, only three activities appeared to exist, but in reality many more activities should be there.

Summarizingly, the following interoperability problems were defined after applying our method. First, it is not clear whether landing or arrival times are included taxi and in- or out-block times. Second, scheduled date time can both be used for the arriving and departing flights. Third, block time can mean both in-block as off-block time. Furthermore, next to differences between the stakeholders, there is also the difference in communicating with the passengers. For the passengers an aircraft is landed after its wheels touched the ground. While some stakeholders consider an aircraft landed after it is in-blocks. Finally, next to these semantic differences, there are many syntactic differences. Whereas one data source was nicely organized and structured, other data sources were rather unorganized and sometimes not even syntactically correct.

When evaluating these results, the case company immediately noticed that some important data were missing. Therefore, another case study was executed with the same three systems, only now the dataset contained the necessary data. Before the other case study was executed, the conceptual method was modified.

4.2 Modifications conceptual method

Several recommendations were generated in order to improve the conceptual method. First, the sequence of the activities in the data preparation phase should not be strict. The sequence depends on the used data. Therefore, the order for these activities should be determined for each dataset individually.

In the modelling phase, an explicit step should be added for comparison purposes. The control flow analysis and performance analysis are executed for each dataset individually. Therefore, whenever these results are not compared between the datasets, no interoperability problems can be determined. Besides, performance analysis is not always useful. This analysis should only be executed whenever there are time dependent activities within the dataset.

In addition, whenever the data are not suited for the RCA steps, the method should skip these steps. An alternative method for these RCA steps was determined through interviews.

The updated conceptual method is visualized in Figure 6.

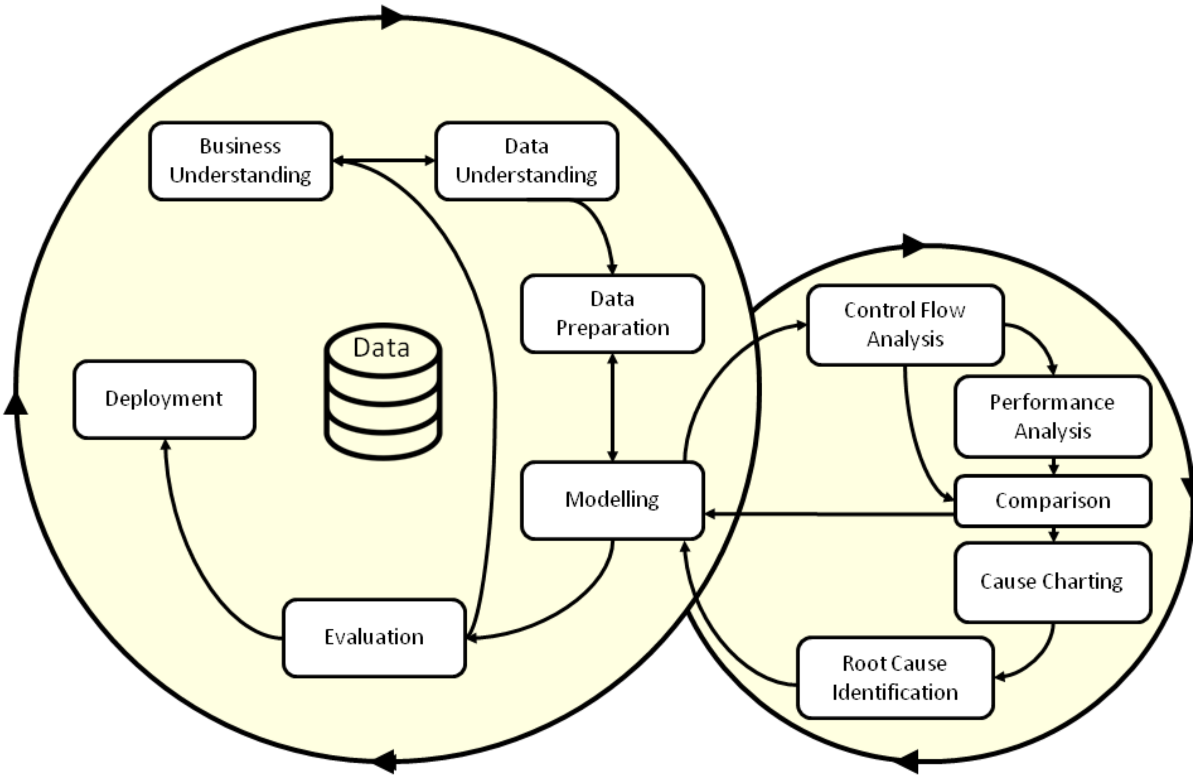


*Figure 6. The **CRI**SP-DM extension for **R**oot **C**ause **A**nalysis of **D**ata **I**nteroperability **P**roblems (CIRCA-DIPS) conceptual method.*

