On the Fragmentation of Process Information: Challenges, Solutions, and Outlook

Han van der Aa[1], Henrik Leopold[1], Felix Mannhardt[2,3], Hajo A. Reijers[1,2]

[1] Department of Computer Science, VU University Amsterdam, Faculty of Sciences, De Boelelaan 1081, 1081HV Amsterdam, The Netherlands
[2] Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands
[3] Perceptive Software, Gooimeer 12, 1411DE Naarden, The Netherlands

Abstract. An organization's knowledge on its business processes represents valuable corporate knowledge because it can be used to enhance the performance of these processes. In many organizations, documentation of process knowledge is scattered around various process information sources. Such *information fragmentation* poses considerable problems if, for example, stakeholders wish to develop a comprehensive understanding of their operations. The existence of efficient techniques to combine and integrate process information from different sources can therefore provide much value to an organization. In this work, we identify the general challenges that must be overcome to develop such techniques. This paper illustrates how these challenges should be and, to some extent, are being met in research. Based on these insights, we present three main frontiers that must be further expanded to successfully counter the fragmentation of process information in organizations.

1 Introduction

Corporate knowledge is recognized as a principle source of sustainable competitive advantage and value creation [3]. *Process knowledge* is a particular type of corporate knowledge that relates to processes, their context, and their execution. It is regarded as valuable corporate knowledge, because it can be used to enhance the performance of business processes and, hence, also of organizations [23]. In order to meet this value proposition, it is vital that process knowledge is available when needed. A significant threat to this required availability is that process knowledge is often fragmented throughout an organization. Some is only available in the form of tacit knowledge held by specialized process participants or domain experts [13]. Other knowledge is explicitly captured as *process information*, but scattered over a plethora of sources, such as documents, models, and systems. This paper focuses on the latter category of process knowledge.

The problems resulting from the *fragmentation of process information* are considerable. First, information contained in different sources potentially contradicts each other. This can result, for instance, in situations where different stakeholders hold different expectations on what a process aims to establish and take mutually counter-productive measures. Second, even if sources do not contradict each other, insights from multiple sources must be combined to obtain a complete understanding of any given process. This may be quite tedious, depending on the types of sources that are attempted to be combined. Imagine, for example, the task of combining the insights for the same process from its work instruction, a listing of business rules, and a graphical model. So-called *process information defragmentation techniques* counter these problems. They automatically detect and resolve conflicts between various sources, and also integrate information from different sources in order to provide more comprehensive process insights.

At this point, we argue that the development of process information defragmentation techniques takes place in a haphazard way. A number of sources, such as *event logs*, are being studied and harvested intensively; others are virtually being ignored, such as *policies* and *spreadsheets*. Moreover, researchers that look into different information sources are hardly aware of each other's work, while that would actually be highly beneficial to further the defragmentation process. A more thorough understanding of the challenges that are involved with the development of process information defragmentation techniques as well as the opportunities seem called for. Hence, it is the goal of this paper to present a systematic view on this field. We achieve this through the identification of the major challenges that process information defragmentation techniques must meet, a description of the state of the art, and a way forward for the advancement of further techniques.

The remainder of this paper is organized as follows. In Section 2, we first consider why information fragmentation exists and the problems that it causes for organizations. Section 3 then elaborates on the types of heterogeneity across information sources that defragmentation techniques must deal with. In Section 4, we illustrate existing approaches that consider these challenges in order to present insights into the current state of the field. Section 5 discusses the main shortcomings of the existing methods and considers future research directions. Finally, Section 6 concludes the paper.

2 Fragmented Information

The fragmentation of process-related information poses considerable problems for organizations, yet its causes are well-understood. In Section 2.1, we consider the factors that drive information fragmentation. Because it may be difficult or even undesirable to root out these factors, we argue that that the better approach is to repair the resulting defragmentation with efficient techniques. We identify two main streams of such techniques in Section 2.2.

2.1 Causes

There are two ways in which information fragmentation can manifest itself: (i) sources contain different information, and (ii) sources use different representation formats.

Different information First, information fragmentation exists because different documents/records/models cover different perspectives on processes. Each type of process documentation can be considered as a simplified description of a process from a certain vantage point: It emphasizes particular aspects and omits details that are not relevant to the party that created it [28]. As such, the process description by the head of the sales department, for instance, may significantly differ from the technical specification used by system administrators. In addition to differing individual perspectives, there are also external sources of process-related information that exist independently from how a company itself characterizes its business processes [38]. Important examples include *policies* or *regulations*, such as Basel III. Clearly, such sources often also contain important pieces of information about business processes, in particular in normative terms [35].

