



*Vorlesung*  
***Methodische Grundlagen des  
Software-Engineering***  
im Sommersemester 2014

Prof. Dr. Jan Jürjens

TU Dortmund, Fakultät Informatik, Lehrstuhl XIV

Teil 2.3: Datenbeschaffung

v. 18.05.2014

1



## 2.3 Datenbeschaffung

[mit freundlicher Genehmigung basierend  
auf einem englischen Foliensatz von  
Prof. Dr. Wil van der Aalst (TU Eindhoven)]

### Literatur:

[vdA11] Wil van der Aalst: **Process Mining: Discovery, Conformance and Enhancement of Business Processes**, Springer-Verlag, 2011.

Unibibliothek (6 Exemplare): <http://www.ub.tu-dortmund.de/katalog/titel/1332248>  
(Bei Engpässen kann eine **Kopiervorlage** der relevanten Ausschnitte zur Verfügung gestellt werden.)

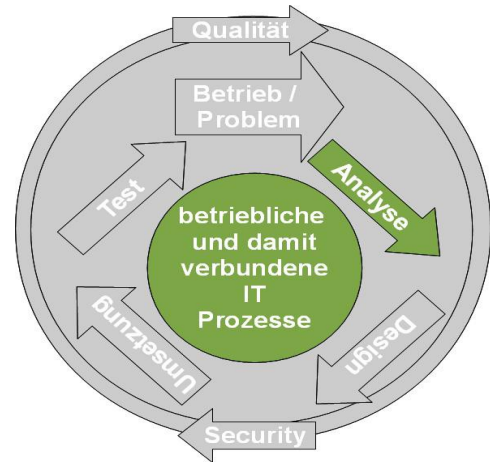
- **Kapitel 4**

- Geschäftsprozessmodellierung

- **Process-Mining**

- Einführung: Process-Mining
- Petrinetze
- Data-Mining
- **Datenbeschaffung**
- Prozessextraktion
- Konformanzanalyse
- Mining: Zusätzliche Perspektiven
- Betriebsunterstützung
- Werkzeugunterstützung
- Analysiere „Lasagne Prozesse“
- Analysiere „Spaghetti Prozesse“
- Kartographie und Navigation
- Epilog

- Modellbasierte Entwicklung sicherer Software





- **Letzter Abschnitt:** Datenbasierte Modellanalyse.
- **Dieser Abschnitt:** „Datenbeschaffung“:
  - Heterogene Datenquellen
  - Event-Logs



- Von heterogenen Datenquellen zu Process-Mining
- Datenspeicherformat XES
- Herausforderungen beim Extrahieren des Event-Logs



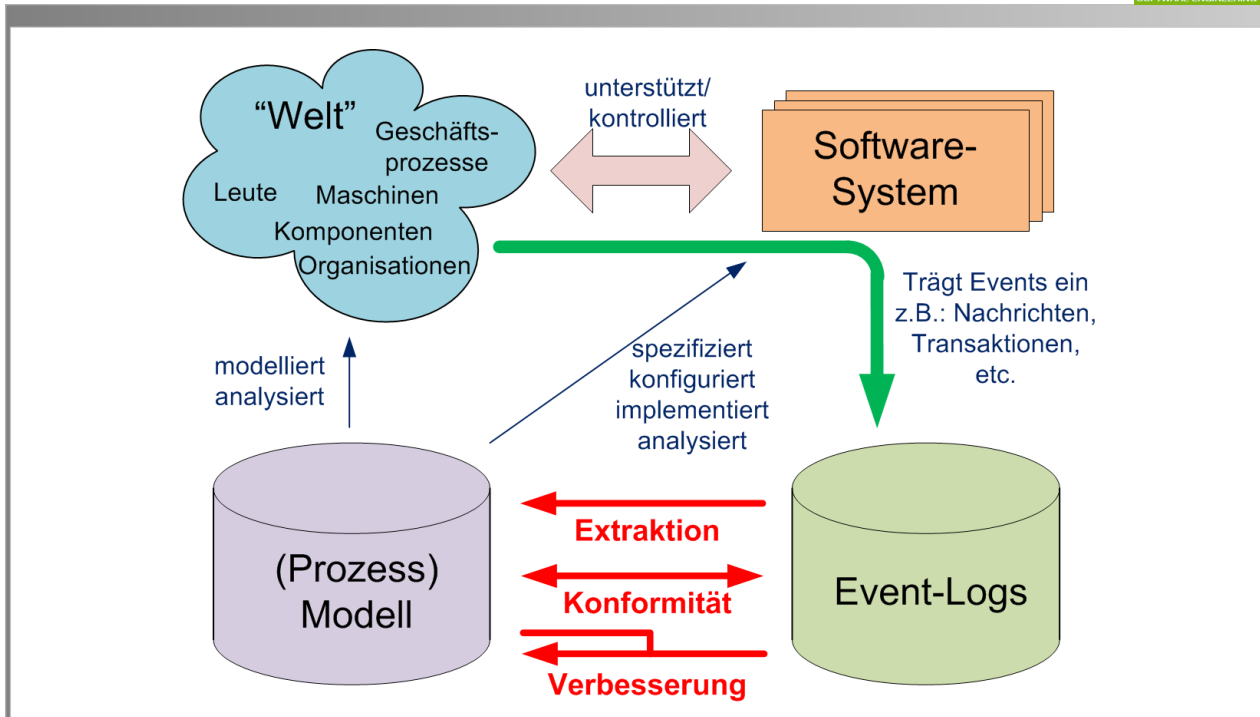
- Was geschah in der **Vergangenheit** ?
- Warum ist es passiert ?
- Was wird vermutlich in der **Zukunft** geschehen ?
- Wann und warum weichen Unternehmen und Leute voneinander ab ?
- Wie kann ein **Prozess** besser kontrolliert werden ?
- Wie kann ein Prozess neu entworfen werden, sodass die **Performanz** gesteigert wird ?

