



Dirk Fahland  
Marco Montali  
Ava Swevels

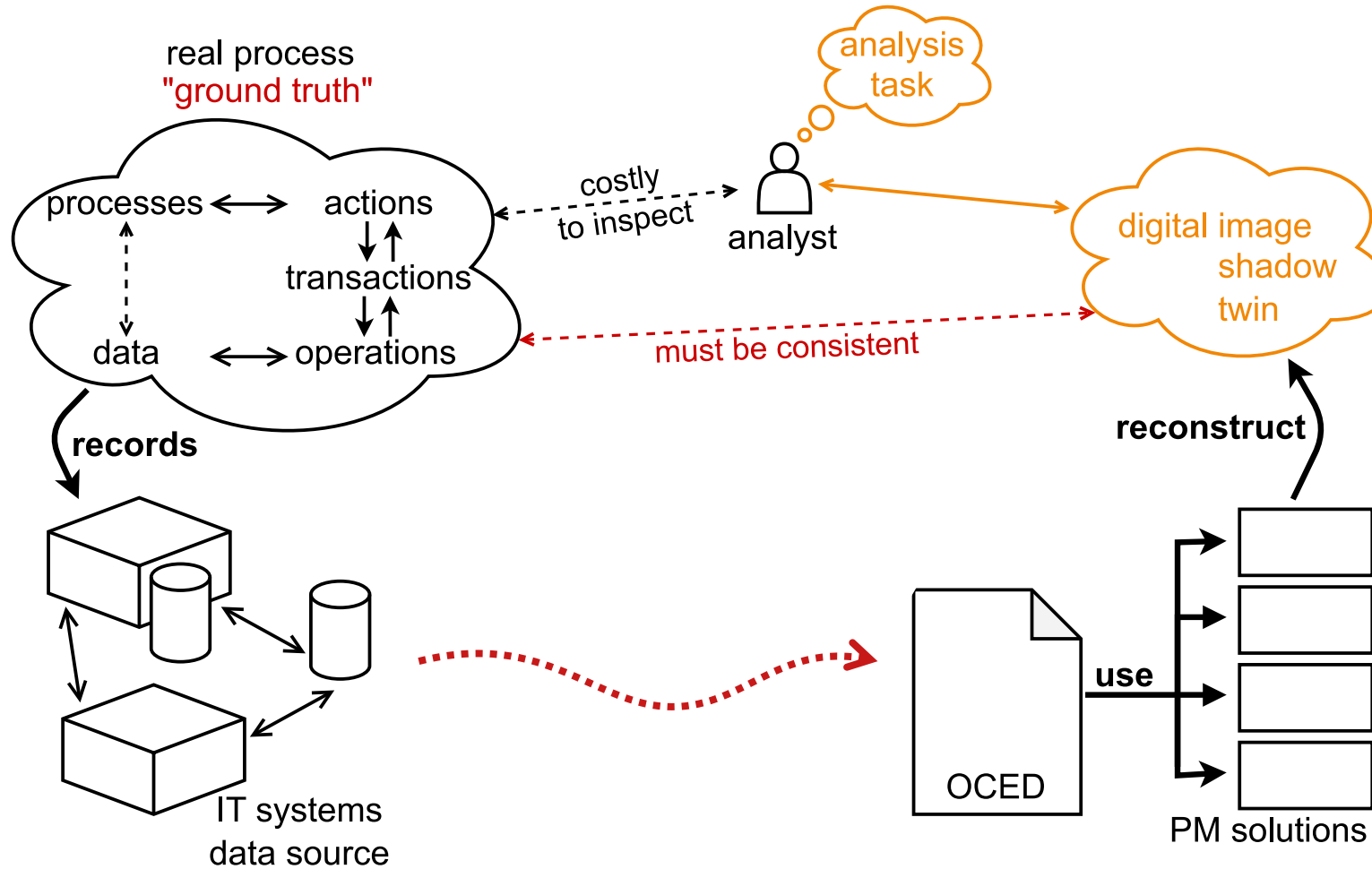


# Implementing OCED in Event Knowledge Graphs

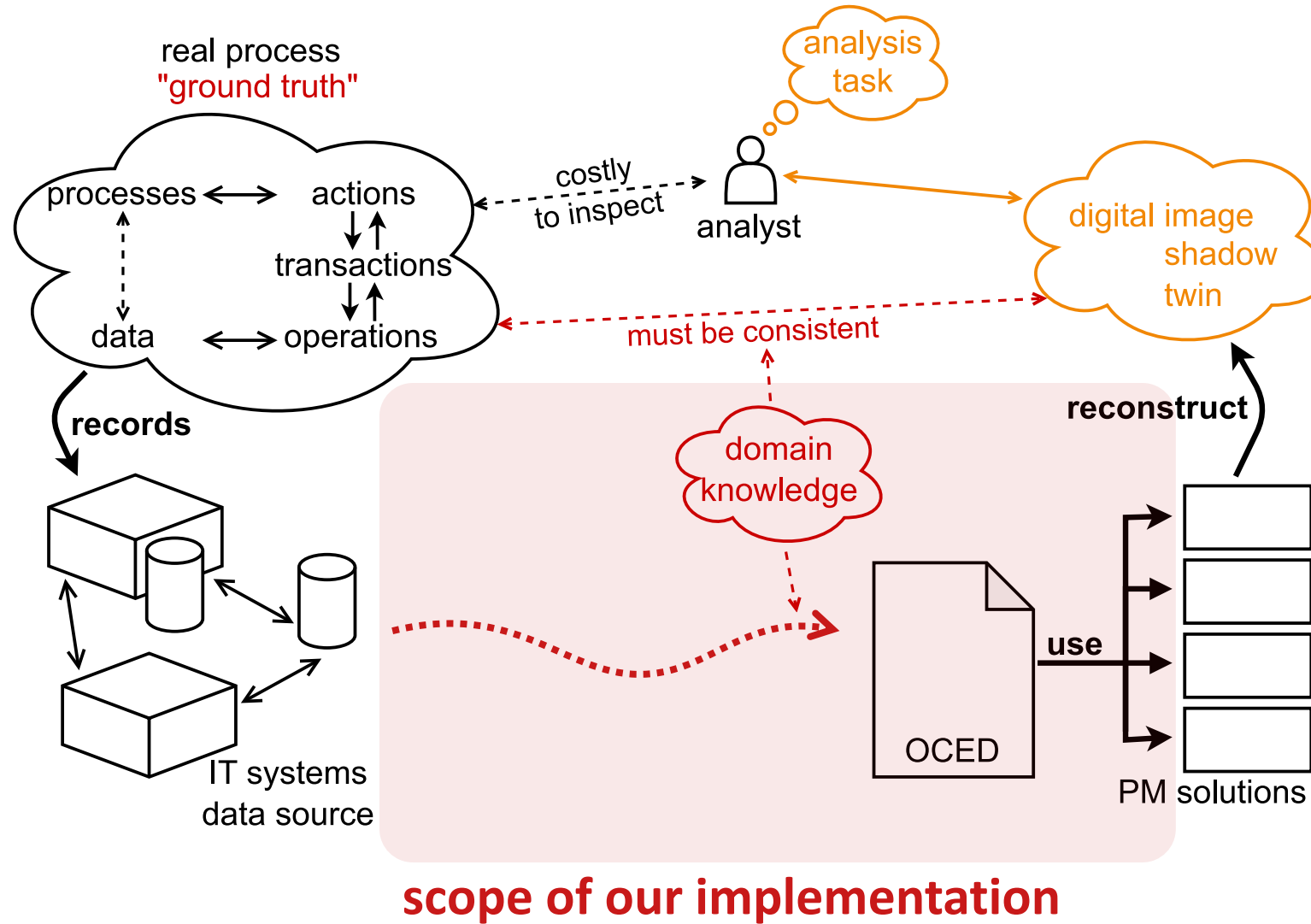


OCED Symposium at 5<sup>th</sup> International Conference on Process Mining (ICPM 2023), Rome, Italy, 26-10-2023