Different representation formats Second, information fragmentation manifests itself through the existence of a variety of different *types of information sources*. This primarily follows from differing preferences of stakeholders about the representation of process-related information. For instance, business professionals, i.e. those who actually conduct the work in the processes, are known not to feel very confident in reading and interpreting *process models* [11]. They rather prefer a verbal description, like a *work instruction*. Some business professionals also use *spreadsheets* for documenting entire business processes [29]. The reason for such a choice is often given by the general familiarity with general purpose tools such as Excel. Business analysts, by contrast, typically do prefer graphical descriptions in the form of processes models [13].

Different representation formats also exist because different types of process information come into existence through different triggers. Information sources such as process models and text documents mainly relate to *build-time* activities of process models, i.e. they are created to characterize business processes. By contrast, the information contained in other sources is created during the *run-time* of a business process. For instance, *event logs* capture execution data generated by users interacting with information systems. Similarly, information recorded in *spreadsheets* support business process execution by providing means to compute certain process steps. Varying information needs and the existence of different types of sources thus result in process information being fragmented over a variety of artifacts, comprising different representation formats.

2.2 Defragmentation tasks

Two major problems result from process information fragmentation that organizations must deal with. First, information captured in different sources can contradict each other, potentially resulting in highly problematic situations. Second, to obtain all available information on a given process, it does not suffice to consult a single source. To counter these issues, we identify two tasks that process information defragmentation techniques must fulfill: (1) detecting and resolving inconsistencies between information sources, and (2) integrating information captured in different sources.

Inconsistency detection and resolution When different information sources relate to the same process, it is crucial to prevent conflicts in terms of contradicting information. Inconsistencies occur especially when documents are being developed independently from each other [40]. For instance, it has been found that different domain experts often operate with diverging assumptions on how a process is running [13, p.158]. This results in information sources that represent multiple, possibly different expectations. The detection of differences is also crucial for comparisons of the desirable behavior of a process with its actual behavior. i.e. through checking how requirements or regulations imposed by regulatory authorities relate to execution behavior, e.g. captured in event logs. Thus inconsistencies must be detected and resolved to prevent possible misinterpretation of process-related information.

Information integration As a result of process information fragmentation, it is not sufficient to consult a single document in order to obtain all information available on a business process. Consider, for instance, two process models related to the same business process. A process model used by managers emphasizes operational details, i.e. *what* has to be done. By contrast, a workflow model representing the actual system implementation focuses on technical details, i.e. *how* the process should be executed [46]. When analyzing this process, it is important to consider the details of both information sources. Otherwise analyses are performed based on incomplete information This can result in sub-optimal decision making and, ultimately, evaporation of valuable corporate knowledge. To ensure that organizations can make optimal decisions, it is therefore worthwhile to integrate the information captured in multiple sources.

3 Challenges

Any process information defragmentation task depends on the ability to relate information from different sources to each other. Without such a relation, it is not possible to check if the information contained in different sources contradicts each other, or to integrate this information into an all-encompassing view. To obtain such relations, we identify two subsequent steps that represent the two main challenges that information defragmentation techniques must always overcome. Figure 1 visualizes these. To automatically compute a relation between process-related information captured in different sources, this information must first be transformed into an interpretable format. In the remainder, we refer to such distilled information as *structured process information*. Second, once structured process information has been obtained from various sources, it must be aligned with each other. Due to differences in terminology and level of detail that may exist between various sources, this often poses a considerable challenge. In the remainder of this section we elaborate on these two challenges.

