Grußworte

Liebe Fachgruppenmitglieder,

Ich freue mich, Sie als Empfänger der neusten Ausgabe des EMISA Forums der Fachgruppe EMISA begrüßen zu dürfen. Der Höhepunkt in diesem Jahr war wieder das Fachgruppentreffen der EMISA, das am 20.-21. Mai 2021 in Kiel angesetzt und als Onlineveranstaltung stattfand. Wir danken Agnes Koschmider und Judith Michael, welche die Leitung übernommen und ein hervorragendes Programm zusammengestellt haben. Insgesamt wurden vier Keynotes beigesteuert und 30 Forschungsvorträge gehalten.

Ein wichtiger Programmpunkt war auch diesmal wieder die Mitgliederversammlung, auf welcher ein neues Leitungsgremium gewählt wurde. Wir danken der GI für die Bereitstellung der Online-Wahlplattform. Somit wird das Leitungsgremium in den Jahren 2022-2024 aus folgenden Personen bestehen:

- Dominik Bork
- Dirk Fahland
- Agnes Koschmider, Sprecherin
- Henrik Leopold
- Judith Michael
- Erik Proper, stellvertretender Sprecher
- Luise Pufahl
- Ulrich Reimer
- Stefanie Rinderle-Ma
- Matthias Weidlich

Jan Mendling und Manfred Reichert standen nach jeweils zwei Amtsperioden als Sprecher nicht mehr zur Wahl. Sie wurden als Fachexperten in das Leitungsgremium gewählt. Nach zwei Amtszeiten als Sprecher möchte ich mich für die Unterstützung und die vertrauensvolle Zusammenarbeit bedanken. Ich freue mich, dass wir in dieser Zeit das Veranstaltungsformat des Fachgruppentreffens erfolgreich überarbeiten konnten. In Zukunft stehen eine ganze Reihe weiterer Veränderungen an. Dies betrifft insbesondere das EMISA Forum. Dies wird die letzte Druckausgabe unserer ehrwürdigen Zeitschrift sein. Wir haben beschlossen, andere Kommunikationskanäle für unsere Mitglieder aufzubauen, und ich freue mich, dass Agnes und Ihr Team sich dieser Aufgabe annehmen werden. Damit rückt insbesondere unser Fachgruppentreffen noch stärker in den Vordergrund. Wir freuen uns, dass Erik Proper die EMISA 2022 am 2.-3. Juni in Luxemburg veranstalten wird, unterstützt von Henrik Leopold als zweiten Programmleiter. Wir freuen uns auf ein Wiedersehen in Luxemburg!

Mit herzlichen Grüßen,

Jan Mendling
(EMISA-Sprecher)

https://www.emisa2022.eu/home

New information technologies such as AI, Digital Twins, Block Chain, Big Data, IoT, etc, enable enterprises to innovate their core activities, from their information systems, via their business processes, to their business models. These innovations pose design and engineering challenges which, in their turn, result in several research challenges for the fields of enterprise modelling and information systems architectures. At the same time, the new information technologies also provide new opportunities to support the work involved in (continuous) design and engineering of enterprises and their information systems. Examples include (IoT based) process and enterprise mining, the use of AI to help in the creation of models, low-code platforms, etc.

EMISA 2021 is the twelfth international workshop in a series that provides a key forum for researchers and practitioners in the field on design methods for information systems. The workshop series emphasises the need for a coherent view on this field, fostering integrated approaches that address and relate all relevant aspects of enterprises (e.g. value propositions, business services, business processes, business rules, information systems, IT infrastructures), cross-cutting concerns (e.g. security, privacy, compliance) across stakeholders, designers, engineers, and domain experts.

Subjects

The workshop is open to a broad range of subjects. Possible topics include, but are not limited to:

- Patterns for enterprise and information systems architectures
- Service and value modelling
- Process modelling and process-aware information systems
- Complex event processing and event-driven architectures
- Domain-specific modeling methods and languages
- Meta-modeling and foundational ontologies
- Theoretical foundations of (enterprise) modelling
- Method engineering
- Quality of modelling methods and languages, and, (architecture) models
- Learning and teaching methods and languages
- Process mining, and enterprise mining in general
- Assisted enterprise and information systems modelling

The workshop is organized by the GI Special Interest Group on Design Methods for Information Systems (GI-SIG EMISA), which provides a forum for researchers from various disciplines who develop and apply methods to support the analysis and design of information systems.

Submission Types

EMISA 2022 calls for submissions in the following categories:

- PhD Research Proposals (short paper of up to 5 pages): There will be a dedicated slot in the program to discuss PhD research proposals including the current status and the further plan of the research work.

- Current Research Talk Proposals (extended abstract of up to 2 page): Proposals for scientific talks of international excellence. Eligible are proposal submissions that are based on published or accepted papers from international conferences or journals.

- Novel Directions Talk Proposals (short paper of up to 5 pages): Proposals for talks that motivate a novel research direction, outline the research gaps to address, and carve out major challenges. These talks shall serve as a stimulus for discussions as part of a dedicated slot in the workshop program.

- Case Reports: EMISA 2022 explicitly welcomes reports on real-world cases involving modelling in an enterprise context. To this end, EMISA 2022 collaboratives with other events (e.g. EMMSAD, ER, CBI, and PoEM) to gather such case reports. See www.models-at-work.org for more details.

Submission

- PhD Research Proposals, Current Research Talk Proposals, and Novel Directions Talk Proposals should be submitted to [to be done].
- For the submission of Case Reports, please follow the instructions on www.models-at-work.org

Publication

- The extended abstracts of the Current Research Talks will be published in the next edition of the EMISA Forum.
- The short papers of PhD Research Proposals and Novel Directions Talk, will be published as an electronic CEUR proceedings volume.
- For the publication channel used for Case Reports, please refer to www.models-at-work.org

Important Dates (to be confirmed)

All dates in 2022:

- Submission: March 6th
- Notification: April 17th
- Camera ready copies for short papers of the PhD Research Proposals and Novel Directions Talks: May 1st
- EMISA 2022: June 2nd and 3rd

Domain models are created and used in many different contexts, including business process management, enterprise engineering, enterprise architecture management, requirements engineering, information systems engineering, software engineering, ontology engineering, etc, across science and industry. These different contexts result in a rich variety of more specific types of domain models, such as enterprise (architecture) models, business process models, information models, class diagrams, value models, reference models, ontologies, knowledge graphs, semantic web specifications, etc. Depending on the context, domain models can be created with the aim of being an (as truthful as possible) representation of the conceptual structure of the domain that is modelled; leading to conceptual models. In addition, to accommodate for specific uses and contexts, domain models may also incorporate "conceptual compromises" which, for instance, result in domain models that lend themselves better for animation, execution, gamification, or automated (logic-based) reasoning. The use of domain models may, sometimes, even go unnoticed since these models do not always take the form of traditional "boxes and lines" diagrams or some other dedicated notation.

The creation, management, and use, of domain models in scientific and industrial practice is done in the context of some expected *Return on Modelling Effort* (RoME). In other words, it is expected that the model will provide a certain value that will offset the costs involved in the creation and maintenance of the model.

To ensure relevance of research into (the many different forms of) domain modelling, it is important to gather insights from the use of domain models in practice. To, indeed, better understand the practical needs for, and use of, domain models, it would be beneficial to have a library of cases in which domain models have played a crucial role. This desire is also shared by existing academic events including (in order of their planning in 2022):

- the GI EMISA (Enterprise Modelling and Information Systems Architectures) workshop,
- the IFIP 8.1 EMMSAD (Exploring Modelling Methods for Systems Analysis and Design) working conference,
- the IEEE CBI (Conference on Business Informatics),
- the ER conference on Conceptual Modelling, and
- the IFIP 8.1 PoEM (Practice of Enterprise Modelling) working conference.

It is the desire of these events to, at least in 2022, include explicit reports on the creation and use of domain models in practice (in science or industry) in a dedicated "Models-at-Work" track.

This shared desire has now resulted in a coordinated effort to gather (and share among the events) case reports regarding the creation and use of models in scientific and industrial practice. The longer term ambition is to create an annual "Models-at-Work" post-proceedings involving the case reports within that year. In line with this, the plan is to have an ongoing call, across the associated events, for such case reports. Each case report will be reviewed in a process similar to the submission of papers to a journal. In other words, submissions can be re-submitted based on improvements required by the reviewers. Once accepted, the case reports will be available to be presented (in virtual or physical mode) at one of the associated events. The allocation of the accepted case reports to the associated events will be done jointly by the authors and the organisers of the involved events.
If successful, the plan is to continue this in the coming years. The resulting library of case reports could be used in (at least) three directions:

- *Science*: challenges from the real world are made explicitly visible to researchers.
- *Practice*: results from research are illustrated (by way of cases) to practitioners.
- *Teaching*: case reports can be used for educational purposes

Requirements on case reports

Case reports involving the use of models in science and/or industry (including NGO's) are welcomed.

Each case report should at least aim to address the following questions:

- Situation & challenge: What was the social and/or technical situation in which the model was created and/or used. What was specifically challenging in the case? What were uncertainties? Were there any social and/or technical complexities?
- Purpose & requirements: What was the intended purpose (and audience) of the model and/or its creation? What were specific requirements on the model?
- Activities & effort: What were the activities involved in creating the model? How much effort (time, budget, people/roles involved, etc) was needed to create the model? What tools and methods were used? How was the validity (in relation to the goal & requirements) of the model managed and assessed?
- Resulting model: What kind of model resulted? A formal model? An implicit, or informal model? Did the model have to include "quality compromises" for strategic/political reasons? Was the developed model a refinement of a standard or published/known model? Was a specific modelling language and/or tool used? If so, which language/tool? If it is allowed to be shared, then include (a summary of) the actual model.
- Return on modelling effort: In line with the intended purpose, what was the expected return on modelling effort? What was the materialised return on modelling effort? Which stakeholder(s) made the investment in modeling, and which stakeholder(s) reaped the benefits?
- Reflection: What were lessons learned? Positive lessons? Negative lessons? Challenges for (future) research?
- Evidence: What sources were used to author the case report? How were they gathered and used in authoring this case report? For instance: Was the case report (co-)authored by the actual practitioners? Did the authors conduct interviews of the practitioners involved? How were these interviews conducted? Or, did the authors conduct a form of "investigative journalism"?

Case reports can be (co)authored by practitioners who participated in the development/use of the actual model(s). However, we are specifically also open to case reports based on e.g. interviews of practitioners involved in the creation/use of models in specific cases, or based on "investigative journalism", conducted by (young / aspiring) researchers.

Organisation

For 2022, the organisation provided below will be used. When successful, this will evolve further.

Editor in chief

Henderik A. Proper, Luxembourg Institute of Science & Technology and the University of Luxembourg, Luxembourg

Deadlines

There are no specific deadlines for submissions. However, to enable inclusion as a case report to a specific associated event, please indicate this at the time of submission, while also respecting following indicative submission timeframes for 2022:

When submitted before march, accepted case reports can be invited for presentation at

- GI EMISA (Enterprise Modelling and Information Systems Architectures) workshop.
- IFIP 8.1 EMMSAD (Exploring Modelling Methods for Systems Analysis and Design) working conference
- IEEE CBI (Conference on Business Informatics).

