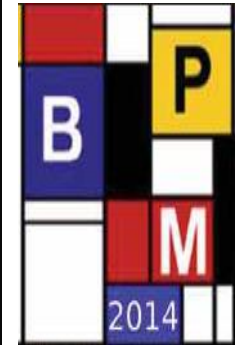


Detecting, Assessing and Mitigating Data inaccuracy-related Risks in Business Processes



Arava Tsoury

atsoury@is.haifa.ac.il

Advisors: Dr. Pnina Soffer

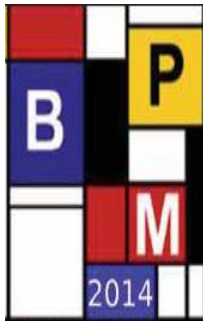
Dr. Iris Reinhartz - Berger

Department of Information Systems

University of Haifa, Israel

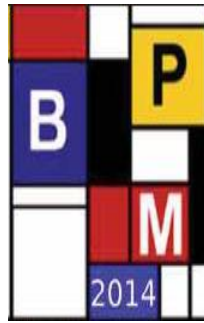


Agenda

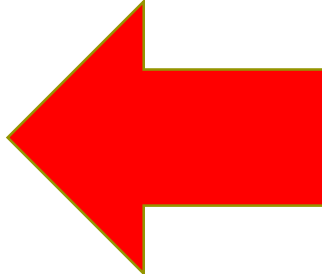


- *Motivation & Background*
- *The State of the Art*
- *Research Questions*
- *Research Method*
- *Expected Contribution*
- *Current Status*
- *Discussion*

Background & Motivation



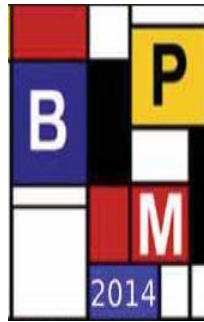
- The management of business processes is commonly supported by process aware information systems storing, using and manipulating data
- Inappropriate process design can lead to negative impacts on organizations
- The accuracy of data has an important role and much influence on the business process
- Inaccurate data can disrupt
 - tasks
 - goals
 - the process flow itself
 - data related elements
 - the outcome of the process



reduce the efficiency and the effectiveness of decision-making that depends on these activities and the related outcomes

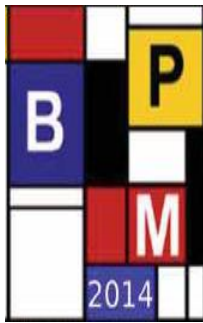
Background & Motivation

- Data inaccuracy is an instance of mismatch between some value or fact in the real world and its representation in the information system

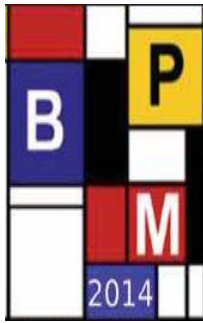


Background & Motivation

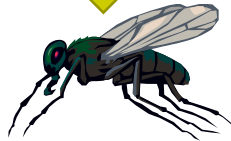
This Is Nancy



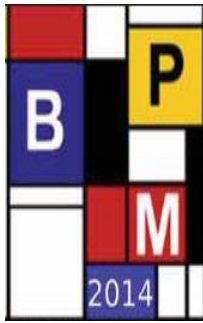
Background & Motivation



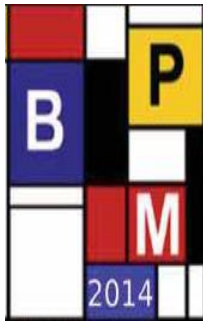
This Is Bob



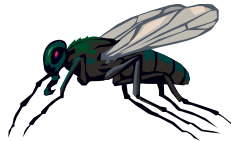
Background & Motivation



Background & Motivation

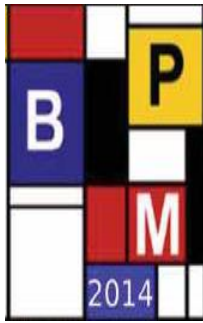


**BZZZZ....
BZZZZZ**



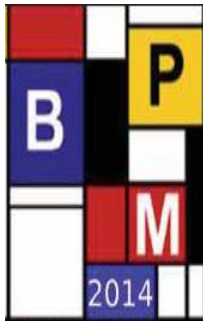
Lets assume that Nancy types in 3
units instead of 30 units....

Background & Motivation



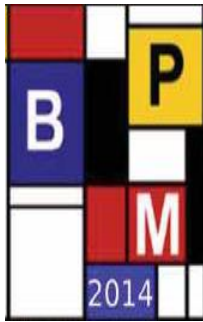
Each time that Daniel approve order request, first, he checks whether there is enough resources, and this is based on previous order requests details...

