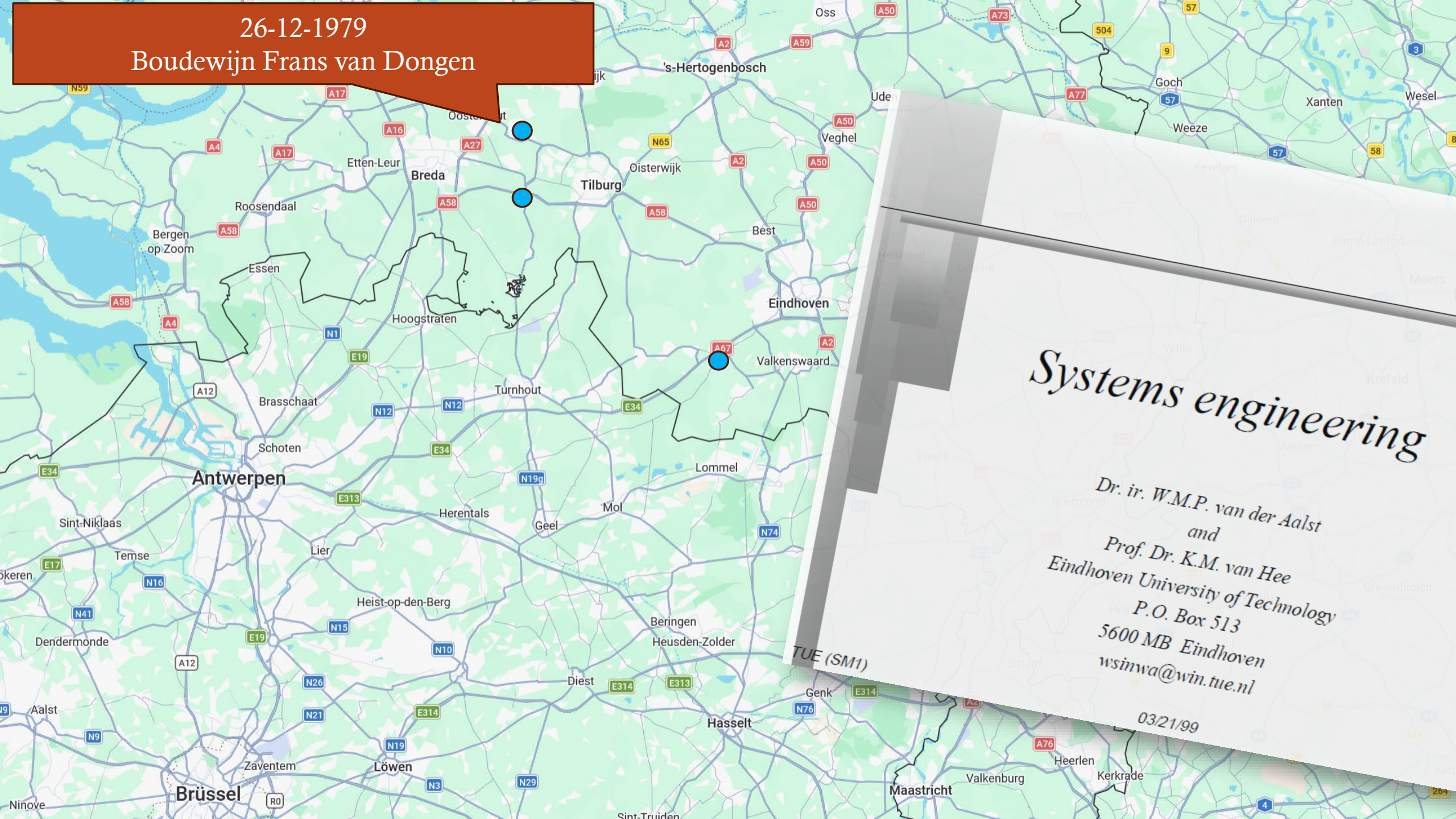
Process Science: The Challenges

Boudewijn van Dongen

Eindhoven University of Technology

TU/e

26-12-1979
Boudewijn Frans van Dongen



Alpha Algorithm:

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\, t \in \sigma \}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\, t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\, t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\, a \rightarrow_L b \wedge \forall_{a_1,a_2 \in A}\, a_1 \#_L a_2 \wedge \forall_{b_1,b_2 \in B}\, b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\, A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

