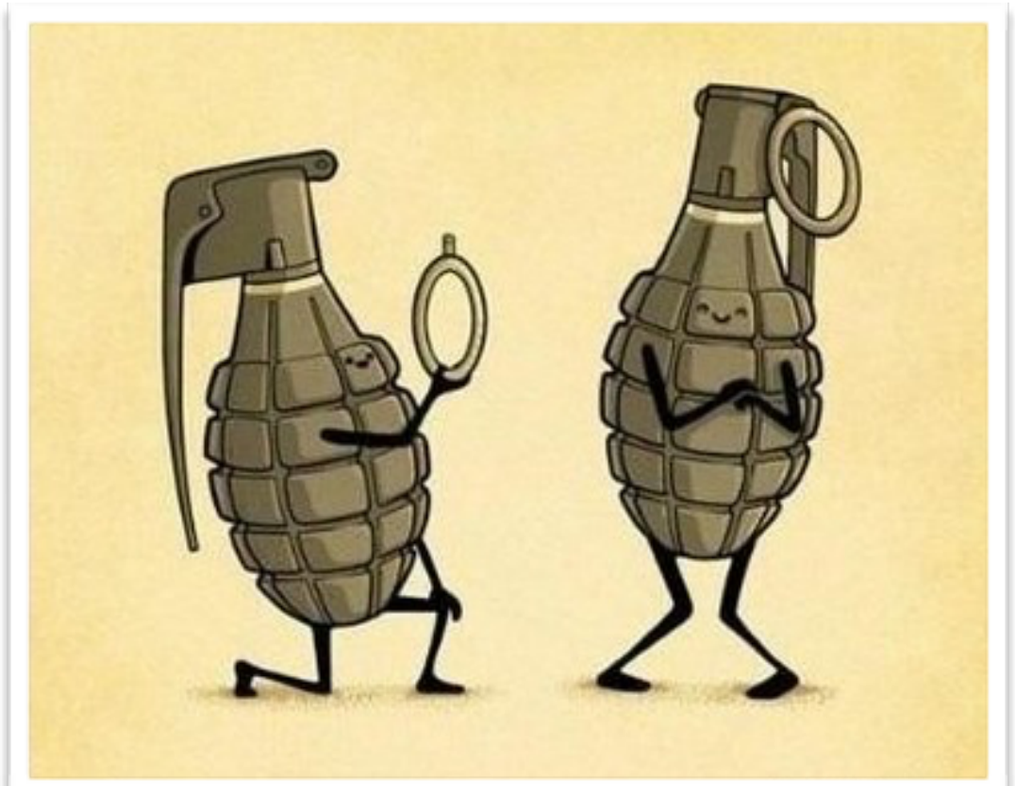
Towards Data-Aware Dynamic Systems



Marco Montali

Free University of Bozen-Bolzano

Our Starting Point

Marrying processes and data is a must if we want to really *understand* how complex dynamic systems operate

Dynamic systems of interest:

- business processes
- multiagent systems
- distributed systems

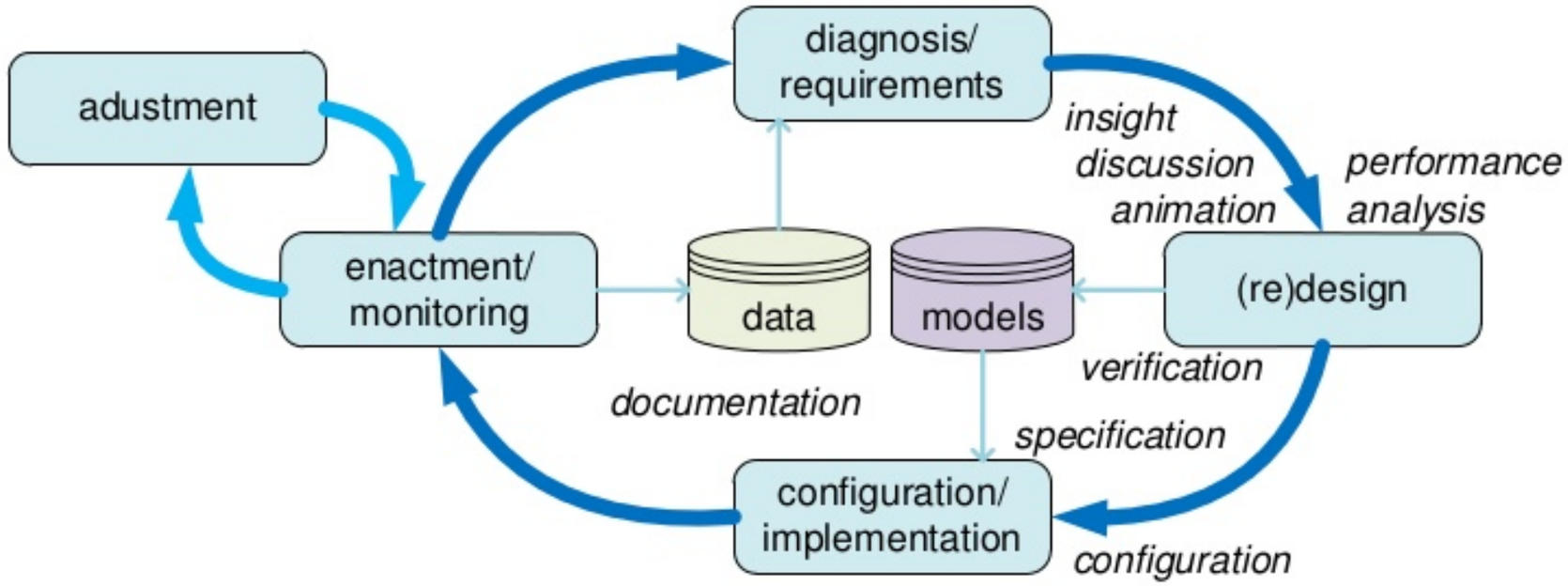
Our Thesis

Knowledge representation and computational logics

can become a swiss-army knife to

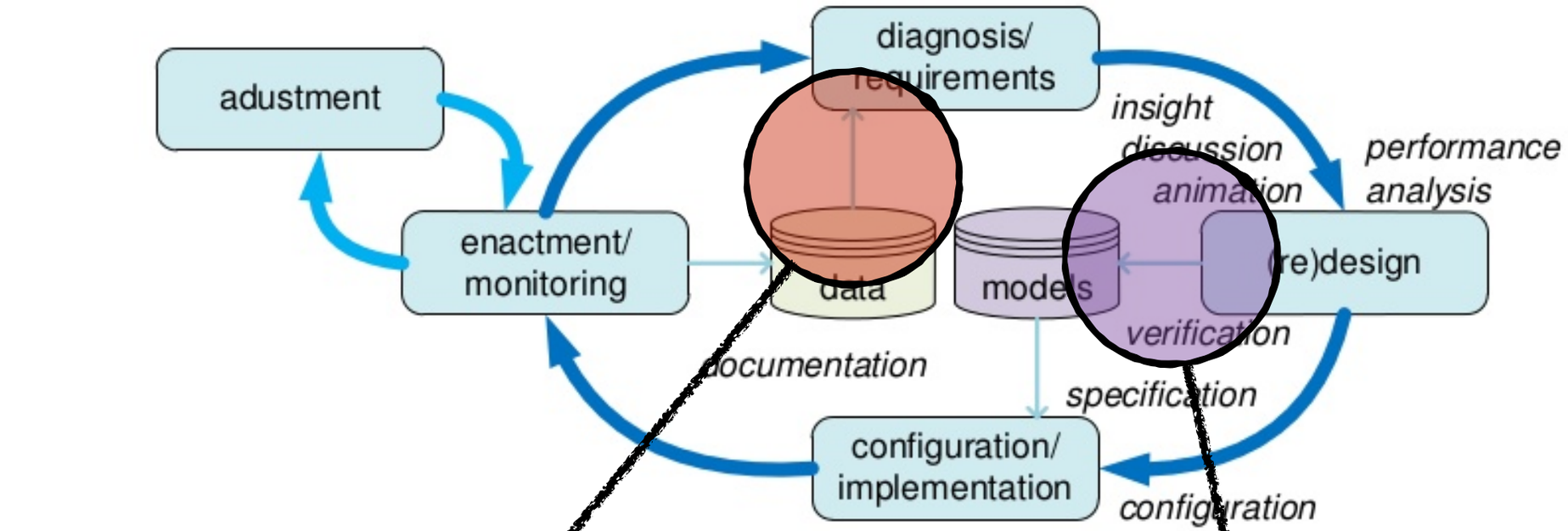
understand data-aware dynamic systems,
and
provide automated reasoning and verification capabilities along their entire lifecycle