When submitted before may, accepted case reports can be invited for presentation at:

- ER conference on Conceptual Modelling.

When submitted before august, accepted case reports can be invited for presentation at:

- IFIP 8.1 PoEM (Practice of Enterprise Modelling) working conference.

Submission
Submissions should not exceed 25 pages (including references, and included appendixes). Within the next weeks, we hope to clarify the publishing platform that will be used for the *Models at Work* series, and the associated layout requirements.

Initial submissions should be mailed as a PDF to: submissions@models-at-work.org

Call for Papers: 14th ZEUS Workshop

February 24-25, 2022 in Hannover Germany
website: https://zeus2022.pi.uni-bamberg.de/

Participation is free of charge!

Objectives

ZEUS focuses on the discussion of fresh ideas, the presentation of work in progress, and the establishment of a scientific network between young researchers in the region.

1. Discuss fresh ideas
We offer a forum to discuss ideas at a level that is more work-in-progress than in a traditional conference. We thereby want to attract especially PhD students in the early phases of their work. Participants can get feedback from outside their group before a submission to a reviewed conference. This makes ZEUS a great opportunity to discuss ideas.

2. Practice scientific work
We see the ZEUS workshop as an opportunity to practice the whole range of scientific work. We do not put the sole focus on the submitted papers themselves, but also on the presentations and the discussions during the workshop. To this end, we hand out a Best Presentation Award since 2010 at the end of the workshop to appreciate high quality presentations.

3. Establish contacts between young researchers in the region
We aim at bringing together young researchers who work in the same geographic and scientific region. This way, we would like to provide an opportunity for people to establish a scientific network that can be intensely used, including mutual visits at affordable costs. The workshop will serve as a platform to present current research ideas and research directions.

Topics

The topics of the ZEUS workshop are centered around service technologies, which include a rich set of facets. The purpose of analysis, synthesis, or simulation of service technologies are as welcome as practical evaluations, use case-driven feasibility studies, or technology adoption models. ZEUS also calls for contributions in the field of Cloud Computing, RESTful services, and microservices.

Topics include, but are not limited to:

- Service lifecycle: analysis, specification, modelling, testing, deployment, execution, monitoring, adaptation
- Patterns, languages, reference models, and model extensions
- Multi-view and multi-perspective engineering (SOA, choreographies, collaborations, conversations, artifact-centric systems)
- Formal methods, models, simulation, and verification
- System architectures for service composition
- RESTful Web services (design aspects, hypermedia, linked data, mashups, conversations)

- Microservices and Nanoservices (architecture, lifecycle, deployment, composition)
- Workflows, business processes, and business decisions (modelling, execution, analysis, mining, as well as papers on blockchains and BPM)
- Complex event processing (correlation, aggregation, transformation, monitoring, extraction)
- Security, compliance, and non-functional requirements and properties
- Cloud-enabled applications, migration to/from the Cloud, Cloud Integration, Serverless Computing
- Containerization, Container Orchestration Systems
- Composable Big Data Analytics Pipelines
- Applications, frameworks, methods, tool demonstrations, and case studies

Submission

We are looking forward to three types of contributions for ZEUS. All papers must be submitted following the instructions at the ZEUS submission site handled by EasyChair: https://easychair.org/conferences/?conf=zeus2022

Results can be presented in talks or tool demonstrations. Submissions will be reviewed by at least three reviewers each in order to assure general fitness regarding content, readability and scope and to give first feedback to the authors. Depending on innovation, technical soundness and presentation clarity, papers may be rejected or accepted as position or workshop papers.

Workshop papers:
Workshop papers are "regular" contributions that describe original solutions in field of ZEUS. These papers must not exceed 6 pages (LNCS style). The 6 pages does not include references, so there is more space for your work. Workshop papers are reviewed according to the call for papers. Accepted papers shall be included in the proceedings and presented at the workshop.

Positions papers:
Position papers should draft a new idea and put it up for discussion at the workshop. Position papers should only be an extended abstract and must not exceed 3 pages (LNCS style) without references. Position papers are briefly reviewed according to the call for papers. The main idea and the relation to existing work should be contained. Accepted papers shall be included in the proceedings.
Position papers allow authors to get early feedback during the workshop, but should not disallow extending the paper to a full paper submitted to a first class conference – even if the position paper is referenced and the delta is explained properly.

Tools demonstrations:
ZEUS also offers a forum to demonstrate implementations of techniques and algorithms in the area of the aforementioned topics to get early feedback and provide interesting insights for the audience. Tool demonstrators are asked to submit a demo script of no more than 3 pages (LNCS style) without references which states how the tool is linked to the call for papers and what to expect during the demonstration.

Important Dates

Submission: January 18, 2022
Notification: February 13, 2022
Camera-ready (pre-proceedings) version: February 22, 2022

Registration: February 20, 2022
Workshop: February 24-25, 2022

Submission Guidelines

Template: LNCS style - https://github.com/latextemplates/LNCS
Workshop Paper: 6 pages excluding references
Position Paper: 3 pages excluding references
Tool Demonstration: 3 pages excluding references

Organization

Steering Committee

- Oliver Kopp, Daimler AG, Germany
- Nico Herzberg, Campeleon, Germany
- Stefan Kolb, Lion[5] GmbH, Germany
- Stephan Haarmann, Hasso Plattner Institute Potsdam, Germany
- Johannes Manner, University of Bamberg, Germany

Program Chair

Johannes Manner, University of Bamberg, Germany

Local Organizer

Daniel Lübke, Digital Solution Architecture GmbH

Contact

E-mail: zeus2022@easychair.org

More Information

ZEUS 2022 Homepage: https://zeus2022.pi.uni-bamberg.de/
ZEUS WS Series: http://zeus-workshop.eu

Drivers and Barriers for Microservice Adoption (Extended Abstract)

Holger Knoche[1] Wilhelm Hasselbring[2]

Abstract: In this research talk, we present the results of a survey on drivers and barriers for microservice adoption among professionals in Germany; published in Enterprise Modelling and Information Systems Architectures (EMISAJ) – International Journal of Conceptual Modeling [KH19].

In addition to overall drivers and barriers, we particularly focus on the use of microservices to modernize existing software. We observe interesting differences between early adopters who emphasize scalability of their Internet-scale systems, compared to traditional companies that emphasize maintainability.

Keywords: Microservice architecture; Software modernization; Microservice adoption

1 Microservices

Microservices are an architectural style for software which currently receives a lot of attention in both industry and academia. Several companies employ microservice architectures with great success, and there is a wealth of blog posts praising their advantages. Especially so-called Internet-scale systems use microservices to satisfy their enormous scalability requirements and to rapidly deliver new features to their users. In addition to scalability [Ha16], microservices may furthermore enable both agility and reliability [HS17]. However, microservices are not only popular with large, Internet-scale systems. Many traditional companies are also considering whether microservices are a viable option for their applications. These companies may have other motivations to employ microservices, and see other barriers which could prevent them from adopting microservices.

2 Drivers and Barriers

As a consequence, many companies are currently considering whether microservices are a viable option for their software systems. However, many of these systems are not *Internet-scale*; instead, they are used by a known, limited, and stable number of users. Therefore, these companies may consider microservices for other reasons than the early adopters. Even more interesting are expected barriers which may prevent these companies from adopting microservices. Several authors warn against considering microservices as viable for every

[1] Kiel University, Software Engineering Group, Kiel hkn@informatik.uni-kiel.de
[2] Kiel University, Software Engineering Group, Kiel hasselbring@email.uni-kiel.de

software system, as there are numerous trade-offs to consider [Ki16; Si16]. Furthermore, these drivers and barriers possibly differ among industry sectors.

Although microservice adoption is discussed extensively in online media, there is yet little research data on the subject. While some studies on microservice adoption in practice exist, many of them have only been conducted with few participants, and several open questions still remain. In order to gain insight into the reasons why *traditional* companies are considering the adoption of microservices, we conducted a survey among software development professionals in Germany. Since many companies already have existing software assets, we furthermore investigated to what extent microservices are perceived as a tool for software modernization, which goals are pursued by introducing microservices into existing software, and how the potential impact on runtime performance and transactionality is rated. A particularly interesting aspect of microservices is that they are considered as a viable means for incrementally modernizing monolithic software applications [KH18].

3 Results

The premier drivers were found to be scalability, maintainability, and time to market, while the skill set of both development and operations staff was identified as the main barrier. In particular, for the adoption of microservices as a means for software modernization, maintainability is the leading driver.

Literatur

[Ha16] Hasselbring, W.: Microservices for Scalability: Keynote Talk Abstract. In: Proceedings of the 7th ACM/SPEC International Conference on Performance Engineering (ICPE). ACM, S. 133–134, 2016.

[HS17] Hasselbring, W.; Steinacker, G.: Microservice Architectures for Scalability, Agility and Reliability in E-Commerce. In: Proceedings of the 2017 IEEE International Conference on Software Architecture Workshops (ICSAW). S. 243–246, 2017.

[KH18] Knoche, H.; Hasselbring, W.: Using Microservices for Legacy Software Modernization. IEEE Software 35/3, S. 44–49, 2018.

[KH19] Knoche, H.; Hasselbring, W.: Drivers and Barriers for Microservice Adoption – A Survey among Professionals in Germany. Enterprise Modelling and Information Systems Architectures (EMISAJ) – International Journal of Conceptual Modeling 14/1, http://doi.org/10.18417/emisa.14.1, S. 1–35, 2019.

[Ki16] Killalea, T.: The Hidden Dividends of Microservices. Communications of the ACM 59/8, S. 42–45, 2016.

[Si16] Singleton, A.: The Economics of Microservices. IEEE Cloud Computing 3/5, S. 16–20, 2016.

Conceptual Modeling in a Digital World: A New Framework and Research Agenda (Extended Abstract)*

Jan Recker [2], Roman Lukyanenko[3], Mohammad Jabbari[4], Binny M. Samuel[5], and Arturo Castellanos[6]

Abstract: In an increasingly digital world, conceptual modeling research is more relevant than ever to the information systems field, but it requires an update with current theory. In [Re21] we develop a new theoretical framework of conceptual modeling to change the assumptions that govern research in this area. Our framework draws attention to the role of conceptual modeling scripts as mediators between physical and digital realities. We identify new research questions about grammars, methods, scripts, agents, and contexts that are situated in intertwined physical and digital realities. We discuss several implications for conceptual modeling scholarship that relate to the necessity of developing new methods and grammars for conceptual modeling, broadening the methodological array of conceptual modeling scholarship, and considering new dependent variables.

Keywords: conceptual modeling, mediation, ontological reversal, digital objects, representation.

1 Introduction

In a world that abounds with digital technologies, modern information systems (IS) no longer only represent reality but also increasingly shape it. This fundamental shift in IS's role has profound implications for the development and use of conceptual modeling (CM) to capture the relevant features of a real-world domain that an IS is intended to support.