TU/e

Discovering Workflow Performance Models from Timed Logs

W.M.P. van der Aalst and B.F. van Dongen

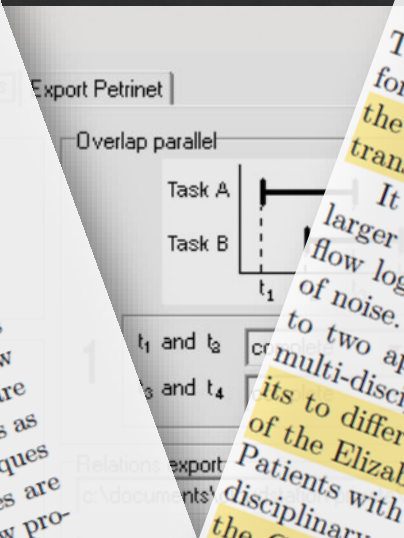
Department of Technology Management, Eindhoven University of Technology
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands.
w.m.p.v.d.aalst@tm.tue.nl

Abstract. Contemporary workflow management systems are driven by explicit process models, i.e., a completely specified workflow design is required in order to enact a given workflow process. Creating a workflow design is a complicated time-consuming process and typically there are discrepancies between the actual workflow processes and the processes as perceived by the management. Therefore, we have developed techniques for discovering workflow models. Starting point for such techniques is a so-called "workflow log" containing information about the workflow process as it is actually being executed. In this paper, we extend our existing mining technique α [4] to incorporate time. We assume that events in workflow logs bear timestamps. This information is used to attribute timing such as queue times to the discovered workflow model. The approach is based on Petri nets and timing information is attached to places. This paper also presents our workflow-mining tool EMiT. This tool translates the workflow log of several commercial systems (e.g., Staffware) to an intermediate dependent XML format. Based on this format the tool mines for causal relations and produces a graphical workflow model expressed in terms of Petri nets.

Key words: Workflow mining, workflow management, data mining

1 Introduction

During the last decade workflow management concepts and systems have been applied in many enterprise information systems such as Staffware, IBM MQSeries, COSA, ... Despite the capabilities for structured business processes, i.e., models describing the life-cycle of ... one can configure these systems to ... management systems many ... Consider for example ... PeopleSoft, Baan ... Despite



This paper presented an approach to extract both a workflow model and performance indicators from timed workflow logs. The approach is supported by the EMiT tool also presented in this paper and has been validated using logs of transactional information systems such as Staffware.

It is important to see the results presented in this paper in the context of a larger effort [4, 20, 28, 29]. The overall goal is to be able to analyze any workflow log without any knowledge of the underlying process and in the presence of noise. At this point in time, we are applying our workflow mining techniques to two applications. The first application is in health-care where the flow of multi-disciplinary patients is analyzed. We have analyzed workflow logs (visits to different specialist) of patients with peripheral arterial vascular diseases at the Elizabeth Hospital in Tilburg and the Academic Hospital in Maastricht. Patients with peripheral arterial vascular diseases are a typical example of multi-disciplinary patients. The second application concerns the processing of fines by the CJIB (Centraal Justitieel Incasso Bureau), the Dutch Judicial Collection Agency located in Leeuwarden. For example fines with respect to traffic violations are processed by the CJIB. However, this government agency also takes care of the collection of administrative fines related to crimes, etc. Through workflow mining we try to get insight in the life-cycle of for example speeding tickets. Some preliminary results show that it is very difficult to mine the flow of multi-disciplinary patients given the large number of exceptions, incomplete data, etc. However, it is relatively easy to mine well-structured administrative processes such as the processes within the CJIB. In both applications we are also trying to take attributes of the cases being processed into account. This way we hope to find correlations between properties of the case and the route through the workflow process.

Acknowledgements The authors would like to thank Eric Verbeek, Ton Weijters, and Laura Maruster for contributing to this work.

Available at www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®
Data & Knowledge Engineering 47 (2003) 237–267
DATA & KNOWLEDGE ENGINEERING
www.elsevier.com/locate/datak
ELSEVIER

Workflow mining: A survey of issues and approaches

W.M.P. van der Aalst [a,*], B.F. van Dongen [a], J. Herbst [b], L. Maruster [a], G. Schimm [c], A.J.M.M. Weijters [a]

[a] Department of Technology Management, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands
[b] DaimlerChrysler AG, Research and Technology, P.O. Box 2360, D-89013 Ulm, Germany
[c] OFFIS, Escherweg 2, D-26121 Oldenburg, Germany