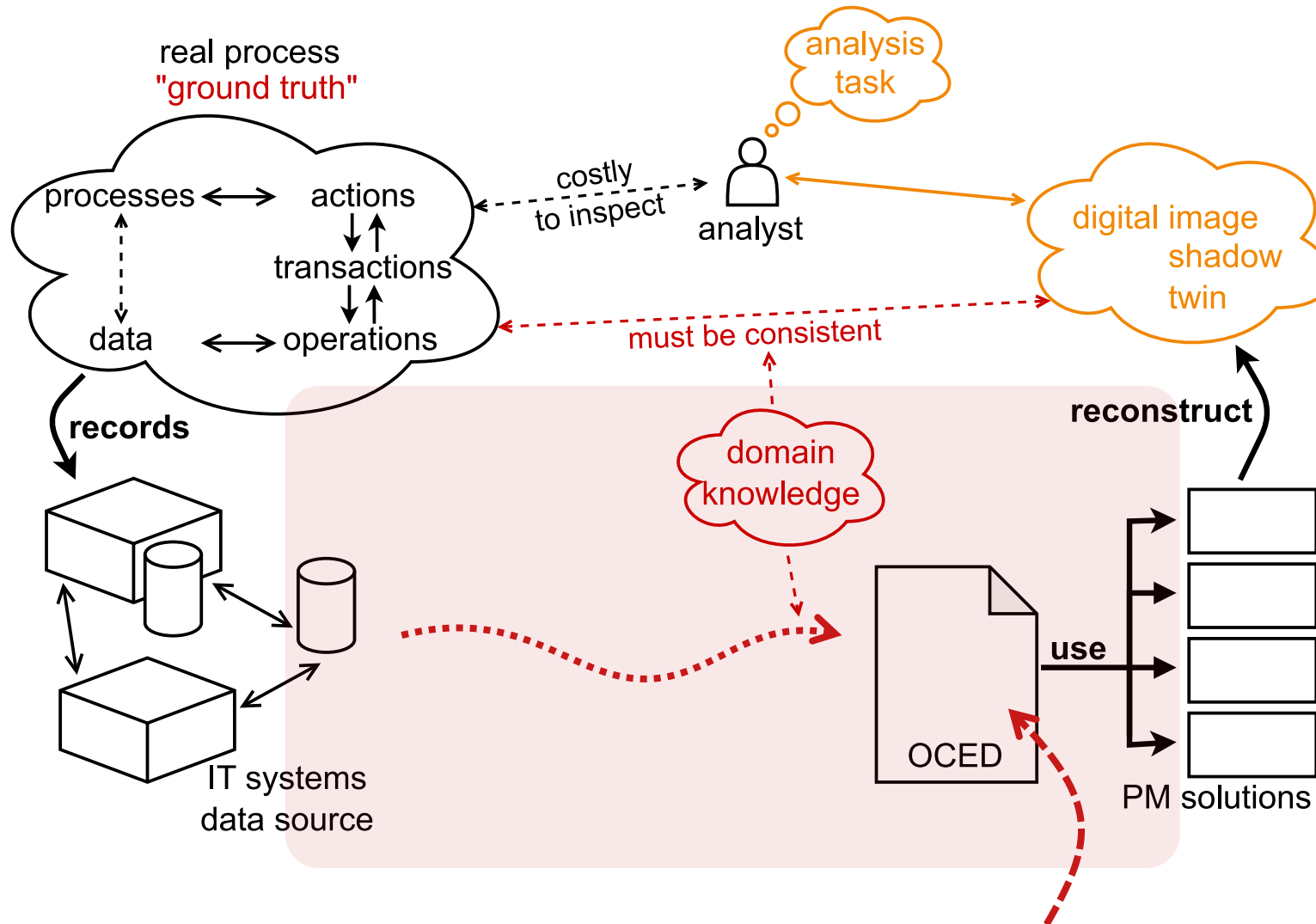
# OCED implementation must do **Source** → **OCED** → **Analysis**



# OCED implementation must do **Source** → **OCED** → **Analysis**



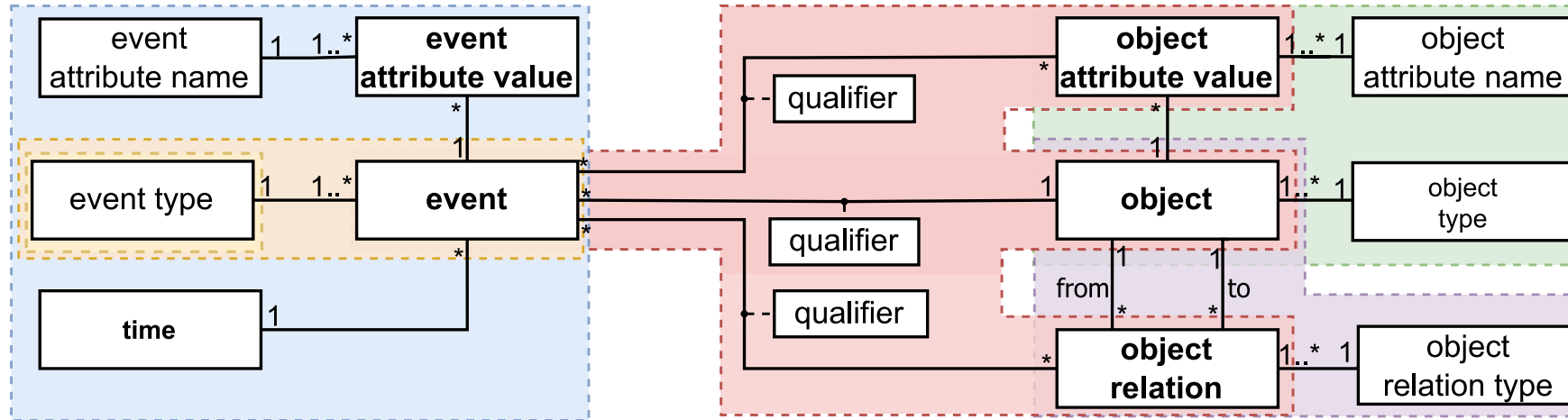
# OCED implementation must do **Source** → **OCED** → **Analysis**



