Behavioral characteristics of Process Models from which the provided event logs were generated.

This documents contains information that can be used by contestants in order to tune their discovery algorithms.

The 10 events logs are generated from 10 different process models. The event logs only record the information about the order with which activities are completed. Therefore, there is no life-cycle transition information and, also, no timestamp information. However, these information types are not relevant for the contest in question.

Each of these models is characterized by the following aspects:

- **Sequences.** Certain activities need to be sequentially executed. For example, when a given activity A occurs, it is eventually followed by a certain activity B in all runs of the process.
- **Exclusive Choices.** Certain process model branches at given decision points are mutually exclusive. For example, a decision point exists between activity A and B. In any run of the process, if activity A is executed, then activity B cannot, or vice versa.
- **Parallel Executions.** Certain branches are "parallel", meaning that they can be completed in any order. For example, if a branch "A followed by B" is parallel to a branch "C followed by D", activities A, B, C and D can be executed in any order with the only constraint that B cannot finish before A and D cannot finish C. For instance, the execution runs <…, A, C, B, D> or <A, C, D, B> are valid whereas <A, D, C, B> is not, with the latter being because D cannot conclude before C concludes.

Each of this model can optionally contain the following characteristics:

- **Loops.** Certain parts of the model can be repeated an arbitrary number of times.
- **Optional Activities.** Certain activities are optional and can be skipped in certain runs of the process.
- **Inclusive Choices.** Within the process, multiple sets of activities are optional, i.e. at least one set should be executed, but multiple sets of activities are also allowed. The difference with an exclusive choice resides on the fact that, in an exclusive choice, exactly one branch is activated; conversely, in an inclusive choice, more than one branch can be activate.
- **Recurrent activities:** Activities can be executed in multiple non-subsequent points during runs of the process.
- **Long-term dependencies**: A decision made at one point in the process can restrict the possibilities at subsequent decision points. For example, at the beginning of a process, a choice is made between an activity A and an activity B. When activity A is chosen, later during any run, an activity

C cannot be executed; if activity B is chosen, activity C can still be executed. In the Petri-net terminology, this corresponds to Petri nets with non-free-choice constructs.

Exclusive choices can be characterized by **balanced** or **unbalanced paths**. If an exclusive-choice is characterized by being balanced, in any run of the process, each mutually exclusive set of activities has equal probability of being chosen. If conversely it is unbalanced, one set has a 90% probability of being chosen and the other sets, together, have 10%, with each of them having the same probability. In the remainder, we generated event logs such that either all decision points are balanced or they are all unbalanced.

With reference to the characteristics above, the processes to which the generated the event logs refer to are as follows:

Optional Characteristics (Always 2 for each process model)		Exclusive-choice decision points	Process model / Event Log
Optional activities	Long-term dependencies	Unbalanced Paths	1*
Optional activities	Loops	Balanced Paths	2*
Optional activities	Inclusive Choices	Unbalanced Paths	3
Loops	Reoccurring activities	Unbalanced Paths	4
Loops	Long-term dependencies	Balanced Paths	5*
Loops	Inclusive Choices	Balanced Paths	6
Inclusive Choices	Reoccurring activities	Balanced Paths	7
Inclusive Choices	Long-term dependencies	Balanced Paths	8
Reoccurring activities	Long-term dependencies	Unbalanced Paths	9*
Optional activities	Reoccurring activities	Unbalanced Paths	10*

The event logs marked in the table above with an asterisk (i.e. *) are characterized by containing 20% of incomplete traces. Those traces are incomplete in the sense that they miss the last events. This is very common in reality because the event log are usually extracted from information systems in which a certain number of process executions are still being carried on.