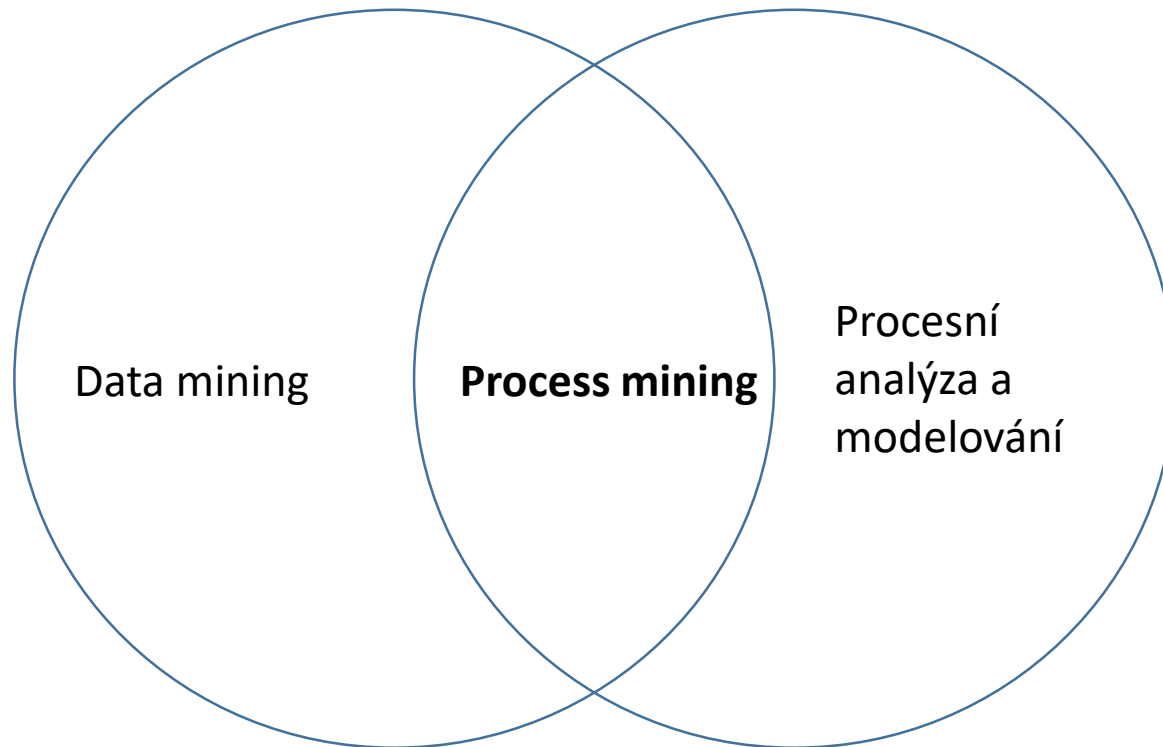


# Process mining

PV165 – Procesní řízení

# Vymezení oblasti



## Příbuzné koncepty

Business Activity Monitoring (BAM)

Complex Event Processing (CEP)

Corporate Performance Management (CPM)

┌──────────────────┐  
Business Intelligence

## Stejný cíl:

Analýza procesů za účelem jejich vylepšení  
a tedy i celkového zvýšení provozního výkonu  
firmy