Received 2 October 2002; received in revised form 29 January 2003; accepted 26 February 2003

Abstract

Many of today's information systems are driven by explicit process models. Workflow manage
systems, but also ERP, CRM, SCM, and B2B, are configured on the basis of a workflow model sp
the order in which tasks need to be executed. Creating a workflow design is a complicated time-c
process and typically there are discrepancies between the actual workflow processes and the
perceived by the management. To support the design of workflows, we propose the use of wo
Starting point for workflow mining is a so-called "workflow log" containing inform
workflow process as it is actually being executed. In this paper, we introduce the
mining and present a common format for workflow logs. Then we discuss the most
and present some of the workflow mining approaches available today.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Workflow mining; Workflow management; Data mining; Petri nets

1. Introduction

During the last decade workflow management technology [2,4,21,35,41] has becom
available. Workflow management systems such as Staffware, IBM MQSeries, COSA, etc

* Corresponding author. Tel.: +31-40-247-4295; fax: +31-40-243-2612.
E-mail addresses: w.m.p.v.d.aalst@tm.tue.nl (W.M.P. van der Aalst), joachim.j.herbst@daimlerchrysler.com
(J. Herbst), gschimm@offis.de (G. Schimm).

© 2003 Elsevier B.V. All rights reserved.

Wil van der Aalst is a full professor of Information Systems and head of the section of Information an
Technology of the Department of Technology Management at Eindhoven University of Technology. He i
also a part-time full professor at the Computing Science faculty at the department of Mathematics and
Computer Science at the same university. His research interests include information systems, simulation, Petri
nets, process models, workflow management systems, verification techniques, enterprise resource planning
systems, computer supported cooperative work, and interorganizational business processes.

Boudewijn van Dongen is
University of Technology
and developed the workfl
Computer Science and M
Technology Managemen

Laura Maruster received her B.S. degree in 1994 and M.S. in
West University of Timisoara, Romania. At present she is a
nology Management of Eindhoven University of Technology
interests include induction of machine learning and statistical
covery.

Joachim Herbst studied c
workflow mining. Since
research interests include
concurrent engineering.

Guido Schimm studied computer science and business economy at the Un
He joined the Oldenburg Institute for Computer Science Tools and System
the business intelligence and knowledge management team. Guido has
workflow projects. Currently, his research interest is focused on theoretical
mentation of workflow mining technologies.

Ton Weijters is associate professor at the Department of Technology Management
versity of Technology (TUE), and member of the BETA research group. Current
application of Knowledge Engineering and Machine Learning techniqu
process mining (ii) fundamental research in the domain of Machine L
is the author of many scientific publications in the mentioned

A Meta Model for Process Mining Data

B.F. van Dongen and W.M.P. van der Aalst*

Department of Technology Management, Eindhoven University of Technology
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands.
{b.f.v.dongen,w.m.p.v.d.aalst}@tue.nl

Abstract. Modern process-aware information systems store detailed information about processes as they are being executed. This kind of information can be used for very different purposes. The term *process mining* refers to the techniques and tools to extract knowledge (e.g., in the form of models) from this. Several key players in this area have developed sophisticated process mining tools, such as Aris PPM and the HP Business Cockpit, that are capable of using the information avail[...] generate meaningful insights.

What most of these commercial process mining tools [...] is that installation and maintenance of the systems [...] effort, and deep knowledge of the underlying inform[...] over, information systems log events in different [...] interface between process-aware information syst[...] ing tools is far from trivial. It is vital to correctly [...] logs recorded by the underlying information sys[...] pose a meta model for event logs. We give the [...] that should be available, both informally and [...] back our meta model up with an XML forma[...] ing framework that is capable of reading [...] approach presented in this paper is very p[...] first step towards and ontological analysis [...]

1 Introduction

Under the umbrella of buzzwords such as [...] and "Business Process Intelligence" ([...] Thumb, InWoLvE, Process Miner, a[...] ARIS PPM, HP BPI, and ILOG JViews) [...] these tools is to extract knowledge from event logs ([...] system or audit trails in a WFM system), i.e., to do *process* [...] The research domain *process mining* is relatively new. A compl[...] of recent process mining research is beyond the scope of this paper. [...] we limit ourselves to a brief introduction to this topic and refer to [...] http://www.processmining.org web page for a more complete ove[...]

The goal of process mining is to extract information about [...] [eve]nt logs. It assumes that it is possible to record events [...]