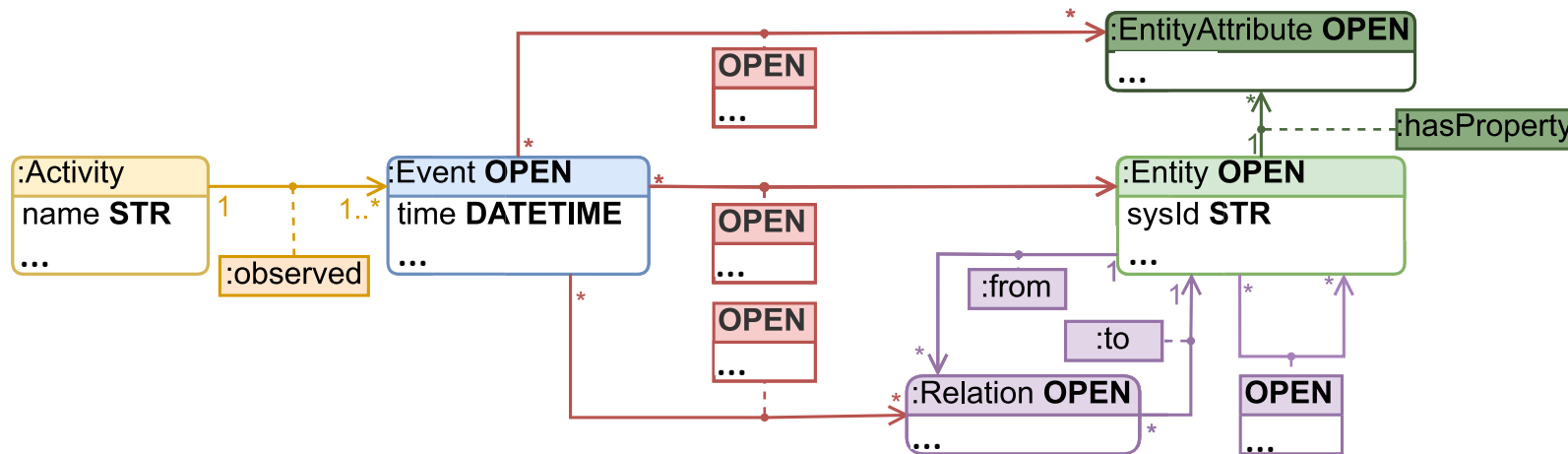
**choice: graph-based data model**

Temporal Property Graphs <> Event Knowledge Graphs

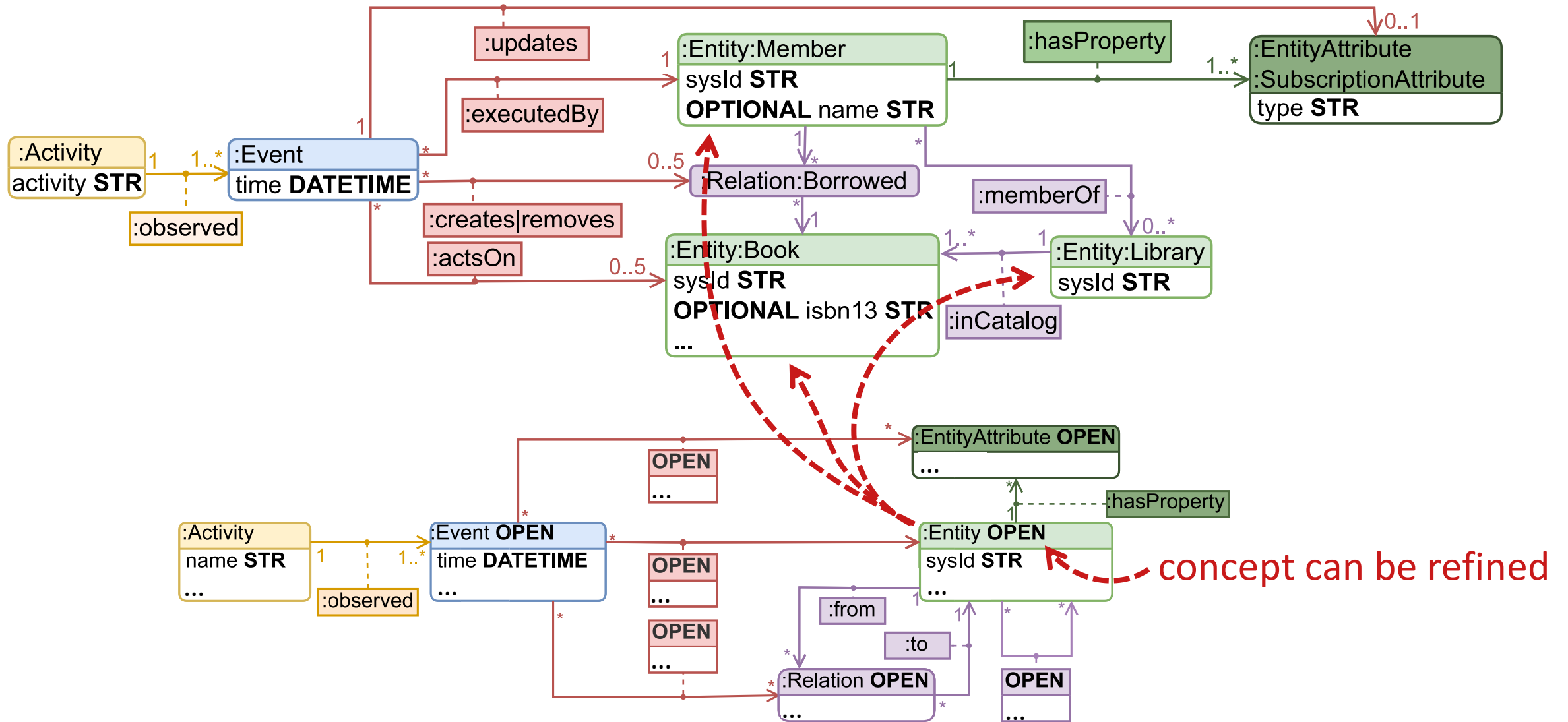
# OCED



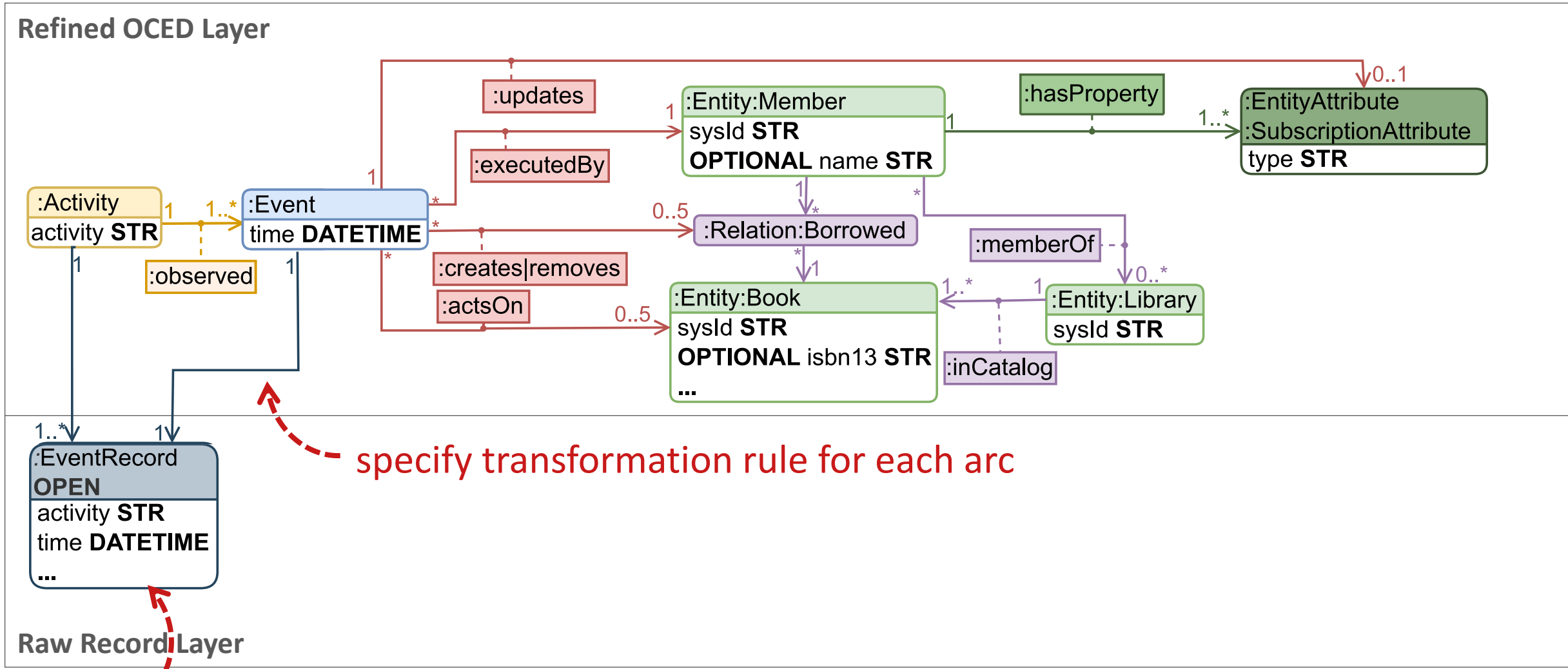
formalized in PG-Schema (graph data model specification language)



# OCED + Domain Data Model



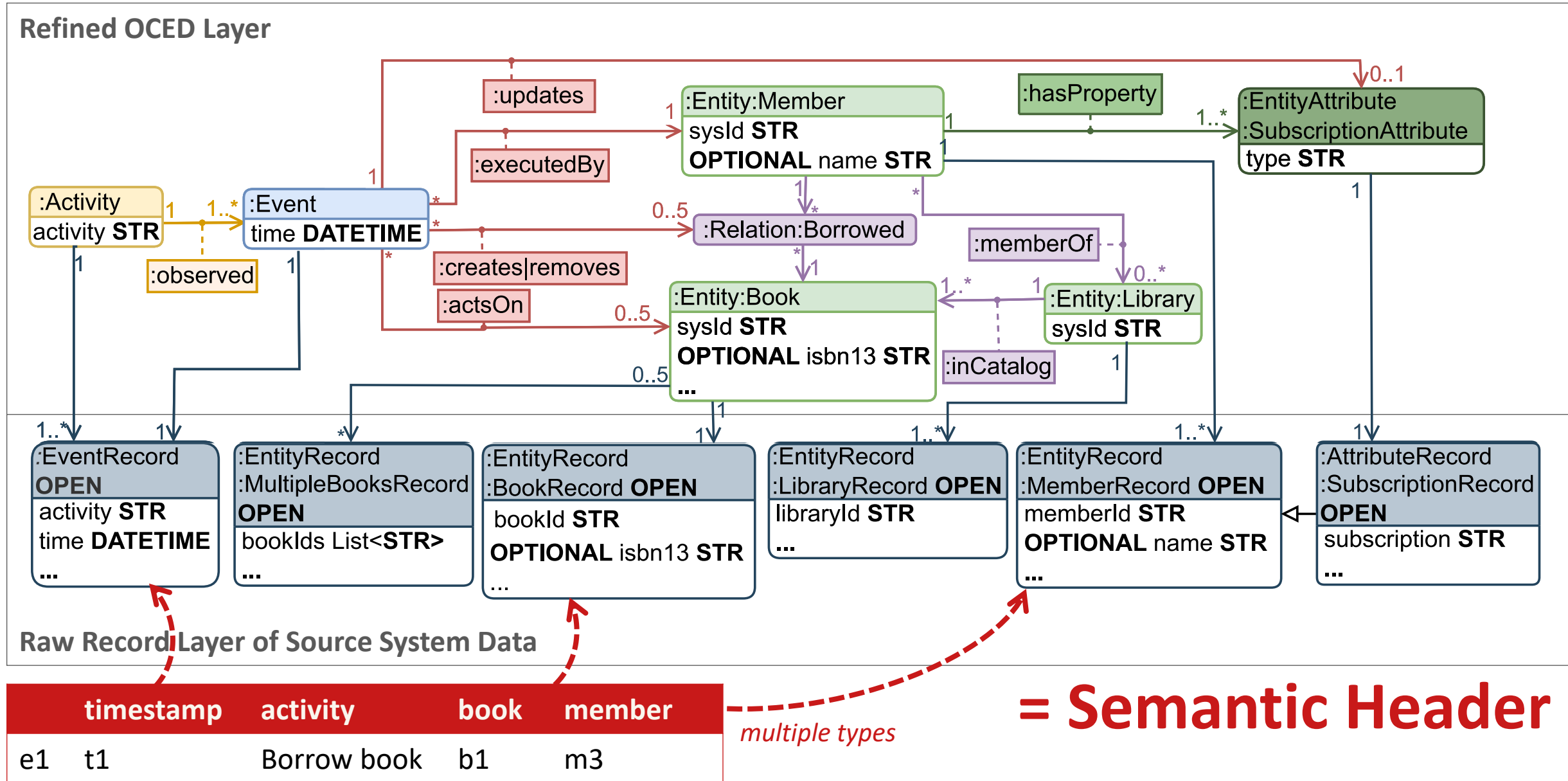
# OCED + Domain Data Model + Transformation



