FONDAZIONE
BRUNO KESSLER

Semantic-based Process Analysis

Mauro Dragoni

Fondazione Bruno Kessler (FBK), Shape and Evolve Living Knowledge Unit (SHELL)

https://shell.fbk.eu/index.php/Mauro_Dragoni - dragoni@fbk.eu

work done in collaboration with

Piergiorgio Bertoli[2], Francesco Corcoglioniti[1], Chiara Di Francescomarino[1], Chiara Ghidini[1], Michele Nori[2], and Marco Pistore[2], and Roberto Tiella[1]

[1]Fondazione Bruno Kessler, Trento

[2]SAYService, Trento

ISWC 2014 – Riva del Garda, Trento

October, 23rd 2014

FONDAZIONE BRUNO KESSLER

Process Analysis

Extracts analytical knowledge about the performances of a business process starting from collected process execution data

Three Challenges

- **Challenge 1**: Combining three different dimensions.
  - D1: the procedural dimension (P)
  - D2: the domain of interest (K)
  - D3: the execution dimension (T)
- **Challenge 2**: Semantic Reasoning
- **Challenge 3**: Scalability

FONDAZIONE
BRUNO KESSLER

Semantic Process Analysis

Employs (SW) techniques that leverage the explicit formalization of the semantics of a business process and the data it manipulates

Our approach / contributions:

- Integrated OWL 2 / RDF model of P + K + T queried with SPARQL

  → *address* ***Challenge 1***

- OWL 2 reasoning for making explicit inferrable knowledge

  → *address* ***Challenge 2***

- Implementation based on SW triplestores

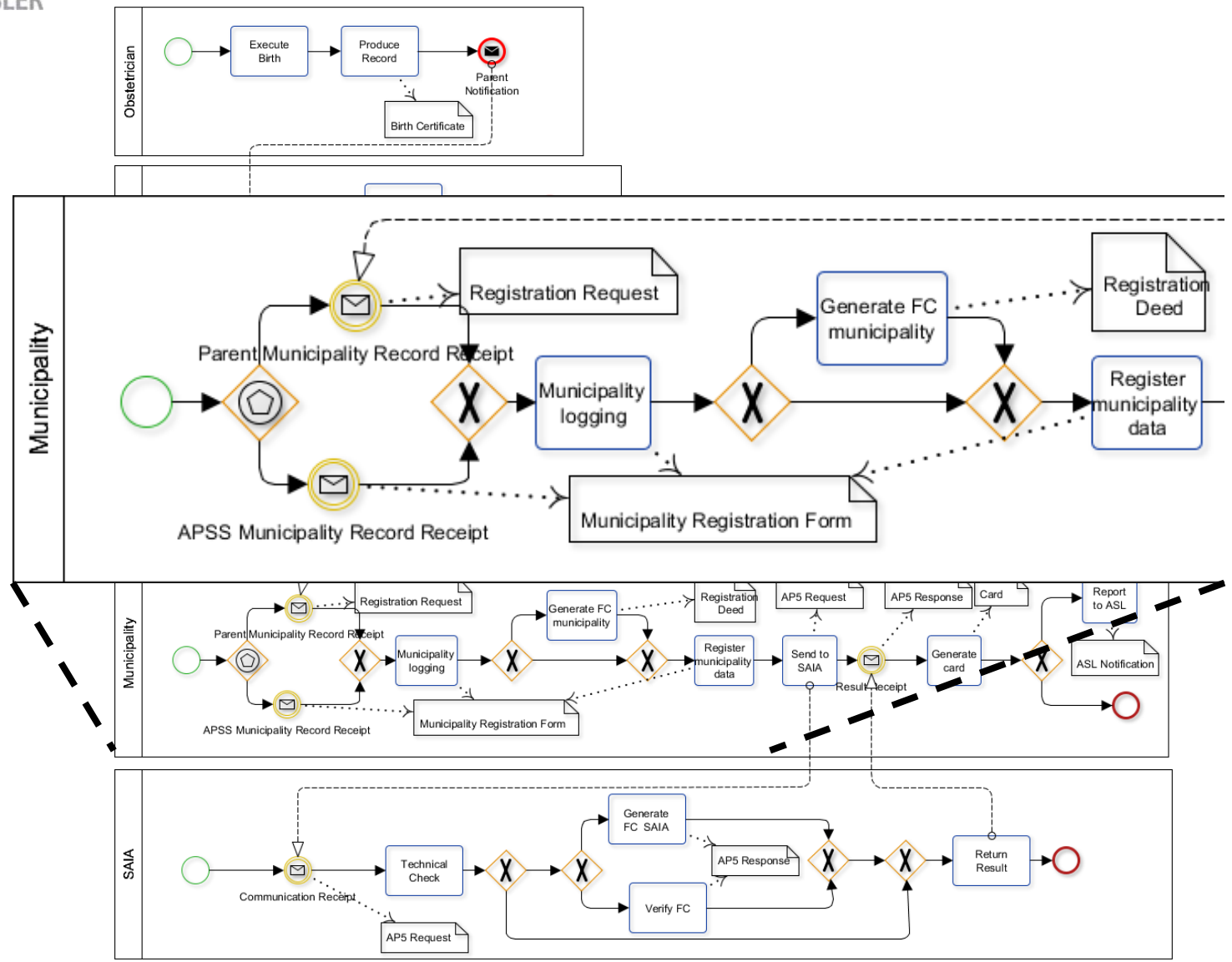
  → *address* ***Challenge 3***

FONDAZIONE
BRUNO KESSLER

Outline

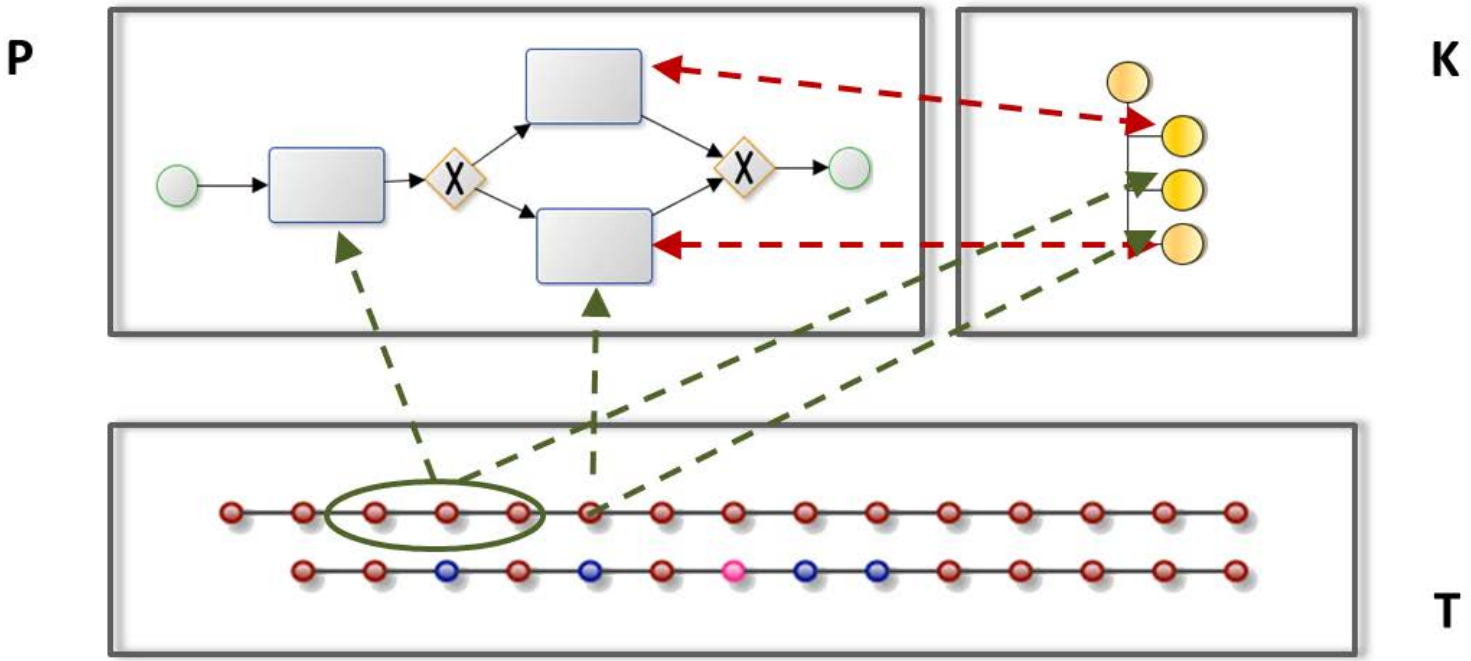
1. Our Use Case
2. The Proposed Model
3. The Architectural Solution
4. Evaluation