[...] supported by the INTEROP Network of Ex[...] [http:/]/www.interop-noe.org)

Interoperability in the ProM Framework

H.M.W. Verbeek[1], B.F. van Dongen[1], J. Mendling[2], and W.M.P. van der Aalst[1]

[1] Department of Technology Management, Eindhoven University of Technology
P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands.
{h.m.w.verbeek,b.f.v.dongen,w.m.p.v.d.aalst}@tm.tue.nl

[2] Vienna University of Economics and Business Administration
Augasse 2-6, 1090 Vienna, Austria
[...]n.mendling@wu-wien.ac.at

[...]S, XESame, and ProM 6

[H.M.]W. Verbeek, Joos C.A.M. Buijs, [B.F. v]an Dongen, and Wil M.P. van der Aalst

[Techn]ische Universiteit Eindhoven
[...] Mathematics and Computer Science
[...]00 MB Eindhoven, The Netherlands
[...]ijs,b.f.v.dongen,w.m.p.v.d.aalst}@tue.nl

1 In[troduction]

Informa[...] terprise[...] vision o[...] are: wo[...] ning sy[...] languag[...] sus on [...] specific [...] in the [...] differen[...] in conc[...]

Unlike [...] lation [...] provide [...] tems ([...] Note th[...]

The IEEE XES Standard for Process Mining: Experiences, Adoption, and Revision

Moe T. Wynn
Queensland University of Technology, AUSTRALIA

Wil van der Aalst
Technical University of Aachen, GERMANY

Eric Verbeek
Eindhoven University of Technology, THE NETHERLANDS

Bruno Di Stefano
Nuptek Systems Ltd., CANADA


The ProM Framework:
A New Era in Process Mining Tool Support
B.F. van Dongen, A.K.A. de Medeiros, H.M.W. Verbeek, A.J.M.M. Weijters,
ProM 4.0: Compre
W.M.P. van der Aalst[1], B.F. van Dong
Alves de Medeiros[1], A. Rozinat[1], V. Rubin
and A.J.M.M. Weijters[1]
[1] Eindhoven University of Technology, Eindhoven, The Netherlan
w.m.p.v.d.aalst@tue.nl
[2] University of Paderborn, Paderborn, Germany
ProM 6: The Process Mining Toolkit
H.M.W. Verbeek, J.C.A.M. Buijs, B.F. van Dongen, W.M.P. van der Aalst
Department of Mathematics and Computer Science,
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
{h.m.w.verbeek,j.c.a.m.buijs,b.f.v.dongen,w.m.p.v.d.aalst}@tue.nl
Abstract. Process mining has been around for a decade, and it has
proven to be a very fertile and successful research field. Part of this suc-
cess can be contributed to the ProM tool, which combines most of the
existing process mining techniques as plug-ins in a single tool. ProM 6
removes many limitations that existed in the previous versions, in par-
ticular with respect to the tight integration between the tool and the
been developed from scratch and uses a completely
ure. The changes were driven by many real-life
ts into the design of process analysis so
of XESame in this toolkit allows
mat without programming.
ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Information Systems 32 (2007) 713–732
www.elsevier.com/locate/infosys
Business pro
n industrial application
M.M. Weijters[a], B.F. van Dongen[a],
g[a,b], H.M.W. Verbeek[a]
P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands
nd Technology, San 31 Hyoja-Dong, Nam-gu,
a
ay 2006
Jr.
mation systems (e.g., WfM, ERP, CRM, SCM, and B2B systems) record business events in so-called
ess process mining takes these logs to discover process, control, data, organizational, and social structures.
gh many researchers are developing new and more powerful process mining techniques and software vendors are
incorporating these in their software, few of the more advanced process mining techniques have been tested on real-life
processes. This paper describes the application of process mining in one of the provincial offices of the Dutch National
Public Works Department, responsible for the construction and maintenance of the road and water infrastructure. Using a
variety of process mining techniques, we analyzed the processing of invoices sent by the various subcontractors and
suppliers from three different perspectives: (1) the process perspective, (2) the organizational perspective, and (3) the case
perspective. For this purpose, we used some of the tools developed in the context of the ProM framework. The goal of this
paper is to demonstrate the applicability of process mining in general and our algorithms and tools in particular.
© 2006 Elsevier B.V. All rights reserved.
Keywords: Process mining; Social network analysis; Workflow management; Business process management; Business process analysis;
Data mining; Petri nets
1. Introduction
Today, many enterprise information systems
store relevant events in some structured form. For
example, Workflow Management Systems (WfMSs)
typically register the start and completion of
activities [1]. ERP systems like SAP log all transac-
tions, e.g., users filling out forms, changing docu-
ments, etc. Business-to-business (B2B) systems log
the exchange of messages with other parties. Call
center packages but also general-purpose CRM
systems log interactions with customers. These
examples show that many systems have some kind
of event log often referred to as "history", "audi
trail", "transaction log", etc [2–5]. The
typically contains information ab
ring to an activity and a c
process instance).
handled,
*Corresponding author. Tel.: +31 40 2474295.
E-mail address: w.m.p.v.d.aalst@tm.tue.nl
(W.M.P. van der Aalst).
0306-4379/$ - see front matter © 2006 Elsevier B.V. All
doi:10.1016/j.is.2006.05.003
4TU.ResearchData
SCIENCE • ENGINEERING • DESIGN
U/e