Complex Systems Lifecycle



picture by Wil van der Aalst

Today…

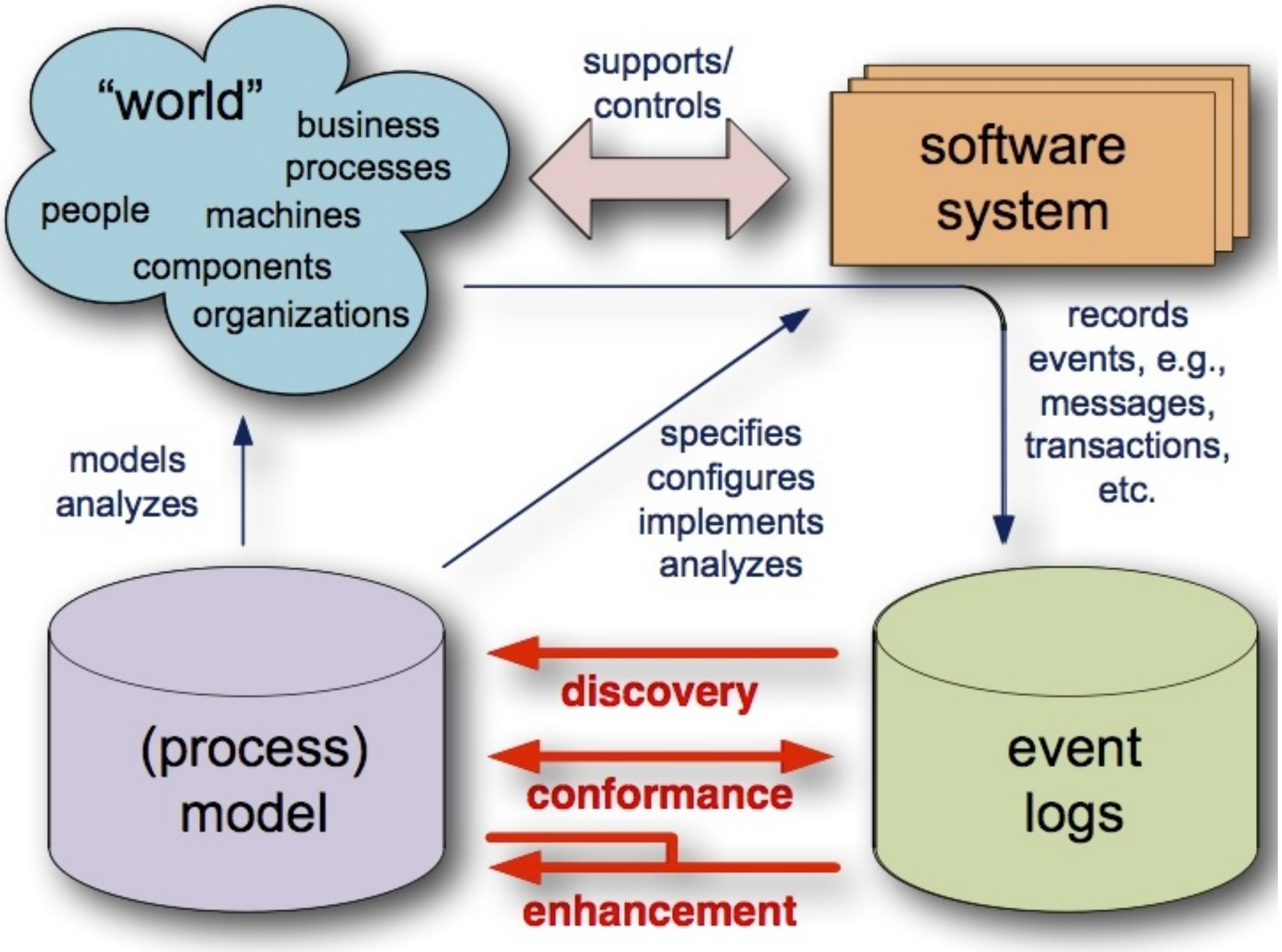


picture by Wil van der Aalst

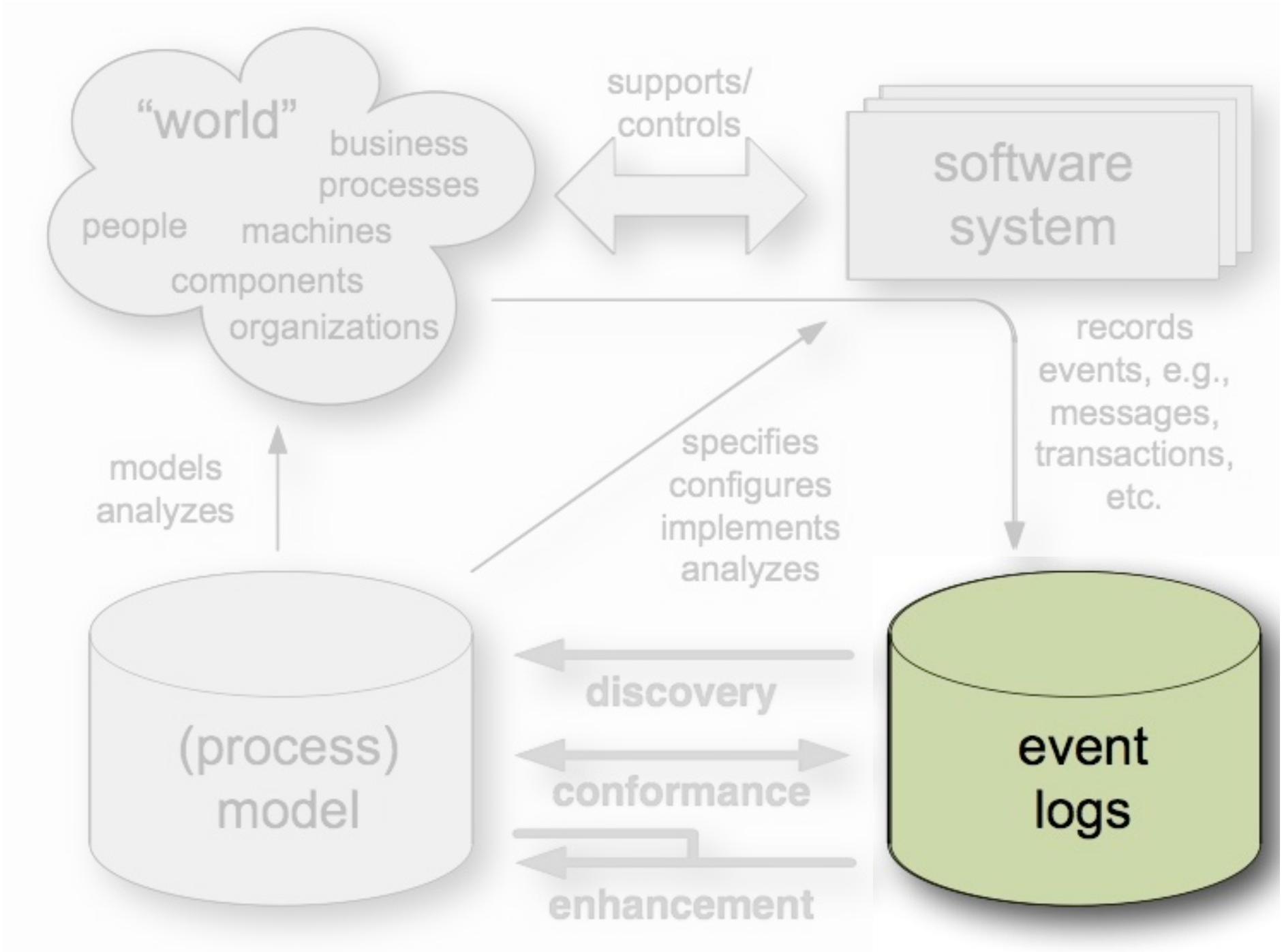
Data preparation for process mining

Commitments

Process Mining



Process Mining



Expected Reality

log

Case ID	Activity	Timestamp	Resource
1	Submit paper	01-07-2015:10.05	Anna
	Invite reviewer	03-07-2015:12.00	Budi
	Get review	10-07-2015:16.10	Rudy
	Acceptance	12-07-2015:15.00	Rudy
	Submit final paper	19-07-2015:19.15	Anna
2	Submit paper	02-07-2015:17.19	John
	Invite reviewer	03-07-2015:12.00	Tiara
	Get review	11-07-2015:15.45	Clara
	Rejection	12-07-2015:15.00	Clara

trace

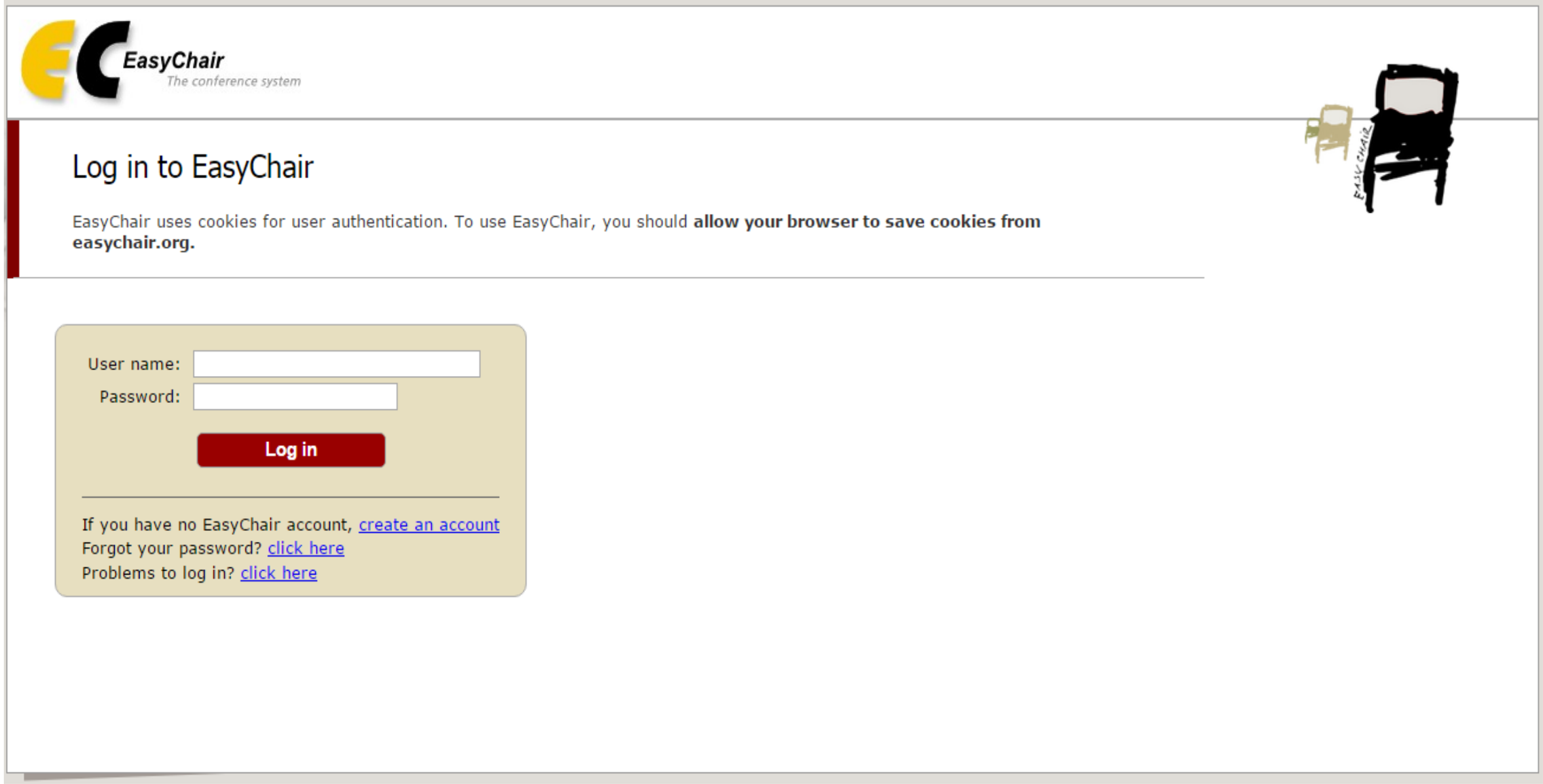
event

Expected Reality

- XES Standard

```
<log xes.version="1.0" xes.features="nested-attributes">
<trace>
   <string key=“concept:name” value=“1” />
   <event>
       <string key=“concept:name” value=“register request” />
       <date key=“time:timestamp” value=“2010-12-30T11:02:00.000+01:00” />
   </event>
</trace>
```