Use case: Birth Management Process



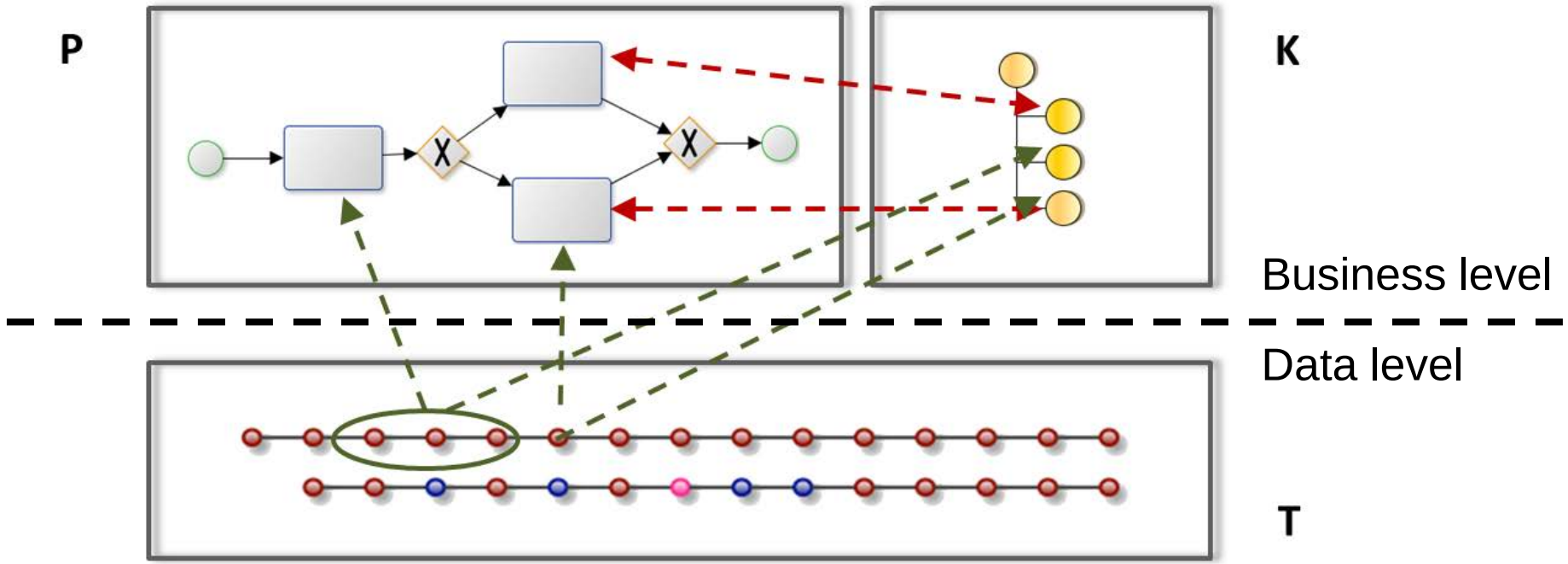
The Proposed Model: an Integrated View

- Reconciliation of knowledge and information related to different dimensions:



The Proposed Model: an Integrated View

- Reconciliation of knowledge and information related to different dimensions:



FONDAZIONE
BRUNO KESSLER

The Integrated Ontological Model

The Integrated Ontological Model

- BPMN Ontology

.