Background & Motivation



- Daniel has approved Nancy's Order request
- He knows that there is a need for only 3 units
- After that, based on this value, he approves (or not) other new requests as well...

Background & Motivation

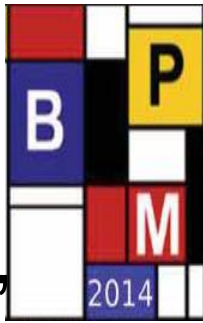


Oh, what have I
done??
It is 30 units...not 3...

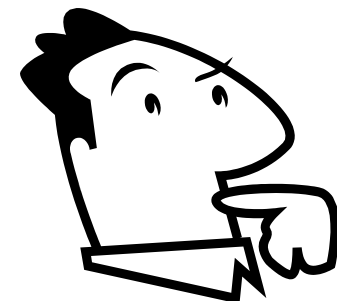
I just need to
make a small
correction



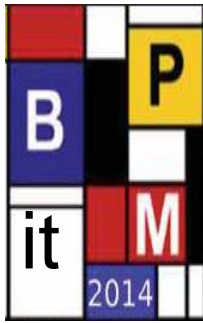
Background & Motivation



- When data inaccuracy is discovered (or not), the consequences might depend on the time of discovery
- This may occur after many actions have been performed based on the incorrect value.
 - Daniel has already approved many order requests
 - After the correction, there is not enough resources...
 - Other scenarios also possible...



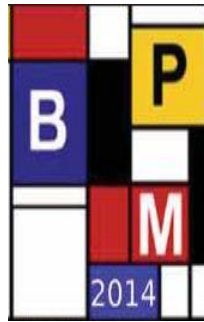
Background & Motivation



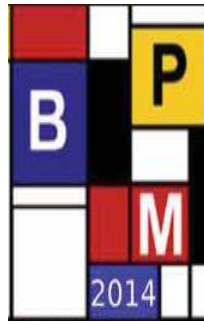
- Discovery time has great importance since it can influence the severity of the risk
 - the amount of harm that can be expected due the use of inaccurate data.
- Risk assessments helps
 - identify the possible causes of harm
 - ensure that potential problems are well understood
 - understand the severity level of the risk
 - determine the necessary actions that are need to be taken in order to reduce risks to a reasonable level.

Background & Motivation

- Metrics for assessing the severity of data items and the robustness of the process design may support decision regarding process redesign while also considering a cost benefit analysis of potential solutions.