6

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.1: S. 95

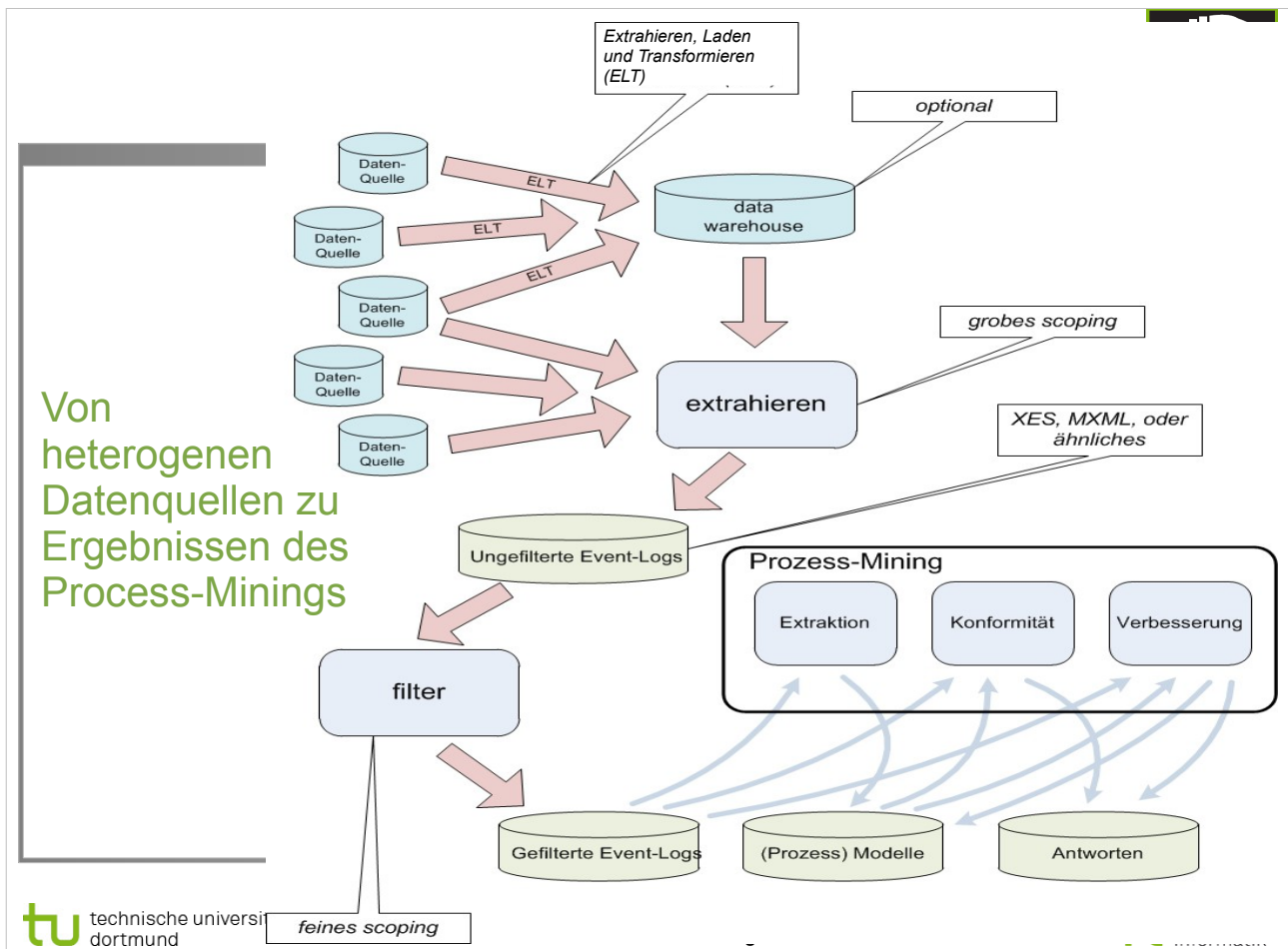


7

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 1.3: S. 9 Fig. 1.4



## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.1: S. 96 Fig. 4.1





- **Prozess** enthält **Fälle** (cases).
- **Fall** besteht aus **Events**, jeden Event genau einem Fall zuordnen.
- **Events** innerhalb eines Falles: **geordnet**.
- Events können **Attribute** haben.
- Beispiele typischer **Attributnamen**: **activity**, **time**, **costs** und **resource**.

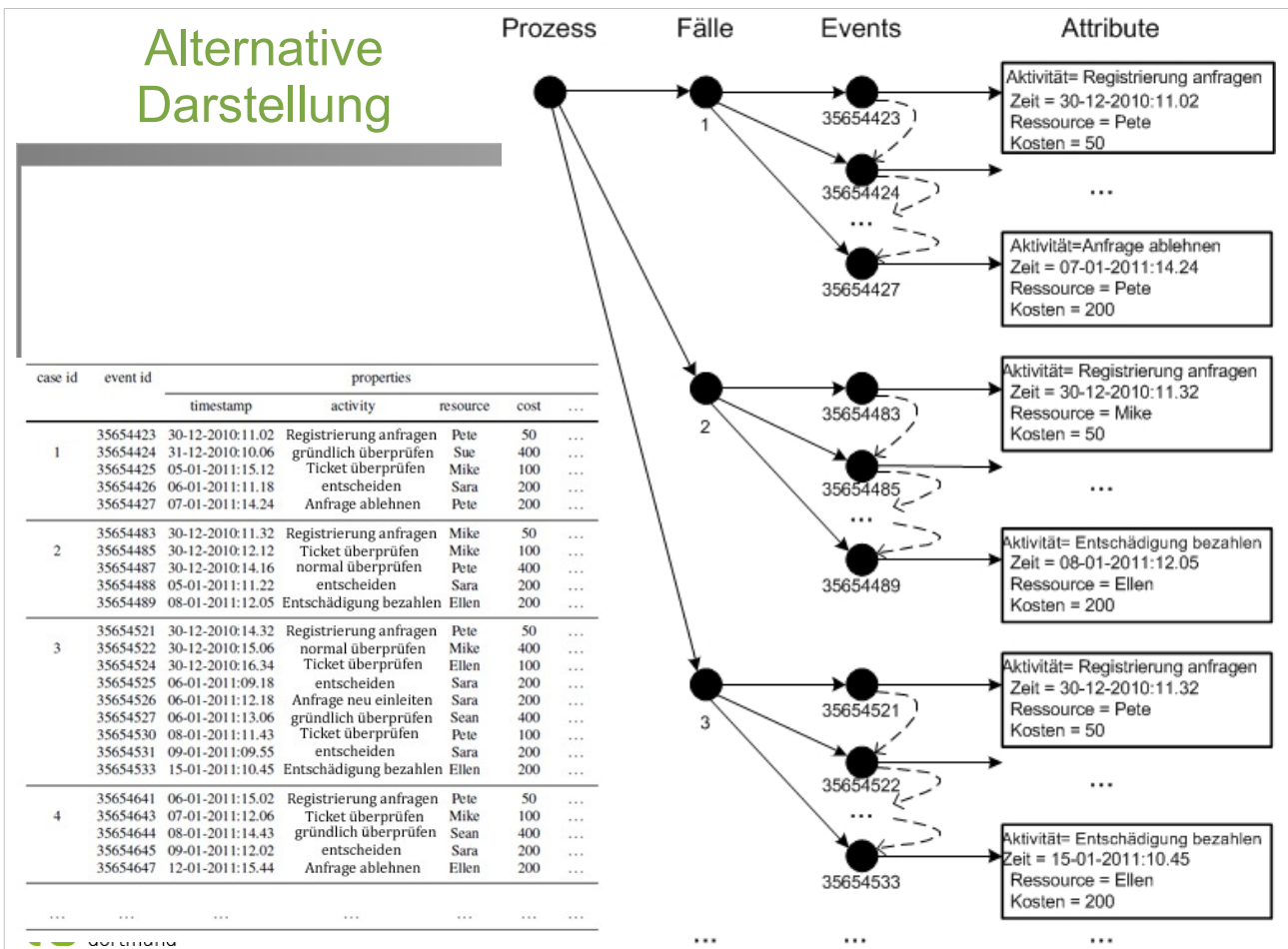
case id	event id	properties				
		timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	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	...
...	...	...	...	...	...	...