Actual Reality



Actual Reality

LOGIN

ID	User
1	Alifah Syamsiyah
2	Marco Montali
3	Diego Calvanese
4	Wil van der Aalst

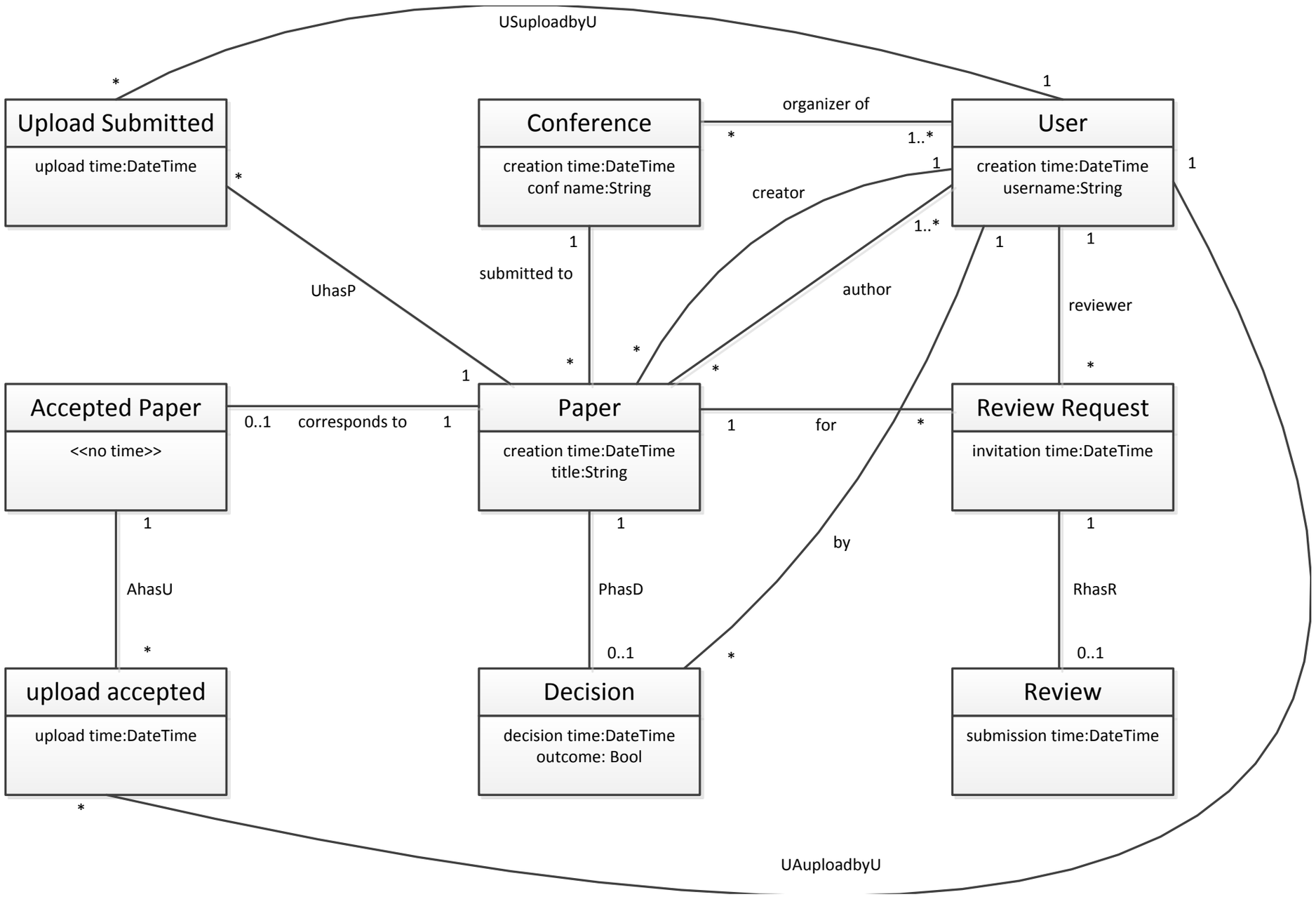
CONFERENCE

ID	Name	Organizer	Time
666	BPM 2015	2	2015-02-14 01:00:00
667	Caise 2015	4	2015-03-06 01:00:00
668	ER 2015	4	2015-03-26 01:00:00
669	EDOC 2015	2	2015-04-05 03:00:00

PAPER

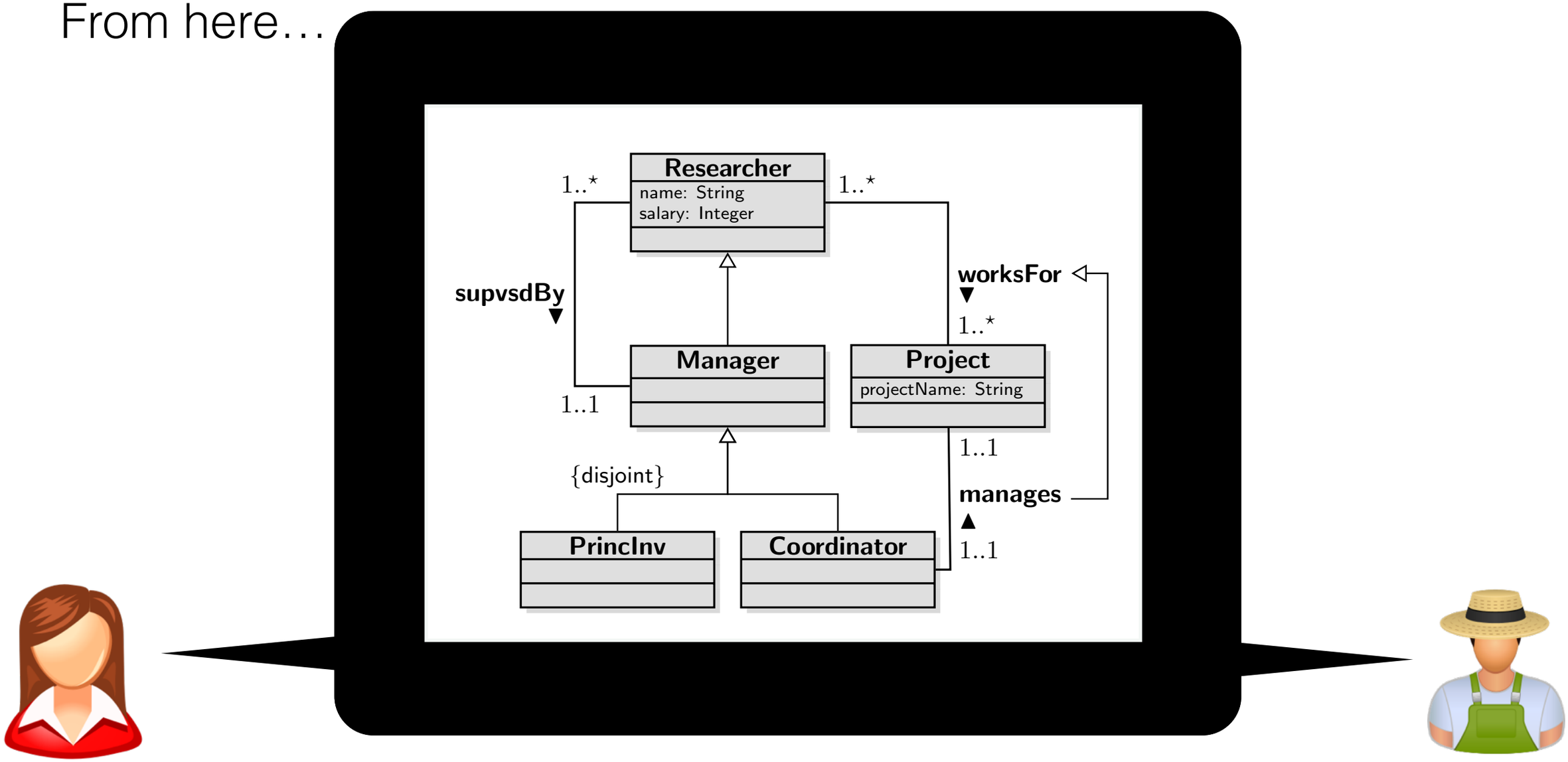
ID	Title	CT	User	Conf	Type	Status
1	Ontop at Work	2015-03-02 15:09:35	1	669	FP	RX
2	A Survey of Web Services	2015-03-02 12:36:01	3	668	SP	RX
3	The Definitive Guide for BPM	2015-03-04 13:36:20	1	666	FP	AB

Understanding Reality…



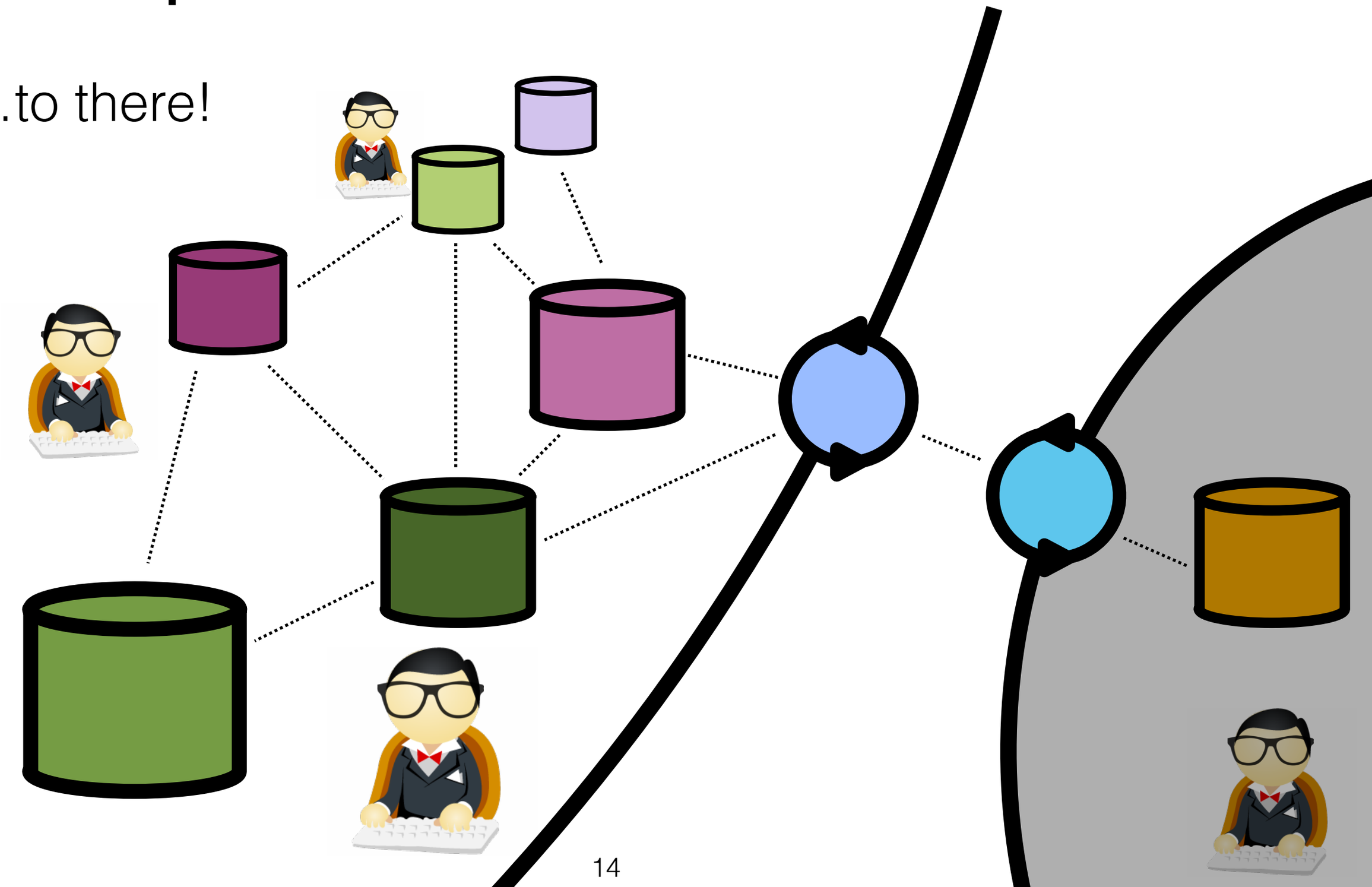
Impedance Mismatch

From here…



Impedance Mismatch

…to there!