.

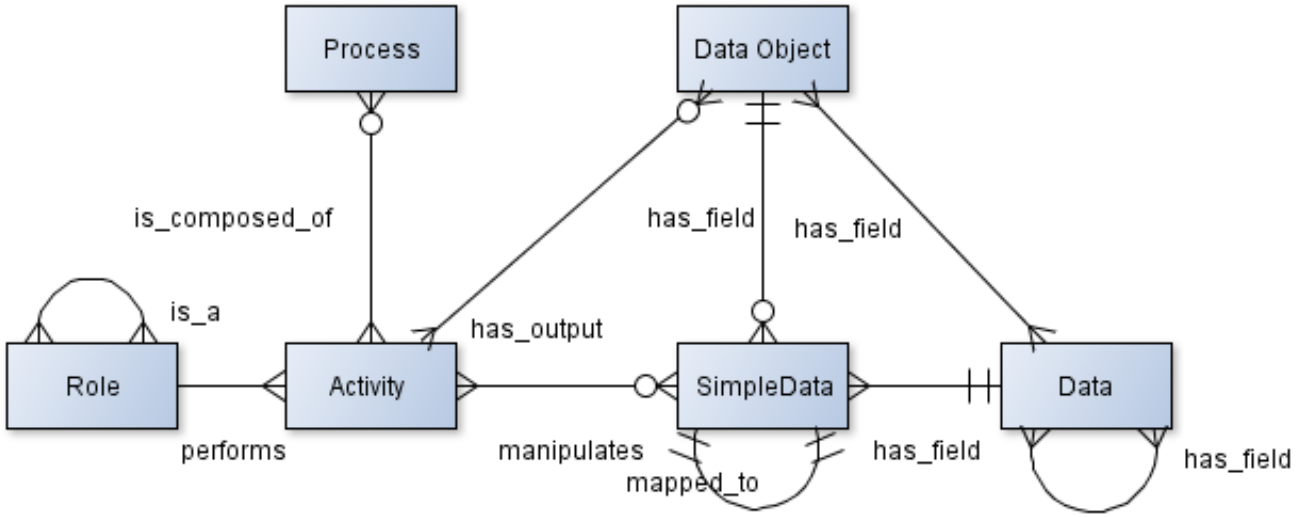
[1] Rospocher, M., Ghidini, C., Serafini, L.: An ontology for the business process modelling notation. In: 8th International Conference on Formal Ontology in Information Systems (FOIS 2014), 22-25 September 2014, Rio de Janeiro, Brazil. (2014)

https://shell-static.fbk.eu/resources/ontologies/bpmn2_ontology.owl

The Integrated Ontological Model

- BPMN Ontology
- Domain Ontology

.

