SIGNPOST- A Semiotic-Based Process Mining Methodology

Fahame Emamjome, Robert Andrews, Arthur ter Hofstede (QUT)

Hajo Reijers (Utrecht University)

Universiteit Utrecht

QUT

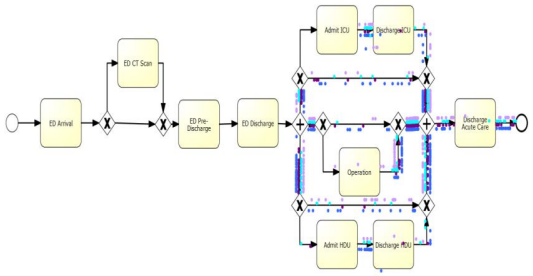
ECIS European Conference on Information Systems JUNE 2020 Marrakech

AIS 2020 Virtual Conference Series
ECIS | PACIS | AMCIS

Process Mining

A branch of data analytics aims to extract knowledge from process-related data, representing attributes of activities carried out during the execution of processes, to be able to discover, monitor or improve those processes.

Phase1: Define Research Questions



Phase 5: Results, insights, improvements

Case ID	Activity	Complete Timestamp
19:115933	Start	10/02/2010 13:45
19:115933	Ambulance transportation	10/02/2010 13:45
19:115933	ED Presentation	10/02/2010 13:55
19:115933	ED Presentation, Triage Category	10/02/2010 13:55
19:115933	Attended to in ED	10/02/2010 13:55
19:115933	ED P Diag: Chest Pain	10/02/2010 13:55
19:115933	Admitted to a Clinical Unit (inpat	10/02/2010 17:44
19:115933	Admitted to Medical	10/02/2010 17:44

Phase 2: Data collection
Phase 3: Data cleaning

Phase 4: Mining and analysis

Problematization

- **Death of Theory**
  - Identifying the actual problems
  - Street light effects
  - The potential of process mining to generalize findings and contribute to theoretical knowledge

- **Epistemic fallacy**
  - Recorded event data is not an accurate representation of reality
  - Collected data is only representing some facets of reality.
  - Cleaned data is not necessarily faithful to reality

"Scientists no longer have to make educated guesses… Instead, they can mine the complete set of data for patterns that reveal effects, producing scientific contributions..." (Prensky, 2009)

"unreproducible random noise" (sturbuck, 2016)

"the 'science' in 'data science' is on average extremely low to non-existent" (Damhof, 2017)

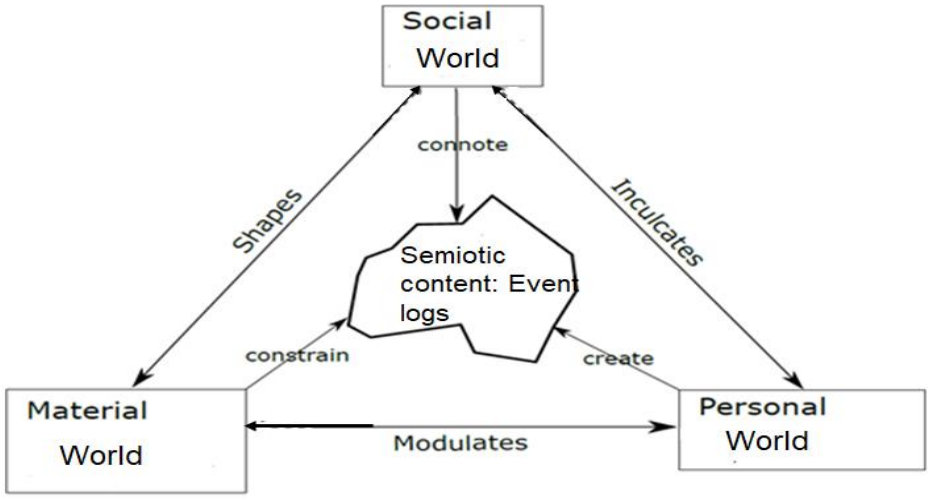
A Semiotic-Abductive Approach: SIGNPOST

Death of Theory → **Abductive Reasoning**

Theories will be used in **formulation of research questions/hypothesis and propositions** as well as **analysis**