LOGIN

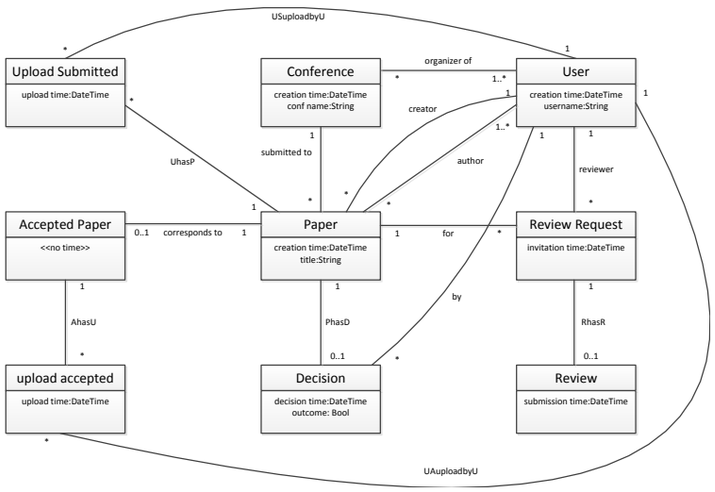
ID	User
1	Alifah Syamsiyah
2	Marco Montali
3	Diego Calvanese
4	Wil van der Aalst

CONFERENCE

ID	Name	Organizer	Time
666	BPM 2015	2	2015-02-14 01:00:00
667	Caise 2015	4	2015-03-06 01:00:00
668	ER 2015	4	2015-03-26 01:00:00
669	EDOC 2015	2	2015-04-05 03:00:00

PAPER

ID	Title	CT	User	Conf	Type	Status
1	Ontop at Work	2015-03-02 15:09:35	1	669	FP	RX
2	A Survey of Web Services	2015-03-02 12:36:01	3	668	SP	RX
3	The Definitive Guide for BPM	2015-03-04 13:36:20	1	666	FP	AB



- How to resolve the “impedance mismatch”?
- How to get a “view” of such data tailored to process mining?

Impedance Mismatch is Really an Issue

Crompton (2008): domain experts **loose too much time to big into data** and turn them into knowledge

- Engineers in the oil/gas industry: 30-70% of their working time spent for **data searching** and **data quality**

Optique

- **Scalable, End-User Access to Big Data)**
- http://optique-project.eu
- Goal: engineer techniques for accessing data through domain ontologies
- Case studies: Statoil, Siemens

Facts on Statoil

- **1000 TB of dati** inside relational DBMSs
- Schemas **not aligned**
- More than **2000 tables**, in a plethora of different DBs
- **900 experts** part of “Statoil Exploration”
  - Up to **4 days to formulate queries and encode them in SQL**

Query Example

Show all norwegian wellbores with some aditional attributes (wellbore id, completion date, oldest penetrated age,result). Limit to all wellbores with a core and show attributes like (wellbore id, core number, top core depth, base core depth, intersecting stratigraphy). Limit to all wellbores with core in Brentgruppen and show key atributes in a table. After connecting to EPDS (slegge) we could for instance limit futher to cores in Brent with measured permeability and where it is larger than a given value, for instance 1 mD. We could also find out whether there are cores in Brent which are not stored in EPDS (based on NPD info) and where there could be permeability values. Some of the missing data we possibly own, other not.

```
SELECT [...]
FROM
db_name.table1 table1,
db_name.table2 table2a,
db_name.table2 table2b,
db_name.table3 table3a,
db_name.table3 table3b,
db_name.table3 table3c,
db_name.table3 table3d,
db_name.table4 table4a,
db_name.table4 table4b,
db_name.table4 table4c,
db_name.table4 table4d,
db_name.table4 table4e,
db_name.table4 table4f,
db_name.table5 table5a,
db_name.table5 table5b,
db_name.table6 table6a,
db_name.table6 table6b,
db_name.table7 table7a,
db_name.table7 table7b,
db_name.table8 table8,
db_name.table9 table9,
db_name.table10 table10a,
db_name.table10 table10b,
db_name.table10 table10c,
db_name.table11 table11,
db_name.table12 table12,
db_name.table13 table13,
db_name.table14 table14,
db_name.table15 table15,
db_name.table16 table16
WHERE [...]
```

```
table2a.attr1='keyword' AND
table3a.attr2=table10c.attr1 AND
table3a.attr6=table6a.attr3 AND
table3a.attr9='keyword' AND
table4a.attr10 IN ('keyword') AND
table4a.attr1 IN ('keyword') AND
table5a.kinds=table4a.attr13 AND
table5b.kinds=table4c.attr74 AND
table5b.name='keyword' AND
(table6a.attr19=table10c.attr17 OR
(table6a.attr2 IS NULL AND
table10c.attr4 IS NULL)) AND
table6a.attr14=table5b.attr14 AND
table6a.attr2='keyword' AND
(table6b.attr14=table10c.attr8 OR
(table6b.attr4 IS NULL AND
table10c.attr7 IS NULL)) AND
table6b.attr19=table5a.attr55 AND
table6b.attr2='keyword' AND
table7a.attr19=table2b.attr19 AND
table7a.attr17=table15.attr19 AND
table4b.attr11='keyword' AND
table8.attr19=table7a.attr80 AND
table8.attr19=table13.attr20 AND
table8.attr4='keyword' AND
table9.attr10=table16.attr11 AND
table3b.attr19=table10c.attr18 AND
table3b.attr22=table12.attr63 AND
table3b.attr66='keyword' AND
table10a.attr54=table7a.attr8 AND
table10a.attr70=table10c.attr10 AND
table10a.attr16=table4d.attr11 AND
table4c.attr99='keyword' AND
table4c.attr1='keyword' AND
```

```
table11.attr10=table5a.attr10 AND
table11.attr40='keyword' AND
table11.attr50='keyword' AND
table2b.attr1=table1.attr8 AND
table2b.attr9 IN ('keyword') AND
table2b.attr2 LIKE 'keyword'% AND
table12.attr9 IN ('keyword') AND
table7b.attr1=table2a.attr10 AND
table3c.attr13=table10c.attr1 AND
table3c.attr10=table6b.attr20 AND
table3c.attr13='keyword' AND
table10b.attr16=table10a.attr7 AND
table10b.attr11=table7b.attr8 AND
table10b.attr13=table4b.attr89 AND
table13.attr1=table2b.attr10 AND
table13.attr20=''keyword'' AND
table13.attr15='keyword' AND
table3d.attr49=table12.attr18 AND
table3d.attr18=table10c.attr11 AND
table3d.attr14='keyword' AND
table4d.attr17 IN ('keyword') AND
table4d.attr19 IN ('keyword') AND
table16.attr28=table11.attr56 AND
table16.attr16=table10b.attr78 AND
table16.attr5=table14.attr56 AND
table4e.attr34 IN ('keyword') AND
table4e.attr48 IN ('keyword') AND
table4f.attr89=table5b.attr7 AND
table4f.attr45 IN ('keyword') AND
table4f.attr1='keyword' AND
table10c.attr2=table4e.attr19 AND
(table10c.attr78=table12.attr56 OR
(table10c.attr55 IS NULL AND
table12.attr17 IS NULL))
```

```
SELECT [...]
FROM
db_name.table1 table1,
db_name.table2 table2a,
db_name.table2 table2b,
db_name.table3 table3a,
db_name.table3 table3b,
db_name.table3 table3c,
db_name.table3 table3d,
db_name.table4 table4a,
db_name.table4 table4b,
db_name.table10 table10b,
db_name.table10 table10c,
db_name.table11 table11,
db_name.table12 table12,
db_name.table13 table13,
db_name.table14 table14,
db_name.table15 table15,
db_name.table16 table16
WHERE [...]
```

```
table2a.attr1='keyword' AND
table3a.attr2=table10c.attr1 AND
table3a.attr6=table6a.attr3 AND
table3a.attr9='keyword' AND
table4a.attr10 IN ('keyword') AND
table4a.attr1 IN ('keyword') AND
table5a.kinds=table4a.attr13 AND
table5b.kinds=table4c.attr74 AND
table5b.name='keyword' AND
(table6a.attr19=table10c.attr17 OR
(table6a.attr2 IS NULL AND
table8.attr4='keyword' AND
table9.attr10=table16.attr11 AND
table3b.attr19=table10c.attr18 AND
table3b.attr22=table12.attr63 AND
table3b.attr66='keyword' AND
table10a.attr54=table7a.attr8 AND
table10a.attr70=table10c.attr10 AND
table10a.attr16=table4d.attr11 AND
table4c.attr99='keyword' AND
table4c.attr1='keyword' AND
```

```
table11.attr10=table5a.attr10 AND
table11.attr40='keyword' AND
table11.attr50='keyword' AND
table2b.attr1=table1.attr8 AND
table2b.attr9 IN ('keyword') AND
table2b.attr2 LIKE 'keyword'% AND
table12.attr9 IN ('keyword') AND
table7b.attr1=table2a.attr10 AND
table3c.attr13=table10c.attr1 AND
table3c.attr10=table6b.attr20 AND
table3c.attr13='keyword' AND
table16.attr5=table14.attr56 AND
table4e.attr34 IN ('keyword') AND
table4e.attr48 IN ('keyword') AND
table4f.attr89=table5b.attr7 AND
table4f.attr45 IN ('keyword') AND
table4f.attr1='keyword' AND
table10c.attr2=table4e.attr19 AND
(table10c.attr78=table12.attr56 OR
(table10c.attr55 IS NULL AND
table12.attr17 IS NULL))
```