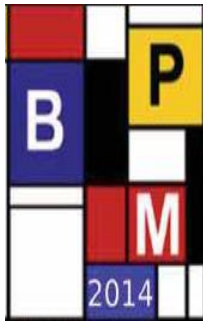


Background & Motivation

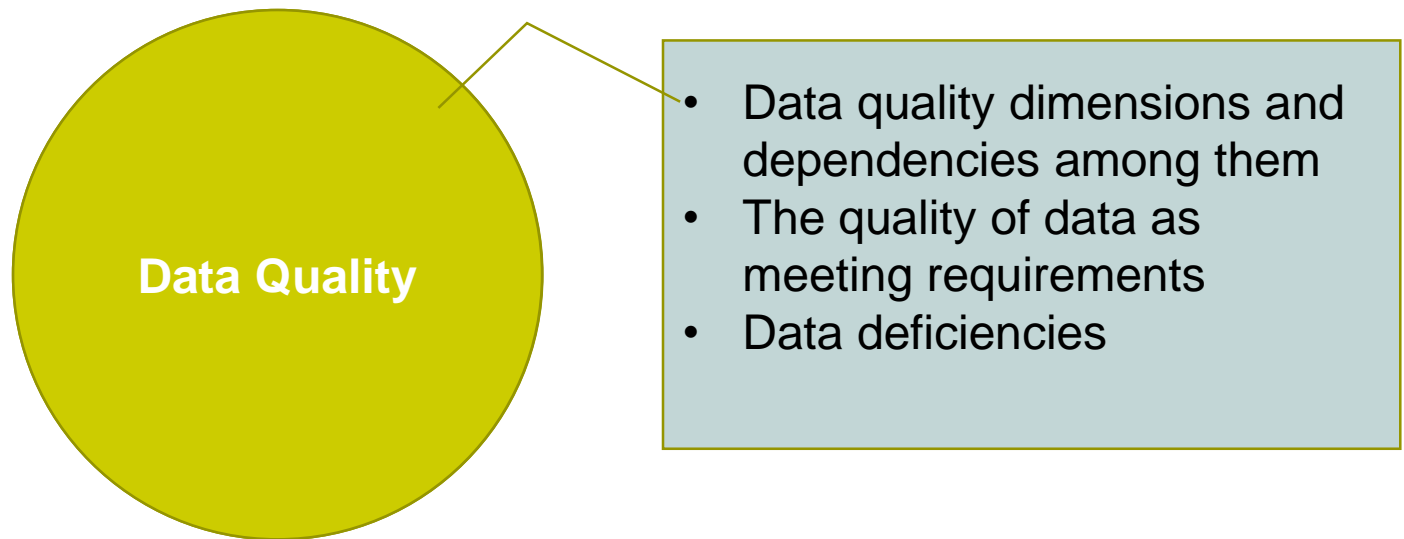


- Fixing the inaccurate data when discovered is not always useful
 - The discovery time may be too late and the correction may be impossible or irrelevant at that point
 - Daniel already approved other requests and exceeded the limit
 - In case Nancy has not discovered her mistake, only 3 units have been delivered to the customer instead of 30
- We concentrate on preventing such situations in the future by suggesting process redesign. for example:
 - Add a control steps before using the value
 - Add a manual actions for verification of the value
 - Define value boundaries based on previous data and cases