4.3 Case study 2

Because the same organizations and the same systems were used for the second case study, the business understanding phase and the data understanding phase were not executed again. Therefore, the second case study started at the data preparation phase. However, we did use a different dataset since the first dataset did not represent the most important concepts of the flight data. This new dataset has a time frame from six days from Summer 2015 of around 4.4 million CSV rows and three columns, adding up to over 13 million data entries.

Again the syntax was not similar for each system, thus the data had to be transformed using the same Python script. After transformation, again the relevant data were selected using Python scripts. The Python script was changed for one of the systems in order to extract the relevant data more thoroughly. After loading the data and cleaning it using the tool, again no missing cases or events were removed.

In the modelling phase the control flow analysis and performance analysis were executed. This time the correct data was extracted from the dataset and the dataset contained the necessary data for the process. Instead of three activities, the process now contained over ten activities, which was a better representation. The process from data preparation to performance analysis did shine light on several data interoperability problems. First, the syntactic differences when analyzing the three different systems. Furthermore, difference were seen in definitions used between the systems. The same activity was called differently or activities had the same name, while representing something differently.

After analyzing the control flow and performance again no bottlenecks were identified. However, there were some activities which had a long duration. But these activities were not bound by time. For example, when scheduling a flight this can be executed weeks before the actual flight takes place. In this case, a long duration between the activities of scheduling and the actual flight is better than a short duration. Each bottleneck that appeared using the tool, could be explained and none of these bottlenecks were real bottlenecks. The drawback is that again, the root cause analysis could not be executed. Although, the root cause phases of the method could not be executed, the root causes of the data interoperability problems could be determined using the knowledge within the company. The problems were often already known.

Summarizingly, the following interoperability differences were defined by this second dataset. First, differences in flight number caused by different standards for airline, air traffic control and airport. Second, different activity names were used. Where some data sources used standardized activity names, others used their own activity names. This is caused by poor collaboration between the stakeholders and not meeting agreements. Third, the interchangeable use of the concepts scheduled, estimated, target, actual and calculated make interoperability unnecessarily complex. These are also concepts which are agreed when defining the activity concepts. Similarly, the different standards for airline, air traffic control and airport also create confusion and make collaboration more complex.

When evaluating the results, the case company did agree with the results. It gave them new insights in the process. They are planning to now use process mining at a regular basis in order to identify possible data interoperability problems and bottlenecks. The sooner these problems are identified, the sooner they can be solved.

Three recommendations were formulated for the case company. Most of the data interoperability issues identified could be explained by the different industry standards which exist. For each of the actors within the networked value constellation an industry standard exists. When one industry standard would be created, these problems would be solved. However, one networked value constellation is in no position to create such a standard. This should be created in collaboration with other airports, airlines and air traffic controls. Furthermore, some interoperability differences could be explained by differences between the actors in the constellation. It is in the interest of each of the actors if the collaboration is efficient and effective. The actors should stress the importance of this collaboration. The last recommendation is to repeat the method. This enables the organization to detect issues or bottlenecks soon and therefore the organization can participate quickly. Luckily, the case company already stated that they intend to start using the method on a regular basis.

5 Conclusion

We identified data interoperability as "the ability to (1) communicate and exchange information; (2) use the information exchanged; (3) access to functionality of a third system". Data interoperability is necessary for survival and to achieve information superiority. Business process mining enables organizations to discover and analyze business processes based on raw event data. It can achieve many advantages for the organization when exploited in a meaningful way. The process diagnostic method is created to make process mining a repeatable process.

Data interoperability problems can be discovered within an aviation information system by combining relevant parts of eight related methods including the CRISP-DM process, e^3 alignment framework, ETL and data linkage. Useful method fragments are identified and selected as basis for the conceptual method proposed in this research. The framework of useful method fragments is extended with method fragments from exploratory data analysis and root cause analysis methods. This brings together all required components with which to uncover the root causes of data interoperability problems.

The process of discovering data interoperability issues can be visualized using either UML, PDD, BPMN or interactive visualizations. We argue that the best solution is to capitalize on the familiarity of the CRISP-DM method and extend this method with the business process mining and root cause analysis steps. We validated the conceptual method using a case study with a follow-up study of an actual aviation information system constellation.