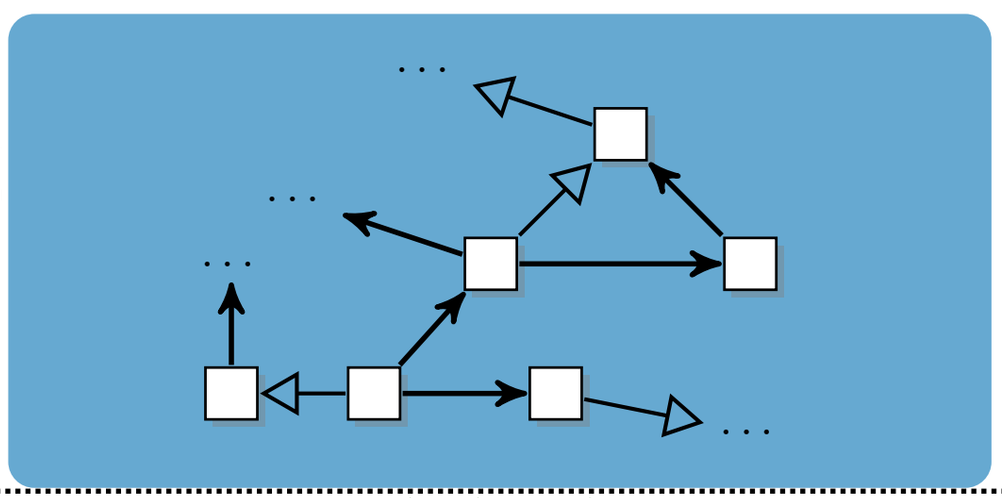
50.000.000

€/year

OBDA

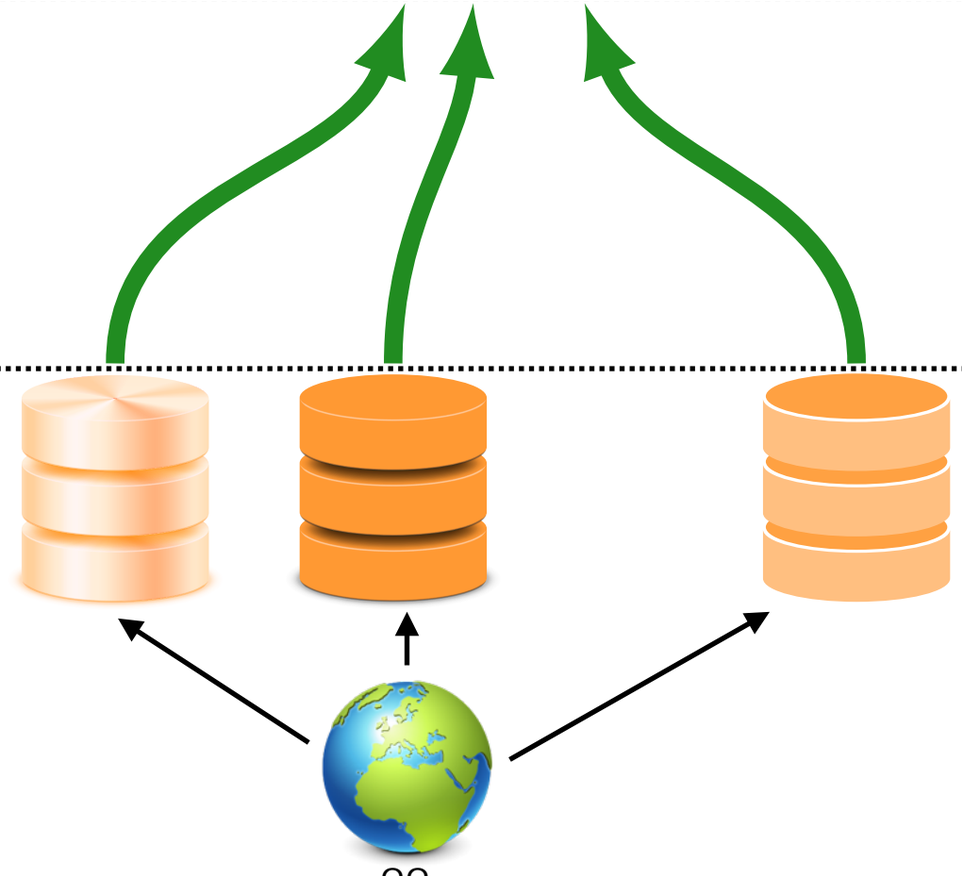
Query

Result



conceptual model

mapping



data sources

ontop

- Open-source OBDA technology developed here at UNIBZ (supervisor: Diego Calvanese)

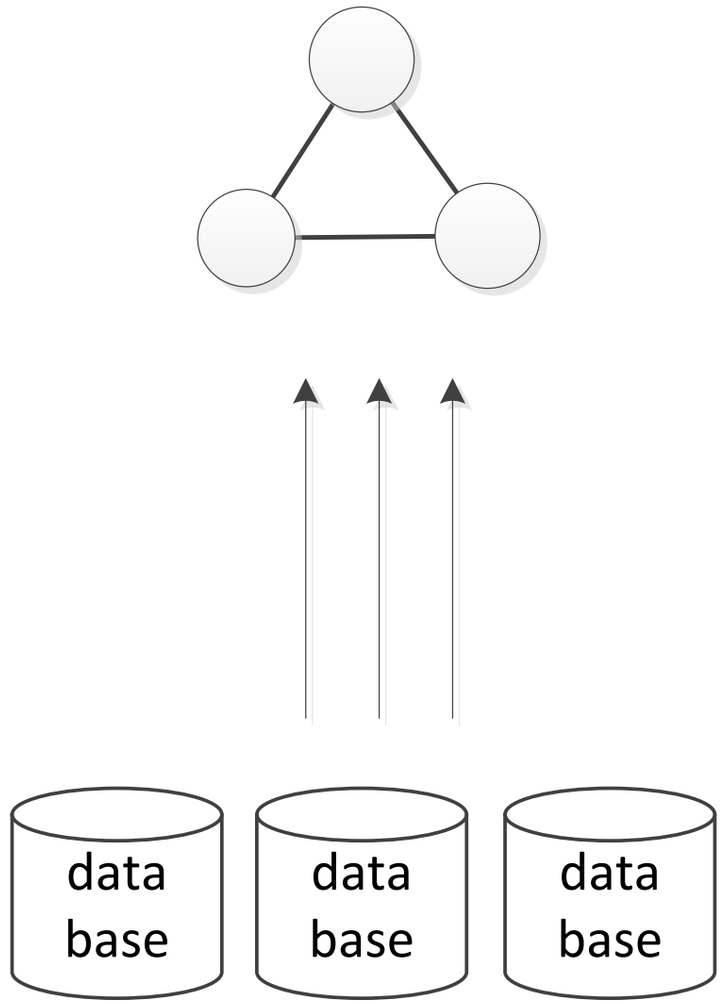
- Fully supports semantic web standards (OWL/ SPARQL)

- Integrates with many different relational DBMSs

- Apache open license

- http://ontop.inf.unibz.it

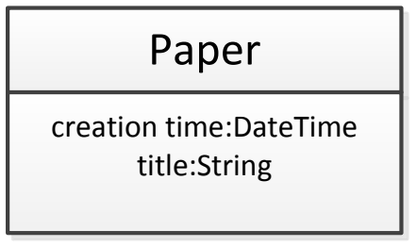
Resolving the Impedance Mismatch



Resolving the Impedance Mismatch

PAPER

ID	Title	CT	User	Conf	Type	Status
1	Ontop at Work	2015-03-02 15:09:35	1	669	FP	RX
2	A Survey of Web Services	2015-03-02 12:36:01	3	668	SP	RX
3	The Definitive Guide for BPM	2015-03-04 13:36:20	1	666	FP	AB



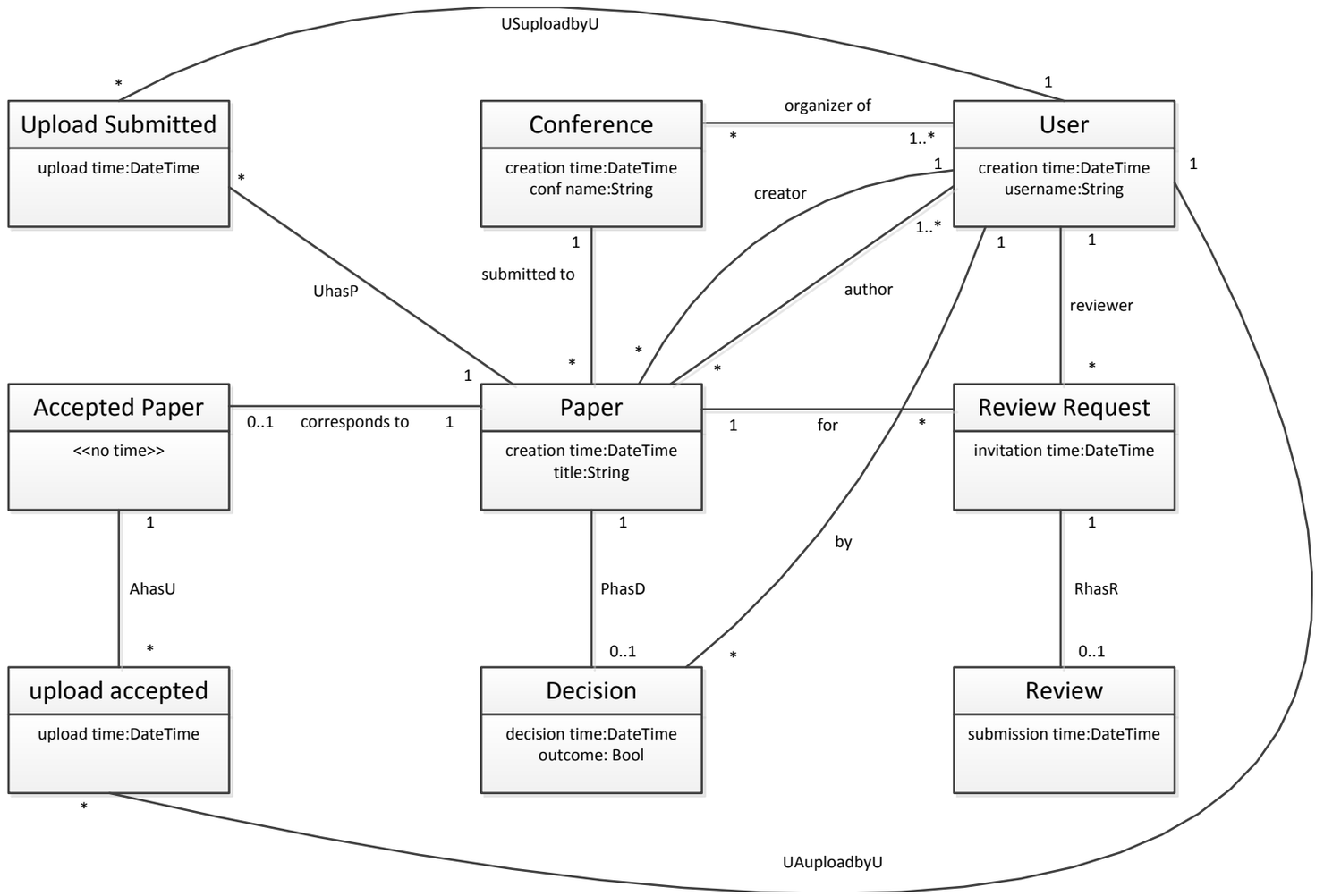

```
mappingId   id1
target      :paper{ID} a :Paper ; :title {Title} ; :creationtime {CT} .
source      select ID, Title, CT
            from Paper
```

What if my DB is Very Nice?