The time is ripe to update CM theory because the traditional assumptions of CM scholarship increasingly under-represent and constrain CM's potential to support the ongoing digitalization of reality. Therefore, the purpose of our paper [Re21] was to offer a new framework for CM scholarship that is fit for the burgeoning digital world.

---

* This extended abstract summarizes our original article "From Representation to Mediation: A New Agenda for Conceptual Modeling Research in a Digital World" that appeared in MIS Quarterly, 45(1), 269-300, 2021.

[2] University of Hamburg, Hamburg Business School, jan.christof.recker@uni-hamburg.de, https://orcid.org/0000-0002-2072-5792
[3] HEC Montreal, Department of Information Technologies, roman.lukyanenko@hec.ca
[4] Queensland University of Technology, School of Information Systems, m.jabbarisabegh@qut.edu.au
[5] University of Cincinnati, Lindner College of Business, samuelby@uc.edu
[6] Baruch College, Zicklin School of Business, Arturo.Castellanos@baruch.cuny.edu

2 A New Framework for Conceptual Modeling in the Digital World

In our paper we reviewed the literature on conceptual modeling to demonstrate that field assumptions in CM research have been largely stable, even though the IS landscape has progressed. Developments in technology and organizing have made CM field assumptions increasingly distant from and poorly fitting with IS reality. These developments limit the contributions and explanatory power of CM research that remains within the realm constructed by the old assumptions. Therefore, a change in CM research is timely and warranted. In response, we propose a novel framework for CM research and update its research agenda in our paper.

Our new Conceptual Modeling in the Digital World (CMDW) framework recognizes that CM occurs within and between physical and digital realities. It situates the *CM script* as the principal CM artifact at the intersection of physical and digital realities. It also recognizes the *CM agent* as any agent, human or digital, that produces or consumes CM scripts. Finally, the CMDW framework offers a new definition of the *CM context* by describing the CM context as the continually changing socio-material setting in which CM occurs.

3 Conclusions

The main value proposition of our CMDW framework is that it repositions CM scholarship in the digital world so it can continue its prominent role in IS scholarship and heighten its significance in an increasingly digital world.

A second core value proposition of our CMDW framework lies in highlighting uncharted prospects of CM research. These prospects can reignite interest in CM research. Our framework promotes a new, proactive agenda that encourages CM scholarship to speak to the broader issues and challenges organizations and society face in a digital world. At the same time, our framework stays connected to traditional CM scholarship, easing the transition between traditional CM research and the uncharted territories. In the paper we discuss three possible territories for CM research as examples for future CM research.

Following the research agenda suggested by the CMDW framework will reveal new limits of CM but it will also increase our confidence in where, how, and why CM is effective and useful. We may even discover that CM holds promises that have not yet been foreseen.

References

[Re21] Recker, J., Lukyanenko, R., Jabbari Sabegh, M. A., Samuel, B. M., Castellanos, A.: From Representation to Mediation: A New Agenda for Conceptual Modeling Research in a Digital World. *MIS Quarterly,* 45(1), 269-300, 2021.

A Consolidated Framework for Implementing Robotic Process Automation Projects (Extended Abstract)[1]

Lukas-Valentin Herm, Christian Janiesch, Alexander Helm, Florian Imgrund, Kevin Fuchs, Adrian Hofmann, and Axel Winkelmann[2]

1 Motivation

Competing in an international market, companies are forced to work faster and more efficiently to stay ahead of the competition. In doing so, they use Business Process Management (BPM) to monitor their business processes, discover shortcomings, and optimize their process execution. Thereby, BPM is also used to automate business processes. However, due to the heavyweight development and deployment of business process automation through traditional BPM software, robotic process automation (RPA) is attracting attention [Sy20].

RPA is a lightweight technology for automating business process tasks across existing off-the-shelf software without using backend application programing interfaces. Instead, RPA uses existing user interfaces, resulting in rapid development for simple, frequent and exception-less processes, leading to a quick return on investment [Sy20]. Despite the potential and high expectations for RPA, the technology has faced some challenges. Although RPA is considered an easy-to-implement technology, experts are still necessary to create reliable business value when RPA is first applied. As a result, it is estimated that between 30% and 50% of initial RPA implementations fail [Ra20].

Many authors such as Syed et al. [Sy20] call for future research to support the methodological support for technology adoption and implementation as well as social-technical implementation. With our research, we aimed at addressing this open issue. Based on a literature review of successful RPA implementation use cases and an expert interview study, we developed a holistic framework for RPA implementation projects.

[1] The original paper won the best paper award at the 18th International Conference on Business Process Management (BPM) 2020 [He20].

[2] All authors are from the Julius-Maximilians-Universität Würzburg, Faculty of Business Management and Economics, Sanderring 2, Würzburg, 97070, Germany, {forename.surname}@uni-wuerzburg.de

2 Framework

Based on the literature reviews, we found different stages in successful RPA project implementations. Through our expert interviews, we were able to validate these findings and add further stages. Subsequently, we merged the stages into a holistic framework, whereby the realization of these stages can be adjusted to the respective company.

Our framework is divided into three phases: initialization, implementation, and scaling. Further, we distinguish between phases performed once per project and stages that are performed continuously, such as different *RPA support processes*. We identified the *identification* of automation need, the *alignment* with business strategy, and *screening* of different (RPA) technologies within the **initialization** phase. Further, there are overlapping stages such as *process selection, RPA software selection*, and *evaluation of business case*, starting in the initialization phase and continuing into the **implementation** phase. In addition, the latter phase should include a *proof of concept* and an *RPA rollout*. Lastly, in the **scaling** phase, the *adaption and scaling* of the accumulated RPA knowledge should be made available for further projects.

3 Evaluation

While in theory various concepts have already been developed, in practice companies still face many issues when implement RPA projects successfully. Thus, our holistic framework can narrow this gap and provide methodological guidance. The interviewees confirm our findings and consider our research as an added value. Due to their diversity regarding their roles, industries, and BPM maturity, we do not estimate further biases within our framework. Likewise, we have already applied our framework at SYSHTEMIS AG, a German software development and IT consulting company and found no significant gaps or conflicts and further cases are in progress [He21].

References

[He20] Herm, L.-V. et al.: A Consolidated Framework for Implementing Robotic Process Automation Projects. In (Fahland, D.; Ghidini, C.; Becker, J.; Dumas, M. Eds.): Proceedings of the 18th International Conference on Business Process Management. Lecture Notes in Computer Science, Vol. 12168, Springer, Seville, pp. 471–488, 2020.

[He21] Herm, L.-V. et al.: Managing RPA Implementation Projects: A Framework Applied at SYSTHEMIS AG. In (Czarnecki, C.; Fettke, P. Eds.): Robotic Process Automation. Berlin, De Gruyter, 2021.

[Sy20] Syed, R. et al.: Robotic Process Automation: Contemporary themes and challenges. In: Computers in Industry, 115, 103162, 2020.

[Ra20] Ravn, C. et al.: Get Ready For Robots: Why Planning Makes The Difference Between Success And Disappointment. https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2016/11/ey-get-ready-for-robots.pdf, accessed: 20/05/2020.

End-User Development of Internet of Things Processes in Augmented Reality (Extended Abstract)*

Ronny Seiger[1]

Abstract: With the increasing number of IoT devices pervading our everyday surroundings there also emerges the desire of interlinking devices and creating processes among them for automation purposes. This "programming" of IoT environments currently requires a high level of technical expertise and abstraction capabilities, which hinders end-users from adapting these new technologies. We present *HoloFlows*–an approach for end-user oriented IoT process development in augmented reality–consisting of a simple modeling language and an intuitive user interface with natural interactions. HoloFlows showed an increased efficiency and user experience as well as steep learning curve for non-experts in a user study and maybe one way of programming the smart home and IoT of the future.

Keywords: Business Processes; Internet of Things; Augmented Reality; End-User Development

1 Introduction and Motivation

The Internet of Things (IoT) increasingly pervades all areas of every day lives. More and more software-controlled and interconnected commodity IoT devices make our homes, offices, factories and hospitals "smart". The programming and linkage of IoT devices in a process-oriented way shows high potential for raising the automation levels and thereby increasing comfort and efficiency in smart spaces [Ja20]. However, IoT devices are usually controlled and managed in isolation of each other, which limits the applicability of Business Process Management (BPM) technology. Moreover, end-users are often overwhelmed when controlling more than one IoT device, especially when it comes to creating automation routines among multiple devices [BV15]. We present the *HoloFlows* approach for enabling the end-user development of processes in IoT by means of augmented reality (AR). *The original article appeared in the Journal on Software and Systems Modeling (SoSyM) [Se21].

2 The HoloFlows Approach

The goal is to provide end-users with means for creating their own simple automation routines in the form of *IoT processes*. Our work consists of the following objectives: 1) create a simple domain-specific language (DSL); 2) visualize physical relations; 3) enable experimentation and feedback; and 4) provide an end-user friendly interface (UI).

DSL: The DSL for IoT processes considers *Sensors* having a *State* and *Actuators* offering *Commands* as first class citizens (cf. Fig. 1a). Different types of *Connections* among these are the basic building blocks for composing IoT processes. We distinguish between the connection types: a) Direct Sensor–Actuator for continuous state transfer; b) Conditional Sensor–Actuator; c) Actuator–Actuator; d) Sensor/Actuator–Gateway–Sensor/Actuator.

[1] University of St.Gallen, Institute of Computer Science, 9000 St.Gallen, Switzerland, ronny.seiger@unisg.ch

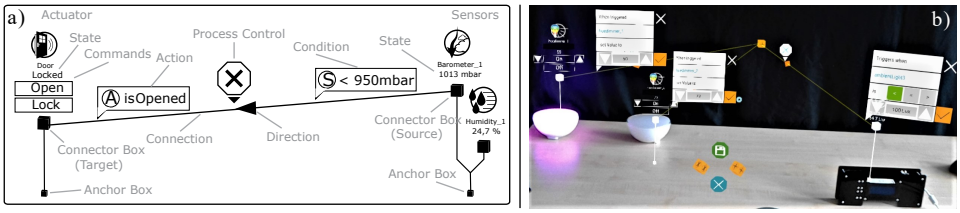


Fig. 1: Elements of the HoloFlows DSL and User Interface (a) with an Example Process in AR (b)

AR-UI: The designed UI projects holographic information above the individual physical IoT devices. It presents the devices' states and enables direct control of their functionality via "airtaps" on virtual buttons. In correspondence with the DSL the user can create the different types of connections by simply "drawing" a virtual wire between the IoT devices that should be part of the process. Depending on the type of connection the user then configures sensor-related conditions and selects the actions to be executed by the actuators (cf. Fig. 1). Once finished, the processes can be directly executed in the IoT environment.

3 Evaluation and Conclusion

We conducted a quantitative study to evaluate the HoloFlows approach. Users were asked to create typical IoT processes of a smart home using HoloFlows and two tools with "classical" desktop-based UIs–namely *Node-RED* for flow-based programming and *Camunda Modeler* for business process-oriented modeling. The comparison of modeling times and TLX scores attests HoloFlows with a high efficiency and relatively low task load, albeit a higher physical effort due to the AR interactions. Using the HoloFlows prototype results in a positive end-user experience achieved by natural interactions in and with the physical smart space.