BPI Challenge 2011

We have prepared a real-life log, taken from a Dutch Academic Hospital. This log contains some 150.000 events in over 1100 cases. **Apart from some anonymization**, the log contains all data as it came from the Hospital's systems. Each case is a patient of a Gynecology department. The log contains information about when certain activities took place, which group performed the activity and so on. Many attributes have been recorded that are relevant to the process. Some attributes are repeated more than once for a patient, indicating that this patient went through different (maybe overlapping) phases, where a phase consists of the combination Diagnosis & Treatment.

Results

Three submissions arrived and were judged by the jury. The jury unanimously ranked them in the following order:

1. J.C. Bose and W.M.P. van der Aalst PDF Analysis of Patient Treatment Procedures
2. F. Caron, J. Vanthienen, J. De Weerdt and B. Baesens PDF Beyond X-Raying a Case-Flow: Adopting Different Focuses on Care-Flow Mining
3. G. Varvaressos ZIP Semantic Process Mining

TU/e

BPI Challenge 2012

We have prepared a real-life log, taken from a Dutch Financial Institute. This log contains some 262.200 events in 13.087 cases. **Apart from some anonymization**, the log contains all data as it came from the financial institute. The process represented in the event log is an application process for a personal loan or overdraft within a global financing organization. The amount requested by the customer is recorded in the case attribute AMOUNT_REQ, which is global, i.e. every case contains this attribute. The event log is a merger of three interrelated subprocesses. The first letter of each task name identifies from which subprocess (**source**) it originated.

The Winner

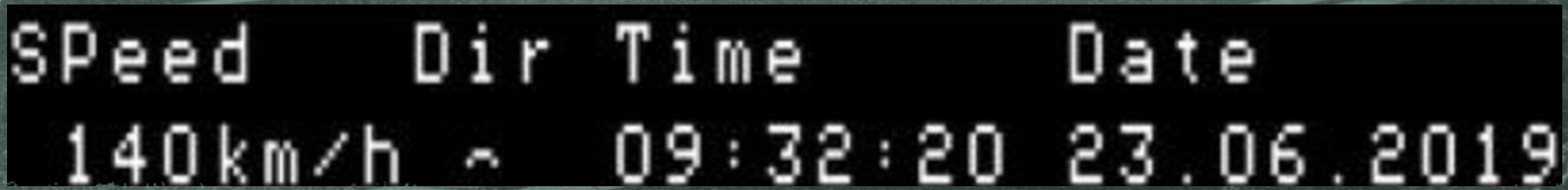
The winner of this year's challenge is:

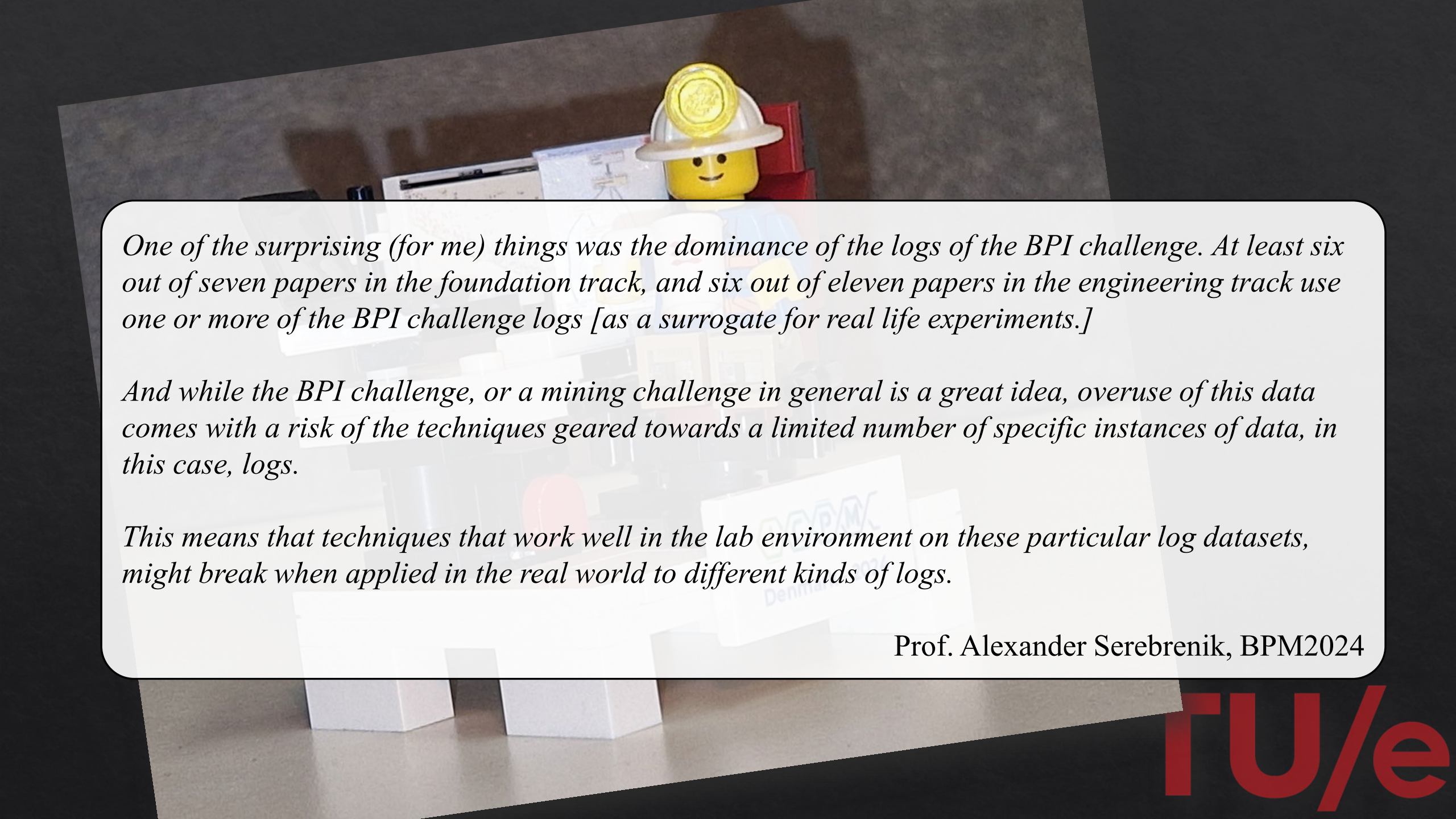
- A.D. Bautista, L. Wangikar and S.M. Kumail Akbar,CKM Advisors, New York, NY, USA

According to the jury: *"Their submission shows a very results-driven method of analyzing, where every analysis seemed to be driven by the motivation to dis/prove a specific hypothesis, related to a concrete and actionable improvement potential in the client company. This results in a successful conversion of analysis results in digestible business level results and recommendations"*

TU/e

	Source company	Country
2011	Academic Hospital	The Netherlands
2012	Financial Institute	The Netherlands
2013	Volvo IT	Belgium
2014	ICT department Rabobank	The Netherlands
2015	Five municipalities	The Netherlands
2016	Unemployment agency UWV	The Netherlands
2017	Financial institute of 2012	The Netherlands
2018	EU Subsidy facilitator	Germany
2019	Coatings and paints company	The Netherlands
2020	TU/e	The Netherlands