- Ontology bootstrapping automatically creates
  - a conceptual model that mirrors 1-1 the relational DB
  - identity mappings
- Useful for “small” case studies

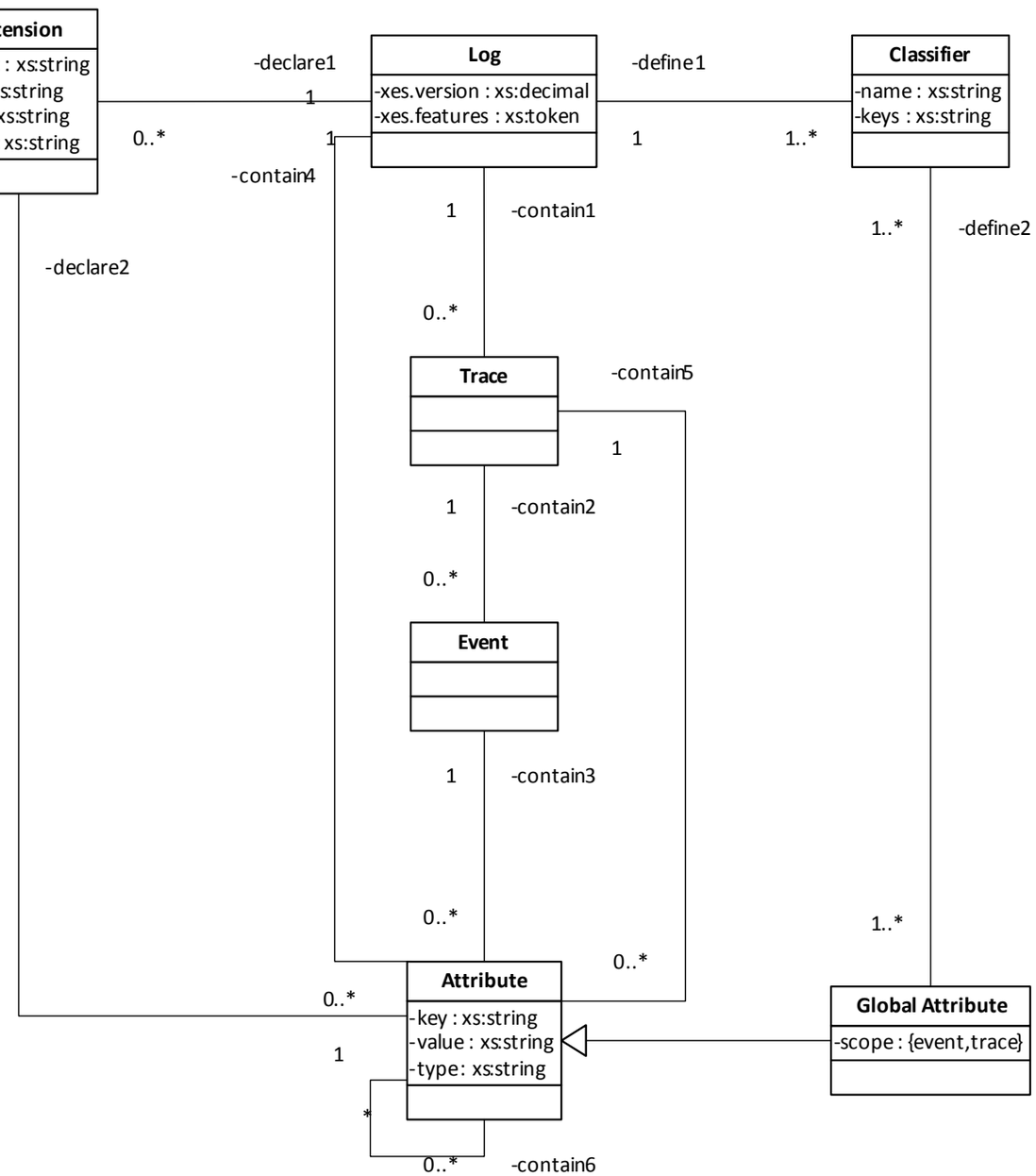
OBDA for Process Mining

- Need to resolve a second impedance mismatch problem!
- From here…



OBDA for Process Mining

- …To there!



OBDA for Process Mining

- From here…

LOGIN

ID	User
1	Alifah Syamsiyah
2	Marco Montali
3	Diego Calvanese
4	Wil van der Aalst

CONFERENCE

ID	Name	Organizer	Time
666	BPM 2015	2	2015-02-14 01:00:00
667	Caise 2015	4	2015-03-06 01:00:00
668	ER 2015	4	2015-03-26 01:00:00
669	EDOC 2015	2	2015-04-05 03:00:00

PAPER

ID	Title	CT	User	Conf	Type	Status
1	Ontop at Work	2015-03-02 15:09:35	1	669	FP	RX
2	A Survey of Web Services	2015-03-02 12:36:01	3	668	SP	RX
3	The Definitive Guide for BPM	2015-03-04 13:36:20	1	666	FP	AB

OBDA for Process Mining

- …To there!

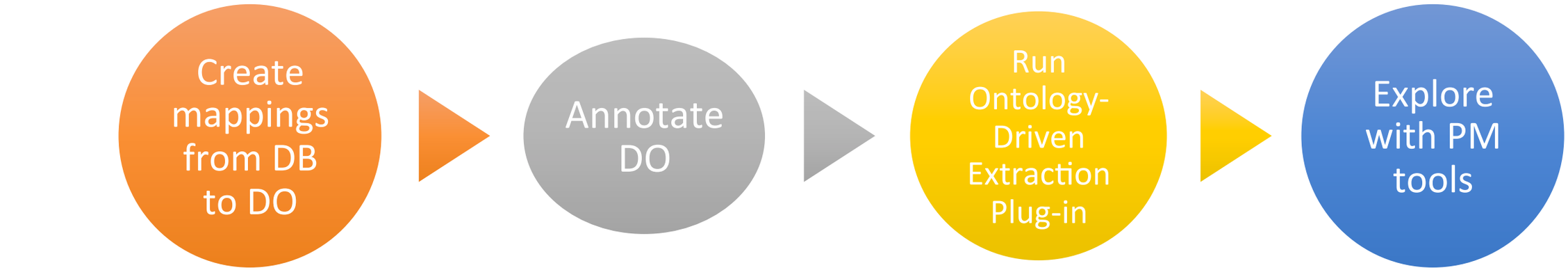
log

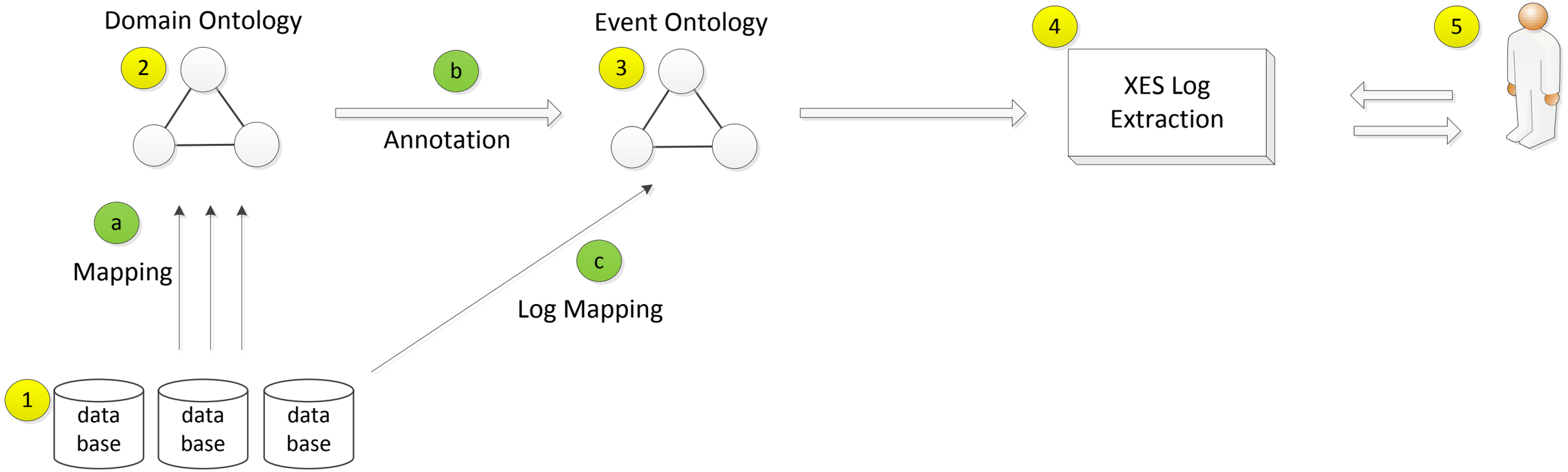
Case ID	Activity	Timestamp	Resource
1	Submit paper	01-07-2015:10.05	Anna
	Invite reviewer	03-07-2015:12.00	Budi
	Get review	10-07-2015:16.10	Rudy
	Acceptance	12-07-2015:15.00	Rudy
	Submit final paper	19-07-2015:19.15	Anna
2	Submit paper	02-07-2015:17.19	John
	Invite reviewer	03-07-2015:12.00	Tiara
	Get review	11-07-2015:15.45	Clara
	Rejection	12-07-2015:15.00	Clara