timestamp	activity	book	member
e1 t1	Borrow book	b1	m3



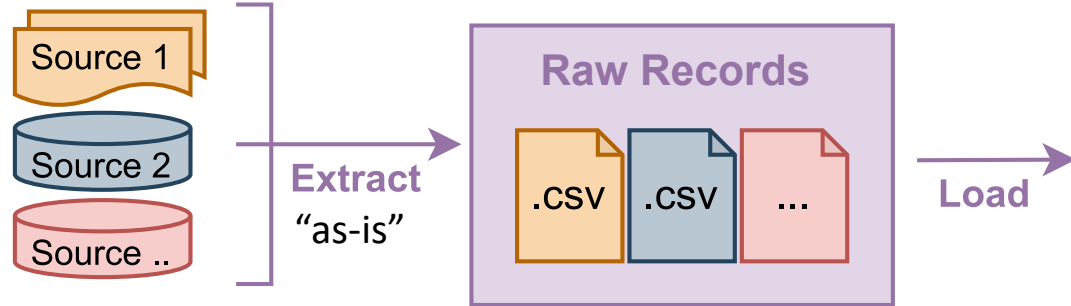
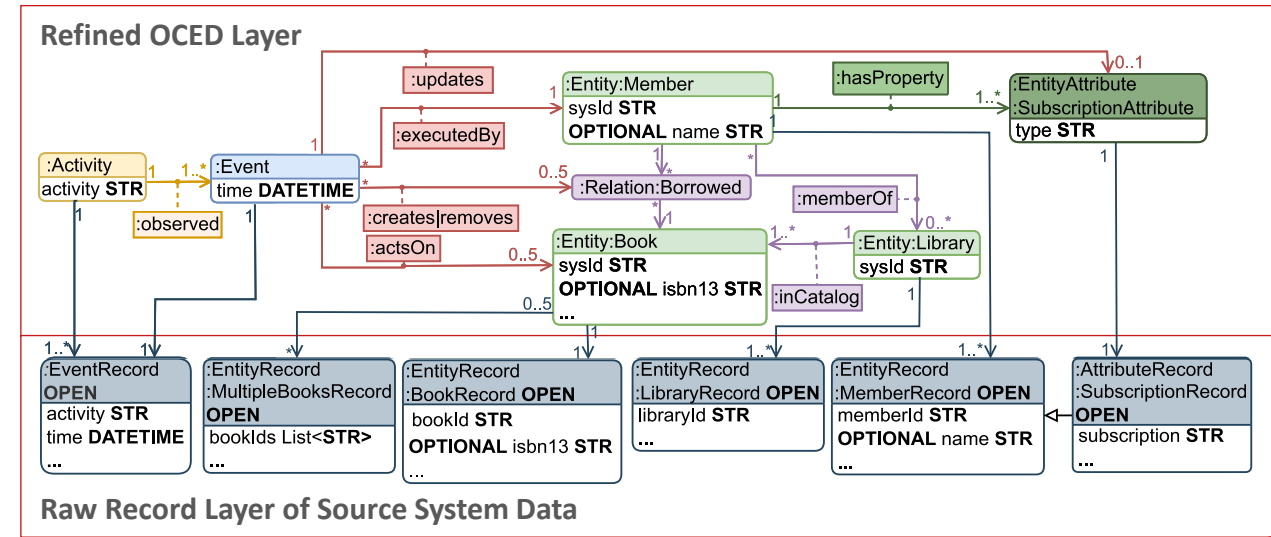
# OCED + Domain Data Model + Transformation



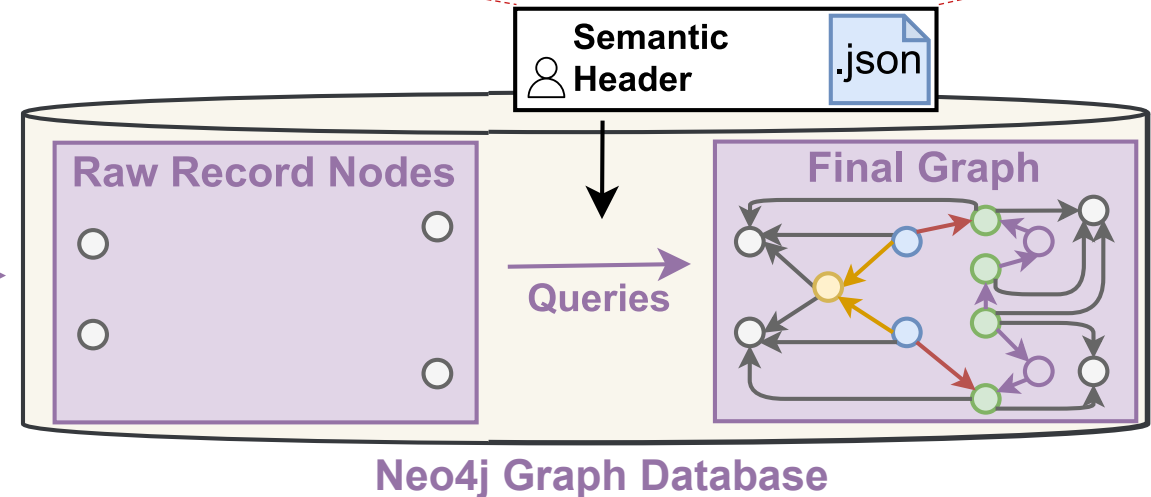


# OCED-PG

- implemented full OCED proposal + extensions (next slides)
- 7 industrial case studies
- 5 BPIC datasets as OCED (defined 5 semantic headers)



- e.g.,
- Events + Object references
- or
- Events
  - Objects
  - Relations



PromG Python library: OCED-PG + OCPM analysis  
<https://github.com/promg-dev>

# OCED-PG: BPIC Challenges

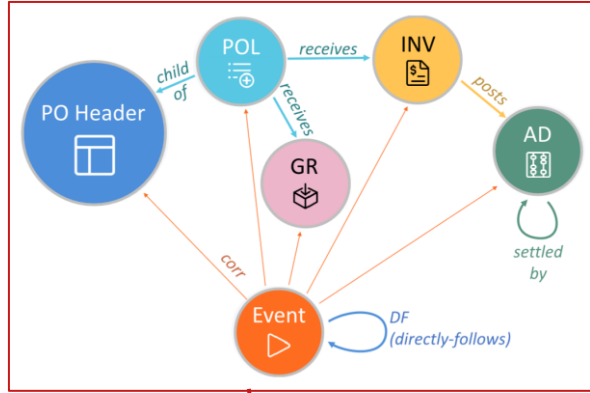
Data set	Source Size (GB)	#node types					#edge types	Memory (GB)	Time (mins)
		:Event	:Activity	:Entity	:Attribute	:Relation			
Library	0.002	1	1	3	1	1	7	1	0.5
BPIC'14	0.08	2	1	7	0	0	11	1	7.7
BPIC'15	0.11	1	1	3	0	0	7	1	4.4
BPIC'16	1.06	4	1	13	0	0	17	5	162.4
BPIC'17	0.29	1	1	5	0	0	4	1	19.7
BPIC'19	0.52	1	1	7	0	0	8	5	31.2
SAP	0.01	1	1	4	0	2	5	1	2.0
Manufacturing	0.03	1	1	4	0	0	1	1	3.7

*naïve queries*  
→ optimization coming  
to improve performance

Semantic Headers available → <https://github.com/promg-dev>  
build EKGs for Object-Centric Analysis of “well-known” real-life datasets