These findings together answer our main research question *How can the root causes of data interoperability problems be discovered using business process mining in an aviation information system?* Our solution is the CIRCA-DIPS method as proposed in *Figure 6* of this research. This method includes the following phases in alignment with its underlying CRISP-DM process: business understanding, data understanding, data preparation, modelling, evaluation and deployment.

6 Discussion and future research

Our research is subject to various limitations. First, our systematic literature review approach depends on the quality of the search results and the keywords. To counter this, the defined keywords have been validated by two experts in the area of data interoperability and aviation information systems. Furthermore, the quality of the search results of Google are validated by several researchers. To mitigate the risk of missing keywords, the snowballing technique was used to include relevant articles which did not appear in the systematic literature review approach. In total, ten articles were included using this extra snowballing technique.

Moreover, process mining (as van der Aalst also suggests) cannot be assumed to identify all possible behaviours. Additionally, no bottlenecks were actually identified using the performance analysis in the modelling phase. Therefore, the root cause analysis steps could not be performed. In other words, the cause charting and root cause identification steps were never validated. Future research should be executed in order to test these steps.

Finally, the activities used in the case study had no start and end time. Therefore, the bottlenecks within activities could also not be identified. Future research should also include activities which do have start and end times.

With this research we have proposed to construct a best-of-breed conceptual method to uncover interoperability problems in big data, interorganizational information systems in the aviation industry, based on standard fragments from knowledge discovery, process mining and root cause analysis, among others. It is our ambition to evaluate our CIRCA-DIPS method within interorganizational information system constellations—such as healthcare—in the near future as well.

References

A. K. Alves De Medeiros, a. J.M.M.W., 2005. Genetic process mining. *Applications and Theory of Petri Nets 2005, ...*, (i), pp.48–69. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.76.4916.

Van der Aalst, W.M.P. et al., 2007. Business process mining: An industrial application. *Information Systems*, 32(5), pp.713–732.

Van der Aalst, W.M.P., 2009. Process Mining. *Database Encyclopedia*.

Van der Aalst, W.M.P. et al., 2012. Process mining manifesto. *BPM 2011 Workshops Proceedings*, pp.169–194.

Van der Aalst, W.M.P., 2012. Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems*, 3(2), pp.235–255.

Van der Aalst, W.M.P. & Dustdar Schahram, 2012. Process Mining Put into Context. *IEEE internet computing*, pp.82–86.

Bach, B., Bozin, B. & Kavsek, D., 1997. Root cause analysis. In *Nuclear Energy in Central Europe*. pp. 251–258.

Berardi, D., Calvanese, D. & De Giacomo, G., 2005. Reasoning on UML class diagrams. *Artificial Intelligence*, 168(1-2), pp.70–118.

Bozkaya, M., Gabriels, J. & Werf, J., 2009. Process Diagnostics: A Method Based on Process Mining. *2009 International Conference on Information, Process, and Knowledge Management*, (1).

Carlis, J. V & Konstan, J. a, 1998. Interactive visualization of serial periodic data. *Focus*, San Franci, pp.29–38.

Chen, D. & Daclin, N., 2006. Framework for enterprise interoperability 1. *Proc. of IFAC Workshop EI2N*, pp.77–88.

Cox, N.J. & Jones, K., 1981. Exploratory data analysis. In *Quantitative Geography*. London: Routledge, pp. 135–143. Available at: http://xa.yimg.com/kq/groups/16412409/1159714453/name/exploratorydataanalysis.pdf.

Dell'Erbaa, M. et al., 2003. Harmonise: A solution for data interoperability. *IFIP Advances in Information and Communication Technology*, 105, pp.433–445.

Derzsi, Z. & Gordijn, J., 2006. A Framework for Business / IT Alignment in Networked Value Constellations 2 Case Study : Wireless Internet Access Provision in Hotels. , (section 4), pp.219–226.

Dobing, B. & Parsons, J., 2006. How UML is used. *XRDS: Crossroads, The ACM Magazine for Students*, 49(5), pp.109–113.

Van Dongen, B.F. et al., 2005. The ProM Framework: A New Era in Process Mining Tool Support. *In Applications and Theory of Petri Nets 2005*, (i), pp.444–454.

Eriksson, H. & Penker, M., 2000. Business Modeling With UML: Business Patterns at Work. *Open Training*, p.12.