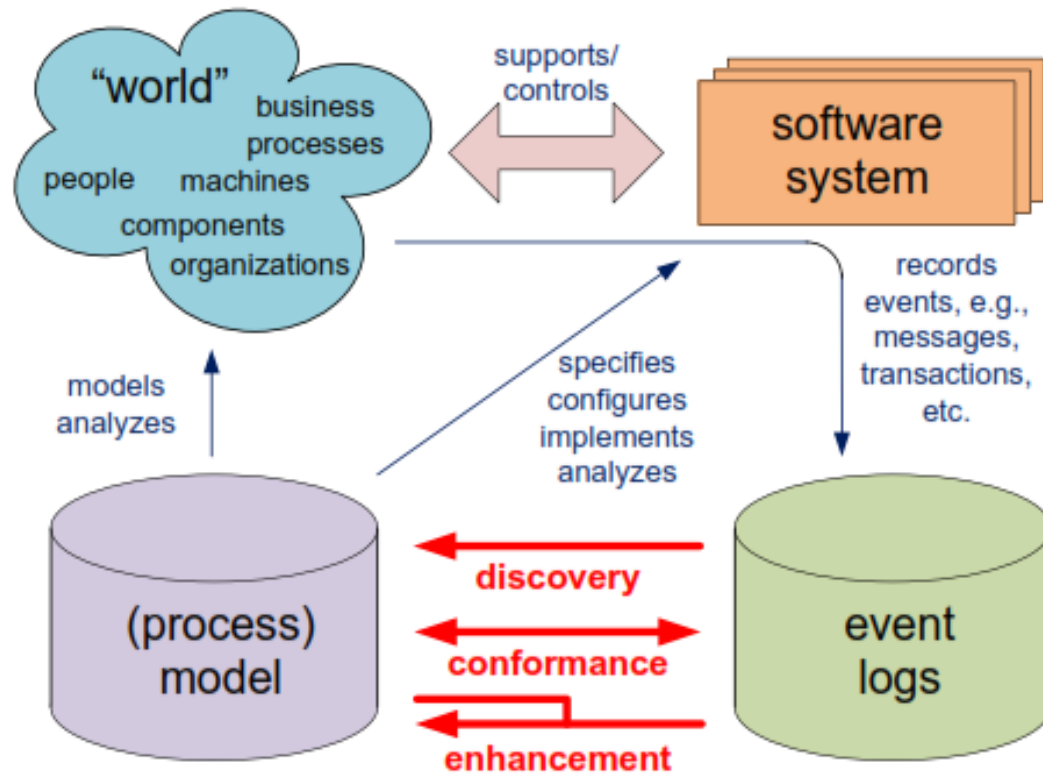
# Task Force on Process Mining

- Rostoucí zájem o dolování z procesů jak ze strany výzkumníků, tak ze strany průmyslu
- Roku 2009 bylo založeno IEEE Task Force on Process Mining, shromažďující SW firmy (HP, IBM, Fujitsu...), konzultační společnosti (Deloitte, Gartner, Excellentia BPM...) i výzkumné instituce
  - Cíle:
    - Zvyšovat povědomí o state-of-the-art process miningu
    - Navrhovat nové techniky, nástroje a hledat nová uplatnění
    - Standardizovat log event data
    - Pořadat workshopy, konference...
    - Publikovat články, knihy...

# Exponenciální růst dat

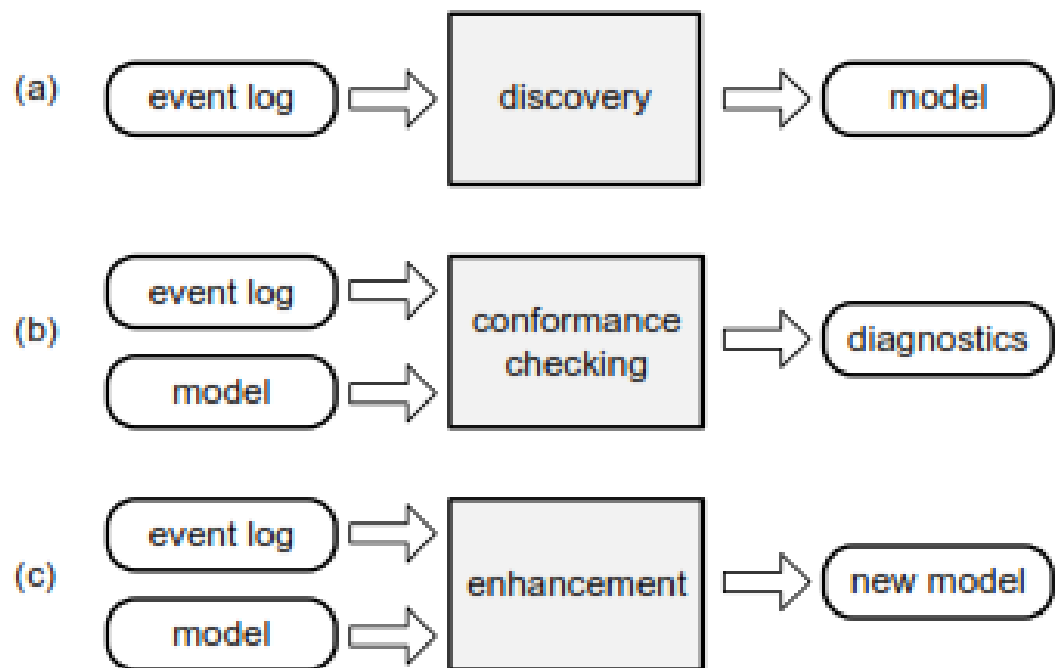
- Digitalizace zasahuje do všech aspektů života (viz Moorův zákon)
  - Informační a kolaborační systémy
  - Senzory
  - Čtečky kódů
  - Čipové karty...
- Automatizovaná analýza událostí
- Klíčový objekt – *event log*
  - Každá aktivita / instance procesu je sekvencí záznamů o událostech, osobách, časech, atd.

# Typy process miningu



- 1) Discovery**  
Identifikace modelu bez předchozí znalosti
- 2) Conformance**  
Porovnání existujícího modelu s tím, jak proces „reálně“ funguje
- 3) Enhancement**  
Rozšíření nebo vylepšení existujícího modelu na základě dat z event logů

# Typy process miningu



BPM, UML, Petriho síť...