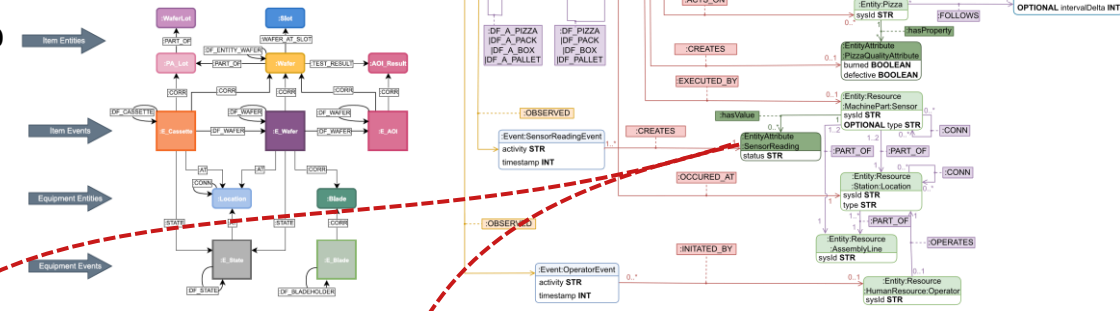
# OCED-PG Industrial Implementations

P2P Auditing

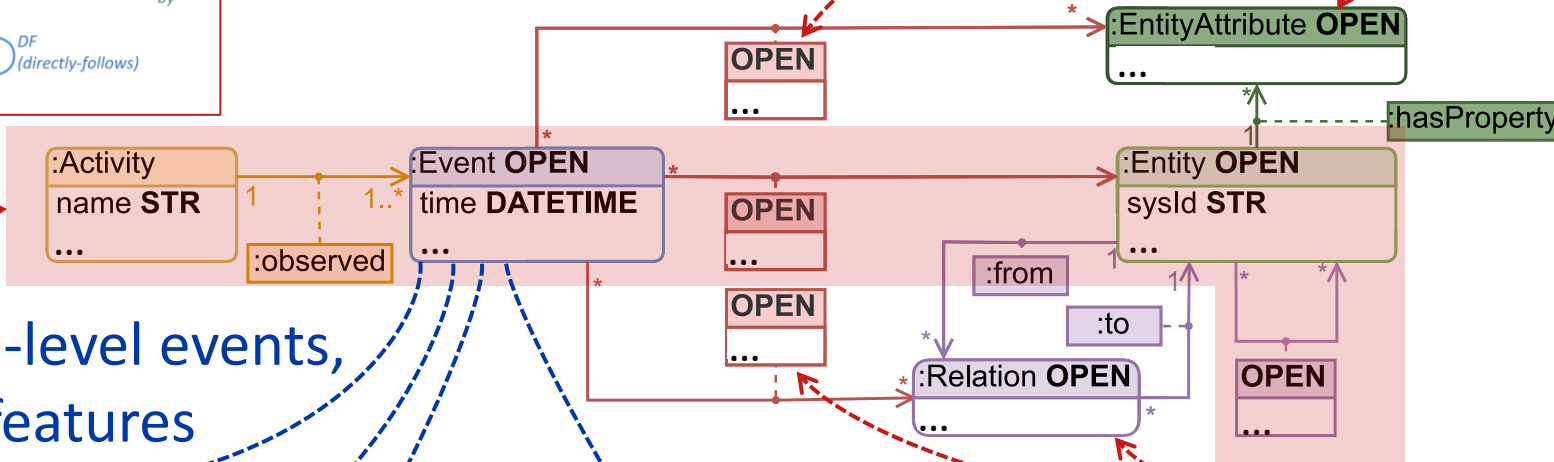


7 industrial case studies  
source > EKG > analysis

Manufacturing

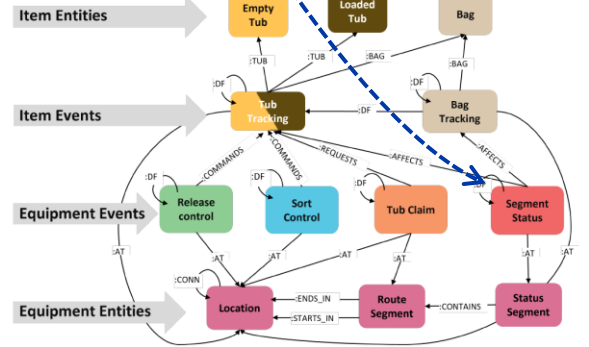
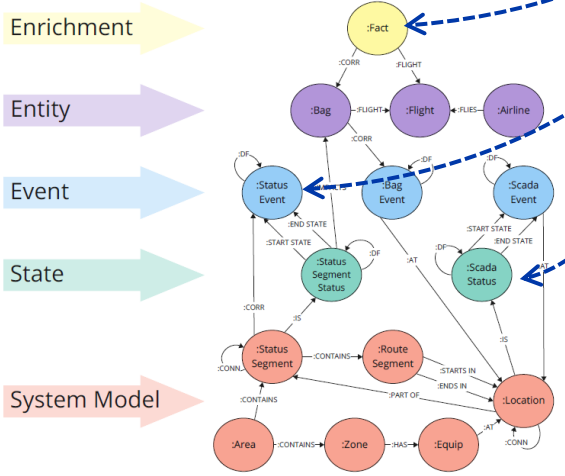


OCED core

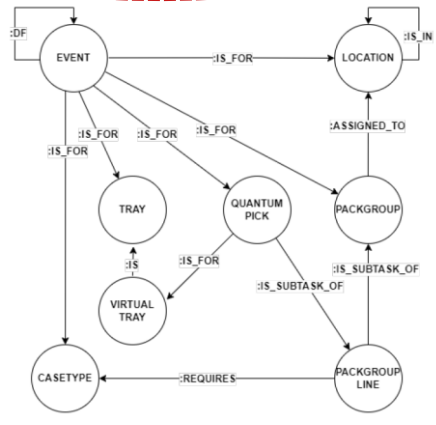


extensions: high-level events, states, analysis features

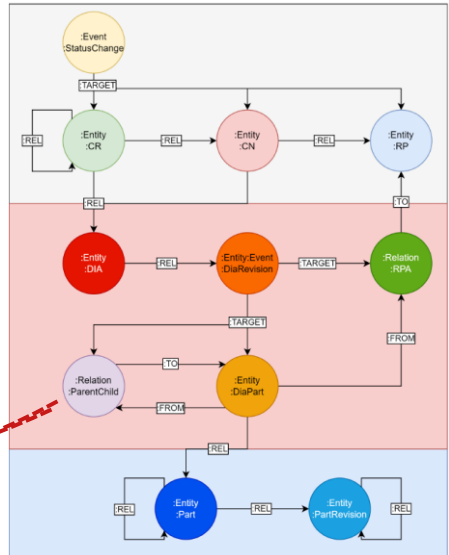
Airport Baggage Handling



Warehouse Automation



Configuration Management



# Demo: BPIC'17

```
37 "attributes": [  
38   {  
39     "name": "activity",  
40     "columns": [  
41       {  
42         "name": "event"  
43       }  
44     ],  
45     "optional": false  
46   },  
47   {  
48     "name": "lifecycle",  
49     "columns": [  
50       {  
51         "name": "lifecycle:transition"  
52       }  
53     ],  
54     "optional": false,  
55     "use_filter": false,  
56     "filter_exclude_values": [  
57       "SUSPEND",  
58       "RESUME"  
59     ]  
60   },  
61   {  
62     "name": "timestamp",  
63     "columns": [  
64       {  
65         "name": "time"  
66       }  
67     ],  
68     "datetime_object": {  
69       "format": "y/M/d H:m:s.nX",  
70       "timezone_offset": "+01"  
71     },  
72     "optional": false  
73   },  
74   {
```

raw records meta-data

```
{  
  "name": "BPIC17",  
  "version": "1.0.0",  
  "records": [  
    "(record:EventRecord {timestamp, activity, lifecycle, eventOrigin, action})",  
    "(record:ApplicationRecord {case})",  
    "(record:ApplicationDetailRecord WHERE record.eventOrigin = 'Application' {case, application})",  
    "(record:WorkflowRecord {case})",  
    "(record:WorkflowCorrRecord WHERE record.eventOrigin = 'Workflow' {case})",  
    "(record:OfferRecord:OfferEventIdRecord WHERE record.eventOrigin = 'Offer' AND record.eventId)",  
    "(record:OfferRecord:OfferIdRecord WHERE record.eventOrigin = 'Offer' AND record.eventId)",  
    "(record:ResourceRecord {resourceId})"  
  ],  
  "nodes": [  
    {  
      "type": "Event",  
      "constructor": [  
        {  
          "prevalent_record": "(record:EventRecord)",  
          "result": "(e:Event {timestamp:record.timestamp, activity:record.activity, lifecycle:record.lifecycle})"  
        }  
      ]  
    },  
    {  
      "type": "Activity",  
      "constructor": [  
        {  
          "prevalent_record": "(record:EventRecord)",  
          "result": "(a:Activity {activity:record.activity, lifecycle:record.lifecycle})",  
          "infer_observed": true  
        }  
      ]  
    }  
  ]  
}
```