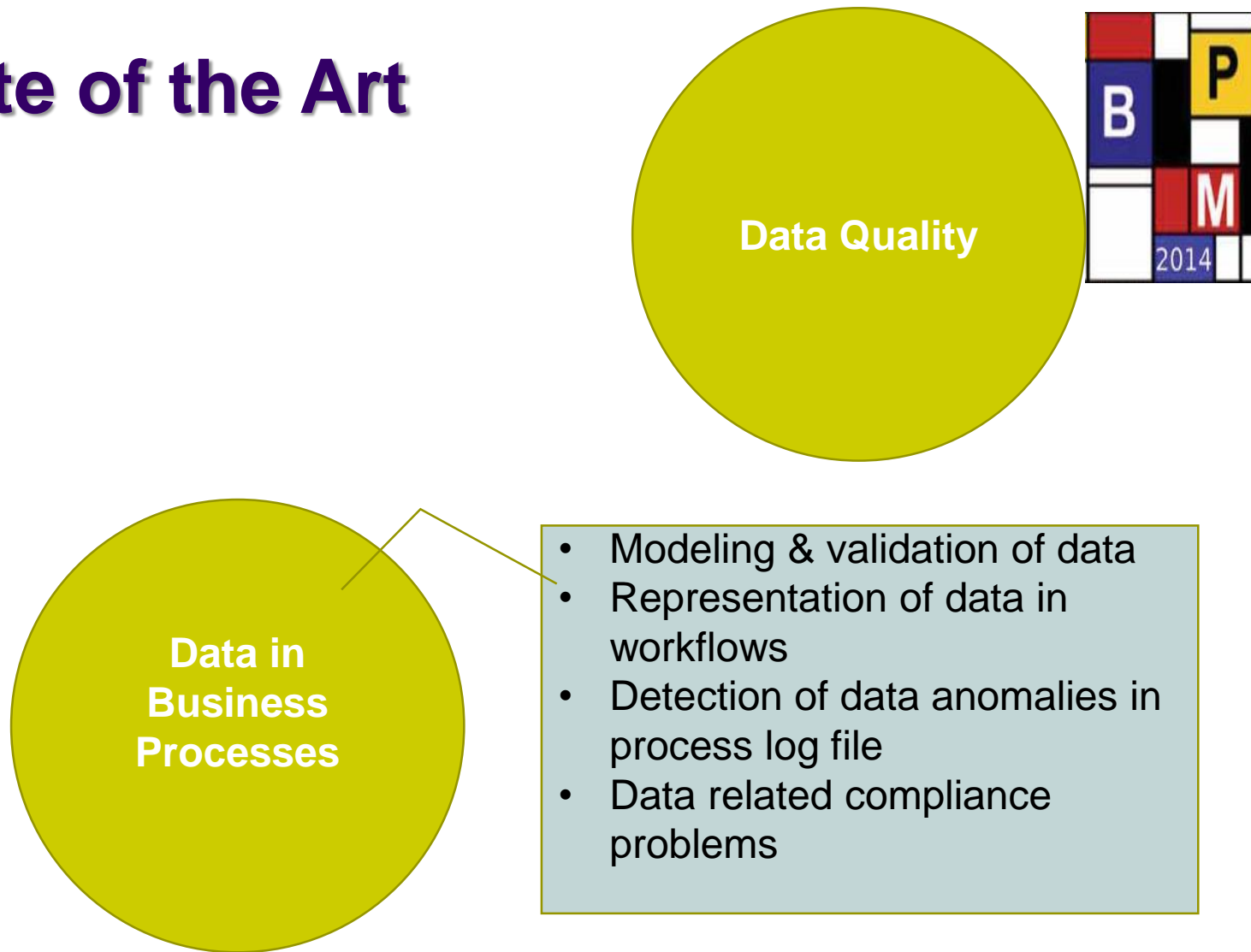
The State of the Art



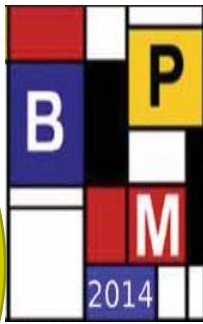
- Dozens of studies were made in the context of data quality in general and data quality in business processes in particular.
- Their purpose is mainly to assure high quality data in information systems and in particular in business processes



The State of the Art



The State of the Art



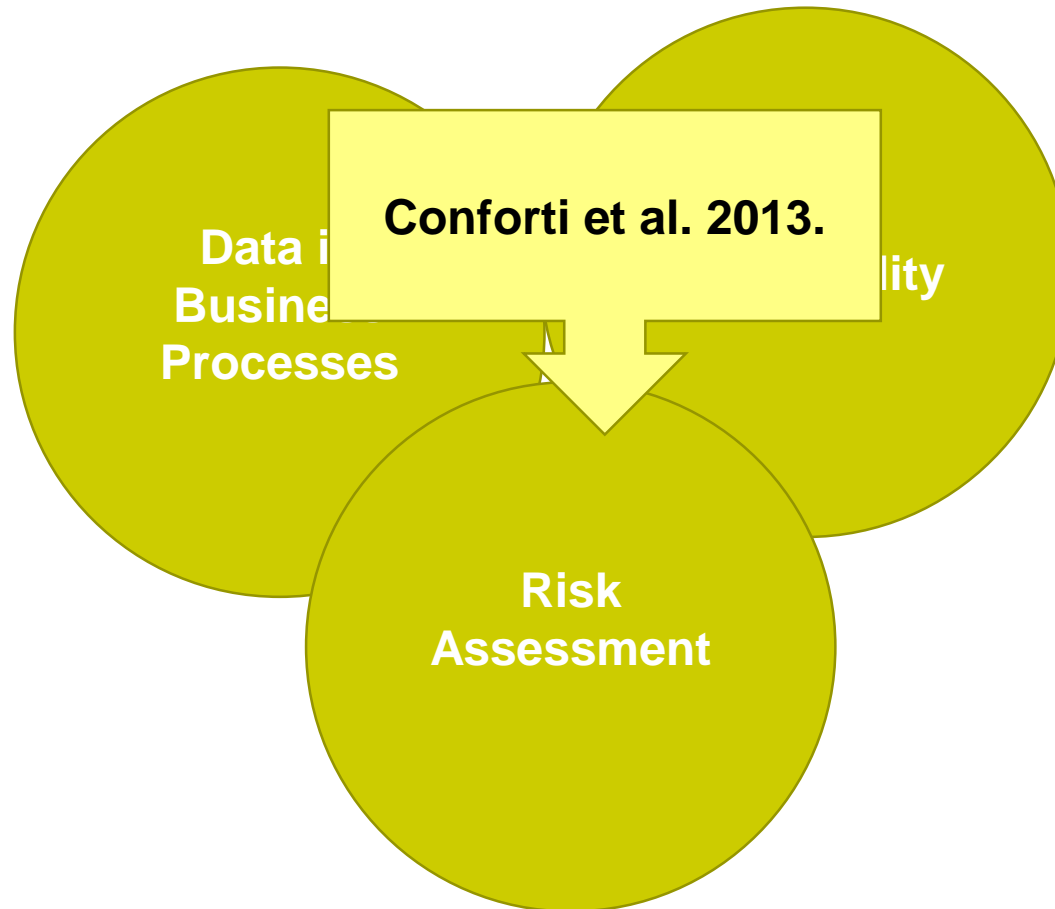
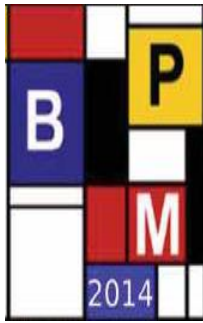
**Data in
Business
Processes**

Data Quality

**Risk
Assessment**

- Managing & analyzing risks in business processes
- Establishing risk mitigation strategies into a business process model
- Identifying risks from logs and other post-execution artifacts

The State of the Art



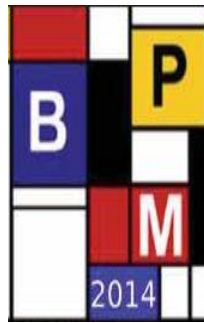
The State of The Art



- Conforti et al. propose a technique that aims at providing risk prediction on run-time which also consider data perspective
 - The Technique combine information on known risks and faults with historical data of the process (from the process log)
 - The Technique support decision making based on probabilities for process's specific risks
-
- Risks are not prevented
 - Decision regarding instance of process needs to be taken when the values likelihood exceeds a tolerance threshold

Research Questions

How to detect, assess and mitigate data inaccuracy risks in business processes?