Diagnostické informace o porovnání mezi modelem a údaji z event logů

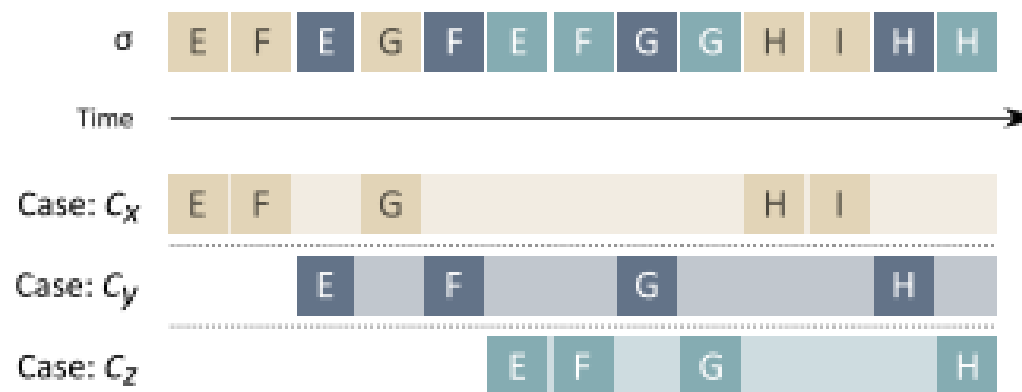
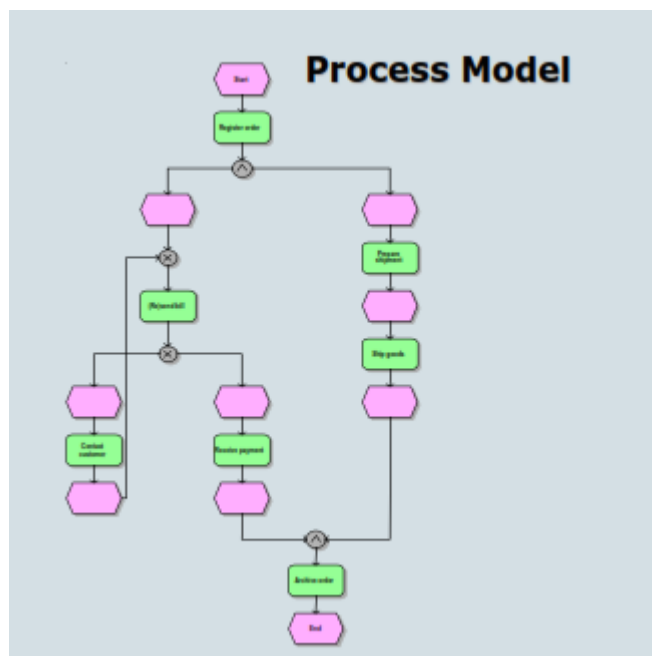
Vylepšený model v dané notaci

# Možné perspektivy process miningu

- Perspektiva procesní - kontrolní toky (Jak?)
- Organizační perspektiva (Kdo?)
- Případová perspektiva (Co?)
- Časová perspektiva

# Možné perspektivy process miningu

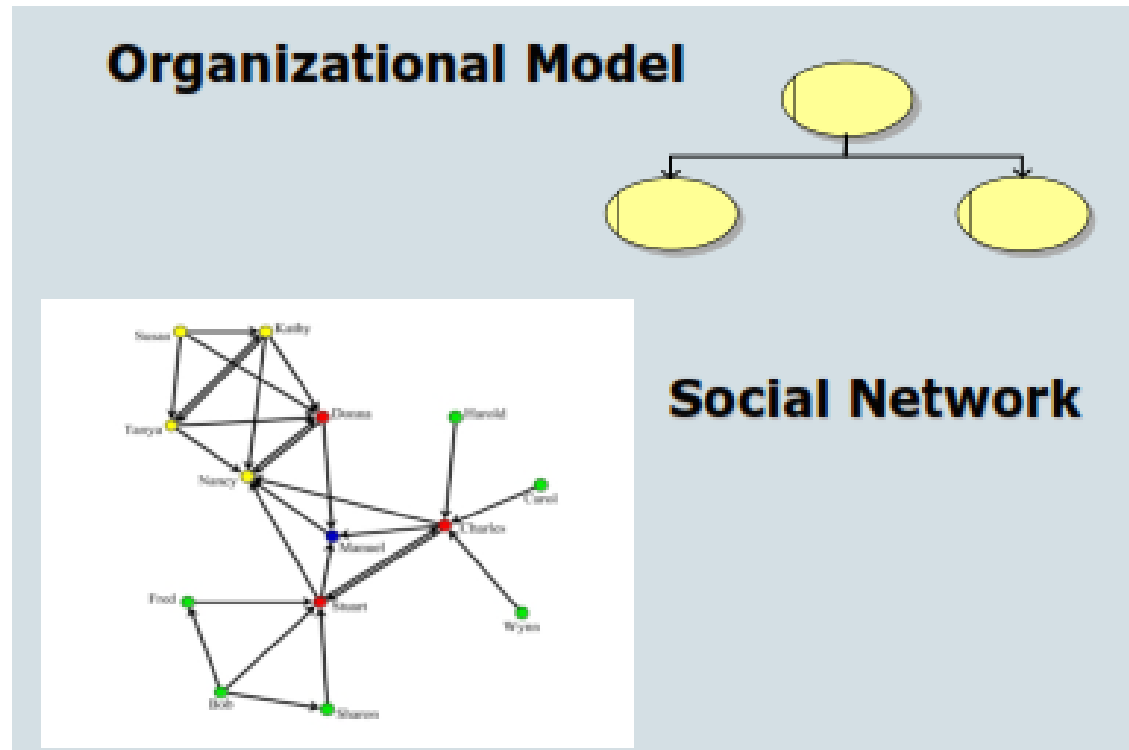
- Perspektiva procesní – aka perspektiva kontrolních toků (control-flows)
  - Detekce vazeb mezi jednotlivými aktivitami z event logů