With HoloFlows we investigated the application and combination of emerging technologies such as IoT and AR with concepts from BPM for process modeling and automation. AR hereby provides a promising new interaction paradigm to facilitate the control and development of IoT devices and processes for end-users as it allows to directly augment the physical world with digital holographic information and means for interaction.

References

[BV15] Barricelli, B. R.; Valtolina, S.: Designing for End-User Development in the Internet of Things. In (Díaz, P.; Pipek, V.; Ardito, C.; Jensen, C.; Aedo, I.; Boden, A., eds.): End-User Development. Springer, Cham, pp. 9–24, 2015.

[Ja20] Janiesch, C.; Koschmider, A.; Mecella, M.; Weber, B.; Burattin, A.; et al.: The Internet of Things Meets Business Process Management: A Manifesto. IEEE Systems, Man, and Cybernetics Magazine 6/4, pp. 34–44, 2020.

[Se21] Seiger, R.; Kühn, R.; Korzetz, M.; Aßmann, U.: HoloFlows: modelling of processes for the Internet of Things in mixed reality. Software and Systems Modeling/, pp. 1–25, 2021.

On the Declarative Paradigm in Hybrid Business Process Representations: A Conceptual Framework and a Systematic Literature Study (Extended Abstract)[1]

Amine Abbad Andaloussi[2], Andrea Burattin[2], Tijs Slaats[3], Ekkart Kindler[2], Barbara Weber[4]

1 Context and Motivation

Process artifacts provide a key instrument to support the enactment and management of business processes. Besides providing a blue-print for process enactment and facilitating several types of analysis and verification techniques, process artifacts can be used for requirement elicitation and process improvement. These activities, in turn, require models that are interpretable by machines but also comprehensible to humans.

In the literature, a wide array of *process artifacts* have been used to describe business processes (e.g. textual descriptions, formal process models, simulation tools). These artifacts are usually expressed using *imperative* or *declarative* languages [Fa09]. The choice of either class of languages depends on the process specifications. Imperative languages are typically used for representing rigid processes where the possible execution paths can be easily elicited and represented in the model. Conversely, declarative languages are commonly used to describe flexible processes as they allow to abstract the individual execution paths and rather emphasize the constraints guiding the overall process [Fa09]. Previous research has shown that an individual modeling language cannot deliver a concise and precise process representation. Indeed, in many processes, some aspects can better be described in an imperative way, while other aspects can better be captured in a declarative way. [RSS13]. Moreover, it has been suggested that the use of a unique artifact (e.g., formal process model) can challenge the comprehension of process stakeholders (e.g., domain experts) who are not familiar with that kind of artifact [An20b]. To address these gaps, authors in the literature have proposed several *"hybrid"* representations combining languages or artifacts. However, there is no conceptual framework allowing to identify, classify and discuss all these hybrid representations in a uniform way. This shortcoming is addressed in this work through a new *conceptual framework* and *systematic literature review* (SLR), setting the building blocks for defining and characterizing hybrid representations.

[1] The original article appeared in Information Systems (IS) Journal [An20a]. Work supported by the Innovation Fund Denmark project EcoKnow (7050-00034A).

[2] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

[3] Department of Computer Science, University of Copenhagen, Denmark

[4] Institute of Computer Science, University of St. Gallen, Switzerland

2 Conceptual Framework and Literature Review

The proposed conceptual framework provides a unified terminology allowing to distinguish *hybrid languages* and *hybrid artifacts*. The former combines existing imperative and declarative languages into a hybrid language used to express a single process artifact, while the latter combines a set of interrelated process artifacts each expressed using a different language [An20a]. Following the definition of hybrid languages and hybrid artifacts, the SLR is conducted to delve further into the characteristics of these two kinds of hybrid representations.

Thirty articles are identified following an extensive literature search. The findings show that existing hybrid representations have emerged within two research lines. In the first research line, hybrid languages are used to provide a concise modeling of the flexible and rigid aspects of business processes, while hybrid artifacts are proposed to facilitate the modeling, comprehension and maintenance of process models. With regards to the second research line, hybrid artifacts are suggested to support the integration of business rules (represented textually) into process models (represented graphically). Moreover, based on the findings, different types of hybrid languages and hybrid artifacts are identified and synthesized into a taxonomy. Therein, hybrid languages are divided into languages providing process models with a *hierarchical structure* (i.e., a different language is used for each sub-process of the model) or process models with a *mixed structure* (i.e., one language combining the syntax of existing imperative and declarative languages is used to represent the entire process). Hybrid artifacts, in turn, are organized into representations *combining a process model and textual annotations* and representations *combining a process model and guided simulations*

Overall, in this study, we have proposed a conceptual framework for defining, classifying and discussing hybrid process representations. This framework was developed based on an SLR and taxonomy meant to provide an overarching understanding of hybrid representations and help researchers to better identify and position their contributions. The outcome of this work will pave the path for new hybrid approaches and offer a sound basis to be systematically updated with up-coming studies.

Bibliography

[An20a] Andaloussi, Amine Abbad; Burattin, Andrea; Slaats, Tijs; Kindler, Ekkart; Weber, Barbara: On the declarative paradigm in hybrid business process representations: A conceptual framework and a systematic literature study. Information Systems, 91:101505, 2020.

[An20b] Andaloussi, Amine Abbad; Zerbato, Francesca; Burattin, Andrea; Slaats, Tijs; Hildebrandt, Thomas T; Weber, Barbara: Exploring how users engage with hybrid process artifacts based on declarative process models: a behavioral analysis based on eye-tracking and think-aloud. Software and Systems Modeling, pp. 1–28, 2020.

[Fa09] Fahland, Dirk; Lübke, Daniel; Mendling, Jan; Reijers, Hajo; Weber, Barbara; Weidlich, Matthias; Zugal, Stefan: Declarative versus Imperative Process Modeling Languages: The Issue of Understandability. In: Proceedings of EMMSAD. Springer, pp. 353–366, 2009.

[RSS13] Reijers, Hajo A; Slaats, Tijs; Stahl, Christian: Declarative modeling–An academic dream or the future for BPM? In: Business Process Management, pp. 307–322. Springer, 2013.

Ontology-Based Process Modelling - Will we live to see it? (Extended Abstract)

Carl Corea[1], Michael Fellmann[2], Patrick Delfmann[3]

Abstract: In theory, ontology-based process modelling (OBPM) bares great potential to extend business process management. Many works have studied OBPM and are clear on the potential amenities, such as eliminating ambiguities or enabling advanced reasoning over company processes. However, despite this approval in academia, a widespread industry adoption is still nowhere to be seen. This can be mainly attributed to the fact, that it still requires high amounts of manual labour to initially create ontologies and annotations to process models. As long as these problems are not addressed, implementing OBPM seems unfeasible in practice. In this work, we therefore identify requirements needed for a successful implementation of OBPM and assess the current state of research w.r.t. these requirements. Our results indicate that the research progress for means to enable OBPM are still alarmingly low and there needs to be urgent work on extending existing approaches.

Keywords: Ontology-Based Process Modelling; Ontologies; Automated Annotations

1 Extended Abstract

In the scope of Business Process Management, *process models* have evolved as central artifacts for the design, enactment and analysis of company processes [We07]. Many modelling languages, such as BPMN[4], are available and have received widespread adaptation in practice. While these standards offer support for the *representation* of company processes, the actual *content* of the model is still the responsibility of the modeller. That is, process models are designed by human modellers, often times also in a collaborative and incremental manner. In this setting, modelling errors can occur frequently [Ri17, Fe15]. For example, humans might accidentally model a non-compliant sequence of activities. Also, as the activities are captured with natural-language descriptions (labels), different views or understandings of modellers can lead to terminological issues in the resulting models, e.g. ambiguities of the prescribed activities or duplicate elements that clutter the model.

To conquer such modelling problems, there is a broad consensus in academia that business process models should be extended with an additional conceptual layer, namely *ontologies* [Fe15]. Ontologies are engineering artifacts that can be used to formally conceptualize a domain of interest [CD17]. As shown in Figure 1, creating *ontology-based process models*

[1] Universität Koblenz-Landau, Institut für Wirtschafts- und Verwaltungsinformatik, ccorea@uni-koblenz.de
[2] Universität Rostock, Lehrstuhl für Wirtschaftsinformatik, michael.fellmann@uni-rostock.de
[3] Universität Koblenz-Landau, Institut für Wirtschafts- und Verwaltungsinformatik, delfmann@uni-koblenz.de
[4] https://www.omg.org/bpmn/

by extending process models with ontologies would thus allow to define unambiguous semantics of process models and create a shared semantic understanding of business processes for humans and machines alike.

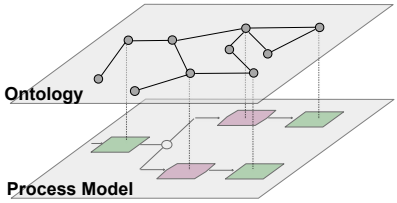


Fig. 1: Exemplary Ontology-Based Process Model.

Many works have discussed the potential benefits of OBPM, e.g. advanced compliance reasoning [Fe15, CD17]. Still, industry adoptions are sparse. Based on reports such as [Ri17], this can be mainly attributed to the problems in creating the ontology-based process models themselves. First, ontologies must initially be created. As this requires a high expertise in knowledge representation, this is currently still a difficult task for companies. Second, even given an ontology, the ontology has to be annotated to the process model. As finding the connections between ontology- and process model elements is a highly complex task, manual annotation can be seen as highly unfeasible in practice [Ri17]. Also, while there have been some approaches for (semi-) automated annotation, the lack of industry adoptions suggests that companies need more support in implementing OBPM.

In this report, we therefore investigate what methods and results are still missing to support companies in OBPM and leverage industry adoption. Here, our contributions are as follows: We identify requirements needed for successfully implementing and maintaining ontology-based process models for companies. Then, we identify the state-of-the-art on OBPM research based on a literature review and assess to which extent current results support the identified requirements. Last, based on the literature analysis, we identify current research gaps and propose a research agenda to guide future research and tool development.

References

[CD17] Corea, C.; Delfmann, P.: Detecting Compliance with Business Rules in Ontology-Based Process Modeling. In: Proceedings of the Int. Conference on Wirtschaftsinformatik, St. Gallen, Switzerland. pp. 226–240, 2017.

[Fe15] Fellmann, M.; Delfmann, P.; Koschmider, A.; Laue, R.; Leopold, H.; Schoknecht, A.: Semantic Technology in Business Process Modeling and Analysis. Part 1: Matching, Modeling Support, Correctness and Compliance. In: Emisa Forum 35. pp. 15–31, 2015.

[Ri17] Riehle, D.; Jannaber, D.; Delfmann, P.; Thomas, O.; Becker, J.: Automatically Annotating Business Process Models with Ontology Concepts at Design-Time. In: International Conference on Conceptual Modeling. Springer, pp. 177–186, 2017.