Gottschalk, P., 2009. Maturity levels for interoperability in digital government. *Government Information Quarterly*, 26(1), pp.75–81. Available at: http://dx.doi.org/10.1016/j.giq.2008.03.003.

Heflin, J.D. & Hendler, J., 2000. Semantic interoperability on the web. *Proceedings of Extreme Markup Languages 2000*, pp.1–15. Available at: http://www.cse.lehigh.edu/~heflin/pubs/extreme2000.pdf\nhttp://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA440535.

Heiler, S., 1995. Semantic interoperability. *ACM Computing Surveys*, 27(2), pp.271–273.

IATA, 2011. *Vision 2050*, Available at: http://www.wbcsd.org/pages/edocument/edocumentdetails.aspx?id=219&nosearchcontextkey=true.

Johnston, H. & Vitale, M., 1988. Creating competitive advantage with interorganizational information systems. *Mis Quarterly*, 12(2), pp.153–165. Available at: http://www.jstor.org/stable/10.2307/248839.

Leith, R.M., Hipel, K.W. & Goertz, H., 1991. Exploratory Data Analysis. *Canadian Water Resources Journal*, 16(1), pp.81–92.

L-anonimized (2008).

Mellor, S.J., 2002. Executable UML: a foundation for model-driven architecture. *Addison-Wesley Professional*, pp.1–9.

Meulendijk, M., Spruit, M., Lefebvre, A., & Brinkkemper, S. (2017). To what extent can prescriptions be meaningfully exchanged between primary care terminologies? A case study of four Western European classification systems. *IET Software, 11*(5), 256–264.

Moher, D. et al., 2009. Preferred reporting items for systematic literature reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), pp.264–269.

Morris, E. et al., 2004. System of Systems Interoperability (SOSI): final report (No. CMU/SEI-2004-TR-004). *Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst.*, (April).

Ouksel, A.M. & Sheth, A., 1999. Semantic Interoperability in Global Information Systems: A brief introduction to the research area and the special section. *SIGMOD record*, 28(1), pp.5–12. Available at: http://dx.doi.org/10.1145/309844.309849.

Pijpers, V. et al., 2013. Using conceptual models to explore business-ICT alignment in networked value constellations. *Requirements Engineering*, 17(3), pp.203–226.

Pokraev, S. et al., 2005. Semantic and pragmatic interoperability: A model for understanding. *CEUR Workshop Proceedings*, 160.

Qamar, R. & Rector, A., 2007. Semantic issues in integrating data from different models to achieve data interoperability. *Studies in health technology and informatics*, 129(Pt 1), pp.674–678.

Rh, E.I., 2001. Towards a principled approach to semantic interoperability. , pp.19–25.

Rooney, J.J. & Van den Heuvel, L.N., 2004. Root Cause Analysis for Beginners. *Quality Progress*, (July), pp.45–53.

Shen, Z., Meulendijk, M., & Spruit, M. (2016). A federated information architecture for multinational clinical trials: STRIPA revisited. *24th European Conference on Information Systems*. Istanbul, Turkey.

Sheth, A., 1999. Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics. *Interoperating Geographic Information Systems*, pp.5–30. Available at: http://link.springer.com/10.1007/978-1-4615-5189-8.

Stasko, J. et al., 2007. Jigsaw: Supporting investigative analysis through interactive visualization. *VAST IEEE Symposium on Visual Analytics Science and Technology 2007, Proceedings*, (March), pp.131–138.

Staugaitis, S.D., 2002. ROOT CAUSE ANALYSIS. , pp.1–42.

Tapscott, D., Ticoll, D. & Lowy, A., 2000. *Digital capita - harnessing the power of business webs*, Boston: Harvard Business School Press.

Veltman, K.H., 2001. Syntactic and semantic interoperability: New approaches to knowledge and the semantic web. *New Review of Information Networking*, 7(1), pp.159–183.

Van de Weerd, I. & Brinkkemper, S., 2008. Meta-Modeling for Situational Analysis and Design Methods. *Handbook of Research on Modern Systems Analysis and Design Technologies and Applications*, pp.35–54.

V-anonimized (2010).

White, S. a, 2004. Introduction to BPMN. *IBM Cooperation*, 2(0), pp.1–11.

Wirth, R. & Hipp, J., 2000. CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, (24959), pp.29–39.

Yang, Q.Z. & Zhang, Y., 2006. Semantic interoperability in building design: Methods and tools. *Computer-Aided Design*, 38(10), pp.1099–1112.