# Možné perspektivy process miningu

- Organizační perspektiva
  - Validace organizační struktury
  - Detekce sociální sítě
    - Umožňuje např. vyhledávání osob podobným profilem práce



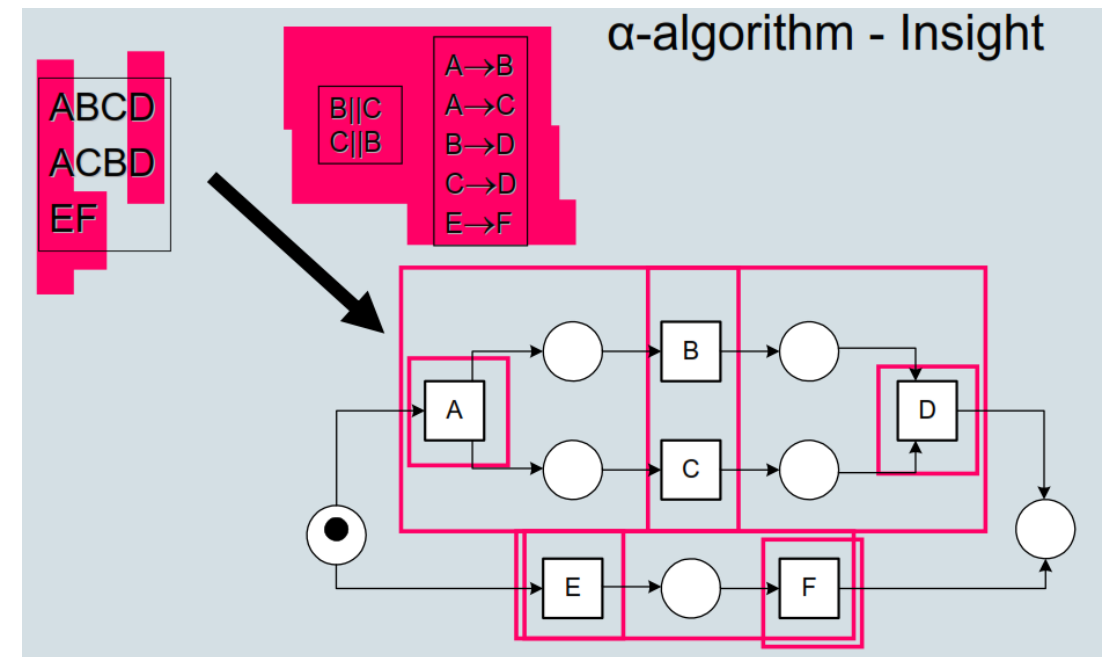
# Možné perspektivy process miningu

- Případová perspektiva
  - Zaměřuje se na konkrétní případy samotné
  - Obvykle je přínosná tehdy, když se v procesu děje něco „zajímavého“
  - Může studovat např. konkrétní cestu v procesu
  - Studuje rozdílné vlastnosti případu v závislosti na klíčových proměnných
    - Jak se mění cesta při velikosti objednávky, výši fakturované částky
    - Jaký vliv mají zdroje a cíle práce (role zadavatele – původce instance)