[We07] Weske, M.: Business Process Management Architectures. Springer, 2007.

Partial Order Resolution of Event Logs for Process Conformance Checking (Extended abstract)[1]

Han van der Aa[2] Henrik Leopold[3,4] Mathias Weidlich[5]

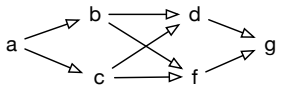
Context. Non-conformance in process execution poses a considerable threat to organizations. Conformance issues occur when the actual behavior of a process deviates from the desired one. Such differences can occur because information systems typically *support* process execution, but do not *enforce* a particular process. Rather, human interaction drives a process, giving people a certain flexibility in the execution of a particular case. To detect conformance issues, techniques for *conformance checking* have been introduced [Ca18]. They strive for the automatic detection of deviations between the observed and desired process behavior by comparing recorded event data to a normative process model.

Problem. A key assumption of most conformance checking techniques is that all events of a case have timestamps that allow to infer a *total* order. However, factors like synchronization issues, manual recording of events, and unreliable data sensing, mean that this assumption is often violated in practice. E.g., in healthcare processes, only the date of certain treatments may be known, but not their specific point in time [Ma13]. As a result of this, events are only partially ordered, meaning that a single trace could actually reflect various possible total orders, i.e., different possible resolutions, as depicted in Figure 1. Since different resolutions can have different conformance results, the conformance of such traces becomes uncertain.

Contribution. Our work aims to still provide conformance results in the presence of such order uncertainty. Its idea is to use information from the entire event log to estimate the probability of each possible resolution of an uncertain trace. For this estimation, we argue that the probability of a specific resolution may be based on order information derived from similar traces contained in the event log. Intuitively, if a possible resolution denotes an execution order that is frequently observed for other traces, this resolution is expected to be more likely than one that denotes a less common sequence. By incorporating these probabilities in our approach, conformance checking is grounded in a stochastic model.

Behavioral models. Estimating resolution probabilities requires a careful selection of the abstraction level based on which traces are compared. This is because, depending on the complexity of a process and the degree of order uncertainty, it may be impossible to

[1] The original article has been published in Decision Support Systems [vdAa20b]
[2] University of Mannheim, Germany. han@informatik.uni-mannheim.de
[3] Kühne Logistics University, Hamburg, Germany. henrik.leopold@the-klu.org
[4] Hasso Plattner Institute, University of Potsdam, Germany
[5] Humboldt-Universität zu Berlin, Germany. matthias.weidlich@hu-berlin.de



$\pi_1 = \langle a, b, c, d, f, g \rangle$
$\pi_2 = \langle a, c, b, d, f, g \rangle$
$\pi_3 = \langle a, b, c, f, d, g \rangle$
$\pi_4 = \langle a, c, b, f, d, g \rangle$

Fig. 1: Partial order and trace resolutions for an uncertain trace $\langle \{a\}, \{b, c\}, \{d, f\}, \{g\} \rangle$.

observe the exact same sequence of activity executions in another trace, unaffected by order uncertainty. We cope with this issue by defining behavioral models that realize different levels of abstraction for the comparison of traces. Particularly, we propose models that determine the likelihood of a resolution based on: (1) fully identical traces, (2) identical sub-traces of length N, and (3) weak-order relations. These models differ in the notion of behavioral regularity that is used for the partial order resolution and, importantly, in their ability to derive information from traces that are also only partially ordered.

Evaluation results. Extensive evaluation experiments, using both real-world and synthetic data, reveal that a model focusing on sub-traces of length 2, i.e., 2-grams, yields the most accurate conformance checking results. When compared to existing baselines, we observe that our approach is more accurate, reducing the average error by 59%. Furthermore, we demonstrate that our approximation method, introduced to cope with the high computational complexity of conformance checking under uncertainty, leads to considerable gains in terms of efficiency, while having a negligible impact on the result accuracy.

Outlook. The discussed work is part of our ongoing research stream of conformance checking in the presence of uncertainty, in which we target three major sources of uncertainty: uncertainty (1) in the process specification [vdAa18], (2) in the mapping from specification to event data [vdAa20], and (3) now also uncertainty in the data itself. For all these directions, we see open research questions related to the perception and quantification of probabilistic results in conformance checking. For the specific direction presented here, we furthermore intend to explore additional behavioral models, involving data beyond the control flow, as well as techniques that use neural networks to learn how order uncertainty shall be resolved.

Bibliography

[Ca18] Carmona, Josep; van Dongen, Boudewijn F.; Solti, Andreas; Weidlich, Matthias: Conformance Checking - Relating Processes and Models. Springer, 2018.

[Ma13] Mans, Ronny S; Van der Aalst, Wil M P; Vanwersch, Rob JB; Moleman, Arnold J: Process mining in healthcare: Data challenges when answering frequently posed questions. In: ProHealth, pp. 140–153. Springer, 2013.

[vdAa18] Van der Aa, Han; Leopold, Henrik; Reijers, Hajo A: Checking process compliance against natural language specifications using behavioral spaces. Inf. Syst., 78:83–95, 2018.

[vdAa20] Van der Aa, Han; Leopold, Henrik; Reijers, Hajo A.: Efficient Process Conformance Checking on the Basis of Uncertain Event-to-Activity Mappings. TKDE, 32(5):927–940, 2020.

[vdAa20b] Van der Aa, Han; Leopold, Henrik; Weidlich, Matthias: Partial order resolution of event logs for process conformance checking. Decision Support Systems, 136:113347, 2020.

Revenue Management Systems as Symbiotic Analytics Systems: Insights from a Field Study (Extended Abstract)

Catherine Cleophas[1], Claudia Schütze[2], Monideepa Tarafdar[3]

The role of algorithms and artificial intelligence for business operation is growing [CWW18]. This development increases the importance of examining and understanding the interplay of human analysts and automated systems in the context of Information Systems (IS). To that end, this paper analyses a field study featuring an airline updating its revenue management (RM) systems and processes.

The role of analysts causes some authors to define RM systems as decision support systems (DSS). However, DSS aim to support and improve decision-making on a managerial level and are primarily used for unstructured and underspecified problems, often of a strategic nature [AP05]. Instead, the vast majority of RM decisions are operational, routine, repetitive, and highly frequent, so no fully manual implementation is conceivable. For example, as implemented at the airline featured in the field study, RM controls ticket offers for more than 6,000 flight departures per day. Analysts monitor bookings, adjust the forecast, set the parameters for optimising offers and overbookings, and override inventory controls. Given this balance of analyst input and algorithmic computation, we term RM an example of symbiotic analytics systems (SAS). The effects of integrating analytics algorithms into the firm's processes and the role of analysts who adjust and overwrite parameters and results require further examination.

Thus motivated, we observe a natural experiment in practical RM in a field study: An airline updates its RM methodology by updating the automated systems and, in consequence, updating the organisation of the analysts working to complement the system. Our analysis follows the multilevel approach of Burton-Jones; Gallivan [BG07]. It considers the job definition on the individual level and system usage and necessary communication effort on the group level. The resulting changes in the analyst organisation mirror two different perspectives on the RM process:

In case 1 (compare Figure 1), the firm implements a functional view of the RM landscape. Analyst jobs are organised sequentially, with a high degree of system-oriented specialisation.

[1] Kiel University, Group Service Analytics, Kiel, Germany cleophas@bwl.uni-kiel.de
[2] RWTH Aachen University, Group Advanced Business Analytics, Aachen, Germany claudia.schuetze@rwth-aachen.de
[3] Lancaster University Management School, Lancaster, United Kingdom m.tarafdar@lancaster.ac.uk

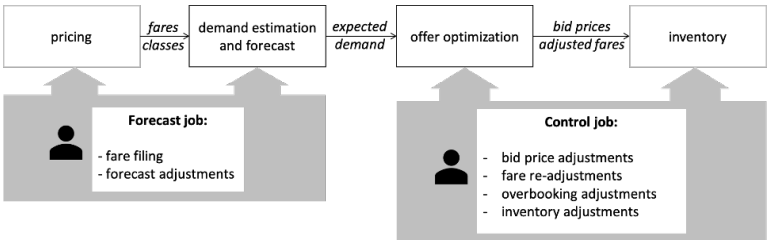


Abb. 1: Case 1

The focus of jobs is on specific RM functions (pricing and forecasting vs optimisation and inventory) and not on the entire process. In case 2, the firm implements the process view: analyst jobs follow a parallel structure along the entire process. All analysts access both predictive and prescriptive analytics systems, with parallelised influences on customer choice and itinerary-levels and on demand and leg-levels.

We find that RM requires configural, heterogeneous system usage. In both cases analyzed, analysts can intervene at various levels: A nudge in the forecast, overwriting optimization parameters for the leg's or itinerary level and finetune overbooking limits at the very end of the process. If in this, RM systems represent further SAS, design choices have to carefully evaluate implications of offering different levels and types of adjustments. We consider further research on other SAS application fields, such as workforce scheduling and inventory management, as necessary to verify this idea.

The cases differ in the implementation of the system landscape and the organizational structure. In the functional view, system units and jobs are specialized in one or a few functions risking that the focus does not lay on the overall process. The process view implements two sub-processes that can work independently. System units and jobs cover a more significant part of the process, emphasising communication. A similarity is as notable as the differences: The firm always considers analysts an inherent part of the system, tasking them with monitoring and giving them the authority to adjust outcomes every step of the process. While we concentrate on a single firm, the relevant software packages' design indicates that this is the case throughout the industry. Thus, the RM system stayed symbiotic, even when updating. Whether this applies to most SAS or is due to specific characteristics of the investigated firm or the RM domain has to be considered by future research.

The perspective on the RM process determines the organizational as well as the system structure. Only when both structures work hand in hand, the collective use of a complex system like the one in RM can contribute to a company's success. If this expectation can be generalized, there arise several caveats for designing SAS. As pointed out in this paper, different job divisions and divisions of responsibility cause different potential fault lines and conflicts in analysts' work. Communication and configural use turned out to be a crucial element of the "human sideöf systems – therefore, supporting these aspects of analyst tasks should be an intrinsic part of systems design.

Literatur

[AP05] Arnott, D.; Pervan, G.: A critical analysis of decision support systems research. Journal of information technology 20/2, S. 67–87, 2005.

[BG07] Burton-Jones, A.; Gallivan, M. J.: Toward a deeper understanding of system usage in organizations: a multilevel perspective. MIS quarterly/, S. 657–679, 2007.

[CWW18] Choi, T.-M.; Wallace, S. W.; Wang, Y.: Big data analytics in operations management. Production and Operations Management 27/10, S. 1868–1883, 2018.

Showing the Advantages of Pull over Push Production with the Aid of Petri Nets (Extended Abstract)

Making Use of the Process-Simulation.Center (P-S.C)

Carlo Simon[1], Stefan Haag[1], Lara Zakfeld[1]

Abstract: The research talk is intended to be about two different modeling and simulation methods using Petri nets to compare push and pull scenarios in production and logistics. The first method called Clock Pulse Simulation facilitates users to observe the behavior of the simulated system over time while the second called Event Triggered Simulation condenses the simulation to the moments state changes occur. Though the first method is more descriptive the second has the advantage of a possibly extremely reduced simulation time.