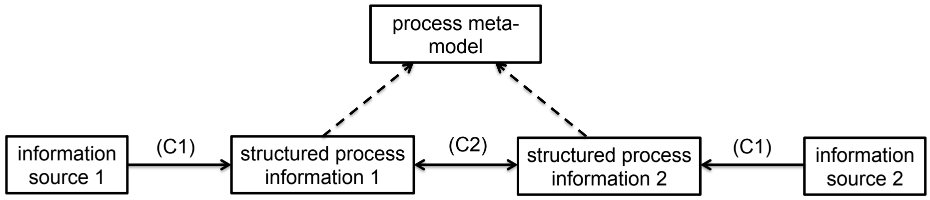

Fig. 1: Two challenges of process information defragmentation

3.1 Challenge 1: Extracting Structured Process Information

In order to integrate or compare process information from multiple sources, it is essential that the information is available in compatible or equal formats. This means that information from any source must be distilled into an interpretable and process-oriented format. While several process meta-models exist (see e.g. [30,34]), we do not argue for the selection of a specific one. This is because the exact requirements differ per situation. We therefore rather say that the distilled information should adhere to a given process meta-model that contains sufficient information for the purpose of a particular defragmentation task. For instance, if a defragmentation technique only considers the behavioral or control-flow perspective of a process, it suffices to be able to express the flow of process elements that are to be executed through, e.g. sequencing, feedback loops, and iteration [34]. In the remainder, we refer to information that adheres to such a process meta-model as *structured process information*.

Extracting structured process information is relatively straightforward for source types with an explicit, process-oriented structure such as *process models* and *business rules*. However, for information sources with a less apparent structuring of its constituting elements, i.e. sources containing *semi-structured data*, this extraction represents a significant challenge. The problem is that semi-structured data can be difficult to interpret in an automated manner [6]. This follows from the lack of an explicit, known, internal representation [8] and because it often cannot be exploited with *traditional* means [19].

Because each type of source has different internal representation of information, all types present their own challenges with regards to the extraction of structured process information. For instance, for textual documents the main challenges revolve around the identification of process model elements and their interrelations contained in a larger text. Which phrases actually refer to the steps that are being taken in a process and which to the actors involved in it? Answering such questions can only be achieved by acquiring some understanding of the textual semantics. For spreadsheets, on the other hand, it is crucial to correctly determine inter-relations between the information stored in different cells. Some cells may refer to the various phases that can be generally distinguished in a process, where other data may refer to execution data for specific cases. Because of the broad variation in type-specific challenges, dedicated extraction techniques for different types of sources must be developed. Yet, it seems crucial to exploit advances that have been made for one source to the fullest extent for harvesting others.

3.2 Challenge 2: Aligning Extracted Process Information

Correlating process information captured from different sources with each other represents the second challenge for defragmentation approaches. *Correspondence links* or *alignments* associate one or more elements from an information source with its corresponding elements in another source [44]. Any technique that checks for inconsistencies between sources, or attempts to integrate information captured in them, depends on such alignments. However, even if process information has been extracted into a structured form from both sources, creating this alignment is often not trivial. Although process analysis techniques often assume or require that extracted process model elements from multiple sources can be directly linked [5], these assumptions are generally not applicable in practice. Especially when information sources have been created by different stakeholders, it is highly likely that they differ with regard to terminology, structure, and abstraction level.

Differing terminology The usage of different vocabularies or naming schemes often leads to terminological differences between sources [43]. As an example, consider the two activities *Evaluate invoice* and *Check bill*. In fact, these labels may refer to the very same task. Both *invoice* and *bill*, as well as *check* and *evaluate*, are synonyms. In order to successfully recognize this, a technique would be required to identify semantic relationships between words.

Differing grammatical structures The previously discussed reasons for terminological variety may also lead to varying grammatical structures. These pertain to sentences in documents as well as the short text labels in process models or spreadsheets, for instance *Invoice creation* and *Create invoice*. In order to successfully relate process concepts with such differing grammatical styles, the different part of speeches such as verbs, nouns, and adjectives must be automatically recognized.

Differing abstraction levels Aside from differing terminology and structure, differences may also arise due to the usage of differing abstraction levels between sources. Figure 2 presents an example of different abstraction levels that must be resolved. While the review of an application comprises a single activity in model (A), it consists of three different activities in model (B). The detection of such one-to-many relationships still represents a considerable challenge for alignment techniques [10].

These three types of differences, together with the extraction of structured process information must be addressed for any information defragmentation task to succeed. Section 4 presents existing techniques that tackle these challenges.