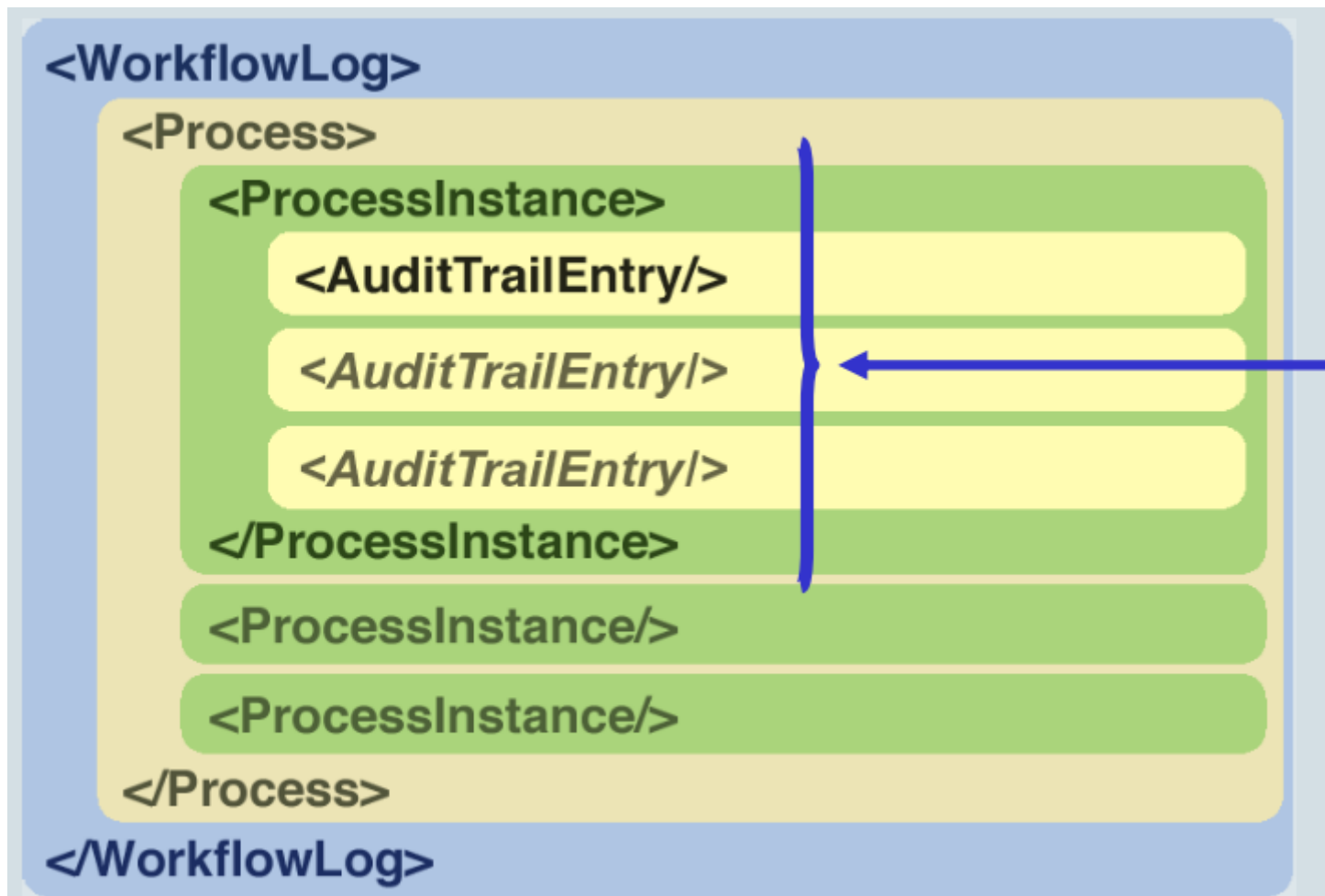


# Možné perspektivy process miningu

- Časová perspektiva
  - Zaměřuje se na časování a frekvenci událostí
  - Vyžaduje timestamps u událostí
  - umožňuje např.
    - Identifikovat bottlenecky
    - predikovat časový rozvoj procesu
    - Detekovat kauzalitu/paralelnost událostí
      - Alfa algoritmus
- “burstiness”



# Event log v XML formátu



Z hlediska process miningu je velmi důležité identifikovat, které aktivity spadají pod tutéž instanci procesu

# Event log v XML formátu

Povinné položky

```
<AuditTrailEntry>  
  <WorkflowModelElement/> Task A </Wf.M.E.>  
  <EventType> complete </EventType>  
  <TimeStamp> 2005-10-26T12:37:33... </TimeStamp>  
  <Originator> John Doe </Originator>  
  <Data>  
    <Attribute name="x"> 1 </Attribute>  
    <Attribute name="y"> whatever </Attribute>  
  </Data>  
</AuditTrailEntry>
```

```
<WorkflowLog>  
  <Process>  
    <ProcessInstance>  
      <AuditTrailEntry/>  
      <AuditTrailEntry/>  
      <AuditTrailEntry/>  
    </ProcessInstance>  
    <ProcessInstance/>  
    <ProcessInstance/>  
  </Process>  
</WorkflowLog>
```