semantic header: transformation rules

# Demo: BPIC'17

```
{
  "type": "Application",
  "constructor": [
    {
      "prevalent_record": "(record:ApplicationDetailRecord)",
      "result": "(a:Entity:Application {sysId: record.case, type:record.applicationType,
      \"infer_corr_from_event_record\": true
    },
    {
      \"prevalent_record\": \"(record:ApplicationRecord)\",
      \"result\": \"(a:Entity:Application {sysId: record.case})\",
      \"infer_corr_from_event_record\": false
    }
  ],
  \"infer_df\": true,
  \"include_label_in_df\": true,
  \"merge_duplicate_df\": true
},
{
  \"type\": \"Workflow\",
  \"constructor\": [
    {
      \"prevalent_record\": \"(record:WorkflowCorrRecord)\",
      \"result\": \"(w:Entity:Workflow {sysId: record.case})\",
      \"infer_corr_from_event_record\": true
    },
    {
      \"prevalent_record\": \"(record:WorkflowRecord)\",
      \"result\": \"(w:Entity:Workflow {sysId: record.case})\",
      \"infer_corr_from_event_record\": false
    }
  ],
  \"infer_df\": true,
  \"include_label_in_df\": true,
  \"merge_duplicate_df\": true
},
}
```

```
Clearing the database.
Importing and creating files
Loading data from BPI_Challenge_2017.csv from batch 0: 100%
Node (e:Event) using (record:EventRecord ) created
Node (a:Activity)using (record:EventRecord ) merged
Node (a:Entity:Application)using (record:ApplicationDetailRecord ) merged
Node (a:Entity:Application)using (record:ApplicationRecord ) merged
Node (w:Entity:Workflow)using (record:WorkflowCorrRecord ) merged
Node (w:Entity:Workflow)using (record:WorkflowRecord ) merged
Node (o:Entity:Offer)Total: took 36.76 seconds
Node (o:Entity:Offer)Completed: : 47it [00:36, [1.28it/s]
Node (r:Entity:Resource)-----
Node (case_aws:Entity)Event 779
Inferring DF over reEntity 222
Discovering multi-obRecord 779
Detecting tasks. Activity 49
TaskInstance 174
TaskAggregation 52
DF 2663
CORR 4445
OBSERVED 953
EXTRACTED_FROM 5792
FROM 70
TO 70
DF_APPLICATION 136
DF_WORKFLOW 490
DF_OFFER 88
DF_RESOURCE 712
DF_CASEAWO 759
DF_TA_Resource 75
DF_CASE_AO 106
DF_TA_CaseAWO 87
DF_CASE_AW 250
DF_CASE_WO 122
CONTAINS 779
DF_TI_Resource 107
DF_TI_CaseAWO 154
-----
```

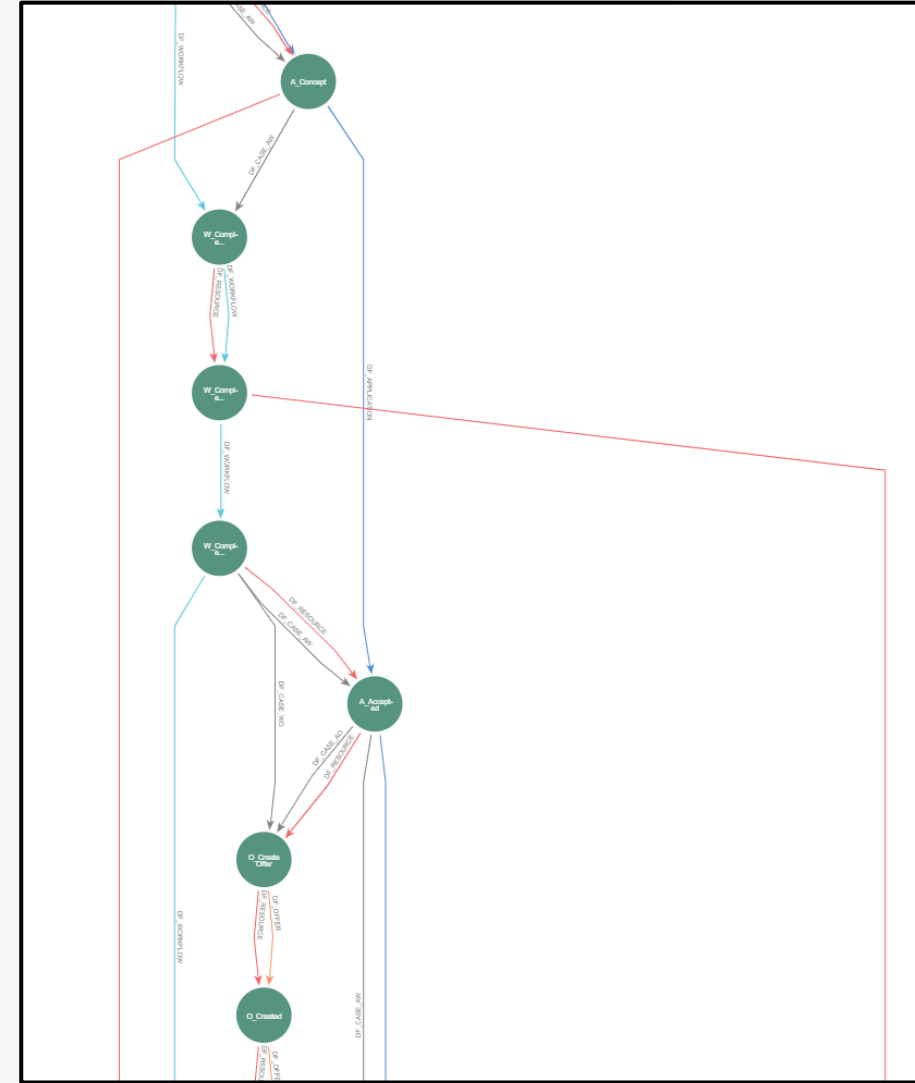
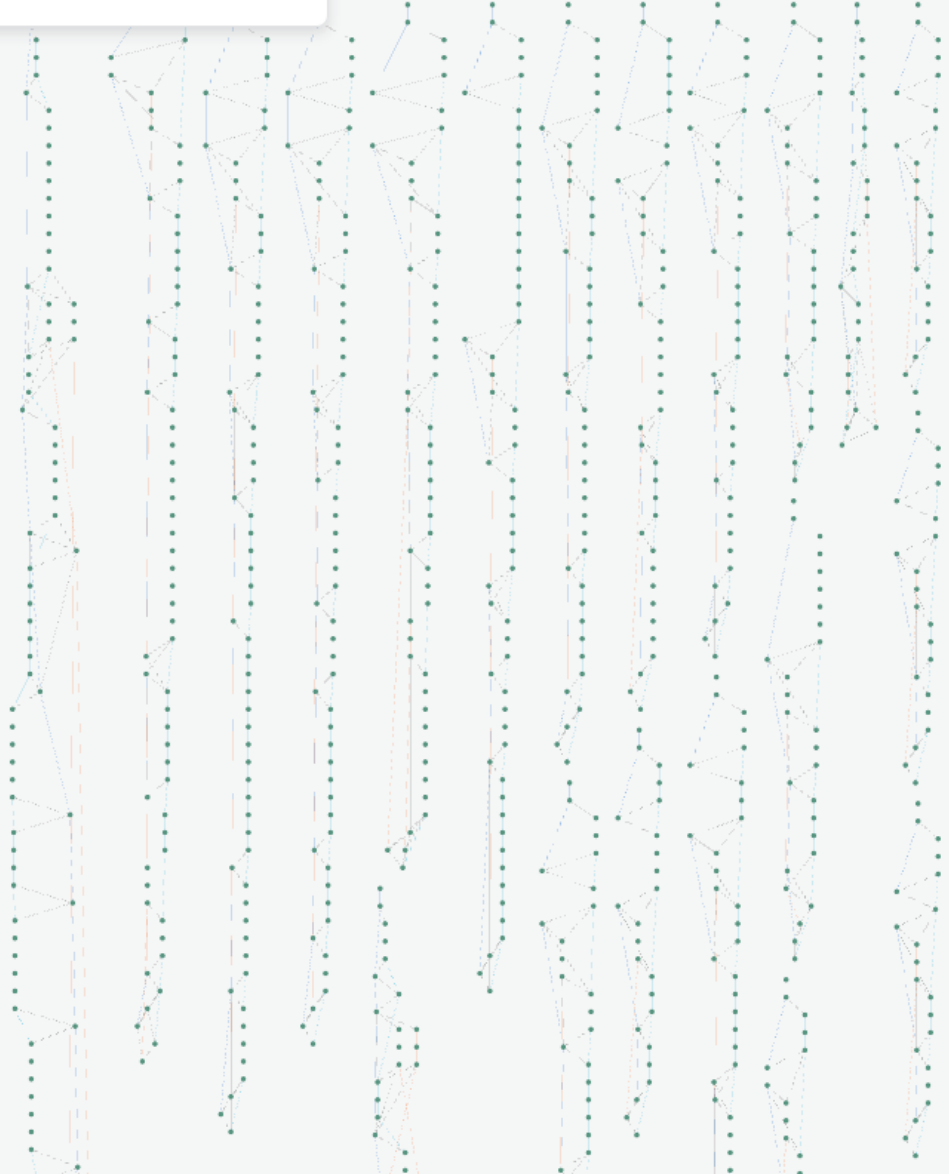
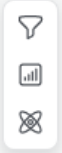
semantic header: transformation rules