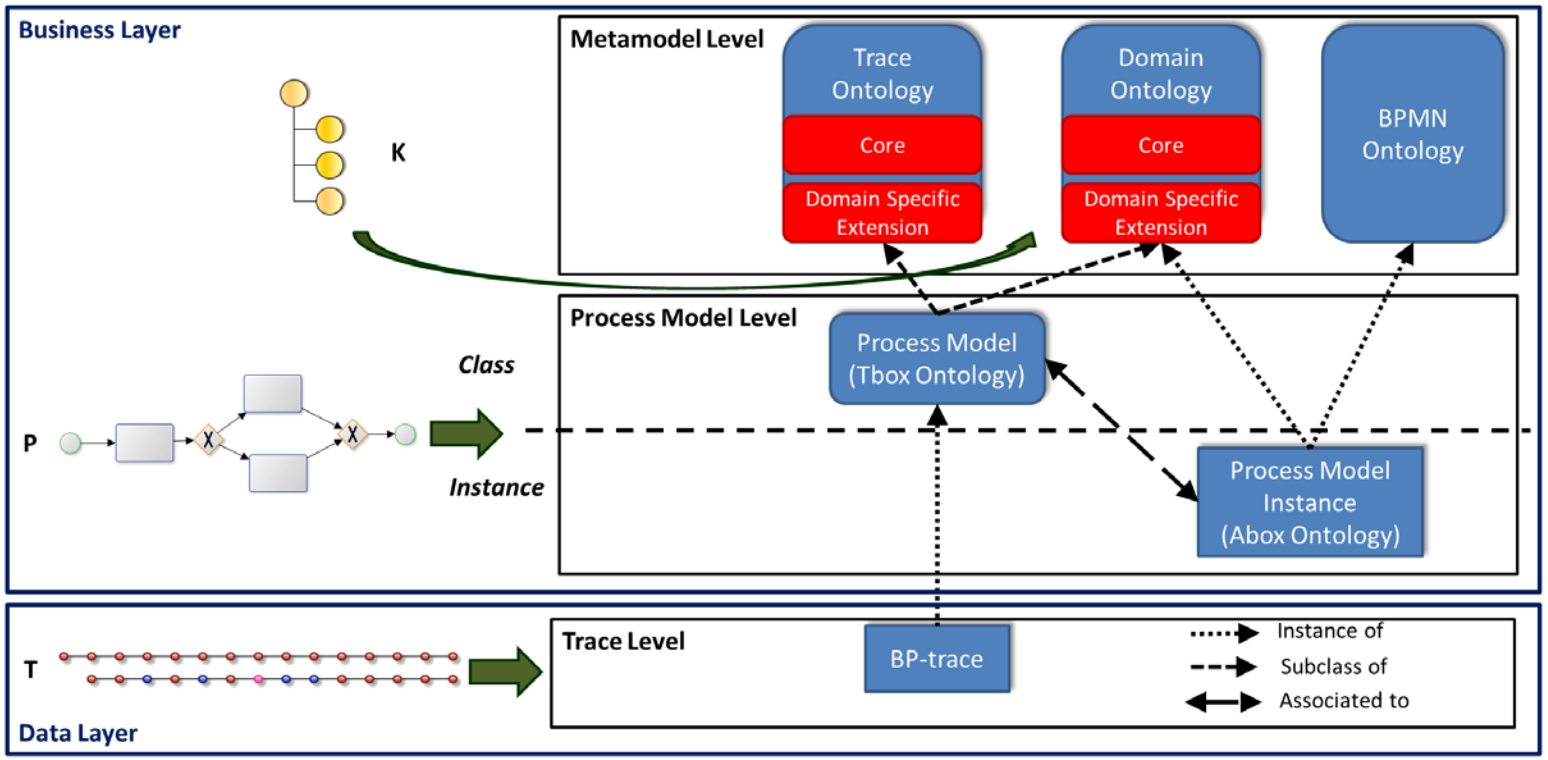


FONDAZIONE
BRUNO KESSLER

The Integrated Ontological Model

- BPMN Ontology
- Domain Ontology
- Trace Ontology

The Integrated Ontological Model



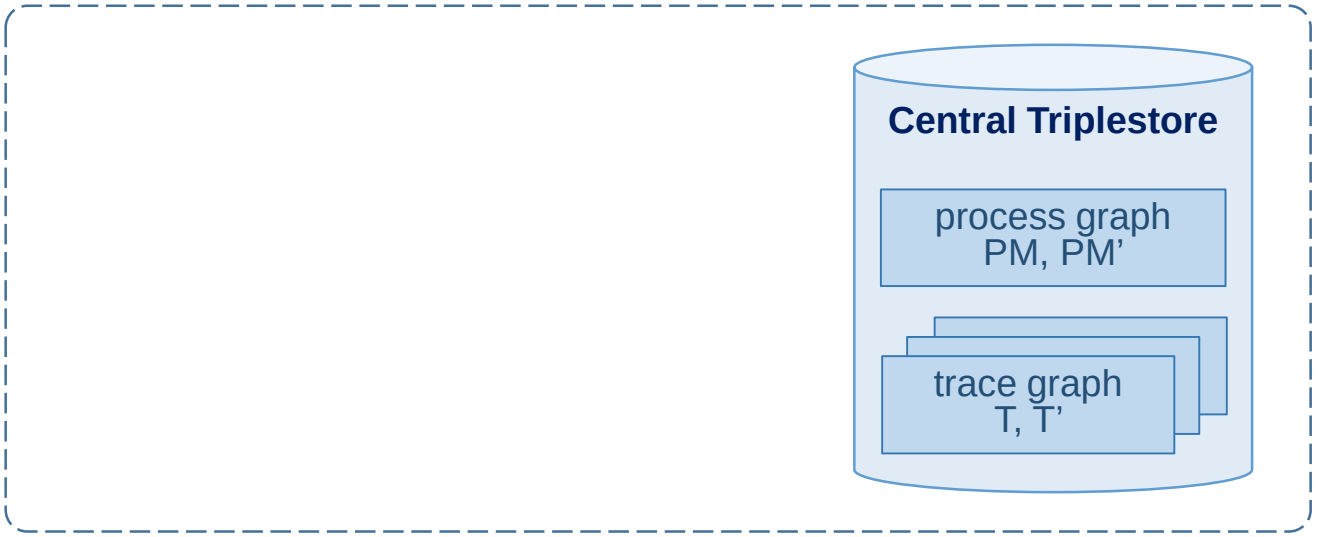
FONDAZIONE
BRUNO KESSLER

The Architectural Solution

- Challenges to cope:
  - collect trace data at fast rate
  - answer to complex queries
- Investigated solution: architecture based on triplestores