# Kvalita event logů

*	Špatná kvalita, záznamy mohou chybět nebo nemusí odpovídat realitě	Ručně psané poznámky, papírová dokumentace
**	Záznam je vedlejší produkt IS, sběr probíhá automaticky, nicméně chybí systematický přístup, je možné IS obejít, tudíž mohou záznamy chybět	Strojové výpisy, tabulky, výstupy nebo chybové záznamy z (libovolných) systému
***	Chybí systematický přístup. Oproti ** je záruka, že záznamy odpovídají realitě. Data je třeba skládat z více zdrojů.	Tabulky, výstupy nebo dokumenty z ERP/CRM systémů, metadata z komunikačních kanálů
****	Systematický a reliabilní přístup – data jsou kompletní a odpovídají realitě. Existuje explicitní podpora pro analýzu procesů a aktivit	Logy z BPM nebo workflow systémů
*****	Sběr je systematický, reliabilní a adresovány jsou i otázky zabezpečení. Záznamy a jejich atributy mají jasnou sémantiku	Sémanticky anotované logy z BPM nebo workflow systémů

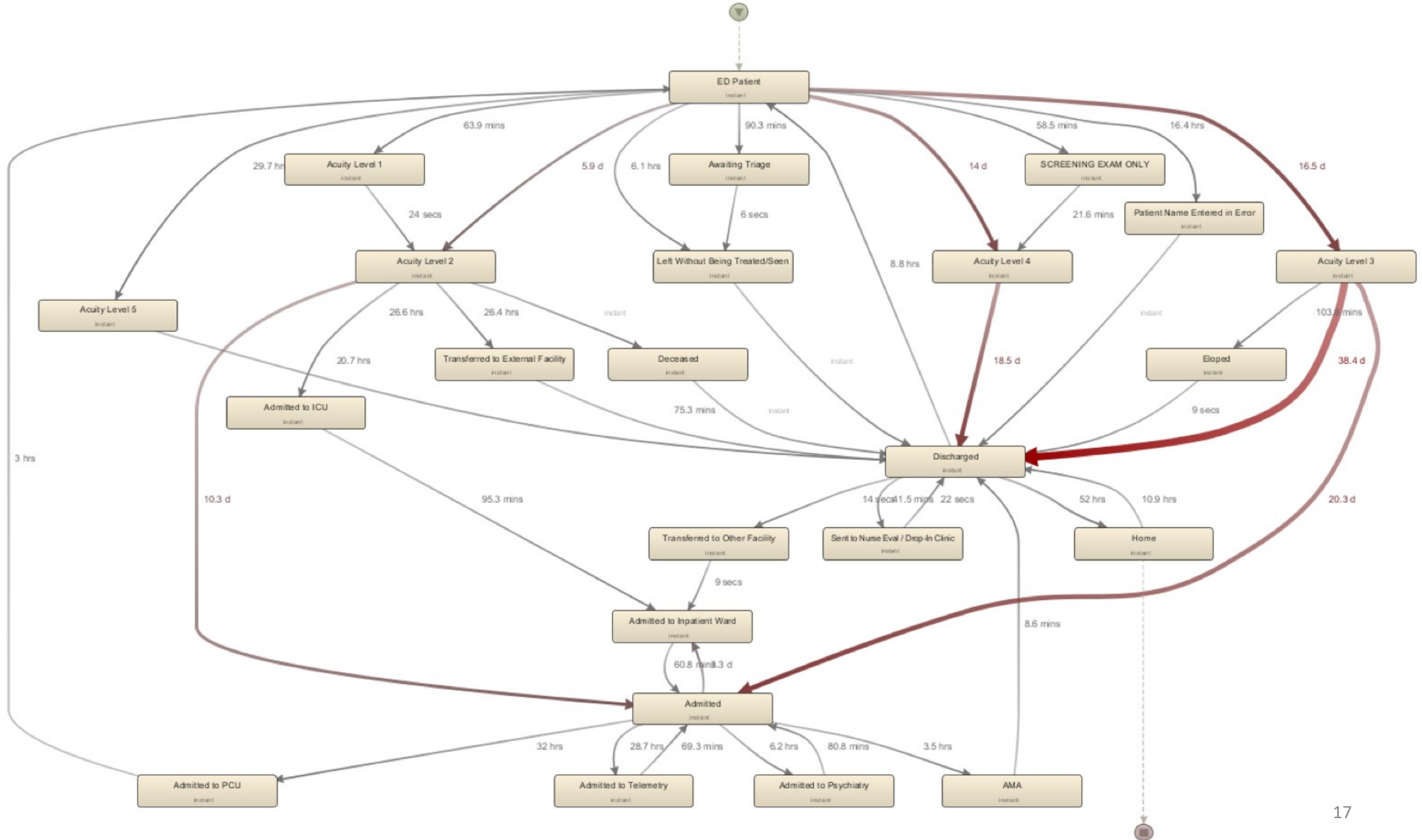
# Výzvy a možnosti

- Data: různé zdroje, nekompletní údaje, různé úrovně granularity, závislost událostí na kontextu
- Změna: „*concept drift*“ – proces se mění zatímco jej analyzujeme (sezónní změny, změny na trhu...)
- Kombinace s jinými typy analýz: např. modelování a simulace
- Vylepšení použitelnosti a porozumění dat laikům: vizualizace, intuitivní rozhraní