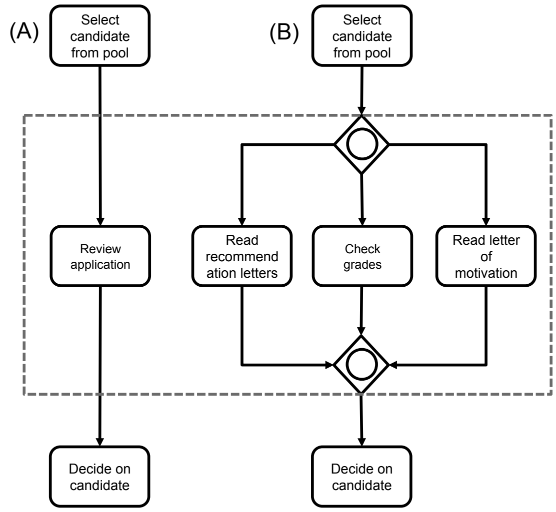

Fig. 2: Processes with different abstraction levels, from [44]

4 Overview

A variety of existing approaches already address the two main challenges of process information defragmentation techniques. In Section 4.1 we consider challenges related to the extraction of structured process information specific to several types of sources, as well as a number of techniques that deal with them. Section 4.2 then considers how existing techniques tackle the issues related to the alignment of structured process information obtained from different sources.

4.1 Extracting Structured Process Information

In this section we illustrate the extraction of structured process information from a selection of source types. The selected sources greatly differ in terms of their structure (or lack thereof) and type of information they can contain. This selection is not intended to be exhaustive. We rather aim to illustrate the broad variety of sources that can contain valuable process information and the particular challenges that must be dealt with in order to extract this information. For each of these source types we first consider the process-relevant information they potentially contain. Second, we consider the major extraction challenges specific to these sources. Finally, we introduce techniques that tackle or attempt to tackle these challenges. Note that the vast majority of these techniques extract structured process information in order to ultimately generate *process models*. This is simply one of the most popular formats to represent process information. Because of their general adherence to a certain process meta-model it is straightforward to conceive of the use of a comparable technique to other formats, e.g. a textual list of the various steps the process is composed of.

Business Rules Business rules represent a crucial and widely used information source for constraint-based process information. They are formal expressions of business policies, regulations and common-sense constraints that make business concerns explicit and traceable [17]. A plethora of different business rules representations formats exist, including the natural language SBVR (Semantics of Business vocabularies and Business rules) format [42], decision tables [12], attribute-relationship diagrams [26], and formal logics such as LTL (Linear Temporal Logic) and first order logic [9]. Business rules management (BRM) can be regarded as complementary to BPM [48]. This is because business rules and processes differ in abstraction levels [26]. Rules are also more practical to be updated, which makes them more suitable for the modeling of dynamic process aspects. For these reasons, business rules are sometimes regarded as the integration link between business and process modeling [28]. Their close relation to business processes make business rules important containers of process information. Recognizing this, several works extract structured process information in the form of process models from various business rule formats.

Existing work translates a variety of business rule representations into process models. Methods that achieve this include ones that convert Attribute Relationship Diagrams [26], decision tables [12], and SBVR rules [17]. Because of their formal nature and close relation to business processes, the distillation of business process elements from business rules to is relatively straightforward. The considered approaches therefore generate high quality results that can be confidently used for defragmentation tasks.

Spreadsheets Spreadsheets are interactive documents for the organization and analysis of data stored in tabular form. Their use is diverse, ranging from inventory administration, to educational applications, to financial modeling [21]. Due to their intuitive usage, they can be created by a large fraction of the working population. For example, Winston [47] estimates that 90% of all analysts in industry perform calculations in spreadsheets, while Panko found that 95% of US firms used them for financial reports [39]. Despite their extensive usage for business critical applications [21], spreadsheets are often ignored in business process management contexts [29]. For example, the business process modeling taxonomy of [16] does not include spreadsheets as a way of modeling processes. Admittedly, Krumnow et al. [29] briefly consider the possibilities of using spreadsheets for modeling, analysis, controlling, and simulation of business processes. Nevertheless, due to the lack of attention for spreadsheets, comprehensive insights into the kind of relevant process information spreadsheets contain and ways to distill this, are missing.

A downside of spreadsheets is that their contents are often not intuitive to comprehend. Users indicate that they experience difficulty understanding spreadsheets created by others [21]. This is problematic, because spreadsheet development and usage often includes multiple people [39]. This results in a situation where many persons involved in the maintenance and usage of spreadsheets do not fully understand the contents. A related result is that spreadsheets have a high risk of undetected errors [22,39]. These issues call for methods that allow users to visualize the contents of spreadsheets in a process-oriented manner. Due to the broad range of usage possibilities of spreadsheets, obtaining process information from them in an automated manner can be considered challenging.