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.2: S. 99 Tabelle 4.1

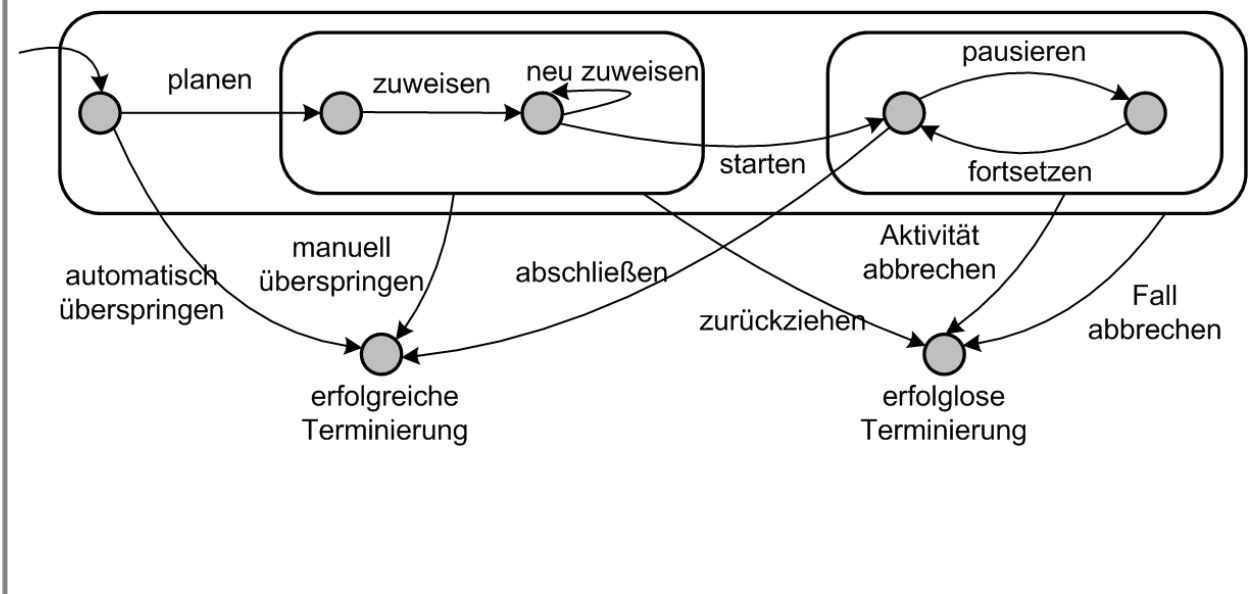
# Alternative Darstellung



## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.2: S. 99 Tabelle 4.1, S. 100 Fig. 4.2

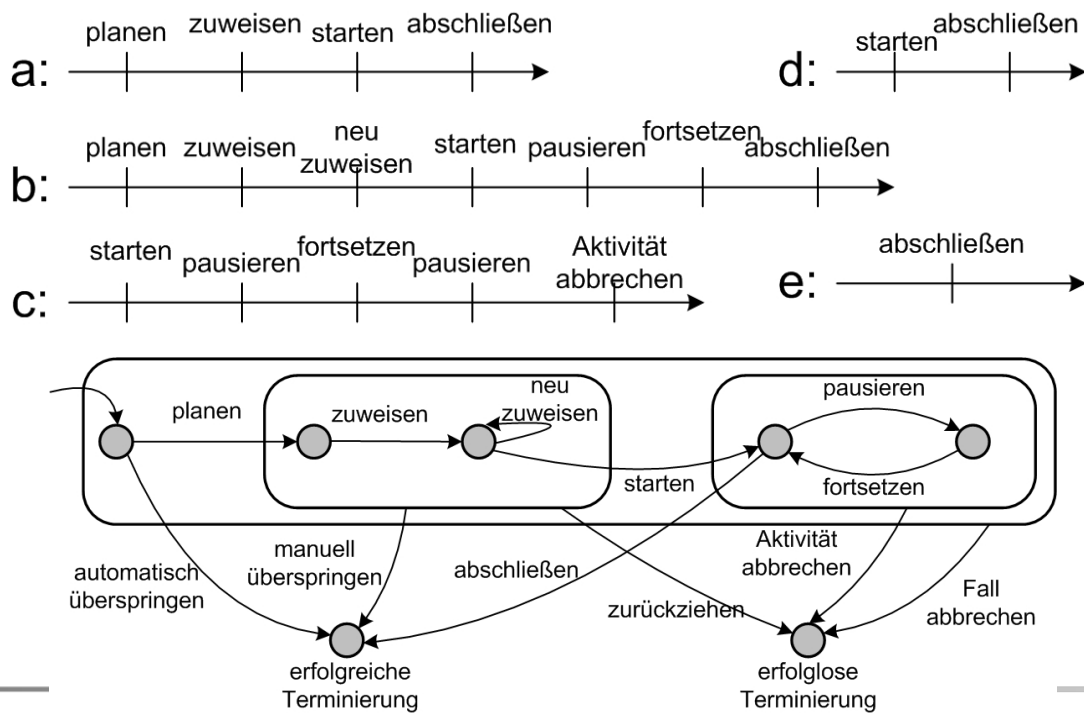


11

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.2: S.101 Fig. 4.3



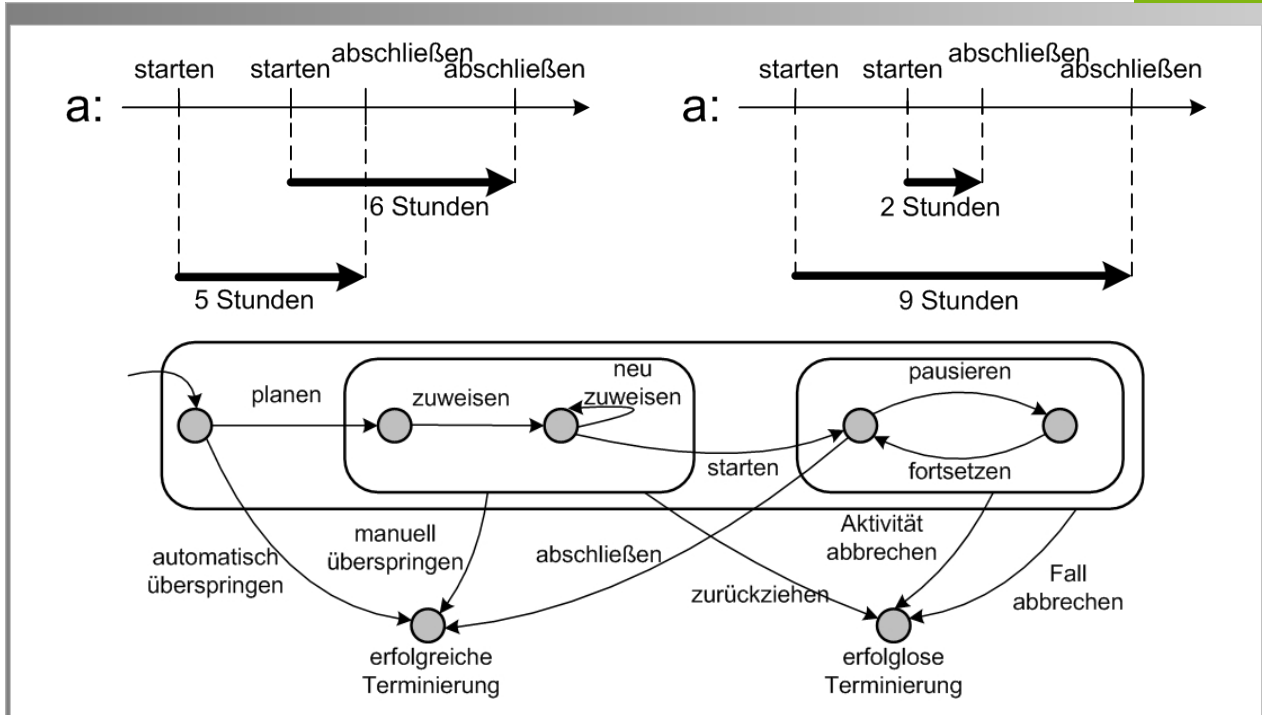
12

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Obige Abbildung: Kap. 4.2: S. 102 Fig. 4.4
- Untere Abbildung: Kap. 4.2: S. 101 Fig. 4.3

# Beispiel-Transaktion: Überlappende Prozessinstanzen



13

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Obige Abbildung: Kap. 4.2: S. 102 Fig. 4.5
- Untere Abbildung: Kap. 4.2: S. 101 Fig. 4.3



- Von heterogenen Datenquellen zu Process-Mining
- **Datenspeicherformat XES**
- Herausforderungen beim Extrahieren des Event-Logs