Detection of
inaccurate data
elements



Assessment of data
inaccuracy-related
risks



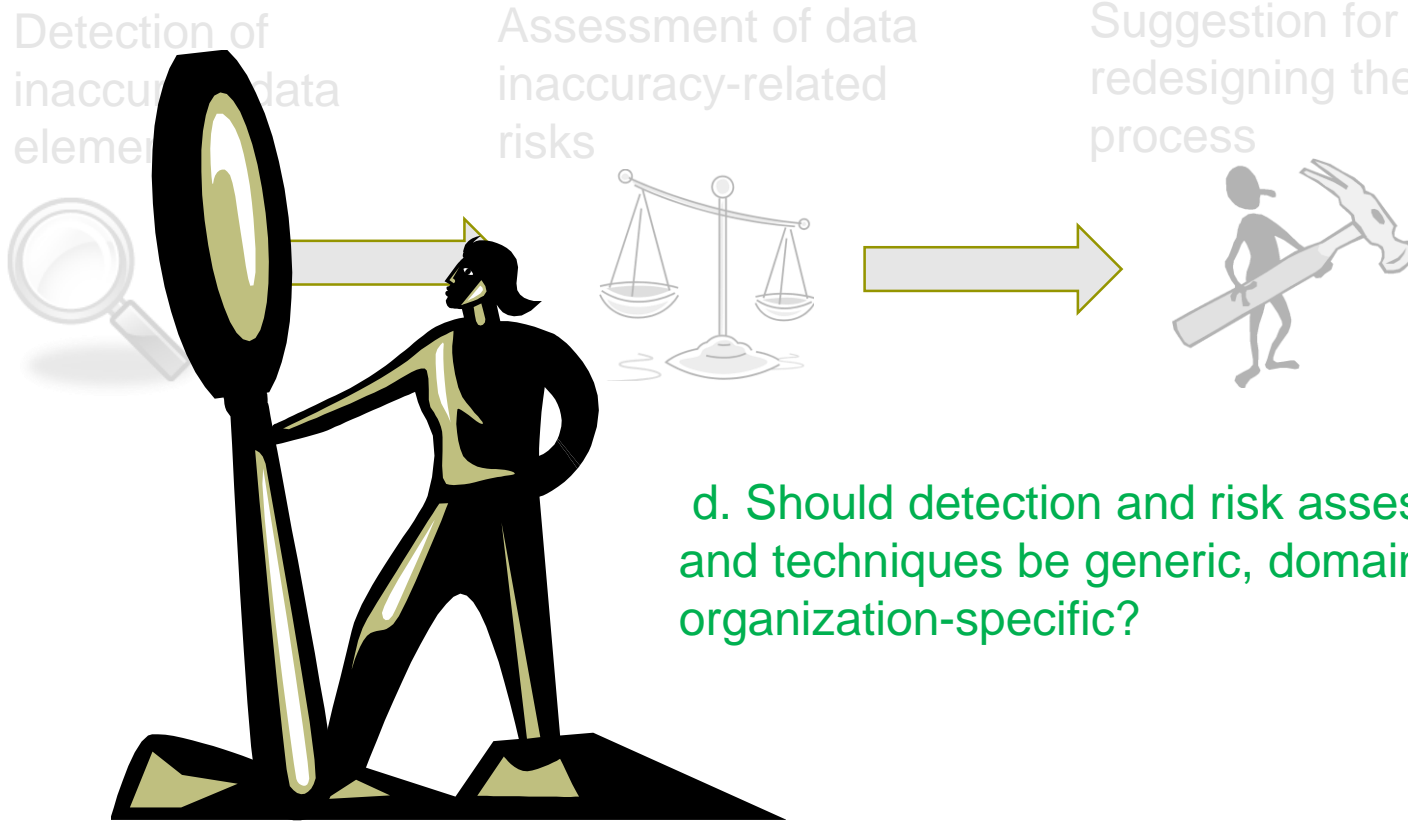
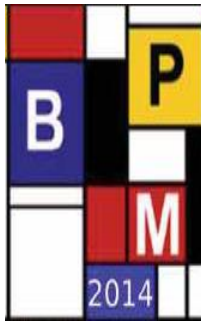
Suggestion for
redesigning the
process



- a. What are the steps and actions that are required in order to detect (systematically and by using maximum automation) data inaccuracy in business processes retroactively (i.e., not in real time)?
- b. What are the risks that are associated with data inaccuracy elements on business processes and how to assess them?
- c. How to reduce data inaccuracy risks by redesigning the business process?