To extract information from spreadsheets, the underlying semantics of the cells and their inter-relations must be understood. For this, there are two main difficulties to overcome: (1) cell type classification, and (2) name resolution. Different cell types exists within spreadsheets. For instance [2] identifies header, footer, data, and filler cells, whereas [20] differentiates between label, data, formula, and empty cells. Once cell types have been successfully identified, appropriate labels or headers must be attached to cells capturing values or formulas. For instance, in the data captured in Figure 3, it is important to recognize that the label "monthly payment" relates to the values €101,25 and €472,56. Although they are popular in practice, approaches that relate spreadsheets to business process concepts are scarce.

	A	B	C	D
1	**loan amount**	**interest (%)**	**term (years)**	**monthly payment**
2	€ 10.000,00	5	10	€ 101,25
3	€ 56.000,00	6	15	€ 472,56

Fig. 3: Excerpt of tabular data typically captured in spreadsheets

A notable exception is the method of [21], which visualizes computations in Excel spreadsheets. The resulting data-flow diagrams can be used to visualize data manipulating processes, and dependencies between data elements. While this work is certainly promising, it focuses on a single type of process information. Due to the wide application range of spreadsheets, it seems paramount to consider the extraction of process information from them in a broader fashion.

Text Documents Text documents are widely used throughout organizations for a variety of purposes. On the one hand, this results from the potential of natural language to freely express almost any type of information. On the other hand, text documents are popular because they can be created and understood by all stakeholders of an organization [4]. In the context of business processes, text documents may include verbal process descriptions and protocols, laws and regulations as well as annotations that provide additional information about a process and its steps.

Despite their wide usage, text documents have an important downside from an analytic point of view: Natural language can be highly ambiguous. Hence, text documents are not always completely clear and easy to understand. As a result, process information captured in more structured representations, such as process models, can be easier and faster interpreted [13]. The problem of ambiguity is amplified when texts are analyzed in an automatic manner. To handle this, recent research has turned towards the use of *natural language process* (NLP) techniques to extract structured process information from text documents.

The issues that must be addressed when extracting structured process information from natural language text are considerable. [15] presents the most important issues found in scientific literature. A seminal problem is the mismatch between syntax and semantics in natural language. This mismatch follows because the syntactic structure of a sentence is not directly linked to its semantics [14]. Another issue that is typical for natural language texts is the usage of anaphoric references or *anaphoras*. Anaphoras are usually pronouns ("he", "her", "it") or determiners ("this", "that") that refer to a previously introduced unit. For instance, the word "it" in the sentence "If it is not available, it is backordered", only receives it meaning if we consider the sentence that precedes it: "If a part is available, it is reserved." Dealing with these and other challenges related to natural language text is complex and prone to errors.

Nevertheless, a variety of methods exist that generate process models from different types of text documents. Examples include methods for generating models from textual process descriptions [15], group stories [18], and use cases [41]. These techniques differ greatly in their applicability. Some (e.g. [18]) require specific domain knowledge, for instance in the form of ontologies, while others, such as [15], are more generally applicable.

Event Data Information systems store data related to the execution of business processes. Contrast to information captured in e.g. process models, business rules, and text documents, which represents desired or supposed process behavior, so-called *event data* represents information about the reality of a process. *Process mining* techniques analyze event data, captured in *event logs*, to discover, monitor, and improve processes [1]. An important use case of process mining is the generation of a process model from an event log. The challenges that these techniques must deal with are very different from the issues faced when extraction process information from other considered sources.

An event log stores information on a case level, where every case represents a single instance of a process execution. To distill event data into the same abstraction level as information obtained from other types of sources, process mining techniques must generalize instance level behavior captured in event logs. This is achieved by *process discovery* techniques. The main issue that process discovery techniques must deal with is the balance between over- and underfitting of process models to the behavior captured in event logs. This is necessary, because event logs can often contain noisy or exceptional behavior, that may obscure the understandability of a generated process model [7].

A plethora of process discovery techniques have been developed that address this problem in many different ways. Due to the widespread attention that process mining has received over the past decade, these techniques can be considered to be far more mature than extraction techniques of other types of information sources. We therefore refer the interested reader to e.g. [1] for broader insights into the different types of process discovery algorithms.

4.2 Aligning Extracted Process Information

Challenge 2 recognizes that to compare or combine process information from different sources, alignments must be created between the process concepts stored in them. So-called *matchers* aim to achieve this in an automated manner. In order to succeed, matchers must address the issues introduced in Section 3.2: differences in terminology, grammatical structures, and abstraction levels.