Although the key simulation results are the same for both methods, their numeric representation varies broadly what made it necessary to think about different visualization techniques such as line diagrams or cord diagrams. The authors observed that the visualization of simulation results, especially in the case of Petri nets, is not investigated sufficiently yet and see a need for further research about this topic. How can a high-level Petri net reachability set be transformed into something that can be interpreted by domain experts?

The approach has been applied to processes taught in a teaching laboratory for logistics. Although this setting is relatively small, everything that occurs in more complex real-world scenarios can be observed within the laboratory as well. The simulation extends the experience the students gain while working on the processes since it overcomes time and volume restrictions that exit in reality. With the aid of the simulation, the students can observe their doing from another perspective and without the necessity to concentrate on the operative steps in the process what boosts the learning effect about the savings of up to 90% internal stock costs when changing from push to pull.

For model development and simulation, a novel web-based high-level Petri net environment called Process-Simulation.Center (P-S.C) was used. Beside the Petri nets, this tool also allows to picture large process structures with the aid of process maps and organizational structures with the aid of an organigramm and a swimlane layout of the nets. Due to a mandator concept, the students can be divided into learning groups. Beyond the described scenario of push and pull processes, meanwhile the tool is also used in various other learning scenarios such as teaching Petri net and Business Process Management in general and is also used by students who conduct their company internships.

This proposal submission bases on published papers from international conferences as explained in the following:

[1] HS Worms, FB Informatik, Erenburgerstr. 19, 67549 Worms, Deutschland, {simon,haag,zakfeld}@hs-worms.de

doi:lnidoi

- The research concerning the two modeling and simulation methods for push and pull processes has been published at the *The Twelfth International Conference on Advances in System Simulation (SIMUL 2020)* and the special track *SIMMaApp* of this conference.[SHZ20, Ha20] The paper on the Event Trigger Simulation has been awarded as one of two best conference papers.
- The functional variety of the P-S.C could be demonstrated at the *34th International European Conference on Modelling and Simulation (ECMS 2020)*.[SH20c]
- Further applications of the P-S.C for teaching have been demonstrated at the *Workshop zur Modellierung in der Hochschullehre (MoHoL 2020)* and have been published additionally in the EMISA Forum.[SH20b, SH20a]

The authors hope to encourage a talk on challenging modeling and simulation approaches and how to teach these approaches to students.

Keywords: Model engineering, Petrinets, Push vs. Pull, Simulation models for Learning and Teaching

1 Acknowledgement

This research was supported by "ProFIL -Programm zur Förderung des Forschungspersonals, Infrastruktur und forschendem Lernen, Hochschule Worms".

Literaturverzeichnis

[Ha20] Haag, S.; Zakfeld, L.; Simon, C.; Reuter, C.: Event Triggered Simulation of Push and Pull Processes. In (Parra, L., Hrsg.): SIMUL 2020 : The Twelfth International Conference on Advances in System Simulation. Porto (Portugal), S. 68–73, 2020.

[SH20a] Simon, C.; Haag, S.: A Case-Study to Teach Process-Aware Information Systems. In: EMISA Forum : Proceedings of the SIG Enterprise Modelling and Information Systems Architectures of the German Informatics Society. Jgg. 40, S. 9–10, 2020.

[SH20b] Simon, C.; Haag, S.: Digitale Zwillinge modellieren und verstehen - Eine Fallstudie zum problembasierten und forschenden Lernen. In (Rehse, J.-R.; Striewe, M.; Ullrich, M., Hrsg.): Workshop zur Modellierung in der Hochschullehre (MoHoL 2020). Wien (Austria), 2020.

[SH20c] Simon, C.; Haag, S.: Simulatable Reference Models To Transform Enterprises For The Digital Age – A Case Study. In (Steglich, M.; Müller, C.; Neumann, G.; Walther, M., Hrsg.): ECMS 2020 : 34th International ECMS Conference on Modelling and Simulation. S. 294 – 300, 2020.

[SHZ20] Simon, C.; Haag, S.; Zackfeld, L.: Clock Pulse Modeling and Simulation of Push and Pull Processes in Logistics. In (Herrmann, F., Hrsg.): SIMMaApp : Special Track at SIMUL 2020 : The Twelfth International Conference on Advances in System Simulation. Porto (Portugal), S. 31–36, 2020.

Toward a Taxonomy of Modeling Difficulties: A Multimodal Study on Individual Modeling Processes (Extended Abstract)[1]

Kristina Rosenthal[2], Stefan Strecker[2]

Abstract: Combining complementary modes of observation of modelers' modeling processes, we study modeling difficulties encountered by modelers while performing a data modeling task. Using the notion of cognitive breakdowns, we identify and confirm five types of modeling difficulties relating to different aspects of data modeling by analyzing audiovisual protocols of the modelers' modeling processes, recordings of modelers' interactions with the employed modeling software tool and survey data of modelers about their own perceptions of modeling difficulties they encountered. The identified types of modeling difficulties motivate a taxonomic theory of modeling difficulties intended to inform design science research on modeling assistance for modelers at different stages of their learning and mastering of conceptual modeling.

Keywords: Conceptual data modeling; Modeling difficulty; Cognitive breakdown

1 Introduction

Conceptual modeling as an activity involves an intricate array of complex cognitive processes and performed actions including goal setting, abstracting, conceptualizing, associating, contextualizing, interpreting and sense-making, judging and evaluating, anticipating and envisioning, visualizing, and, in group settings, communicating, discussing and agreeing. Prior research suggests that data modelers at all stages of mastering conceptual data modeling face specific difficulties but surprisingly little is known about the reasoning of modelers with respect to the modeling difficulties they encounter, which difficulties they encounter, and how to mitigate or even overcome these difficulties by tailored modeling assistance. The multimodal study presented in [RS19] integrates complementary modes of observation of modeling processes to identify modeling difficulties eight beginning modelers face while performing a data modeling task using a modeling software tool. We use the concept of cognitive breakdowns [NS72] to identify modeling difficulties in verbal (think aloud) protocols and complement difficulty identification by visually inspecting recordings of modeler-tool interactions as well as video recordings of individuals' modeling processes. We then complement difficulty identification by surveying these individuals about performing the modeling and about the difficulties they perceived.

[1] The work summarized in this extended abstract is published as [RS19].

[2] University of Hagen, Enterprise Modelling Research Group, Universitätsstr. 41, 58084 Hagen, Germany {kristina.rosenthal,stefan.strecker}@fernuni-hagen.de

2 Findings and Discussion

Our analysis leads us to identify five types of modeling difficulties the beginning modelers face while performing the data modeling task, relating to different aspects of constructing conceptual data models, i.e., entity types, relationship types, attributes, and cardinalities. Our findings suggest that the majority of difficulties encountered by the participants relates to modeling relationship types with modelers especially facing difficulties with regard to developing sensible identifiers for relationship types.

The findings presented in [RS19] serve as a starting point for developing a more elaborate taxonomy of modeling difficulties over the course of multiple studies, in the sense of a classification or taxonomic theory (following, e.g., [Gr06]). The taxonomy, in turn, is intended to serve as theoretical foundation for design science research on developing tailored tool assistance for modelers at different stages of their learning and mastering of conceptual modeling. A number of further studies is needed to deepen our understanding of distinct modeling difficulties, the modelers' corresponding reasoning, presumed causes for encountered difficulties and remedies for mitigating their occurrences. In a recent comparison with findings we obtained from studying eight modeling processes of experienced modelers [RSP20], we find considerable overlaps in modeling difficulties experienced and non-experienced data modelers encounter, i.e., a majority of difficulties relating to modeling relationship types, especially to developing identifiers for relationship types, besides clear differences in difficulties related to determining cardinalities and fostering model integrity.

References

[Gr06] Gregor, S.: The Nature of Theory in Information Systems. Management Information Systems Quarterly 30/3, pp. 611–642, 2006.

[NS72] Newell, A.; Simon, H. A.: Human problem solving. Prentice-Hall, Englewood Cliffs, NJ, 1972.

[RS19] Rosenthal, K.; Strecker, S.: Toward a Taxonomy of Modeling Difficulties: A Multi-Modal Study on Individual Modeling Processes. In (Krcmar, H.; Fedorowicz, J.; Boh, W. F.; Leimeister, J. M.; Wattal, S., eds.): Proceedings of the 40th International Conference on Information Systems, ICIS 2019, Munich, Germany, December 15-18, 2019. Association for Information Systems, 2019.

[RSP20] Rosenthal, K.; Strecker, S.; Pastor, O.: Modeling Difficulties in Data Modeling - Similarities and Differences Between Experienced and Non-experienced Modelers. In (Dobbie, G.; Frank, U.; Kappel, G.; Liddle, S. W.; Mayr, H. C., eds.): Proceedings of the 39th International Conference on Conceptual Modeling, ER 2020, Vienna, Austria, November 3-6, 2020. Vol. 12400. Lecture Notes in Computer Science, Springer, pp. 501–511, 2020.

BERMUDA: Towards Maintainable Traceability of Events for Trustworthy Analysis of Non-process-aware Information Systems

Vlad P. Cosma[1,2][0000−0001−8022−6402], Thomas T. Hildebrandt[1][0000−0002−7435−5563], and Tijs Slaats[1][0000−0001−6244−6970]

[1] Copenhagen University, Copenhagen 2200, Denmark
{vco,hilde,slaats}@di.ku.dk
[2] KMD ApS, Ballerup 2750, Denmark
vco@kmd.dk

Abstract. We propose a method supported by a prototype tool for tracing the relation between domain events and data recorded in non-process-aware information systems via the system and user interfaces. The method contributes to the involvement of domain experts in the tasks of understanding the available data, the domain and the trustworthy and responsible extraction of events from information systems carried out in the first phase of process mining projects.

Keywords: Domain Event · Database · Interface · Information System

1 Motivation

BERMUDA (Business Event Relation Map via User-interface to Data for Analysis) is a method to document and maintain the link between domain events and the data in the information system via the user interface. It facilitates cooperative work spanning the expertise domains of the domain experts using the information system, engineers and data scientists, while still maintaining the necessary barriers in place to protect values, rights and intellectual property. Unlike event abstraction [4] and related work on automatic or semi-automatic event extraction [1,3] we approach the problem in a human-centric manner and make explicit the association between high-level and low-level events, we also do not assume the existence of an event log prior to applying BERMUDA. A similar procedure in [2] focuses on process instance and event level decisions taken when building the event log.