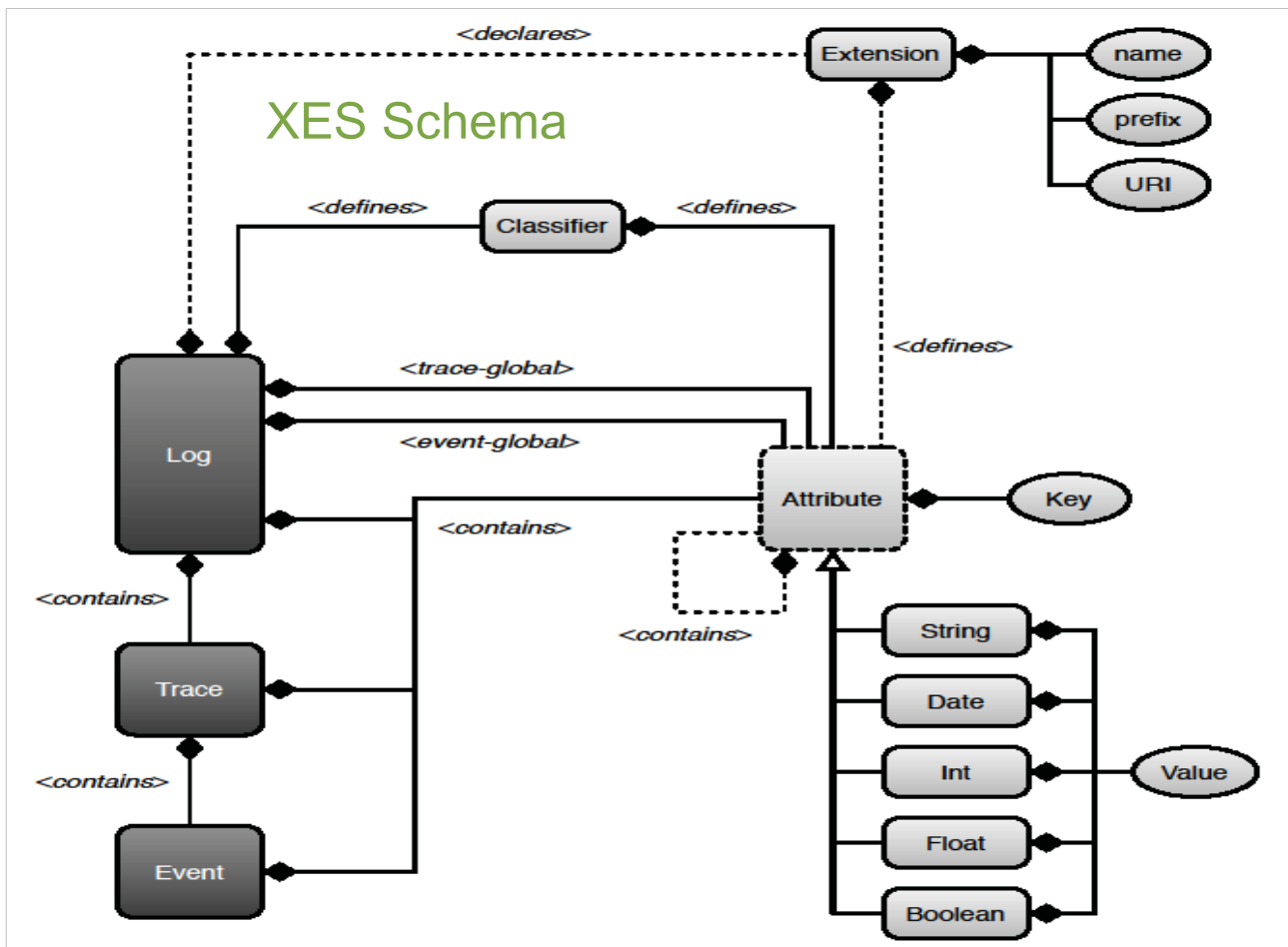
- **Standard-Datenspeicherformat für Event-Logs.**
- Siehe [www.xes-standard.org](http://www.xes-standard.org).
- Von IEEE Arbeitsgruppe für Process-Mining übernommen.
- Vorgänger: MXML und SA-MXML.
- Von Tools wie ProM (ab Version 6), Nitro, XESame und OpenXES unterstützt.
- ProMimport unterstützt MXML.

15

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.3: S. 107 letzter Abschnitt – S. 108 2. Abschnitt



**Literatur:**

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.3: S. 109 Fig. 4.7



**Event-Log** besteht aus:

- **Traces** (Prozessinstanzen)
  - **Events**

**Standarderweiterungen:**

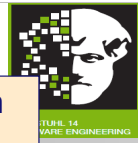
- concept (zur Namensgebung)
- lifecycle (für Transaktionseigenschaften)
- org (unternehmerische Perspektive)
- time (Zeitstempel)
- semantic (Referenzen zu Ontologie)

17

**Literatur:**

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.3: S. 110 unterer Abschnitt & S. 112 1. Abschnitt



```
<log xes.version="1.0" xes.features="nested-attributes" openxes.version="1.0RC7">
  <extension name="Concept" prefix="concept" uri="http://code.deckfour.org/xes/concept.xesext"/>
  <extension name="Semantic" prefix="semantic" uri="http://code.deckfour.org/xes/semantic.xesext"/>
  <extension name="Time" prefix="time" uri="http://code.deckfour.org/xes/time.xesext"/>
  <extension name="Organizational" prefix="org" uri="http://code.deckfour.org/xes/org.xesext"/>
  <extension name="Lifecycle" prefix="lifecycle" uri="http://code.deckfour.org/xes/lifecycle.xesext"/>
  <global scope="trace">
    <string key="conceptname" value="1" />
  </global>
  <global scope="event">
    <string key="conceptname" value="INVALID" />
    <string key="lifecycle:transition" value="complete" />
  </global>
  <classifier name="MXML Legacy Classifier" keys="conceptname lifecycle:transition"/>
  <classifier name="Event Name" keys="conceptname"/>
  <classifier name="Resource" keys="org:resource"/>
  <string key="source" value="d_more_data.zip" />
  <string key="conceptname" value="standard" />
  <string key="lifecycle:transition" value="standard" />
  <string key="description" value="Simulated process" />
  <trace>
    <string key="conceptname" value="1" />
    <string key="description" value="Simulated process instance" />
    <event>
      <string key="org:resource" value="Mike" />
      <date key="time:timestamp" value="2006-01-01T00:00:00.000+01:00" />
      <string key="conceptname" value="invite reviewers" />
      <string key="lifecycle:transition" value="start" />
    </event>
    <event>
      <string key="org:resource" value="Mike" />
      <date key="time:timestamp" value="2006-01-06T00:00:00.000+01:00" />
      <string key="conceptname" value="invite reviewers" />
      <string key="lifecycle:transition" value="complete" />
    </event>
  </trace>
</log>
```

jeder Trace hat einen Namen

jedes Event hat einen Namen und eine Transition

Beginn vom Trace (Prozessinstanz)

classifier = Name + Transition

Name des Trace

Ressource

Zeitstempel

Name des Events (Aktivitätsname)

Transition

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.3: S. 110-111 Fig. 4.8



- Von heterogenen Datenquellen zu Process-Mining
- Datenspeicherformat XES
- Herausforderungen beim Extrahieren des Event-Logs