+

Epistemic Fallacy → **Semiotic**



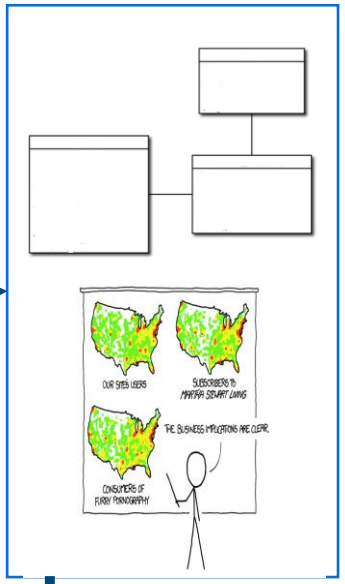
Adapted from Mingers and Willcocks 2014

SIGNPOST Methodology

SIGNPOST Methodology

- **Re-formulate** stakeholders questions
- Develop hypothesis and propositions

- **Collect data** about process participants, organization context and technologies in use (semiotic elements)
- Develop conceptual data models based on available event data and related meta-data

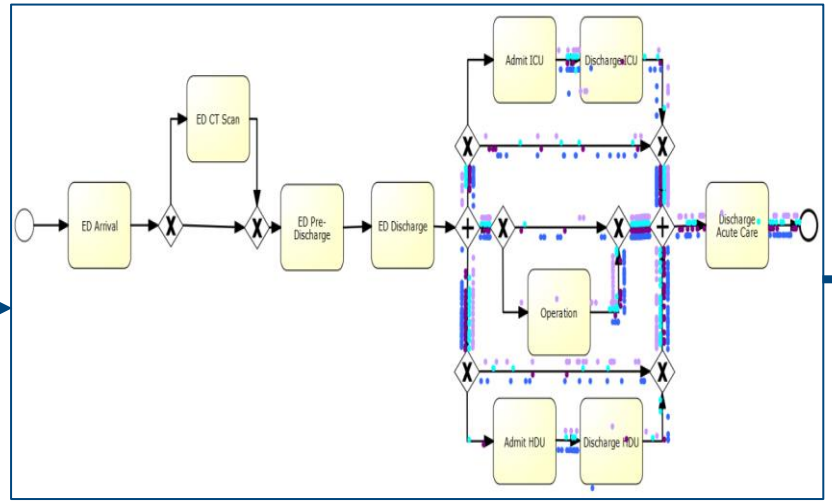


- **Clean data** using two approaches:
  - Prognosis
  - Diagnosis

Case ID	Activity	Complete Timestamp
19:115933	Start	10/02/2010 13:45
19:115933	Ambulance transportation	10/02/2010 13:45
19:115933	ED Presentation	10/02/2010 13:55
19:115933	ED Presentation, Triage Category	10/02/2010 13:55
19:115933	Attended to in ED	10/02/2010 13:55
19:115933	ED P Diag: Chest Pain	10/02/2010 13:55
19:115933	Admitted to a Clinical Unit (inpat	10/02/2010 17:44
19:115933	Admitted to Medical	10/02/2010 17:44

Process Mining

- Automated process discovery
- Conformance checking
- Performance analysis



Analysis and Evaluation:

- Investigate the research questions/proposition/hypothesis
- Position findings in relation to the existing theories and domain knowledge

Conclusion

- We identified two problems in existing process mining methodologies:
  - **Lack of attention to existing theories and potential of the field in contribution to theory building**
  - **Epistemic fallacy**
- We proposed an initial high level methodology to address these problems based on principles of abduction and semiotic

Future Work

- Evaluation and further development of the guidelines through application in real world process mining case studies.

- Further investigation of the potential of process mining to build theories and answer research questions in IS domain.

Thank you

f.emamjome@qut.edu.au