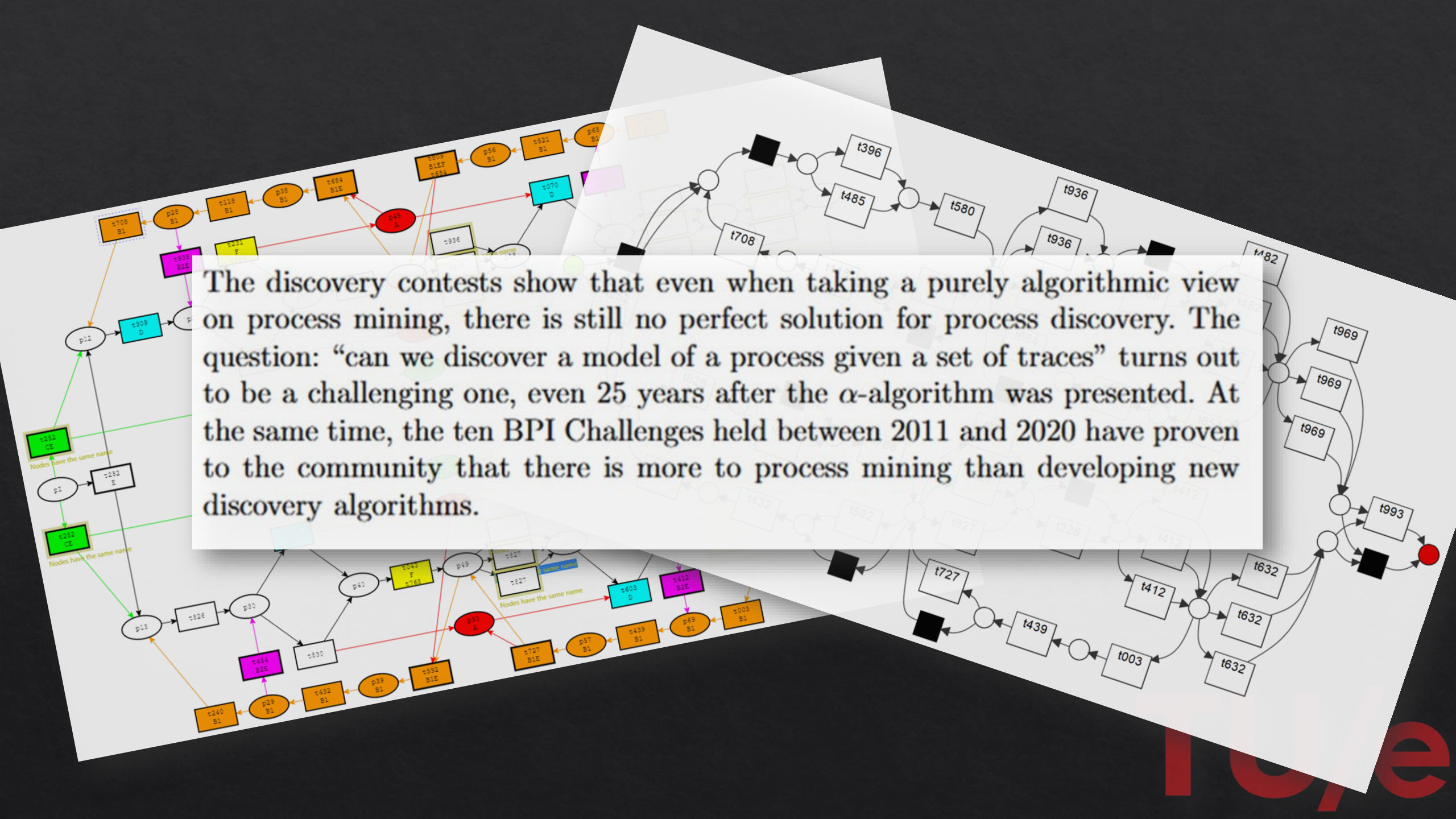
One of the surprising (for me) things was the dominance of the logs of the BPI challenge. At least six out of seven papers in the foundation track, and six out of eleven papers in the engineering track use one or more of the BPI challenge logs [as a surrogate for real life experiments.]