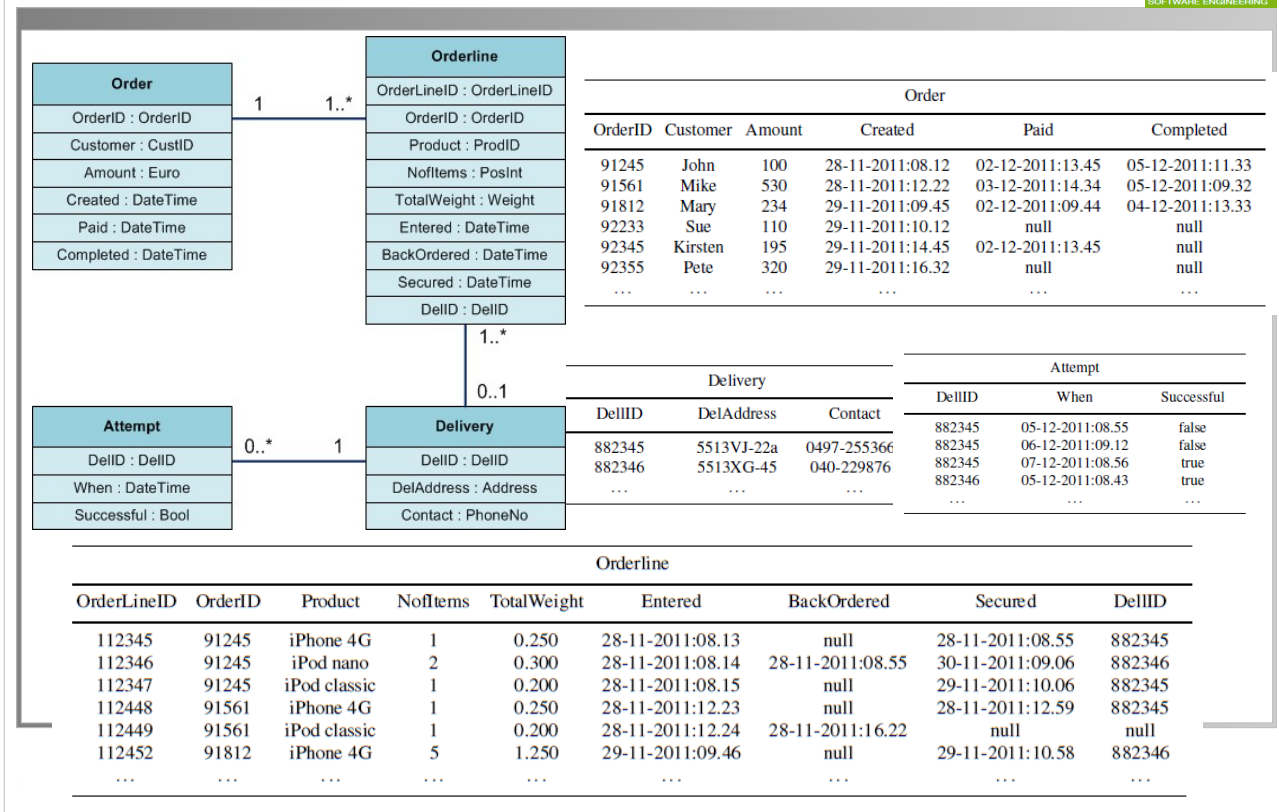
- **Korrelation:** Events in Event-Log: nach Fällen gruppiert.
  - Nicht trivial: setzt **Korrelation** der **Events untereinander** voraus.
- **Zeitstempel:** **Events** pro Fall **ordnen**.
  - Probleme: **nur Datum, unterschiedliche Uhren, verzögertes Loggen.**
- **Snapshots:** Fälle ggf. **über** Dauer der **Aufnahme hinweg** aktiv.  
Z.B.: Fall vor Beginn des Event-Logs gestartet.
- **Scoping:** Welche Tabellen berücksichtigen ?
- **Granularität:** Events in Event-Log: andere Granularität als für Endnutzer relevante Aktivitäten.

20

## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.3: S.113-114



## Literatur:

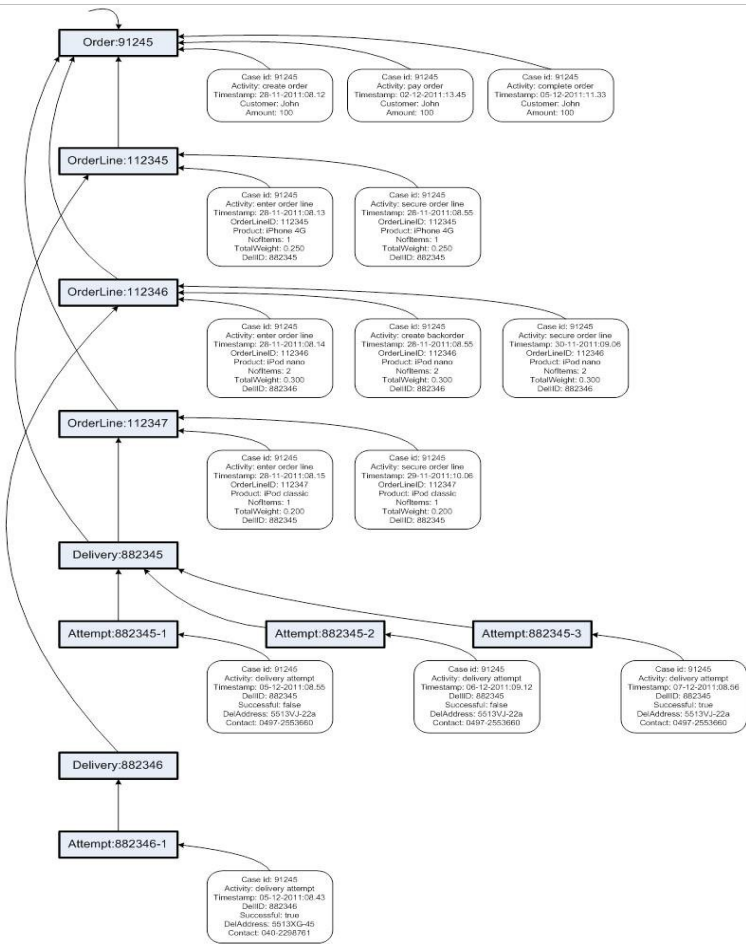
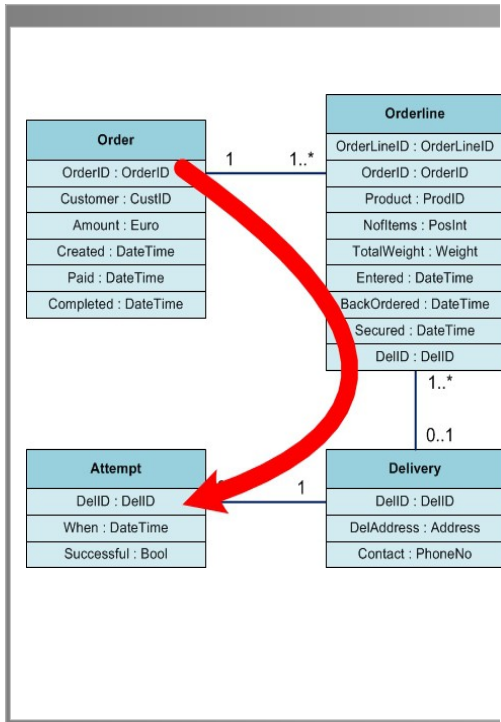
Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.4: S. 115-117, Fig. 4.9, Tabelle 4.2, 4.3, 4.4, 4.5



- **Problem:** Wir möchten aus den vorher gezeigten Tabellen ein Event-Log erstellen.
- d.h. wir möchten aus den vier Tabellen (Order, Orderline, Delivery, Attempt) eine „CaseID“ Tabelle ableiten.
- diese „CaseID“ Tabelle sollte dann Zeitstempel beinhalten, die dann auch als Events bzw. Aktivitäten in Prozess Diagrammen modelliert werden können
- vier Arten von Fällen möglich: Order, Orderline, Delivery, Attempt
- wir betrachten hier als Beispiel die **Bestellung** (Order), d.h. wir können an diesem Event-Log sehen, welche Events mit einer Bestellung in Verbindung stehen.