To deal with terminological differences, e.g. *Evaluate invoice* and *Check bill*, semantic similarity (i.e. synonymous relations) between words must be identified. In prior work, this turned out to be a major challenge in business process contexts despite the existence of frequently employed tools that support the identification of synonyms [31]. This is because popular tools, such as WordNet [37], do not contain the full range of vocabulary that occurs in organizational documents. Hence, many terms cannot be related simply because there are not part of the available taxonomy. Therefore, better strategies to measure semantic similarity have to be identified in the future. Promising directions may, for instance, include co-occurrence based techniques such as DISCO [27]. They are based on text corpora and, hence, can be trained on any large text collection.

Business process concepts captured in different information sources may also exhibit differences in labeling or grammatical styles. To recognize that labels such as *Invoice creation* and *Create invoice*, relate to the same activity, *parts-of-speech*, e.g. verbs and objects, must be identified in these labels. While there are precise techniques available to analyze full natural language sentences (see e.g. [25]), techniques for analyzing short process model text labels have not yet reached the same level of maturity [32]. Current solutions are based on rules and heuristics, rather than statistical models. These models, however, are notably successful in generic natural language parsers [24]. Hence, it might be a promising direction to transfer those concepts from standard natural language processing tools to the analysis of short text labels.

Differences in abstraction levels used to describe process concepts pose the third class of problems for the creation of alignments. To take on these situations, matchers must recognize situations where a concept in one information source represents a sub-part of a larger concept contained in another source. Although there are taxonomies such as the MIT process handbook [36] or WordNet [37] that define part-of-relationships, their applicability is again limited by the low coverage of the variety of terms that are contained in business documents. Hence, also in this direction novel solutions are required. To work around the lack of appropriate taxonomies, some matchers, e.g. [31,33,45], turn towards the context in which activities occur. Namely, they exploit the structural relations between process concepts in order to match activities based on similar neighboring elements. Consider, for instance, the two models contained in Figure 2, the relation between the middle activity in model (A) and the three activities in model (B), becomes very clear when the context of the activities is considered. This is because both sets of activities are preceded and succeeded by similar (even equal) activities.

Despite the incorporation of aforementioned, sophisticated techniques to deal with the challenges of process alignment, the quality of matchers still leaves room for improvement. This is illustrated by the results of the *Process Model Matching Contest 2013*, in which seven matchers from different research groups tried to solve the same matching problem [10]. The best technique only achieved an f-measure of 0.45. This highlights that the alignment of process concepts is a very relevant and open challenge that impedes successful information defragmentation.

5 Discussion

The previous sections outlined the major challenges to be met, and how existing techniques are meeting these to support information defragmentation. To advance the state of these techniques, we identify three main frontiers where further developments are needed.

First, many of the extraction methods considered in Section 4.1 represent preliminary approaches that leave plenty of room for improvement. On the one hand, this results from the limited scope that methods take into consideration. The majority of methods focus only on control-flow information, while other perspectives, e.g. organizational and data perspectives, are largely ignored. On the other hand, the quality of the results obtained through existing techniques is in some cases low. This is especially the case for sources with a less explicit structure, such as text documents and spreadsheets, that pose considerable information extraction challenges. As a result, existing techniques yield sub-optimal results. We therefore argue that in many cases it is necessary to improve upon the quality of existing methods for extracting structured process information from semi-structured sources.

Quality issues also manifest themselves with respect to approaches to align extracted process information from various sources. As seen in Section 4.2, even matchers designed to deal with information from homogeneous source types, i.e. process model matchers, achieve far from perfect results. We therefore argue to further develop the quality of existing matchers as a second frontier, despite the sophisticated techniques already incorporated in them.

Finally, we note the lack of comprehensive insights into process information sources that are available within organizations. To overcome this, it is important to identify which sources are available within organizations, and how these can be used for process information defragmentation. The information sources considered in Section 4.1 represent a set of important and widely available source types, but they merely provide an illustration of the variety of sources that contain process information. Furthermore, even for these known sources, comprehensive insights into their usability for information defragmentation are missing. Particularly, source types with a broad range of uses, namely text documents and spreadsheets, can contain process-related information in a large variety of forms. This lack of insights impedes the results achievable by process information defragmentation since important information may remain unconsidered.