2 Approach

Our method relies on so-called BERMUDA triples **(e,i,d)** recording the relation between respectively a domain event **e**, an interface element **i** of the system in which the domain event is registered in the information system and the location

of the resulting data element **i** in the information system. There are three roles involved in the recording of BERMUDA triples: Data scientist (or analyst), domain expert and system engineer. As guidance towards applying our method we recommend following these steps:

1. **Domain to user interface**. For each domain event **e**, the domain experts record an association **(e,i)** between the domain event **e** and an (user or system) interface element **i**.
2. **User interface to data**. Through code inspection or simulation, system engineers develop the correct database query **d** to extract the data recording the event **e** created via the interface element **i** resulting in a triple **(e,i,d)**.
3. **Triples to event log**. The data scientist merges and refines the database queries and creates the initial version of the event log. The event log entries are enriched with extra attributes that hold a reference to the domain event, the user interface/description and the data source from where the entry originated.

3 Evaluation

In order to qualitatively evaluate the usefulness of the method we conducted two semi-structured interviews, one with a case worker acting as a domain expert and another with a data scientist. From our own experience and the interviews we observed that the method encourages a more structured approach towards event log construction. It also promotes collaboration in order to more efficiently record knowledge gathered from domain experts and system engineers. Both interview participants indicated that the method and tool could be valuable both in training new data scientists and new case workers.

References

1. Baier, T., Rogge-Solti, A., Weske, M., Mendling, J.: Matching of events and activities - an approach based on constraint satisfaction. In: Frank, U., Loucopoulos, P., Pastor, Ó., Petrounias, I. (eds.) The Practice of Enterprise Modeling. pp. 58–72. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)
2. Jans, M., Soffer, P.: From relational database to event log: Decisions with quality impact. In: Business Process Management Workshops. Lecture Notes in Business Information Processing, vol. 308, pp. 588–599. Springer (2017)
3. de Murillas, E.G.L., Reijers, H.A., van der Aalst, W.M.P.: Case notion discovery and recommendation: automated event log building on databases. Knowl. Inf. Syst. **62**(7), 2539–2575 (2020)
4. van Zelst, S., Mannhardt, F., de Leoni, M.: Event abstraction in process mining: literature review and taxonomy, vol. 6, pp. 719–736. Springer New York, New York, NY (2021). https://doi.org/10.1007/s41066-020-00226-2

A Generalizable Approach for Determining The Sensitivity of A Trace within An Event Log

Ryan Hildebrant, Zhenyu Zhang, and Shangping Ren
Department of Computer Science, San Diego State University
{rhildebrant, zzhang, sren}@sdsu.edu

Abstract. In this Research-In-Progress work, we present a potentially generalizable approach that can determine the sensitivity of an attribute or a set of attributes associated with an event in a given event log. This approach is based on the concept of an equivalence class that a given event or trace may form and associate its sensitivity with the size of its equivalence class. For a given event, different equivalent classes can be formed based on different attributes, the proposed approach provides researchers with a more granular tool to apply group based privacy to event logs.

Keywords: K-anonymity Privacy · Process Mining

1 Introduction

Privacy in process mining has been cited as a critical component for process mining applications [1]. Various researchers have presented work that focuses on group-based anonymization techniques and recognize the importance of attributes linked to individual events [2] [3]. However, all of their work focuses on either events or traces and define unique notations to support their approaches, respectively. Because of this, it is difficult to determine the appropriated approach for a given event log. It also makes it more challenging to integrate multiple approaches that apply privacy at different levels. In contrast, we present a generalizable approach that is based on the equivalence classes that a given event may form and associate sensitivities with the size of the equivalent classes.

2 Determining Equivalent Classes and Sensitivity

We use an example to explain for a given event log, how different equivalent classes may be formed and how the event sensitives are related to the cardinalities of the equivalent classes. Assume we have the following event log, where each event entry is represented by `<eventId, traceId, actName, [attr1, ···, attrn]>`. We use * to denote the wildcard which can be of any value.

eventID	traceID	actName	Attributes
e1	1	print	[John, 2:00pm, scripps, *, *]
e2	1	deposit	[Brad, 2:01pm, scripps, $100, Teller]
e3	1	print	[Brad, 2:04pm, scripps, *, *]
e4	2	withdraw	[Alice, 8:00am, SDSU, $200, Teller]
e5	3	withdraw	[Alice, 8:00am, SDSU, $200, Teller]
e6	4	print	[Bob, 2:00pm, scripps, *, *]
e7	4	deposit	[Alice, 2:01pm, scripps, $100, ATM]
e8	4	print	[John, 2:04pm, scripps, *, *]
e9	5	print	[John, 2:00pm, scripps, *, *]
e10	5	deposit	[John, 2:01pm, scripps, $100, ATM]
e11	5	print	[John, 2:04pm, scripps, *, *]

For the given event log, we can form different equivalence classes based on event attribute relations and traces. For instance, consider the event $< print >$. We can form an equivalent class that has the same attribute value of `John` for the `customer` attribute and `scripps` for the `location`. In this case, the equivalent class contains four elements, i.e., $e1, e8, e9$, and $e11$. We can also form an equivalent class that only considers the `location` attribute `scripps`, which has six elements, i.e., $e1, e3, e6, e8, e9$, and $e11$. Therefore, the sensitivity of `John` in terms of his activity location of "scripts" is 4/6. Additionally, we can compare the other attribute values with respect to the event performed by `Bob` and `Brad`, respectively, which is $\frac{1}{6}$. By comparing all of these values, we can conclude that `Bob` and `Brad` are more sensitive than `John` in terms of their activity locations.

This notion of equivalent class and sensitivity can also be applied to traces. Consider traces σ_1, σ_4, and σ_5. If we define our equivalent class on traces that have the same attributes values as `timestamp`, `location`, and `amount`, the three traces then belong to the same equivalent class. In fact, for the given event log, the cardinality of the equivalent class that has the same values of `timestamp`, the `location` and the `amount` is three. We can form an even tighter relation on a subset of these traces, i.e., requiring *timestamp, location, amount, provider* values to be the same, which result in an equivalent class with only two elements, i.e., $\{\sigma_4, \sigma_5\}$. Hence, the sensitivity of `provider` information in a trace with respect to *timestamp, location, amount* is $\frac{2}{3}$. On the other hand, we can change the `provider` value to $[Teller]$ and get a more sensitive result, $\frac{1}{3}$. This result is so sensitive that we can single out a trace within this log. Therefore, an attacker who knows these specific attribute values combinations of `timestamp`, `location`, `amount`, and `provider` could directly identify a set of individuals associated to σ_1.

3 Conclusion

We have briefly shown how we can use equivalent classes to define groups and provided a uniform representation that supports techniques to improve group-based privacy. Our future work is to formalize the notation and apply it to existing work, such as [2] and [3], and compare the results in terms of sensitivity (which quantitatively measures privacy). In addition, we will study how to modify raw event logs to maximize the sensitivity of all information in the log and at the same time minimize the impact on the structural change of event model resulted from the modified log.

References

1. Wil Van Der Aalst. Data science in action. In *Process mining*, pages 3–23. Springer, 2016.
2. Stephan A Fahrenkrog-Petersen, Han van der Aa, and Matthias Weidlich. Pretsa: event log sanitization for privacy-aware process discovery. In *2019 International Conference on Process Mining (ICPM)*, pages 1–8. IEEE, 2019.
3. Majid Rafiei, Miriam Wagner, and Wil MP van der Aalst. Tlkc-privacy model for process mining. In *International Conference on Research Challenges in Information Science*, pages 398–416. Springer, 2020.

Utility-aware Event Log Anonymization for Privacy-Preserving Process Mining*

Gamal Elkoumy[1], Alisa Pankova[2], and Marlon Dumas[1]

[1] University of Tartu, Tartu, Estonia
{gamal.elkoumy,marlon.dumas}@ut.ee
[2] Cybernetica, Tartu, Estonia
alisa.pankova@cyber.ee

Abstract. Process mining techniques enable organizations to analyze business process execution captured in event logs in order to identify opportunities for improving their operational performance. Oftentimes, event logs may contain private information, especially those involving customer-facing processes. In such cases, organizations need to deploy Privacy-Enhancing Technologies (PETs) to hide such private information. They find themselves in a situation where they either need to accept utility loss over achieving lower risk or accepting higher risk over achieving lower utility of the analysis. The ongoing research is investigating a method to estimate the minimum possible re-identification risk that can be achieved to anonymize an event log while keeping the utility loss below a given threshold. Among many available PETs, differential privacy stands out for its ability to prevent predicate singling out attacks and its composable privacy guarantees. Differential privacy injects noise based on a parameter called epsilon. The ongoing research is investigating a method to determine the epsilon value to be used when disclosing logs in terms of utility loss.

1 Introduction

Process Mining is a family of techniques to analyze event logs generated by enterprise information systems to help organizations identify opportunities to enhance their business processes' efficiency, compliance, and quality [2]. The primary input of a process mining technique is an event log, consisting of a collection of event records. Each record contains a reference to a process instance (case identifier), a reference to an activity label, and at least one timestamp. Besides, each event may contain other attributes, such as the resource (e.g., worker) who performed the activity.

Event logs may contain sensitive information about individuals. For example, an adversary may use a prefix/suffix of process execution, combined with timestamps, to perform a re-identification attack [4] and single out individuals. Data

* Work funded by European Research Council (PIX project) and by EU H2020-SU-ICT-03-2018 Project No.830929 CyberSec4Europe (http://cybersec4europe.eu).

regulations, such as GDPR[3], require that organizations put in place mechanisms to protect information about individuals when processing a dataset.

Privacy-Enhancing Technologies (PETs), such as k-anonymity and differential-privacy, protect datasets disclosure, including event logs [4]. Among existing PETs, differential privacy stands out for its ability to prevent predicate singling out attacks and its composable privacy guarantees [3]. Differential-privacy works by injecting noise to the event logs, based on a parameter called ϵ [5]. Small ϵ values result in injecting more noise to the event log in order to provide lower possible re-identification risk [5]. That noise injection leads to a loss of accuracy (a.k.a. utility loss). With such a cost of PETs, organizations find themselves between a hammer and an anvil. They either buy lower re-identification risk due to disclosing their event logs for a high utility loss, or accept higher re-identification risk as a price for a low utility loss.

The ongoing research is investigating a method to determine the minimum possible ϵ value that can be used to anonymize an event log while keeping the utility loss below a given threshold. Specifically, the method proposed in the ongoing research is addressing the problem:

Given an event log L, and given a maximum utility loss λ, generate an anonymized event log L' such that the disclosure risk is the minimum, and the difference between L and L' does not exceed λ.

We tackle the above problem by defining two utility loss metrics: the trace variant frequency error, which is the mean absolute percentage error between the frequencies of the traces variants in the input log L and the anonymized log L$'$, and the temporal error, which is the mean absolute percentage error between activities relative times in the input log L and the anonymized log L$'$.

This research adopts an attack model that mitigates singling out an individual based on any prefix or suffix of their trace. To achieve that, we group the traces that have common prefixes and suffixes in the log and analyze each group's frequency and time differences. Given that analysis, we quantify the minimum ϵ that keeps the utility loss below λ. One representation of event log that can provide such grouping is the Deterministic Acyclic Finite State Automata [1], wherein every prefix or suffix shared by multiple traces is represented once.