# Instanz: Bestellung

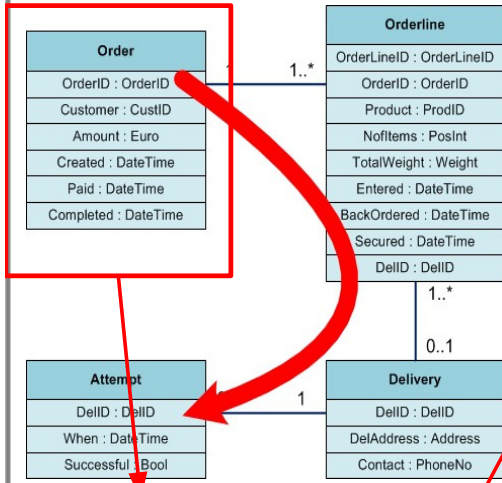


## Literatur:

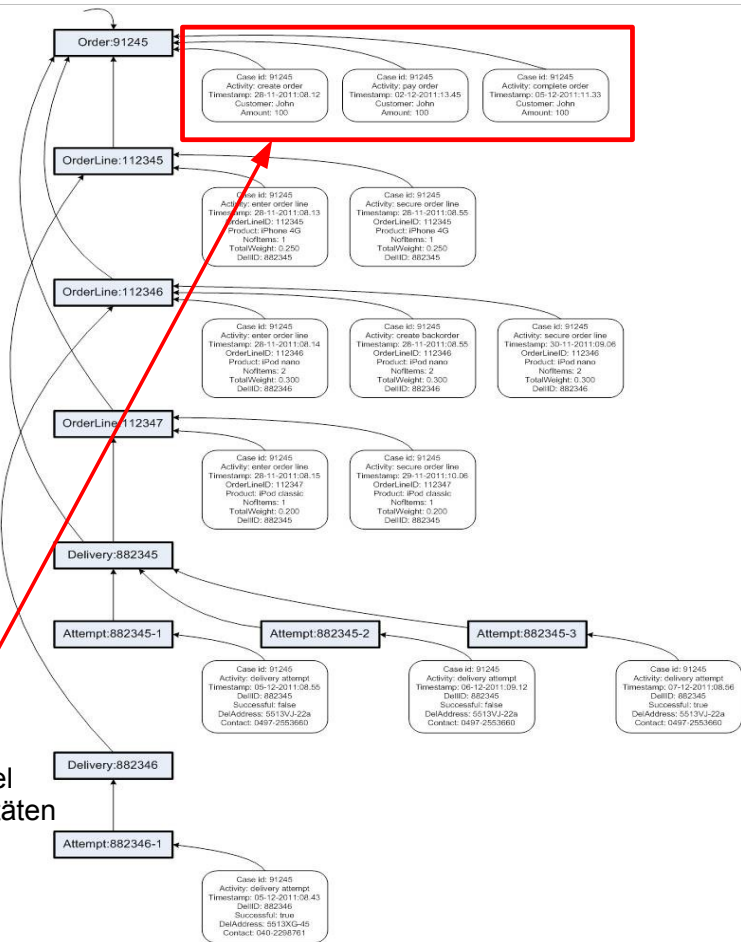
Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.4: S. 115 Fig. 4.9 & S. 119 Fig. 4.10

# Instanz: Bestellung



- Tabelle Order hat drei Zeitstempel
- Würden wir nur diese drei Zeitstempel betrachten, könnten wir nur drei Aktivitäten aus den Tabellen extrahieren.



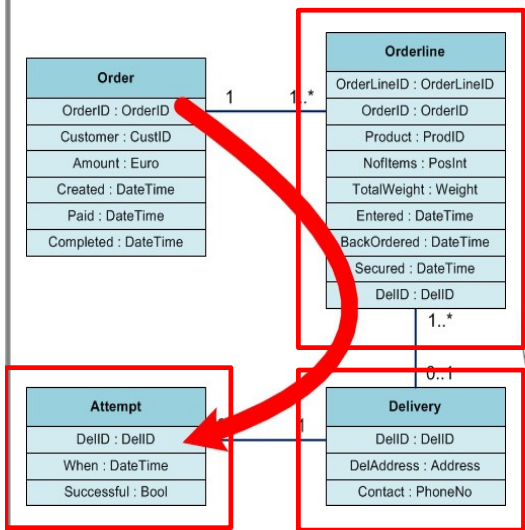
## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

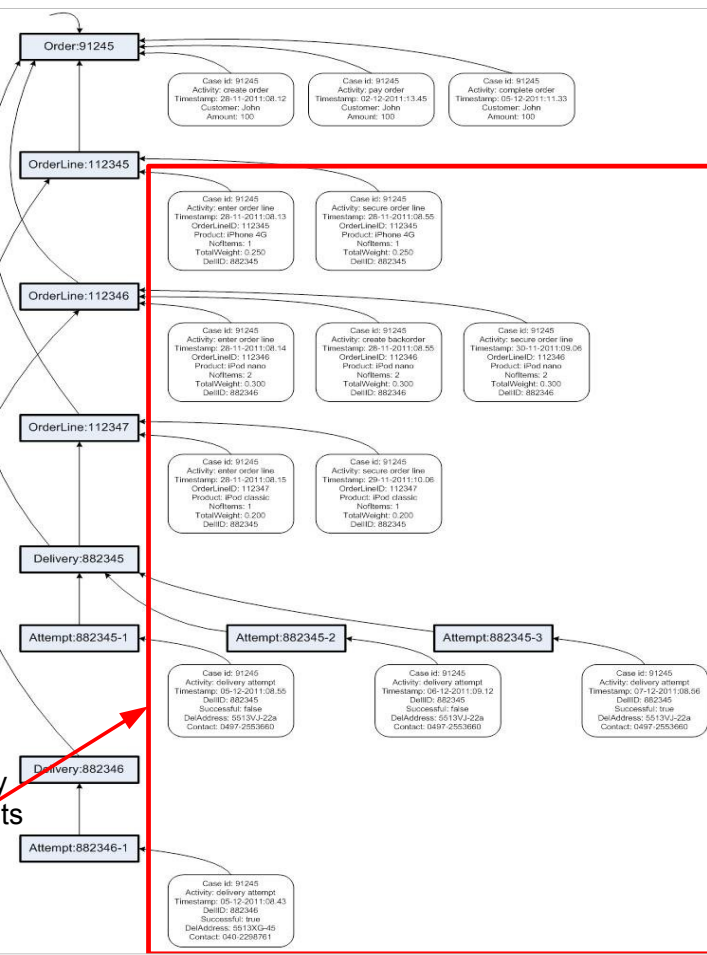
- Kap. 4.4: S. 115 Fig. 4.9 & S. 119 Fig. 4.10



# Instanz: Bestellung



- Verknüpfen wir die Tabellen (Order mit Orderline, Orderline mit Delivery, Delivery mit Attempt) so erhalten wir weitere Events in weiteren Ebenen