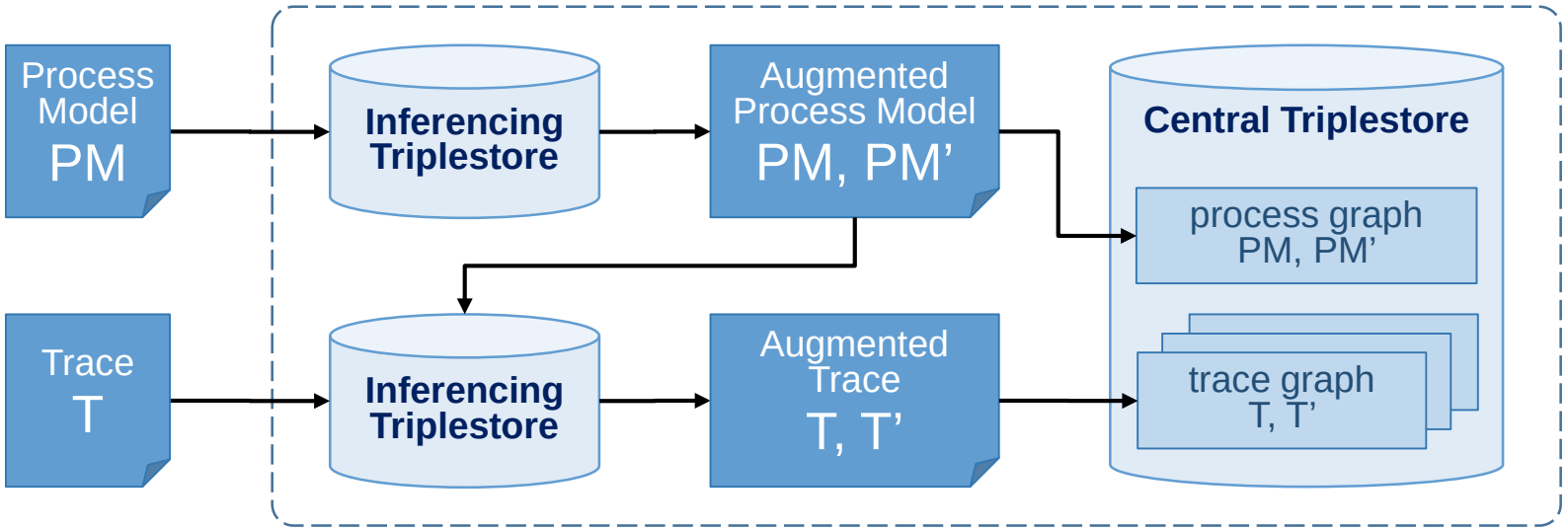
FONDAZIONE
BRUNO KESSLER

How data are organized?



- Process model and traces are stored in separated graphs
- Both explicit and implicit (inferred) data are stored

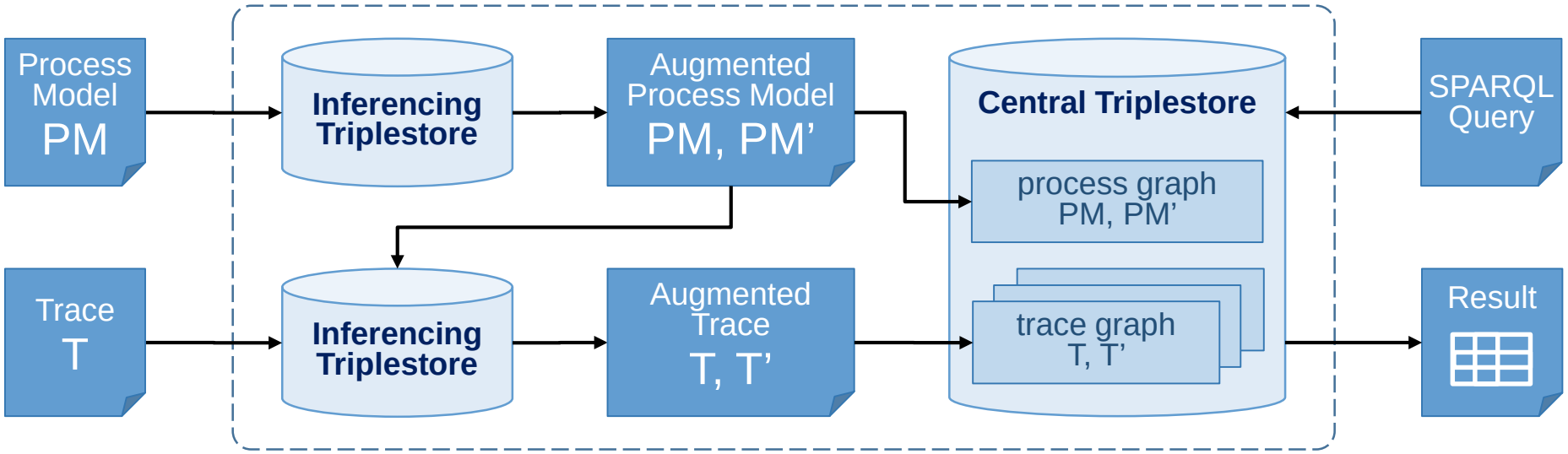
How data are populated?



- Process model (defined at design time) is stored once per all offline
- Trace update operation occurs every time a new piece of information is available
- Separate triplestores are used in order to allow different optimizations based on their purpose

FONDAZIONE BRUNO KESSLER

How data are retrieved?