6 Conclusions

The fragmentation of process information is a major issue that impedes the ability of organizations to obtain value from their documented process knowledge. Through the insights presented in this work, we have demonstrated the relevance of extraction and matching techniques to tackle problems caused by process information fragmentation. Despite the existence of various techniques to tackle these issues, there is a window of opportunity to improve upon the state of art.

We have noted that the problems related to information fragmentation in business process contexts have not yet been investigated in a systematic manner. Therefore, with this work we aimed to structure the challenges that impede process information defragmentation techniques. This resulted in actionable steps that any defragmentation technique faces. To further advance this field, we identified several major development frontiers. First, we argue for the importance of improved and new techniques for the extraction and alignment of process information from various sources. Aside from those directions, an important driver for future research is the lack of a comprehensive overview of relevant information sources. In the end we hope that this work sets the stage for future efforts to resolve the fragmentation of documented knowledge on business processes.

References

1. Van der Aalst, W.: Discovery, conformance and enhancement of business processes. Springer (2011)
2. Abraham, R., Erwig, M.: Header and unit inference for spreadsheets through spatial analyses. In: Visual Languages and Human Centric Computing, 2004 IEEE Symposium on. pp. 165–172. IEEE (2004)
3. Alavi, M., Leidner, D.E.: Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. MIS quarterly pp. 107–136 (2001)
4. Allweyer, T.: BPMN 2.0: introduction to the standard for business process modeling. BoD–Books on Demand (2010)
5. Baier, T., Mendling, J.: Bridging abstraction layers in process mining by automated matching of events and activities. In: Business Process Management, pp. 17–32. Springer (2013)
6. Blumberg, R., Atre, S.: The problem with unstructured data. DM REVIEW 13, 42–49 (2003)
7. Buijs, J., Van Dongen, B., Van der Aalst, W.: Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. International Journal of Cooperative Information Systems 23(01) (2014)
8. Buneman, P., Davidson, S., Fernandez, M., Suciu, D.: Adding structure to unstructured data. In: Database Theory—ICDT'97, pp. 336–350. Springer (1997)
9. Caron, F., Vanthienen, J., Baesens, B.: Comprehensive rule-based compliance checking and risk management with process mining. Decision Support Systems 54(3), 1357–1369 (2013)
10. Cayoglu, U., Dijkman, R., Dumas, M., Fettke, P., Garcıa-Banuelos, L., Hake, P., Klinkmüller, C., Leopold, H., Ludwig, A., Loos, P., et al.: The process model

matching contest 2013. In: 4th International Workshop on Process Model Collections: Management and Reuse (PMC-MR'13) (2013)
11. Chakraborty, S., Sarker, S., Sarker, S.: An exploration into the process of requirements elicitation: a grounded approach. Journal of the Association for Information Systems 11(4), 1 (2010)
12. De Roover, W., Vanthienen, J.: On the relation between decision structures, tables and processes. In: On the Move to Meaningful Internet Systems: OTM 2011 Workshops. pp. 591–598. Springer (2011)
13. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of business process management. Springer (2013)
14. Fillmore, C.J.: The case for case. (1967)
15. Friedrich, F., Mendling, J., Puhlmann, F.: Process model generation from natural language text. In: AISE. pp. 482–496. Springer (2011)
16. Giaglis, G.M.: A taxonomy of business process modeling and information systems modeling techniques. International Journal of Flexible Manufacturing Systems 13(2), 209–228 (2001)
17. Goedertier, S., Vanthienen, J.: Declarative process modeling with business vocabulary and business rules. In: On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. pp. 603–612. Springer (2007)
18. Gonçalves, J., Santoro, F.M., Baiao, F.A.: Business process mining from group stories. In: 13th International Conference on Computer Supported Cooperative Work in Design. pp. 161–166. IEEE (2009)
19. Herbst, J., Karagiannis, D.: An inductive approach to the acquisition and adaptation of workflow models. In: Proceedings of the IJCAI. vol. 99, pp. 52–57 (1999)
20. Hermans, F., Pinzger, M., van Deursen, A.: Automatically extracting class diagrams from spreadsheets. In: ECOOP 2010–Object-Oriented Programming, pp. 52–75. Springer (2010)
21. Hermans, F., Pinzger, M., van Deursen, A.: Supporting professional spreadsheet users by generating leveled dataflow diagrams. In: Proceedings of the 33rd International Conference on Software Engineering. pp. 451–460. ACM (2011)
22. Hesse, R., Scerno, D.H.: How electronic spreadsheets changed the world. Interfaces 39(2), 159–167 (2009)
23. Jung, J., Choi, I., Song, M.: An integration architecture for knowledge management systems and business process management systems. Computers in industry 58(1), 21–34 (2007)
24. Jurafsky, D., Martin, J.H.: Speech & language processing. Pearson Education India (2000)
25. Klein, D., Manning, C.D.: Accurate unlexicalized parsing. 41st Meeting of the Association for Computational Linguistics pp. 423–430 (2003)
26. Kluza, K., Nalepa, G.J.: Automatic generation of business process models based on attribute relationship diagrams. In: Business Process Management Workshops. pp. 185–197. Springer (2014)
27. Kolb, P.: Disco: A multilingual database of distributionally similar words. Proceedings of KONVENS-2008, Berlin (2008)
28. Kovacic, A.: Business renovation: business rules (still) the missing link. Business process management journal 10(2), 158–170 (2004)
29. Krumnow, S., Decker, G.: A concept for spreadsheet-based process modeling. In: Business Process Modeling Notation, pp. 63–77. Springer (2010)
30. La Rosa, M., Reijers, H.A., Van Der Aalst, W., Dijkman, R.M., Mendling, J., Dumas, M., García-Bañuelos, L.: Apromore: An advanced process model repository. Expert Systems with Applications 38(6), 7029–7040 (2011)