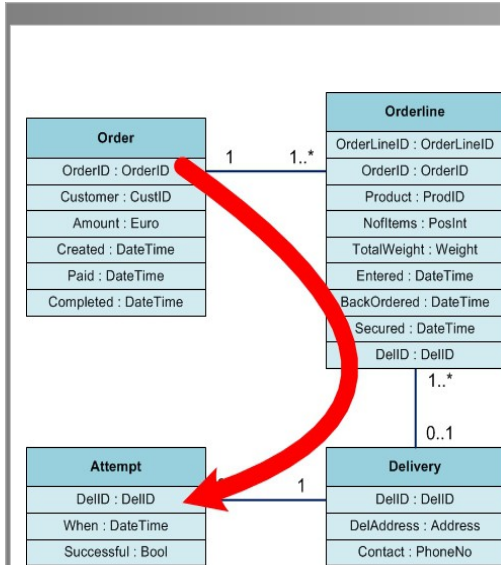


## Literatur:

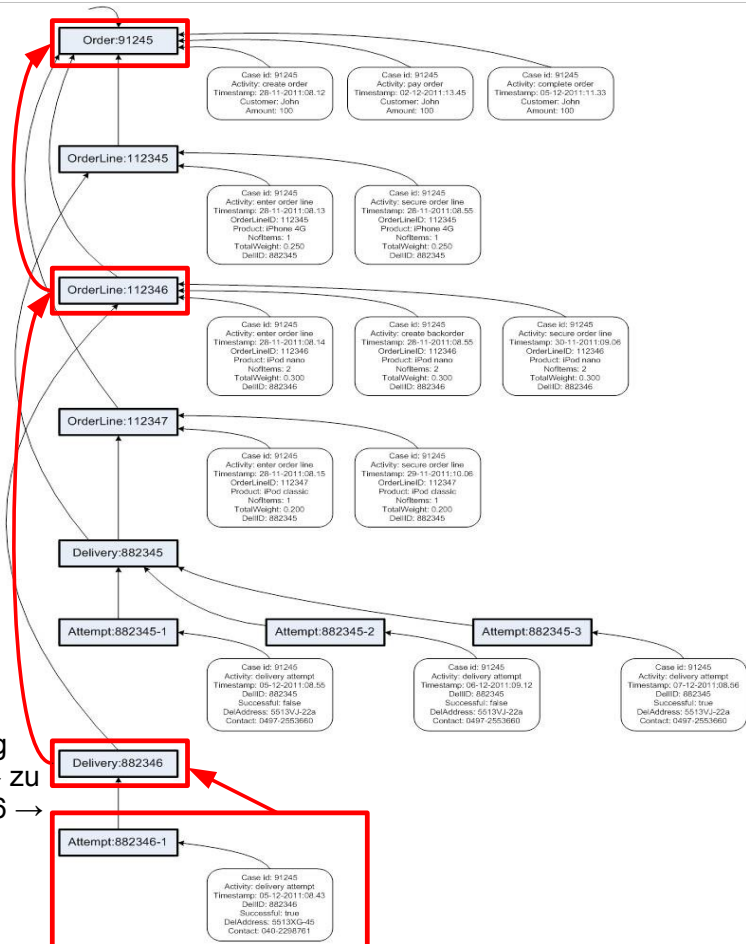
Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.4: S. 115 Fig. 4.9 & S. 119 Fig. 4.10

# Instanz: Bestellung



• Diese sind indirekt mit der Bestellung verbunden (z.B. Attempt: 882346-1 → zu Delivery: 882346 → Orderline: 112346 → Order: 91245)

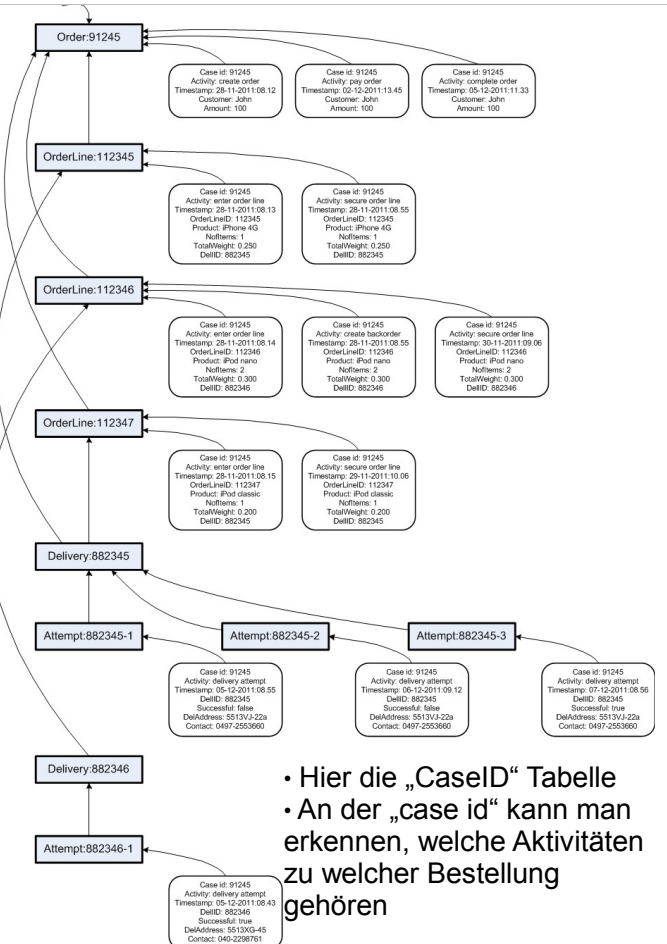


## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.4: S. 115 Fig. 4.9 & S. 119 Fig. 4.10

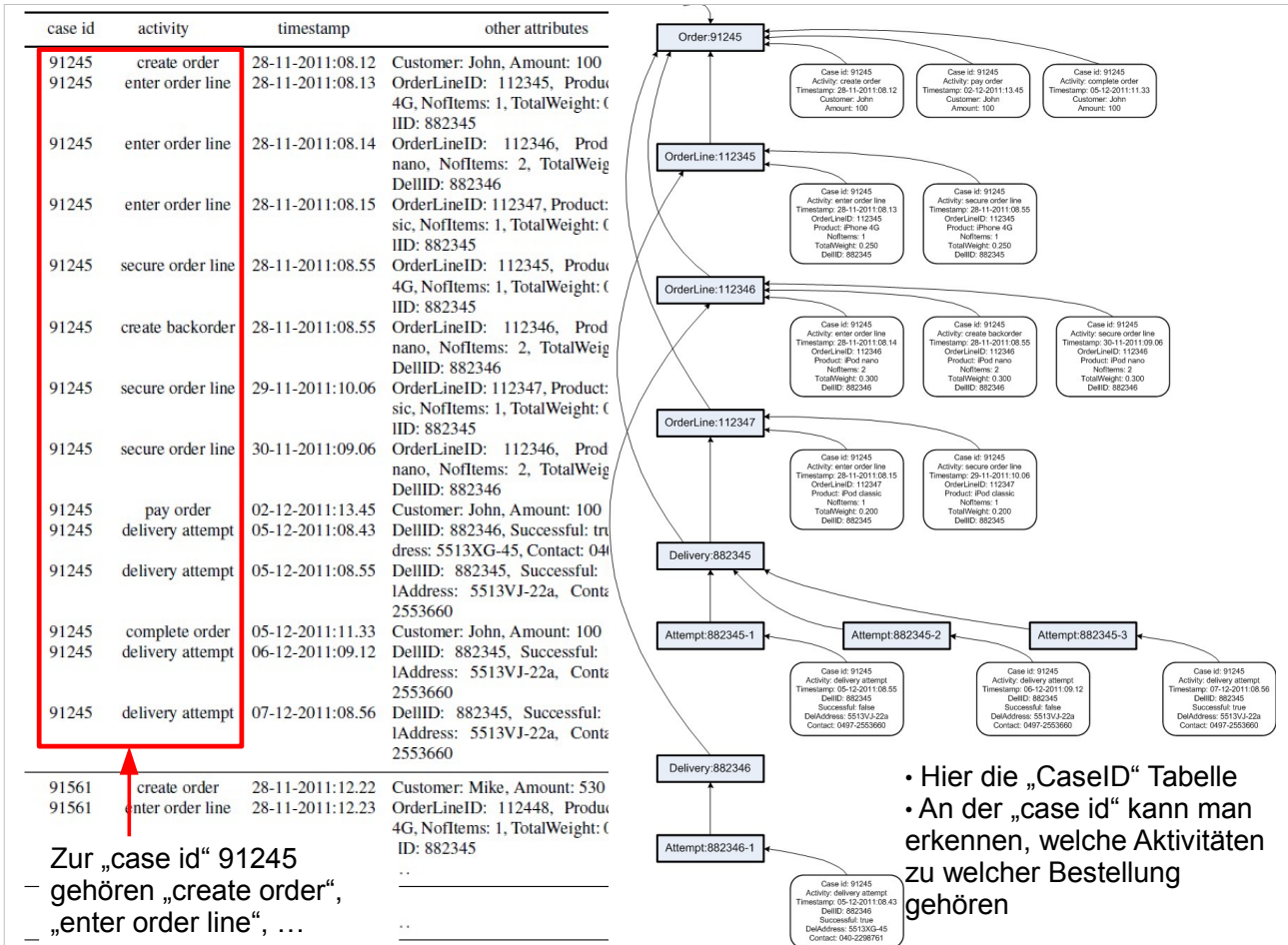
case id	activity	timestamp	other attributes
91245	create order	28-11-2011:08.12	Customer: John, Amount: 100
91245	enter order line	28-11-2011:08.13	OrderLineID: 112345, Product: 4G, NofItems: 1, TotalWeight: 0
91245	enter order line	28-11-2011:08.14	OrderLineID: 112346, Product: nano, NofItems: 2, TotalWeight: 0
91245	enter order line	28-11-2011:08.15	OrderLineID: 112347, Product: classic, NofItems: 1, TotalWeight: 0
91245	secure order line	28-11-2011:08.55	OrderLineID: 112345, Product: 4G, NofItems: 1, TotalWeight: 0
91245	create backorder	28-11-2011:08.55	OrderLineID: 112346, Product: nano, NofItems: 2, TotalWeight: 0
91245	secure order line	29-11-2011:10.06	OrderLineID: 112347, Product: classic, NofItems: 1, TotalWeight: 0
91245	secure order line	30-11-2011:09.06	OrderLineID: 112346, Product: nano, NofItems: 2, TotalWeight: 0
91245	pay order	02-12-2011:13.45	Customer: John, Amount: 100
91245	delivery attempt	05-12-2011:08.43	DellID: 882345, Successful: true
91245	delivery attempt	05-12-2011:08.55	DellID: 882345, Successful: false
91245	complete order	05-12-2011:11.33	Customer: John, Amount: 100
91245	delivery attempt	06-12-2011:09.12	DellID: 882345, Successful: true
91245	delivery attempt	07-12-2011:08.56	DellID: 882345, Successful: true
91561	create order	28-11-2011:12.22	Customer: Mike, Amount: 530
91561	enter order line	28-11-2011:12.23	OrderLineID: 112448, Product: 4G, NofItems: 1, TotalWeight: 0
...	...	...	...
...	...	...	...



## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.4: S.119 Fig. 4.10, S.120 Tabelle 4.6



## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

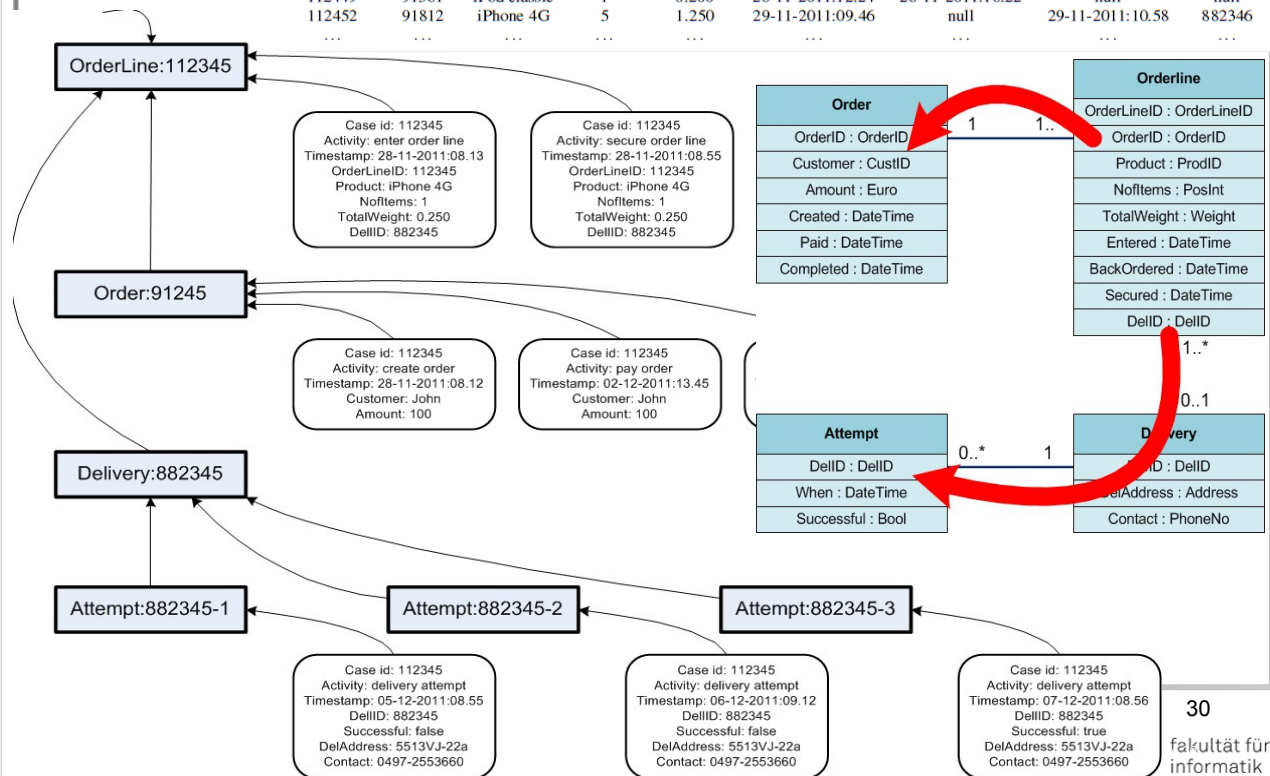
- Kap. 4.4: S.119 Fig. 4.10, S.120 Tabelle 4.6



- Alternativ hätte man auch ein Event-Log erstellen können, das sich an der „**Orderline**“ orientiert
- Der **Datensatz** ist derselbe, nur wird er anders dargestellt.
- d.h. es gibt mehrere „**Views**“ auf den selben Datensatz, ausgehend davon auf welche Tabelle man den Fokus setzt.

# Instanz: Orderline

Orderline									
OrderLineID	OrderID	Product	NoItems	TotalWeight	Entered	BackOrdered	Secured	DelIID	
112345	91245	iPhone 4G	1	0.250	28-11-2011:08.13	null	28-11-2011:08.55	882345	
112346	91245	iPod nano	2	0.300	28-11-2011:08.14	28-11-2011:08.55	30-11-2011:09.06	882346	
112347	91245	iPod classic	1	0.200	28-11-2011:08.15	null	29-11-2011:10.06	882345	
112448	91561	iPhone 4G	1	0.250	28-11-2011:12.23	null	28-11-2011:12.59	882345	
112449	91561	iPod classic	1	0.200	28-11-2011:12.24	28-11-2011:16.22	null	null	
112452	91812	iPhone 4G	5	1.250	29-11-2011:09.46	null	29-11-2011:10.58	882346	
...	...	...	...	...	...	...	...	...	



## Literatur:

Wil van der Aalst: Process Mining: Discovery, Conformance and Enhancement of Business Processes

- Kap. 4.4: S.121 Fig. 4.11, S.15 Fig. 4.9



- Nicht nur **syntaktisches Problem**.
- Verschiedene **Blickwinkel** möglich.
- Wichtig:
  - Richtigen Instanzbegriff auswählen.
  - Events ordnen.
  - Events auswählen.



## In diesem Abschnitt:

- Von heterogenen Datenquellen zu Process-Mining.
- Event Logs, verschiedene Ansichten.
- Überlappende Instanzen.
- Attribute.
- Datenspeicherformat XES.
- Herausforderungen beim Extrahieren.
- Event-Logs vs. Tabellen und Instanzen.

## Im nächsten Abschnitt:

- Prozessextraktion ( $\alpha$ -Algorithmus).