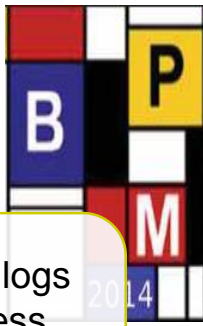
Research Questions

How to detect, assess and mitigate data inaccuracy risks in business processes?



d. Should detection and risk assessment patterns and techniques be generic, domain-specific, or organization-specific?

Research Methodology



Explore

- Explore patterns and impacts that may reflect data inaccuracy situations in event logs
- Check and confirm with domain experts or employees which involved in the process
- Explore the impacts on the process in terms of risks: what are the impacts of these items on the process, its outcomes and the impacts on the organization.

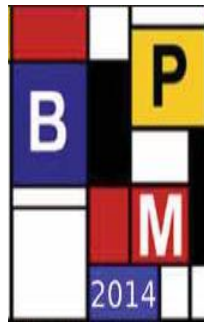
Build

- A method, which will include
 - Patterns of data inaccuracies
 - Patterns of risks
 - Guidelines for using these patterns
- Simulation tool that will support decision making in the context of the process redesign

Evaluate

- Evaluate the method by referring to different criteria such as performance, efforts, usefulness and usability
- The method will be evaluated by reapplying it to the initial inputs of the case studies and by simulations
- Final evaluation of the method will be made through an extensive new case study in an organization

The Envisioned Method



Method Users



Process experts



Consultants



Employees

Process assets



Process event logs



Database



Process documentation



Process models

Method

Detect inaccurate data elements



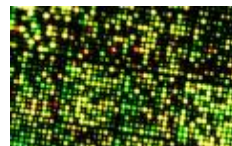
Assesses Risks



Suggestions for process redesign



Method assets



Inaccurate data detection patterns



Tools



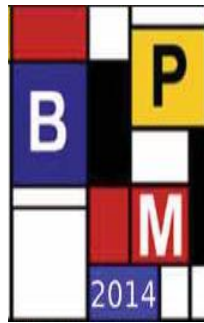
Risk patterns



Simulation tool

The Envisioned Method

Something is not right... Lets check the process design...



Method Users

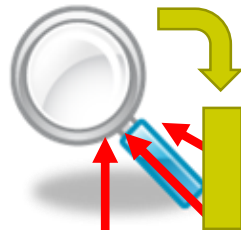


Process assets



Method

Detect inaccurate data elements



Assesses Risks

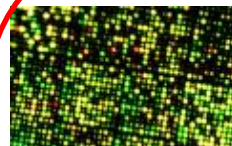
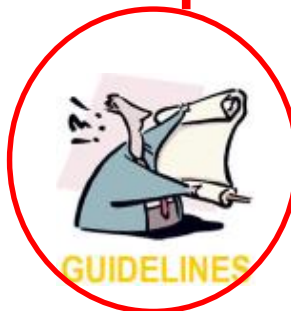