31. Leopold, H., Niepert, M., Weidlich, M., Mendling, J., Dijkman, R., Stuckenschmidt, H.: Probabilistic optimization of semantic process model matching. In: Business Process Management, pp. 319–334. Springer (2012)
32. Leopold, H., Smirnov, S., Mendling, J.: On the refactoring of activity labels in business process models. Information Systems 37(5), 443–459 (2012)
33. Ling, J., Zhang, L., Feng, Q.: Business process model alignment: An approach to support fast discovering complex matches. In: Enterprise Interoperability VI, pp. 41–51. Springer (2014)
34. List, B., Korherr, B.: An evaluation of conceptual business process modelling languages. In: Proceedings of the 2006 ACM symposium on Applied computing. pp. 1532–1539. ACM (2006)
35. Ly, L.T., Rinderle-Ma, S., Knuplesch, D., Dadam, P.: Monitoring business process compliance using compliance rule graphs. In: On the Move to Meaningful Internet Systems: OTM 2011, pp. 82–99. Springer (2011)
36. Malone, T.W., Crowston, K., Herman, G.A.: Organizing business knowledge: the MIT process handbook. MIT press (2003)
37. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM 38(11), 39–41 (1995)
38. Möhring, M., Schmidt, R., Härting, R.C., Bär, F., Zimmermann, A.: Classification framework for context data from business processes. In: Workshop on Business Process Management and Social Software (2014)
39. Panko, R.R.: What we know about spreadsheet errors. Journal of Organizational and End User Computing (JOEUC) 10(2), 15–21 (1998)
40. Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. the VLDB Journal 10(4), 334–350 (2001)
41. Sinha, A., Paradkar, A.: Use cases to process specifications in business process modeling notation. In: Web Services (ICWS), 2010 IEEE International Conference on. pp. 473–480. IEEE (2010)
42. Team, S., et al.: Semantics of business vocabulary and rules (sbvr). Tech. rep., Technical Report dtc/06–03–02 (2006)
43. Wache, H., Voegele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hübner, S.: Ontology-based integration of information-a survey of existing approaches. In: IJCAI-01 workshop: ontologies and information sharing. vol. 2001, pp. 108–117. Citeseer (2001)
44. Weidlich, M., Barros, A., Mendling, J., Weske, M.: Vertical alignment of process models–how can we get there? In: Enterprise, Business-Process and Information Systems Modeling, pp. 71–84. Springer (2009)
45. Weidlich, M., Dijkman, R., Mendling, J.: The icop framework: Identification of correspondences between process models. In: Advanced Information Systems Engineering. pp. 483–498. Springer (2010)
46. Weidlich, M., Weske, M., Mendling, J.: Change propagation in process models using behavioural profiles. In: Services Computing, 2009. SCC'09. IEEE International Conference on. pp. 33–40. IEEE (2009)
47. Winston, W.: Executive education opportunities millions of analysts need training in spreadsheet modeling, optimization, monte carlo simulation and data analysis. OR MS TODAY 28(4), 36–39 (2001)
48. Zoet, M., Versendaal, J., Ravesteyn, P., Welke, R.: Alignment of business process management and business rules (2011)