References

1. Daciuk, J., Mihov, S., Watson, B.W., Watson, R.E.: Incremental construction of minimal acyclic finite-state automata. Computational linguistics **26**(1), 3–16 (2000)
2. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A., et al.: Fundamentals of business process management, vol. 1. Springer (2013)
3. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. **9**(3-4), 211–407 (2014)
4. Elkoumy, G., Fahrenkrog-Petersen, S.A., Sani, M.F., Koschmider, A., Mannhardt, F., von Voigt, S.N., Rafiei, M., von Waldthausen, L.: Privacy and confidentiality in process mining - threats and research challenges. CoRR **abs/2106.00388** (2021)
5. Mannhardt, F., Koschmider, A., Baracaldo, N., Weidlich, M., Michael, J.: Privacy-preserving process mining. BISE **61**, 595–614 (2019)

[3] http://data.europa.eu/eli/reg/2016/679/oj

Conceptualizing a Log Generator for Privacy-aware Event Logs

Kay Kaczmarek[1], Agnes Koschmider[2]

1 Introduction

Privacy-preserving process mining will gain importance in the future due to the mandatory implementation of protection guidelines and laws applicable when handling (personal) data. Mannhardt et al. [Ma19] pointed to two general privacy challenges that have to be addressed within process mining: technological privacy challenges (related to the design of privacy-by-design or privacy-by-default approaches) and organizational privacy challenges. Even if technical solutions exist to bridge these challenges, still the reluctance remains to share data and to trust those solutions. Process mining techniques uncover operational processes of companies and pose the risk to re-identify private data whose access has to be fully protected. Several privacy-preserving process mining techniques have been already suggested. For a comprehensive overview we refer to Elkoumy et al. [El21].

The intention of this paper is to suggest a log generator for privacy-aware event logs. The benefit of the log generator is to build synthetic event logs based on an original event log and to apply process mining on the synthetic log aiming to obtain the same results like with the original event log. Additionally, it should be allowed to sample the synthetic event log and to quantify the privacy risk for different event log sizes. From a technical perspective, we aim to use a training data set to learn constructing accurate and possibly generalized data sets. The trained model should allow to generate new data without retaining training data. Then the generated data is converted into an event log format. Statistical methods and conformance checking are used to evaluate the quality of the event log. Finally, the risk of re-identification is quantified using the approach of von Voigt et al. [Vo20].

2 Approach

Fig. 1 shows the process of synthetic, privacy-aware event log generation. First, the data must be prepared for the machine learning technique or for use as training data respectively. So far we have been using Recurrent Neural Networks with LSTM and Time-series Generative

[1] Kiel University, Kiel, Germany stu96465@mail.uni-kiel.de
[2] Kiel University, Group Prcoess Analytics, Kiel, Germany ak@informatik.uni-kiel.de

Adversarial Network. The synthetically generated data set can be transformed directly into an event log using PM4PY. The evaluation of the quality of the synthetically generated data is compared with the original data using several methods. We tried T-Distributed Stochastic Neighbor Embedding, Principal component analysis (PCA) and conformance checking. The re-identification risk is assessed in interaction with the quality evaluation.

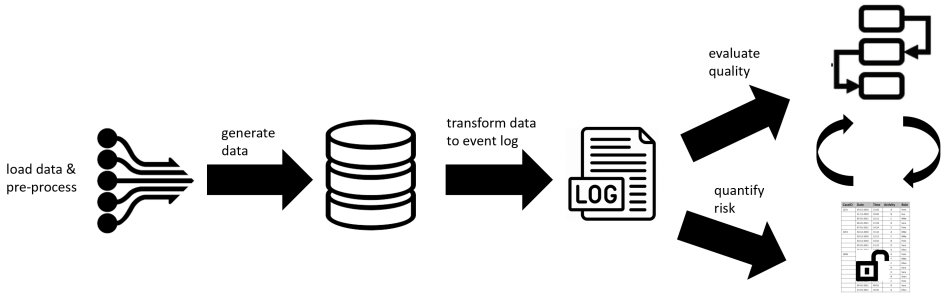


Fig. 1: Process discovery approach for location sensor event data.

3 Challenges

The implementation of a synthetic, privacy-aware event log generation has several challenges. It is difficult to efficiently learn synthetic sequential data using machine learning techniques. There is still a high manual effort required to generate the event log from diverse data types processed by the ML techniques. The generation of an event log is also limited how machine learning techniques learn and the input they expect (e.g., it is challenging to parse varying size of sequences).

Literatur

[El21] Elkoumy, G.; Fahrenkrog-Petersen, S. A.; Sani, M. F.; Koschmider, A.; Mannhardt, F.; von Voigt, S. N.; Rafiei, M.; von Waldthausen, L.: Privacy and Confidentiality in Process Mining - Threats and Research Challenges. CoRR abs/2106.00388/, 2021.

[Ma19] Mannhardt, F.; Koschmider, A.; Baracaldo, N.; Weidlich, M.; Michael, J.: Privacy-Preserving Process Mining - Differential Privacy for Event Logs. Bus. Inf. Syst. Eng. 61/5, S. 595–614, 2019.

[Vo20] von Voigt, S. N.; Fahrenkrog-Petersen, S. A.; Janssen, D.; Koschmider, A.; Tschorsch, F.; Mannhardt, F.; Landsiedel, O.; Weidlich, M.: Quantifying the Re-identification Risk of Event Logs for Process Mining - Empiricial Evaluation Paper. In: CAiSE. Bd. 12127. LNCS, Springer, S. 252–267, 2020.

Sprecher der mit der EMISA assoziierten Arbeitskreise und Fachgruppen

Prof. Dr. Michael Fellmann *AK Semantische Technologien im Geschäftsprozessmanagement* Universität Rostock Institut für Informatik Albert-Einstein-Str. 22 18057 Rostock Tel: +49 - 0381 498-7440 E-Mail: michael.fellmann@uni-rostock.de	**Prof. Dr. Günther Specht** *Arbeitskreis Grundlagen von Informationssystemen* Universität Innsbruck Databases and Information Systems Technikerstrasse 21a A-6020 Innsbruck Tel: +43 512 507 53220 E-Mail: guenther.specht@uibk.ac.at
Horst Kremers *Comittee on Data for Science and Technology (CODATA-Germany)* Postfach 200548 D-13515 Berlin Tel: +49 - 172 3211738 E-Mail: office@horst-kremers.de	**Prof. Dr. Markus Nüttgens** *AK Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten* Universität Hamburg WISO Fakultät, Wirtschaftsinformatik Von-Melle-Park 9 D-20146 Hamburg Tel: +49 – 40-428382792 E-Mail: markus.nuettgens@wiso.uni-hamburg.de

Fachexperten der EMISA

Prof. Dr. Jörg Desel FernUniversität in Hagen Lehrgebiet Softwaretechnik und Theorie der Programmierung D-58084 Hagen Tel: +49 - 2331-987 2609 E-Mail: joerg.desel@fernuni-hagen.de	**Prof. Dr. Dr.hc. Heinrich Mayr** Universität Klagenfurt Institut für Angewandte Informatik Universitätsstraße 65-67 A-9020 Klagenfurt am Wörthersee Tel: +43 - 463 2700 3732 E-Mail: Heinrich.Mayr@aau.at
Prof. Dr. Gottfried Vossen Universität Münster Lehrstuhl für Informatik Leonardo-Campus 3 D-48149 Münster Tel. +49- 251 83-38151 E-Mail: vossen@uni-muenster.de	**Prof. Dr. Andreas Oberweis** Karlsruher Institut für Technologie Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB) Kaiserstraße 89 D-76133 Karlsruhe Tel.: +49 - 721 608 44516 E-Mail: oberweis@kit.edu
Prof. Dr. Mathias Weske Universität Potsdam Hasso-Plattner-Institut für Softwaresystemtechnik Prof. Dr. Helmert-Str. 2-3 D-14482 Potsdam Tel: +49 (0)331-5509191 E-Mail: mathias.weske@hpi.uni-potsdam.de	**Dr.-Ing. Hansjürgen Paul** Institut Arbeit und Technik Forschungsschwerpunkt WISDOM Munscheidstraße 14 D-45886 Gelsenkirchen Tel. +49-209-1707229 E-Mail: paul@iat.eu
Prof. Dr. Stefan Strecker FernUniversität in Hagen Enterprise Modelling Research Group Universitätsstr. 41 D-58097 Hagen Tel: +49 2331 987 4430 E-Mail: stefan.strecker@fernuni-hagen.de	

EMISA-Leitungsgremium

Das im Mai 2018 gewählte EMISA-Leitungsgremium besteht aus den Sprechern der assoziierten Arbeitskreise und Fachgruppen (siehe vorige Seite) sowie folgenden gewählten Mitgliedern:

Prof. Dr. Jan Mendling
(Sprecher des Leitungsgremiums, Redaktion EMISA FORUM)

Humboldt-Universität zu Berlin
Institut für Informatik
Rudower Chaussee 25
D-12489 Berlin

E-Mail: jan.mendling@hu-berlin.de

Prof. Dr. Agnes Koschmider
(Stellv. Sprecherin)

Christian-Albrechts-Universität zu Kiel
Institut für Informatik
Arbeitsgruppe Wirtschaftsinformatik
(Process Analytics)
Hermann-Rodewald-Str. 3
24118 Kiel

E-mail: ak@informatik.uni-kiel.de

Dr. Dirk Fahland

TU Eindhoven
Architecture of Information Systems Group
NL-5600 MB Eindhoven
E-Mail: d.fahland@tue.nl

Dr. Henrik Leopold

Kuehne Logistics University
Großer Grasbrook 17
D- 20457 Hamburg

Universität Potsdam,
Hasso-Plattner-Institut für Softwaresystemtechnik
Prof. Dr. Helmert-Str. 2-3
D-14482 Potsdam
E-Mail: henrik.leopold@the-klu.org

Dr. Judith Michael

RWTH Aachen University
Department of Computer Science
Ahornstraße 55
D-52074 Aachen

E-Mail: michael@se-rwth.de

Prof. Dr. Eric Proper

Luxembourg Institute of Science and Technology
IT for Innovative Services department
5, avenue des Hauts-Fourneaux
L-4362 Esch-sur-Alzette

e.proper@acm.org

Prof. Dr. Ulrich Reimer

FHS St. Gallen
Hochschule für Angewandte Wissenschaften
Institut für Informations- und Prozess-Mgmt.
Rosenbergstrasse 59
CH-9000 St. Gallen

E-Mail: ulrich.reimer@fhsg.ch

Prof. Dr. Stefanie Rinderle-Ma

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Wirtschaftsinformatik und
Geschäftsprozessmanagement

E-Mail: stefanie.rinderle-ma@tum.de

Prof. Dr. Manfred Reichert

Universität Ulm
Fakultät f. Ingenieurwissenschaften u. Informatik
Institut für Datenbanken und Informationssysteme
James-Franck-Ring, Geb. O27
D-89069 Ulm

E-Mail: manfred.reichert@uni-ulm.de

Prof. Dr. Matthias Weidlich

Humboldt-Universität zu Berlin
Institut für Informatik
Rudower Chaussee 25
D-12489 Berlin

E-Mail: matthias.weidlich@hu-berlin.de