trace

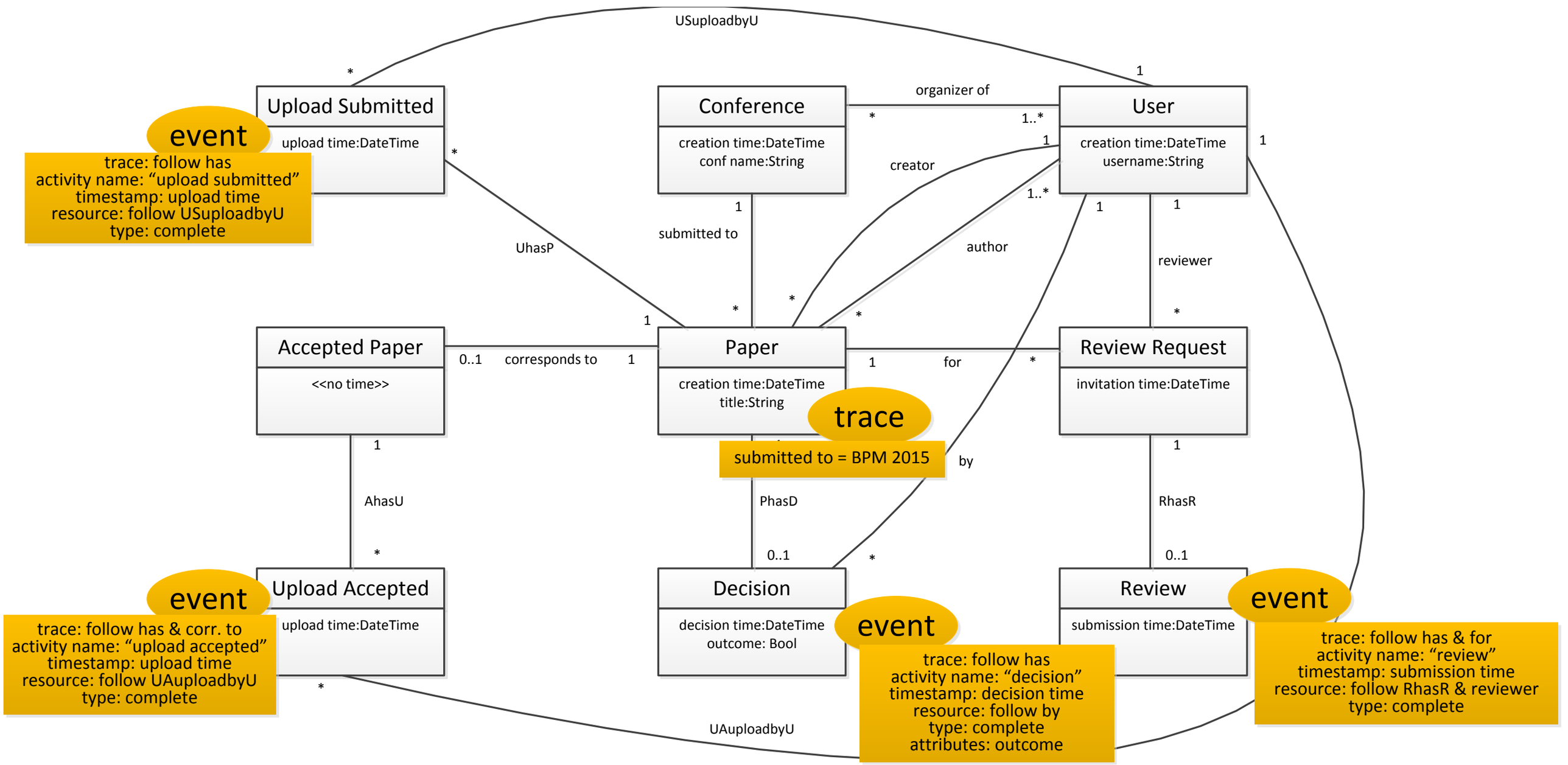
event

Our Framework

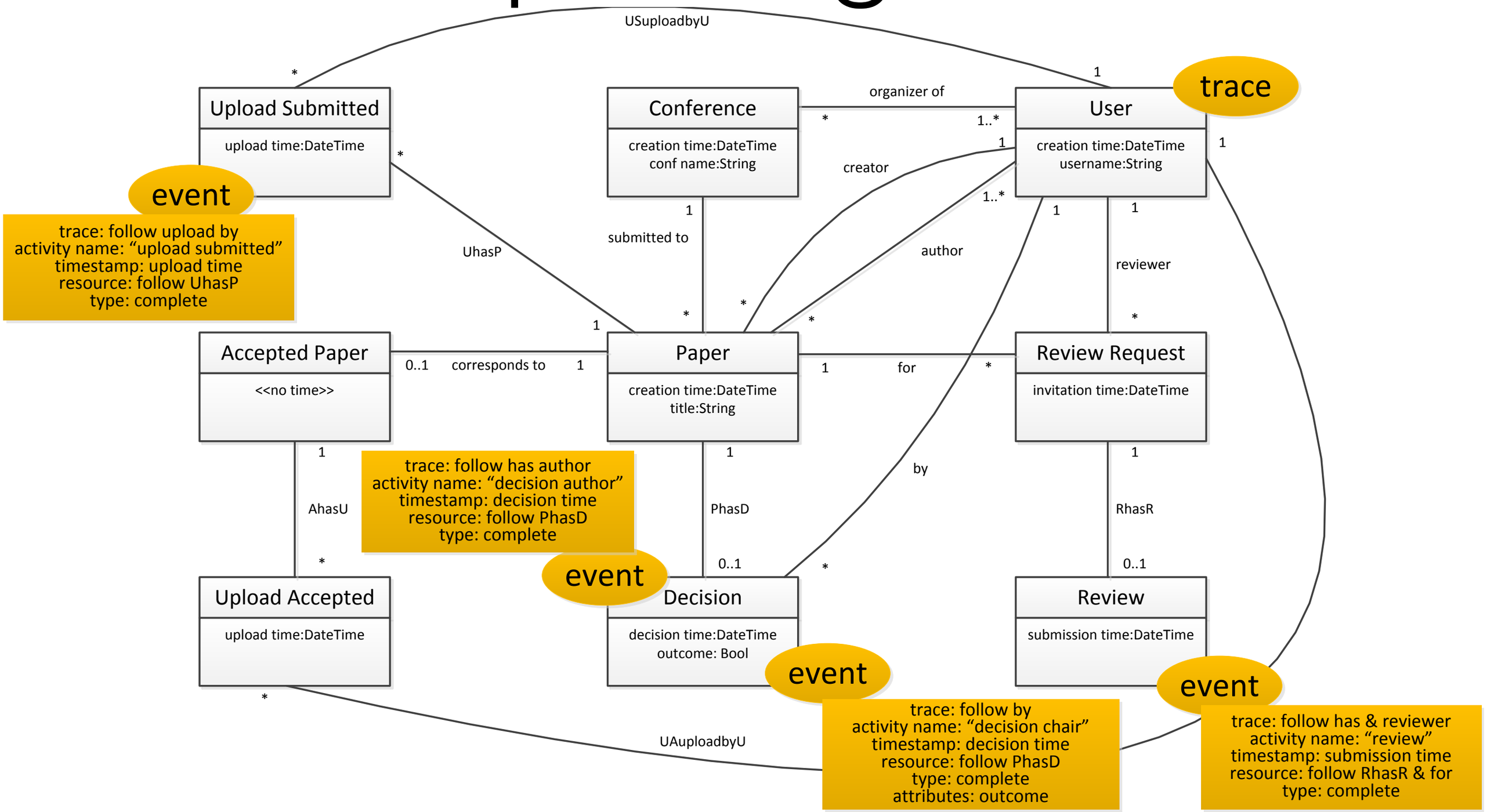




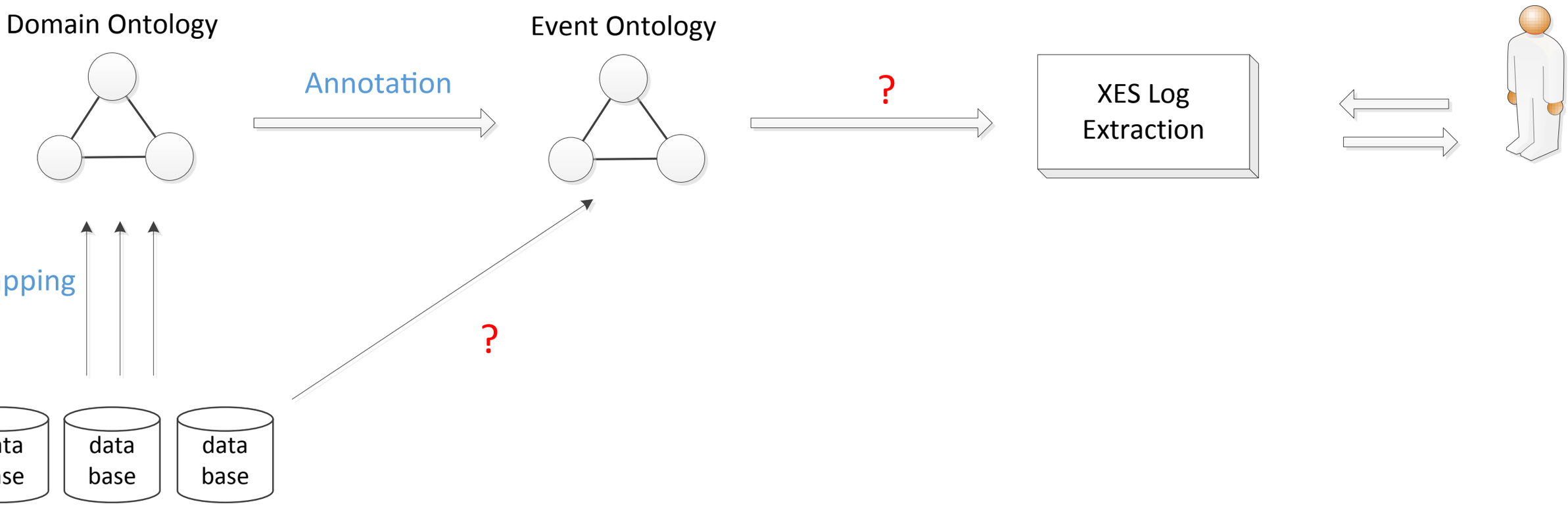
Log Annotations



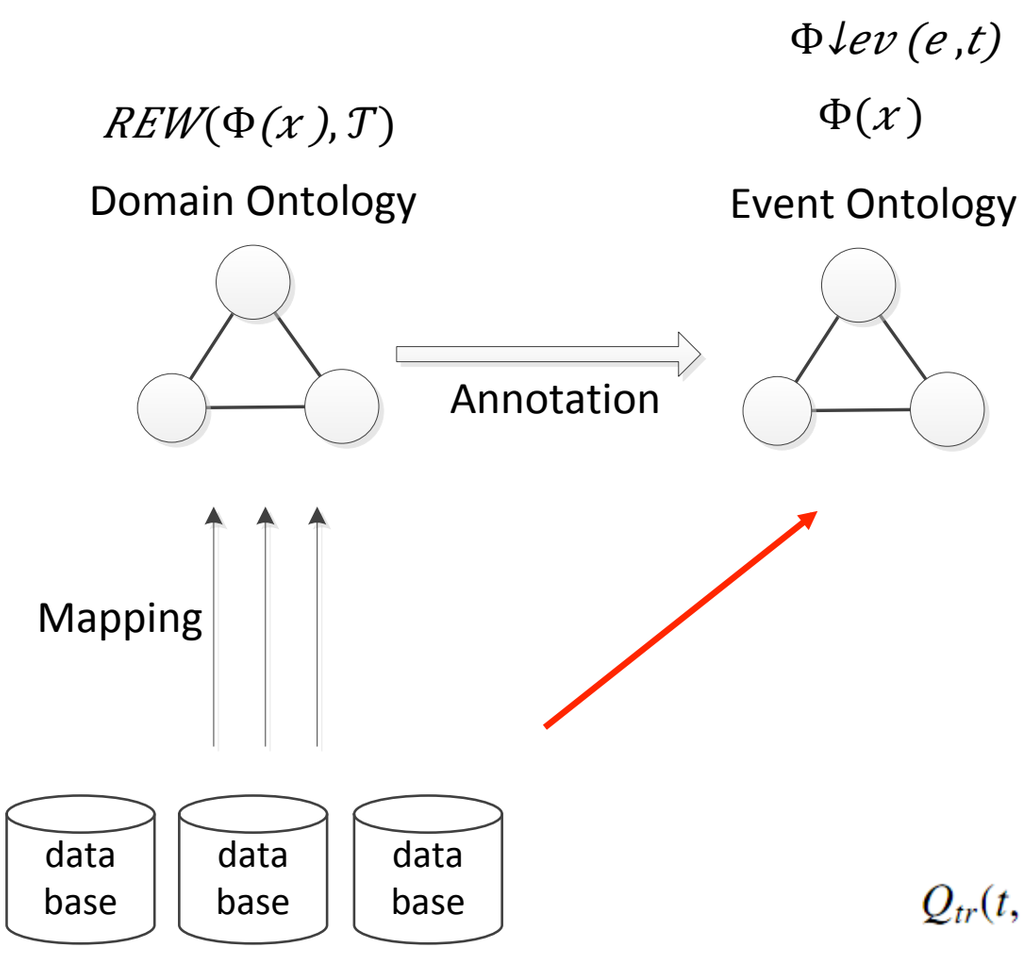
Multiple Log Views



Two Issues


Domain Ontology
Annotation
Event Ontology
?
XES Log Extraction
Mapping
?
data base
data base
data base