Suggestions for process redesign



Data inaccuracies situations

Method assets



Inaccurate data detection patterns



Tools

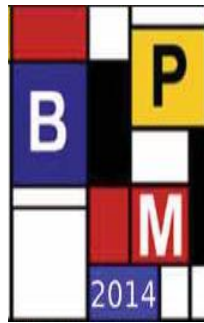


Risk patterns



Simulation tool

The envision Method



Method Users



Process experts



Consultants



Employees

Process assets



Process event logs



Database



Process documentation



Process models

Method

Detect inaccurate data elements



Data inaccuracies situations

Assesses Risks



Risks Assessments

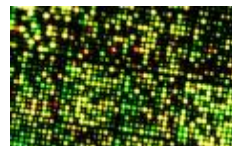
Suggestions for process redesign



Method assets



GUIDELINES



Inaccurate data detection patterns



Tools

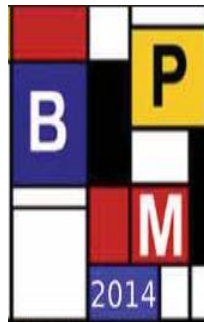


Risk patterns



Simulation tool

The envision Method



Method Users



Process experts



Consultants



Employees

Process assets



Process event logs



Database



Process documentation



Process models

Method

Detect inaccurate data elements



Data inaccuracies situations

Assesses Risks



Risks Assessments

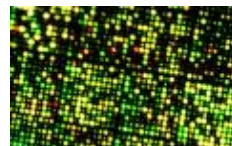
Suggestions for process redesign



Method assets



GUIDELINES



Inaccurate data detection patterns



Tools



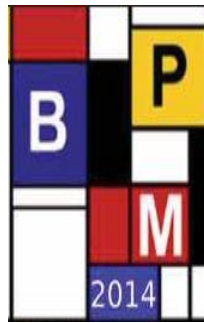
Risk patterns



Simulation tool

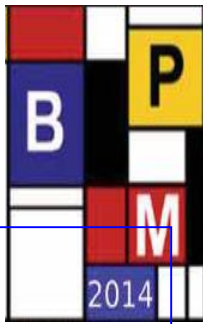
Current Status

- Initial phase of exploration
 - performing a number of case studies in organizations of different sizes and market segments



Market Segment	Size
Automatic pool cleaners	Large
Motion systems	Medium
Medical Devices	Small
Public services	Medium
Civilian protection	Large
Defense systems	Large
Academic college	Medium
Food & Beverage solutions	Large

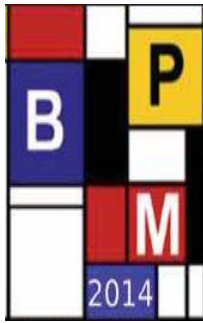
Current Status



- Collecting event logs and mining them to identify exceptional paths in the actual process
- Mining and manipulating the actual databases
- Collecting and analyzing process related artifacts such as models and documents
- Performing interviews with different stakeholders involved in these processes

- Report
 - List of specific data items suspected in inaccuracy and how they were detected (manually/ automatically)
 - Process paths suspected in being related to data inaccuracy and how they were detected (manually/ automatically)
 - Quantification of the frequency of data inaccuracy situations
 - Evaluation of the effect on performance indicators (e.g., cost, time).
 - Understanding the potential risks

Current Status



- Data analysis will use automated tools
 - As a guiding principle, we look for discrepancies in the process

Process mining techniques and algorithms

- Detect unusual cases in the process
- Detect discrepancies between the log and the process model

Database querying and filtering AND Data mining techniques

- Detect discrepancies between the log and the database
- Searching abnormal data

Performing statistics on the actual data

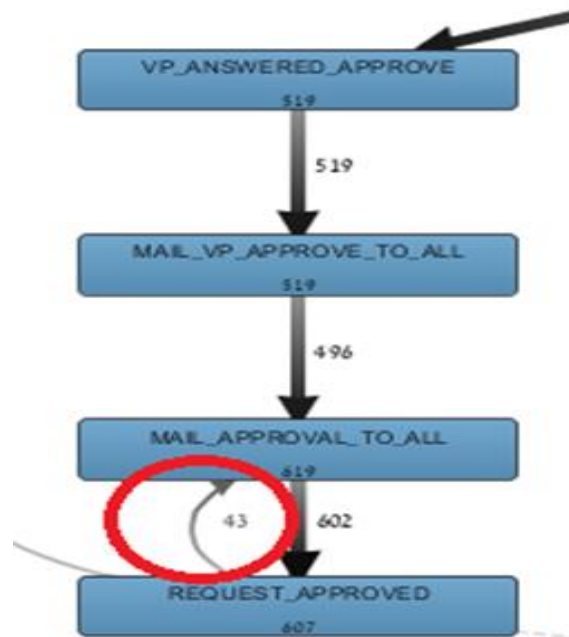
- Searching abnormal data with respect to the domain conventional values

- Triangulation with qualitative analysis of the human interviews

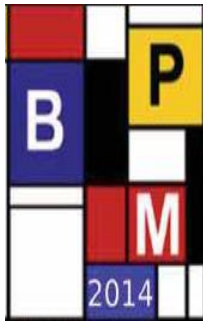


Current Status

- For example, in a case study we preformed, in one of the discovered process models, we found at a certain step, 43 cases (7%) going back to the previous step (loop), as opposed to the "normal" flow taken by the rest of the cases.

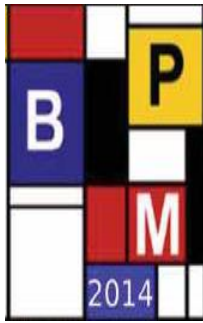


Current Status



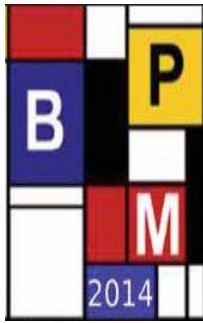
- We questioned relevant employees and found out that something is wrong...
- Using the case ID for querying the database, we found these records had non-existing values for a specific attribute. They had to loop and be corrected for the process to continue.