# Případová studie: zdravotnictví

- Data byly použity z e-healthcare systému (cca 5000 případů – ošetření pacienta)
- Kromě samotné analýzy dat byly provedeny také rozhovory s lékaři, řídicími lékaři, vrchními sestrami, atd.

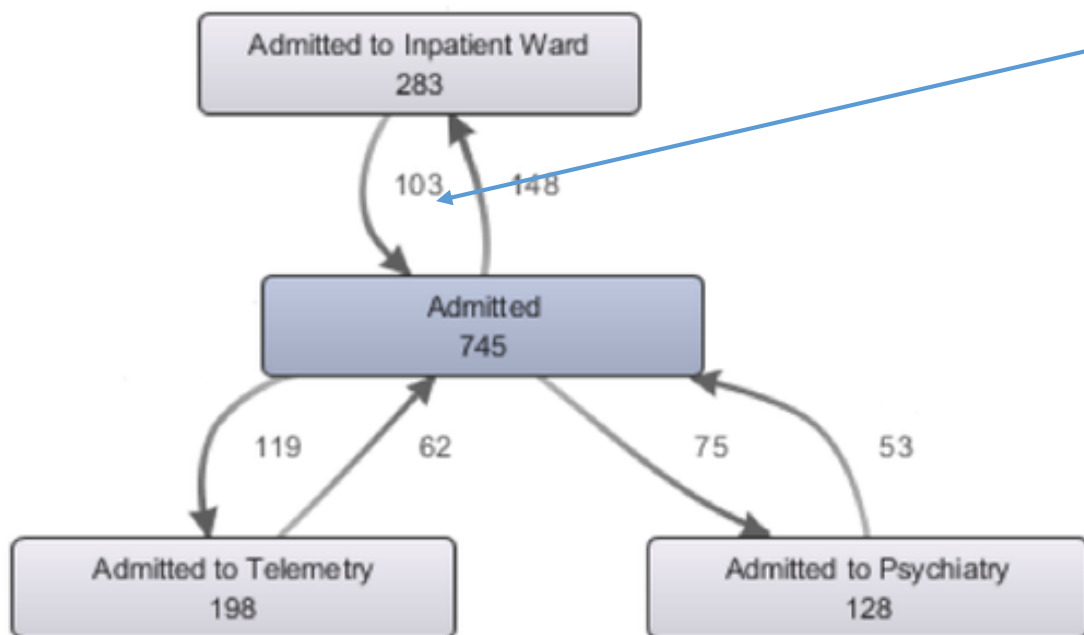




# Případová studie: zdravotnictví

- Kromě standardních procesů se v datech objevilo i několik případů, kdy průchod pacienta „procesem“ byl nestandardní
- Tyto nestandardní situace jsou možností pro zlepšení procesů a zvýšení kvality péče o pacienty