# Demo BPIC'17

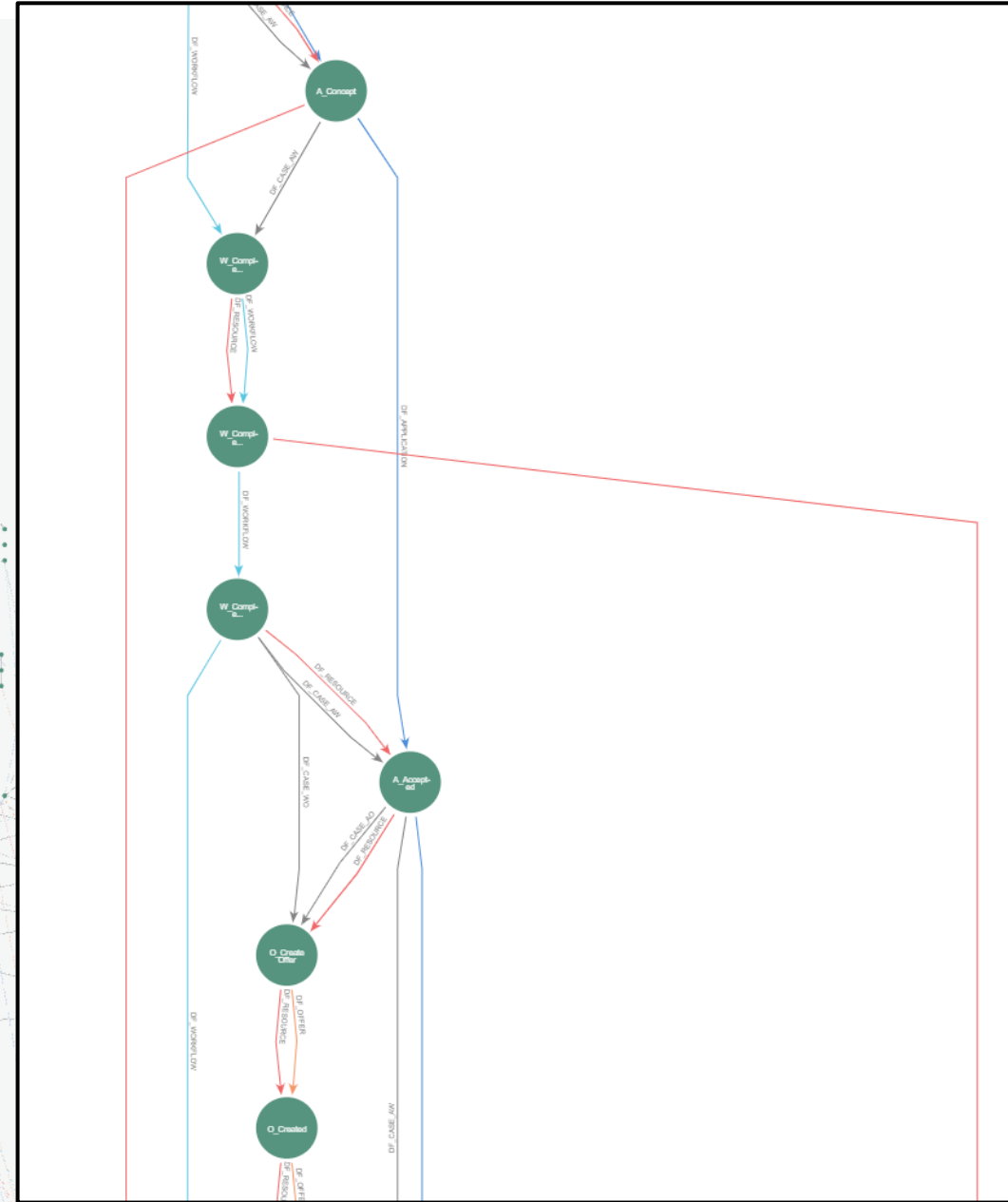
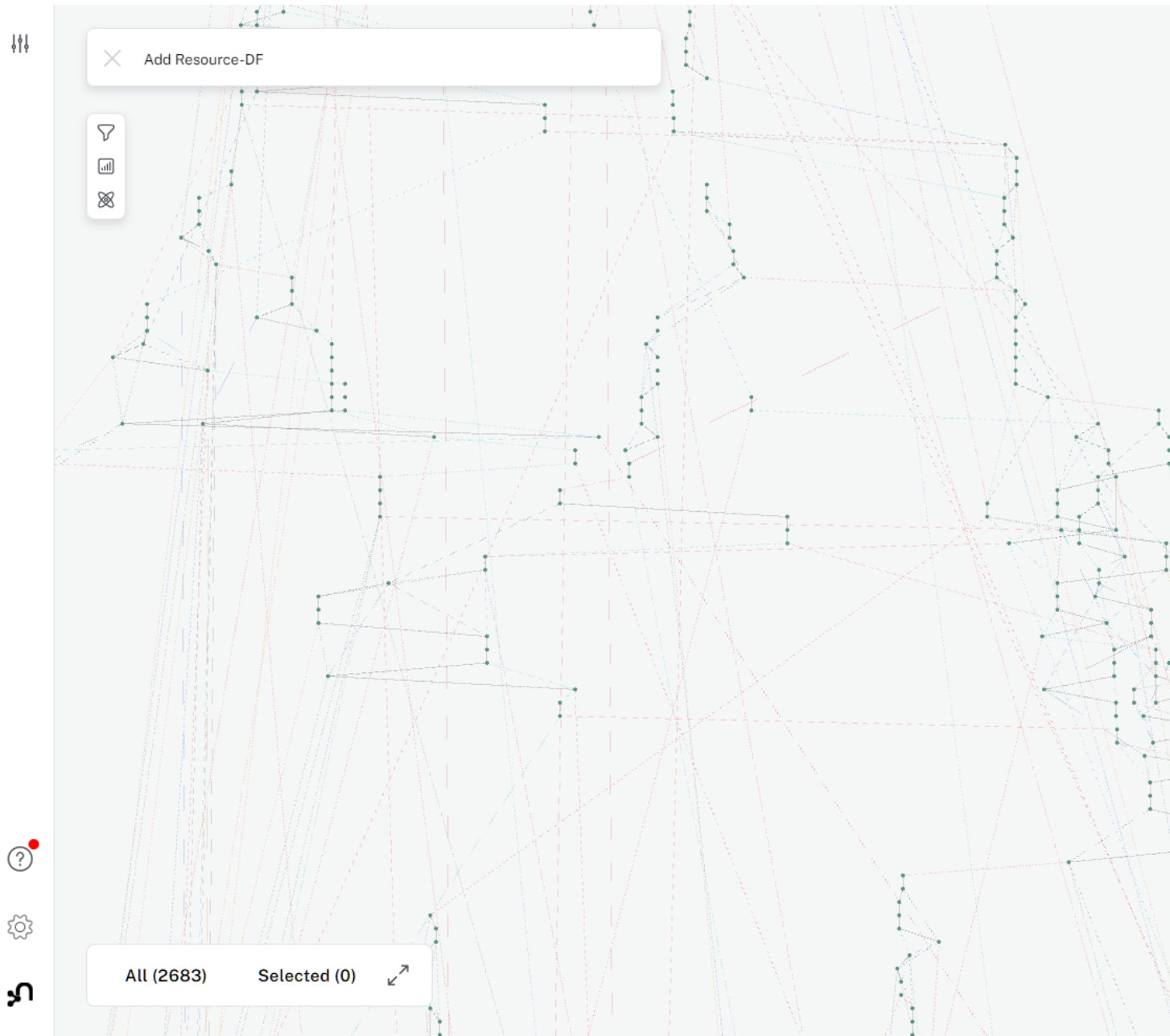
```
MATCH (e:Event) -[df {type:"DF"}]-> (e2:Event)  
RETURN e,df,e2
```