- Queries performed by using SPARQL 1.1
- SPARQL aggregates turned out to be useful for analytical queries
- We introduced SPARQL extension mechanism

FONDAZIONE
BRUNO KESSLER

Evaluation - 1

- Process **P**: 4 pools, 19 activities, 11 domain objects, 19 events, 14 gateways, 54 sequence flows, 6 message flows.
- Domain ontology **K**:
  - 379 classes covering 28 activities and 12 data objects;
  - average of ~25 fields for each data object;
  - max field level-depth 4;
  - 5 properties.
- Set of execution traces **T**:
  - average of ~10 events for each trace;
  - average of 2040 triples for each trace;
  - further 1260 triples can be inferred.

FONDAZIONE BRUNO KESSLER

Evaluation - 2

Query	Description	P	K	T	Inference
Q.1	Average time per process execution spent by a specific municipality.			X	
Q.2	Total number of Registration Request documents filled from January, 1st, 2014.		X	X	
Q.3	Percentage of times in which the flow followed is the one which passes first through the APSS pool and then through the municipality one.	X		X	
Q.4	Number of cases and average time spent by each public office involved in the birth management procedure for executing optional activities.	X	X	X	X
Q.5	Number of times in which the municipality sends to SAIA a request without FiscalCode.	X	X	X	X
Q.6	Last event of trace TraceID.			X	
Q.7	Average time spent by trace TraceID.			X	
Q.8	Does the trace TraceID pass through the activity labeled with “PresentAtTheHospital”?	X		X	

FONDAZIONE BRUNO KESSLER

Evaluation - 3

Traces	Stored triples			Storing		Querying	
	Asserted	Inferred	Total	Throughput	Total time	Avg. Time Q.4	Avg. Time Q.8
1500	3062349	1895471	4957820	37.89 trace/min	2426.88 s	324 ms	41.4 ms
10500	21910269	13057464	34967773	37.41 trace/min	16851.21 s	881.4 ms	26.2 ms
42000	87503538	52045200	139548738	37.34 trace/min	67537.95 s	4510.0 ms	105.0 ms

FONDAZIONE BRUNO KESSLER

Evaluation - 3

Traces	Stored triples			Storing		Querying	
	Asserted	Inferred	Total	Throughput	Total time	Avg. Time Q.4	Avg. Time Q.8
1500	3062349	1895471	4957820	37.89 trace/min	2426.88 s	324 ms	41.4 ms
10500	21910269	13057464	34967773	37.41 trace/min	16851.21 s	881.4 ms	26.2 ms
42000	87503538	52045200	139548738	37.34 trace/min	67537.95 s	4510.0 ms	105.0 ms

Daily, weekly, and monthly load.

FONDAZIONE BRUNO KESSLER

Evaluation - 3

Traces	Stored triples			Storing		Querying	
	Asserted	Inferred	Total	Throughput	Total time	Avg. Time Q.4	Avg. Time Q.8
1500	3062349	1895471	4957820	37.89 trace/min	2426.88 s	324 ms	41.4 ms
10500	21910269	13057464	34967773	37.41 trace/min	16851.21 s	881.4 ms	26.2 ms
42000	87503538	52045200	139548738	37.34 trace/min	67537.95 s	4510.0 ms	105.0 ms

Throughput independent of the load

Daily, weekly, and monthly load.

Evaluation - 3

Traces	Stored triples			Storing		Querying	
	Asserted	Inferred	Total	Throughput	Total time	Avg. Time Q.4	Avg. Time Q.8
1500	3062349	1895471	4957820	37.89 trace/min	2426.88 s	324 ms	41.4 ms
10500	21910269	13057464	34967773	37.41 trace/min	16851.21 s	881.4 ms	26.2 ms
42000	87503538	52045200	139548738	37.34 trace/min	67537.95 s	4510.0 ms	105.0 ms

Daily, weekly, and monthly load.

Throughput independent of the load

Time required for queries is acceptable for real-time usage

FONDAZIONE
BRUNO KESSLER

Lessons Learned

- Divide-et-impera approach for inference to scale
- Usability feedbacks and expertise requirements

FONDAZIONE
BRUNO KESSLER

Mauro Dragoni

https://shell.fbk.eu/index.php/Mauro_Dragoni
dragoni@fbk.eu