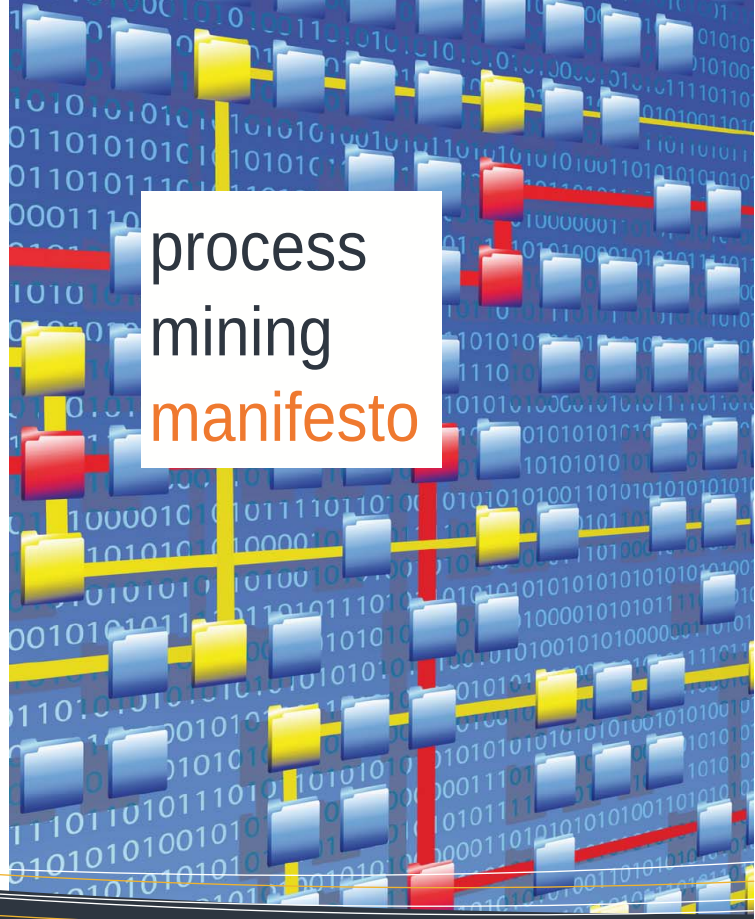
	Challenges
C1	**Finding, Merging, and Cleaning Event Data** When extracting event data suitable for process mining several challenges need to be addressed: data may be distributed over a variety of sources, event data may be incomplete, an event log may contain outliers and events at different level of granularity.
C2	**Dealing with Complex Event Logs Having Diverse Characteristics** Event logs may be extremely large making them difficult to handle whereas other event logs are so small that not enough data is available to make reliable conclusions.
C3	**Creating Representative Benchmarks** Good benchmarks consisting of example data sets and representative quality criteria are needed to compare and improve the various tools and algorithms.
C4	**Dealing with Concept Drift** The process may be changing while being analyzed. Understanding such phenomena is of prime importance for the management of processes.
C5	**Improving the Representational Bias Used for Process Discovery** A more careful and refined selection of the representational bias is needed to ensure high-quality process mining results.
C6	**Balancing Between Quality Criteria such as Fitness, Simplicity, Precision, and Generalization** There are four competing quality dimensions: (a) fitness, (b) simplicity, (c) precision, and (d) generalization. The challenge is to find models that score good in all four dimensions.
C7	**Cross-Organizational Mining** Some organizations work together to handle process instances (e.g., supply chain partners) or organizations are executing essentially the same process while sharing experiences, knowledge, or a common infrastructure. The analysis of event logs originating from multiple organizations provides several challenges.
C8	**Providing Operational Support** Process mining is not restricted to off-line analysis and can also be used for online operational support. Three operational support activities can be identified: detect, predict, and recommend..
C9	**Combining Process Mining With Other Types of Analysis** The challenge is to combine automated process mining techniques with other analysis approaches (optimization techniques, data mining, simulation, visual analytics, etc.) to extract more insights from event data..
C10	**Improving Usability for Non-Experts** The challenge is to hide the sophisticated process mining algorithms behind user-friendly interfaces that automatically set parameters and suggest suitable types of analysis.
C11	**Improving Understandability for Non-Experts** The user may have problems understanding the output or is tempted to infer incorrect conclusions. Process mining esults should be presented using a suitable representation and the trustworthiness of these results should be indicated.