Events with DF relations

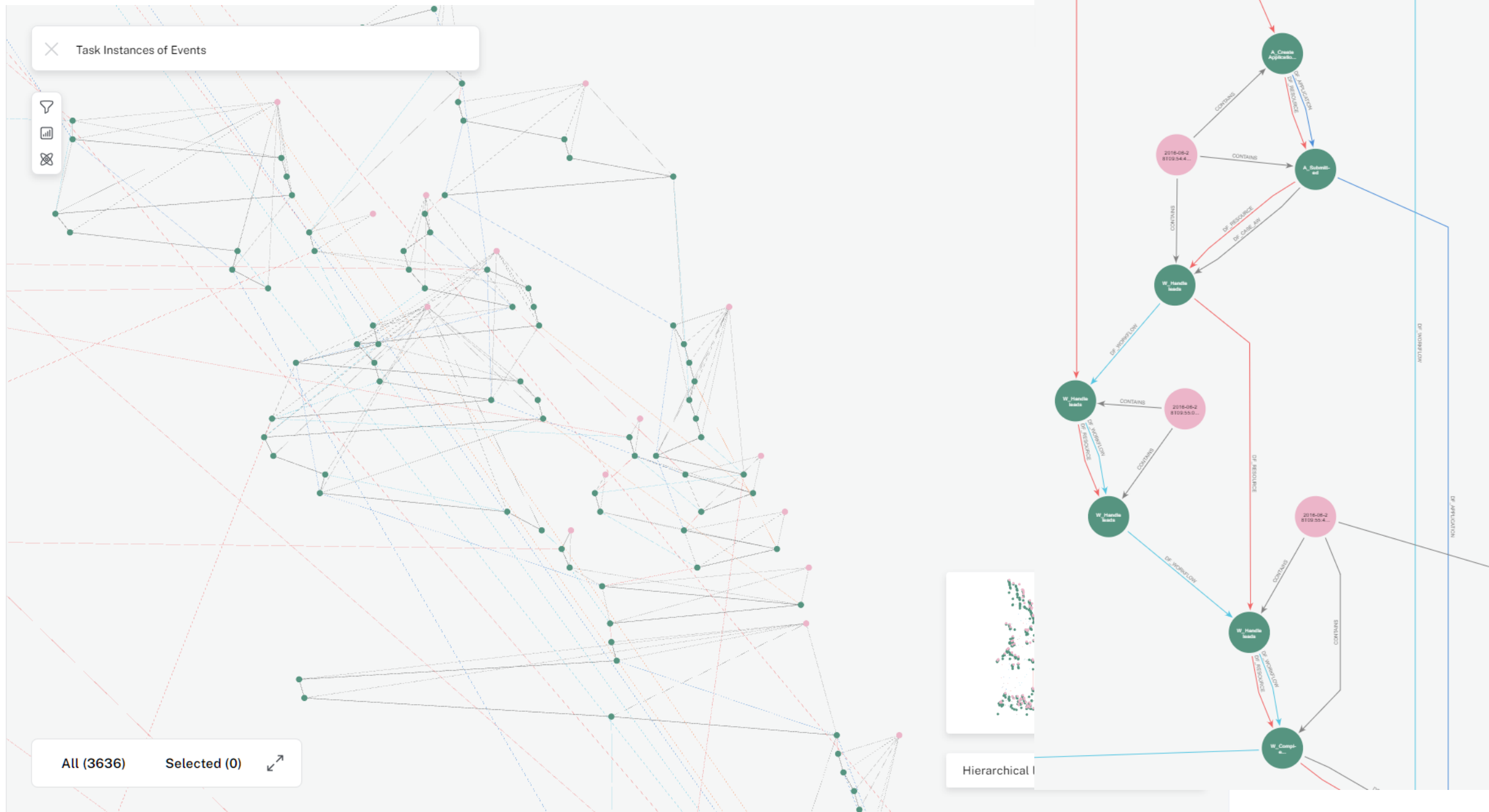




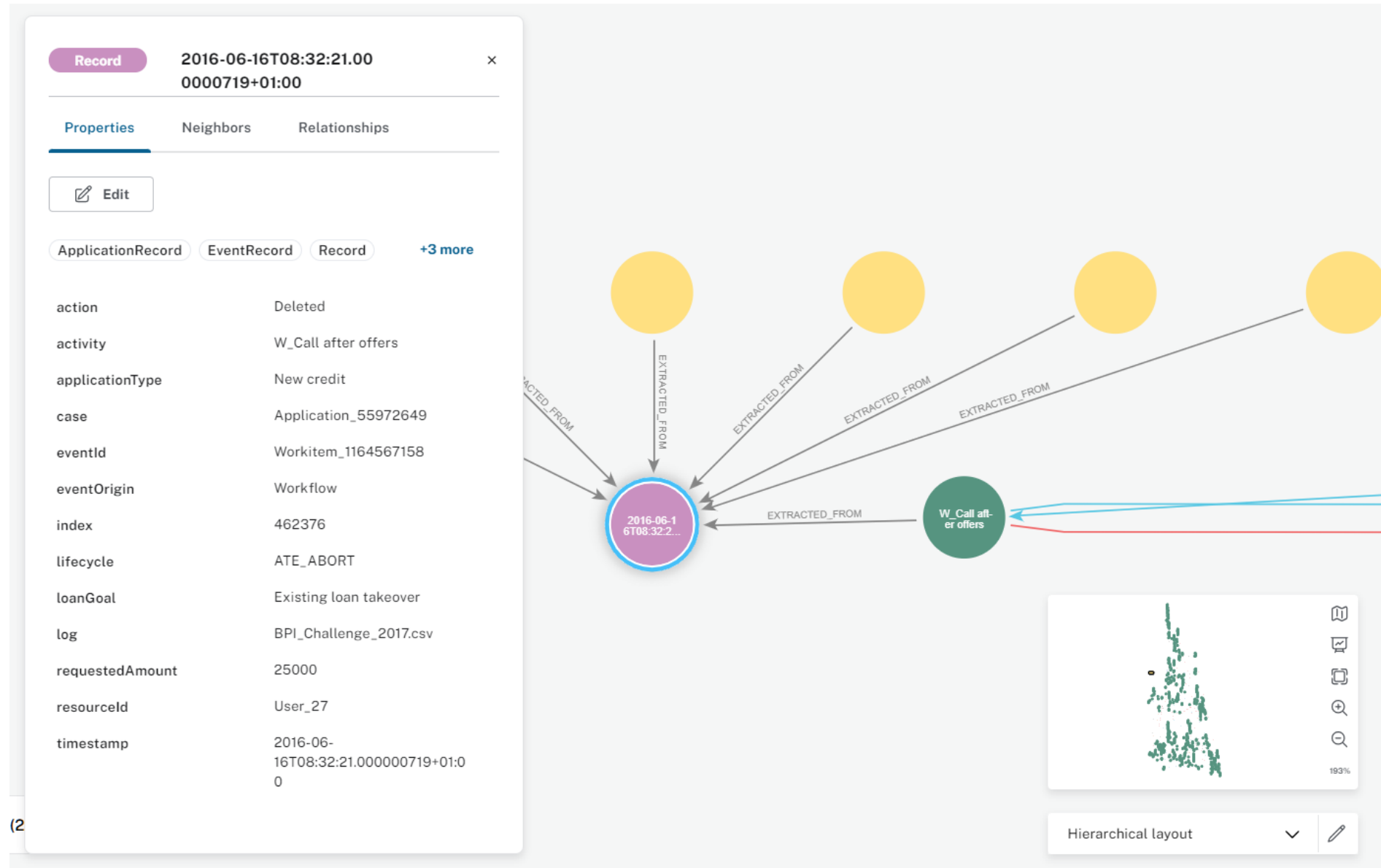
# Demo: BPIC'17 – add resource behavior



# Demo: BPIC'17 – extend: aggregate to tasks



# Demo: BPIC'17 – Data Lineage for OCED



# OCED-PG

- **implemented full OCED:**  
7 industrial case studies, 5 public BPIC datasets as OCED
- **domain knowledge for OCED construction as concrete artifact:**  
**ontologies > standardize semantic header?**
- **start pragmatic, start now,** but:  
more **semantic depth of OCED > more research** to do it right:
  - **qualifiers on what is added/delete exactly**
  - **relate to neighboring communities:** data management, semantic web, temporal graphs, efficient storage, query languages, ...