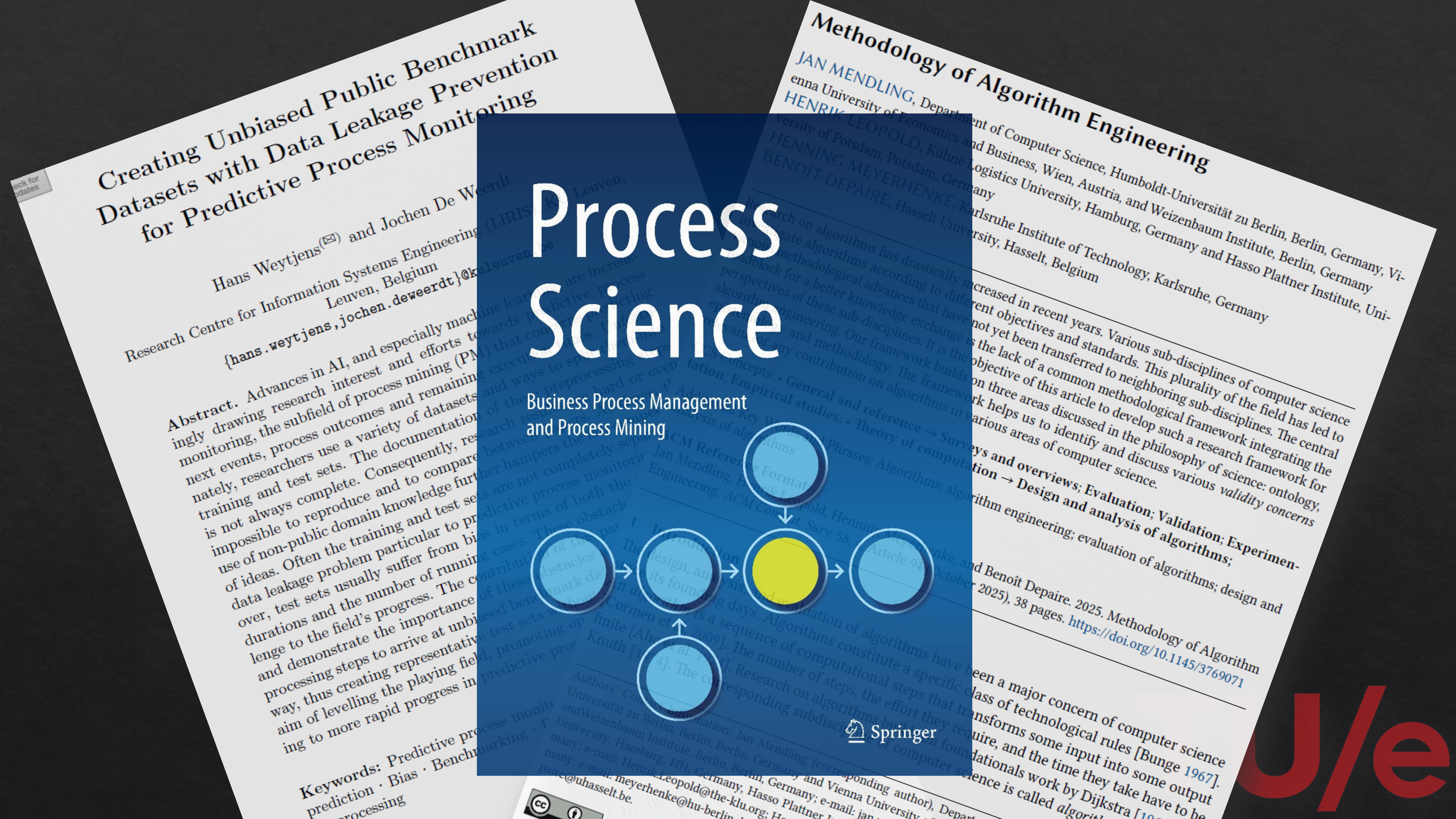
And while the BPI challenge, or a mining challenge in general is a great idea, overuse of this data comes with a risk of the techniques geared towards a limited number of specific instances of data, in this case, logs.

This means that techniques that work well in the lab environment on these particular log datasets, might break when applied in the real world to different kinds of logs.

Prof. Alexander Serebrenik, BPM2024

TU/e


The discovery contests show that even when taking a purely algorithmic view on process mining, there is still no perfect solution for process discovery. The question: “can we discover a model of a process given a set of traces” turns out to be a challenging one, even 25 years after the α-algorithm was presented. At the same time, the ten BPI Challenges held between 2011 and 2020 have proven to the community that there is more to process mining than developing new discovery algorithms.


Creating Unbiased Public Benchmark Datasets with Data Leakage Prevention for Predictive Process Monitoring
Hans Weytjens and Jochen De Weerdt
Process Science
Business Process Management and Process Mining
Springer
Methodology of Algorithm Engineering
JAN MENDLING
HENRIK LEOPOLD
HENNING MEYERHENKE
BENOÎT DEPAIRE

The Challenges of Process Science

- Combining formal and theoretical analysis with engineering research and rigorous research methods
- Stepping away from using BPI Challenges to validate academic work
- Transitioning from case-centric event data to object centric, embracing the multi-faceted nature of process data
- Dealing with bloating, blurring and blasting of generative AI in a time where students seem to believe anything AI says is true
- Dealing with the delicate balance between privacy and accessibility of data for research

TU/e