process mining manifesto

The IEEE Task Force on Process Mining was established in 2009 and has members representing software vendors (Pallas Athena, Software AG, Futura Process Intelligence, HP, IBM, Infosys, Fluxicon, Businesscape, Iontas/Verint, Fujitsu, Fujitsu Laboratories, Business Process Mining, and Stereologic), consultancy firms/end users (ProcessGold, Business Process Trends, Gartner, Deloitte, Process Sphere, Siav SpA, BPM Chile, BWI Systeme GmbH, Excellentia BPM, and Rabobank), and research institutes (TU/e, University of Padua, Universitat Politècnica de Catalunya, New Mexico State University, IST - Technical University of Lisbon, University of Calabria, Penn State University, University of Bari, Humboldt-Universität zu Berlin, Queensland University of Technology, Vienna University of Economics and Business, Stevens Institute of Technology, University of Haifa, University of Bologna, Ulsan National Institute of Science and Technology, Cranfield University, K.U. Leuven, Tsinghua University, University of Innsbruck, and University of Tartu).

Get the process mining manifesto from www.win.tue.nl/ieeetfpm/

IEEE task force on process mining

IEEE task force on process mining

www.win.tue.nl/ieeetfpm/

Starting point for process mining is an event log

SEQUENTIALLY RECORDED EVENTS ARE USED TO DISCOVER, CHECK, AND EXTEND PROCESS MODELS

Each event refers to an activity (i.e., a well-defined step in the process) and is related to a particular case (i.e., a process instance). Event logs may store additional information such as the resource executing or initiating an activity, the timestamp of an event, or data elements recorded with an event.

Discovery

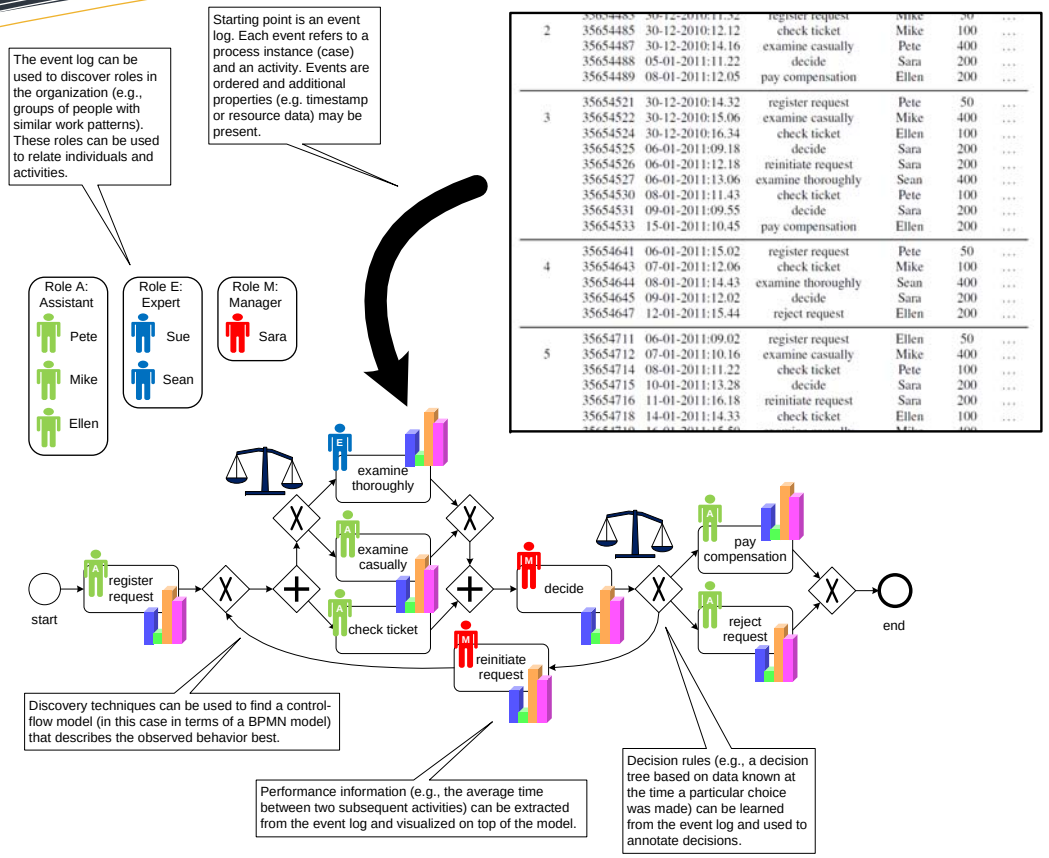
Process discovery techniques take an event log and produce a model without using any other a-priori information. There are dozens of techniques to extract process models from raw event data.