Excepted Contribution



- The proposed research will help detect data inaccuracy spots and assess its consequences (the possible risks).
- It will help estimate the efforts and the changes required in order to redesign a business process.

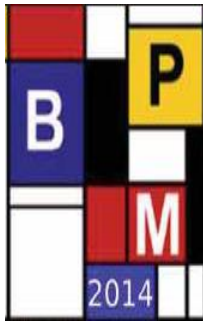




Acknowledgment

This research is partially supported by the Israel Science Foundation under grant 856/13.

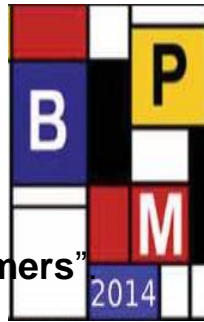




Thank You Very Much!!

Comments and Questions, please!

Data Quality



1. Wang, R., Y., and Strong D., M., “**Beyond accuracy: what data quality means to data consumers**”. Journal of Management Information Systems, Volume 12 Issue 4, March 1996, pp. 5-33
2. Batini, C., Cappiello, C., Francalanci, C., Maurino., A, “**Methodologies for Data Quality Assessment and Improvement.**” ACM Computing Surveys (CSUR), vol 41, issue 3, 2009.
3. Barone, D., Stella, F., and Batini, C., “**Dependency discovery in data quality**”. Advanced Information Systems Engineering, Lecture Notes in Computer Science Volume 6051, 2010, pp 53-67
4. Heravizadeh, M., Mendling, J., and Rosemann, M. “**Dimensions of business processes quality**”. Business Process Management Workshops. Lecture Notes in Business Information Processing Volume 17, 2009, pp. 80-9
5. Wand, Y. and Wang, R. Y., “*Anchoring data quality dimensions in ontological foundations*”. Communications of the ACM 39, 11, pp. 86-95, 1996
6. Wang, R., Y., and Strong D., M., “**Beyond accuracy: what data quality means to data consumers**”. Journal of Management Information Systems, Volume 12 Issue 4, March 1996, pp. 5-33
7. Agmon, N. and Ahituv N. “**Assessing Data Reliability in an Information System**”. Journal of Management Information Systems. Vol. 4, No. 2, Database Management (Fall, 1987), pp. 34-44



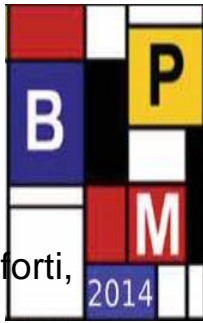
Data in business processes



1. Sadiq, S., Orlowska, M., Sadiq, W. and Foulger, C. “Data Flow and Validation in Workflow Modeling”. ADC '04 Proceedings of the 15th Australasian database conference - Volume 27, pp. 207-214, 2004
2. Soffer, P., “Mirror, mirror on the wall, can I count on you at all? Exploring data inaccuracy in business processes” , Enterprise, Business-Process and Information Systems Modeling, Lecture Notes in Business Information Processing Volume 50, 2010, pp 14-25
3. Bagchi, S., Xue Bai ; Kalagnanam, J. “Data Quality Management Using Business Process Modeling” Services Computing, 2006. SCC '06. IEEE International Conference on. 2006
4. Rozinat, A., Jong, I.S.M. de, Gunther, C.W. and van der Aalst, W.M.P. “Process Mining of Test Processes: A Case Study”. Internal Report, BETA publicatie : working paper, No. 220). Eindhoven: Technische Universiteit Eindhoven, 36 pp, 2007
5. De Leoni, M. and van der Aalst, W.M.P. “Aligning Event Logs and Process Models for Multi-Perspective Conformance Checking: An Approach Based on Integer Linear Programming”. 11th International Conference, BPM 2013, Beijing, China, August 2013. pp. 113-129



Risk Assessment in Business Processes



1. S. Suriadi, S., Weiß, B., Winkelmann, A., ter Hofstede, A., Wynn, M., Ouyang, Adams, M.J., Conforti, R., Fidge, C., La Rosa, M. And A. Pika. "Current research in risk-aware business process management - overview, comparison, and gap analysis". BPM Center Report BPM-12-13, BPMcenter.org, 2012.
2. Conforti, R., de Leoni, M., La Rosa, M. and van der Aalst, W.M.P. "Supporting Risk-Informed Decisions during Business Process Execution", Advanced Information Systems Engineering, Lecture Notes in Computer Science Volume 7908, 2013, pp 116-132