# Případová studie: zdravotnictví

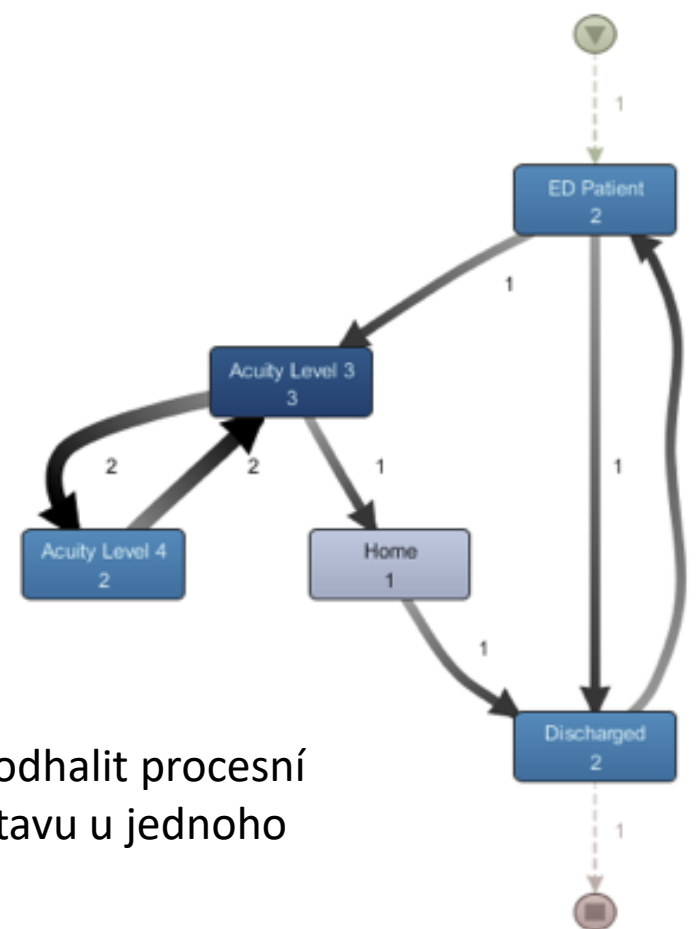


Odhalení cyklu v procesu, u něhož se očekává lineární anebo větvící se struktura

→ Buďto chyba personálu anebo špatně nastavený proces

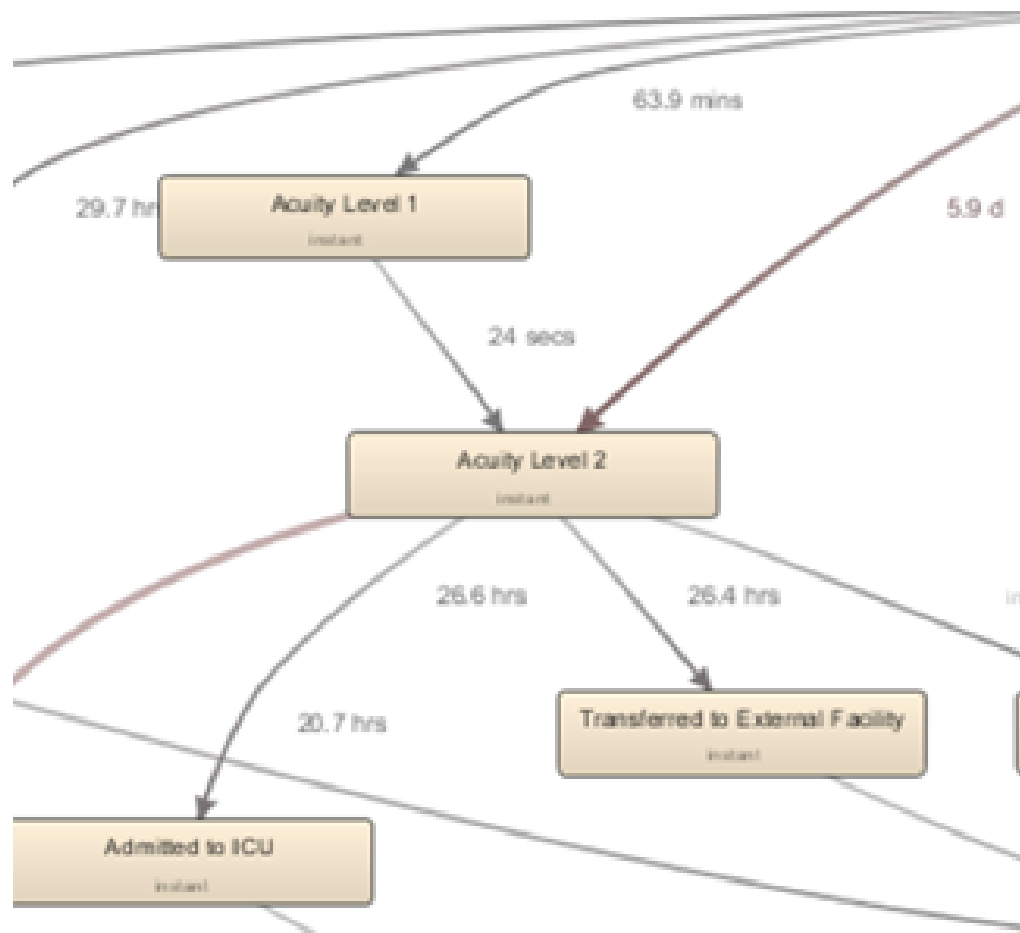
# Případová studie: zdravotnictví

U většiny pacientů dojde 4x ke změně stavu (přiját, propuštěn, úroveň akutnosti (1...n))



Analýza odchylek může odhalit procesní chyby – např. 10 změn stavu u jednoho pacienta

# Případová studie: zdravotnictví



Analýza časových sledů může odhalit situaci, kdy sekvence některých událostí jsou uloženy v systému v reálném čase – trvají řádově minuty nebo hodiny  
(pacient byl fyzicky přesunut z bodu A do B)

a jiné jsou výsledkem administrativního úkonu – trvají řádově sekundy  
(sestra uložila data do systému)  
Tyto události jsou patrně prováděny až ex-post, což může být prostor pro chybu