Conformance

Conformance checking techniques compare an existing process model with an event log of the same process. The comparison shows where the “real process” deviates from the modeled process.

Enhancement

By relating observed events to activities in a process model it is possible to extend the model with additional perspectives (e.g., time, data, and resources).



2	35654485	30-12-2010:12.12	check ticket	Mike	100	…
	35654487	30-12-2010:14.16	examine casually	Pete	400	…
	35654488	05-01-2011:11.22	decide	Sara	200	…
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	…
3	35654521	30-12-2010:14.32	register request	Pete	50	…
	35654522	30-12-2010:15.06	examine casually	Mike	400	…
	35654524	30-12-2010:16.34	check ticket	Ellen	100	…
	35654525	06-01-2011:09.18	decide	Sara	200	…
	35654526	06-01-2011:12.18	reinitiate request	Sara	200	…
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400	…
	35654530	08-01-2011:11.43	check ticket	Pete	100	…
	35654531	09-01-2011:09.55	decide	Sara	200	…
	35654533	15-01-2011:10.45	pay compensation	Ellen	200	…
4	35654641	06-01-2011:15.02	register request	Pete	50	…
	35654643	07-01-2011:12.06	check ticket	Mike	100	…
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400	…
	35654645	09-01-2011:12.02	decide	Sara	200	…
	35654647	12-01-2011:15.44	reject request	Ellen	200	…
5	35654711	06-01-2011:09.02	register request	Ellen	50	…
	35654712	07-01-2011:10.16	examine casually	Mike	400	…
	35654714	08-01-2011:11.22	check ticket	Pete	100	…
	35654715	10-01-2011:13.28	decide	Sara	200	…
	35654716	11-01-2011:16.18	reinitiate request	Sara	200	…
	35654718	14-01-2011:14.33	check ticket	Ellen	100	…

	Guiding principles
GP1	**Event Data Should Be Treated as First-Class Citizens** Events should be trustworthy, i.e., it should be safe to assume that the recorded events actually happened and that the attributes of events are correct. Event logs should be complete, i.e., given a particular scope, no events may be missing. Any recorded event should have well-defined semantics. Moreover, the event data should be safe in the sense that privacy and security concerns are addressed when recording the event log.
GP2	**Log Extraction Should Be Driven by Questions** Without concrete questions it is very difficult to extract meaningful event data. Consider, for example, the thousands of tables in the database of an ERP system like SAP. Without questions one does not know where to start.
GP3	**Concurrency, Choice and Other Basic Control-Flow Constructs Should be Supported** Basic workflow patterns supported by all mainstream languages (e.g., BPMN, EPCs, Petri nets, BPEL, and UML activity diagrams) are sequence, parallel routing (AND-splits/joins), choice (XOR-splits/joins), and loops. Obviously, these patterns should be supported by process mining techniques.
GP4	**Events Should Be Related to Model Elements** Conformance checking and enhancement heavily rely on the relationship between elements in model and events in the log. This relationship may be used to “replay” the event log on the model. Replay may be used to reveal discrepancies between event log and model (e.g., some events in the log are not possible according to the model) and is used to enrich the model with additional information extracted from the event log (e.g., bottlenecks are identified by using the timestamps in the event log).
GP5	**Models Should Be Treated as Purposeful Abstractions of Reality** A model derived from event data provides a view on reality. This view should provide a purposeful abstraction of the behavior captured in the event log. Given an event log, there may be multiple views that are useful.
GP6	**Process Mining Should Be a Continuous Process** Given the dynamical nature of processes, it is not advisable to see process mining as a one-time activity. The goal should not be to create a fixed model, but to breathe life into process models such that users and analysts are encouraged to look at them on a daily basis.

The **Process Mining Manifesto** Lists Six **Guiding Principles** and Eleven **Challenges**

The six guiding principles can be used to avoid common mistakes when applying process mining in real-life settings.

Despite the broad applicability of process mining there are still important challenges that need to be addressed.

The main challenges are listed in the manifesto produced by the IEEE Task Force on Process Mining.