Synthesis of Log Mapping



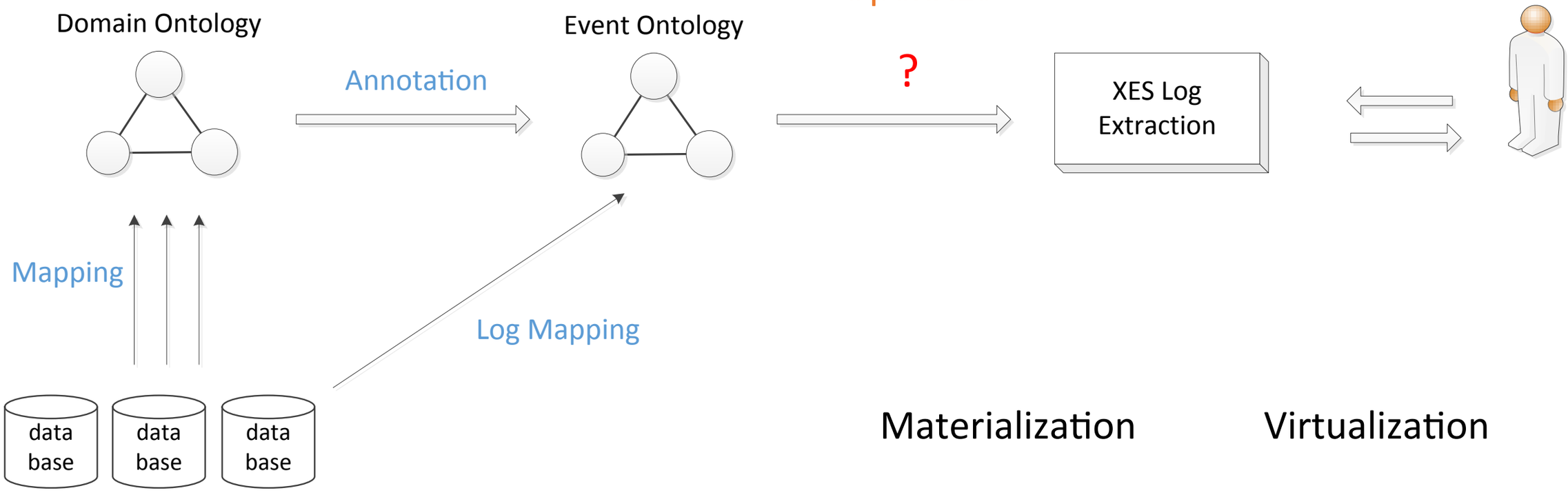
$$Q_{ev}(\vec{e}, t) \rightsquigarrow \left\{ \; TcontainsE(\mathsf{tr}(t), \mathsf{ev}(\vec{e})) \; \right\}$$

$$Q_{tr}(t, l) \rightsquigarrow \left\{ \begin{array}{l} LcontainsT(\mathsf{log}(l), \mathsf{tr}(t)), TcontainsA(\mathsf{tr}(t), \mathsf{attr}(t)), \\ key_{Attribute}(\mathsf{attr}(t), \texttt{"concept:name"}), \\ value_{Attribute}(\mathsf{attr}(t), t), type_{Attribute}(\mathsf{attr}(t), \texttt{"literal"}). \end{array} \right\}$$

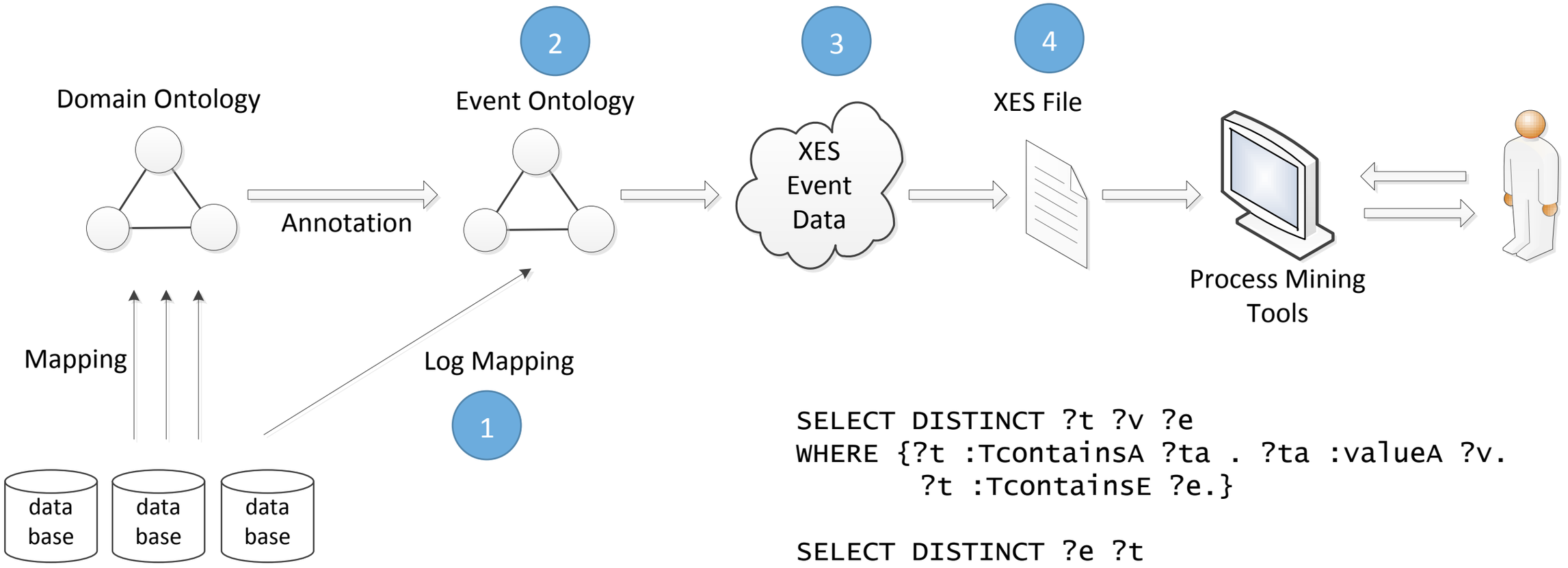
$$\mathcal{Q}(x) = UNFOLD(\, REW(\Phi(x), \mathcal{T}), \mathcal{M} \,)$$

$$\mathcal{Q}{\downarrow}ev\,(e,t) = UNFOLD\,(REW\,(\Phi{\downarrow}ev\,(e,t), \mathcal{T}), \mathcal{M})$$

XES Log Extraction

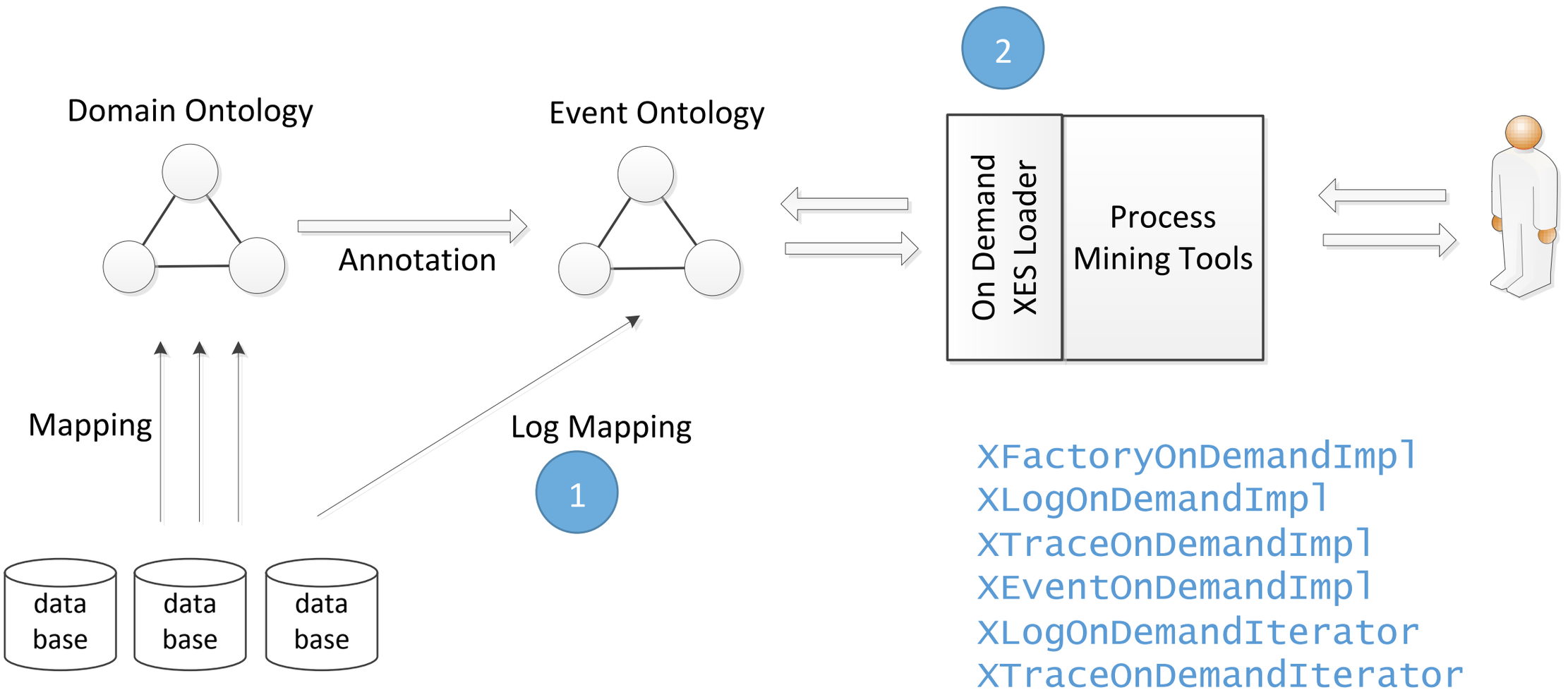


Materialized Log



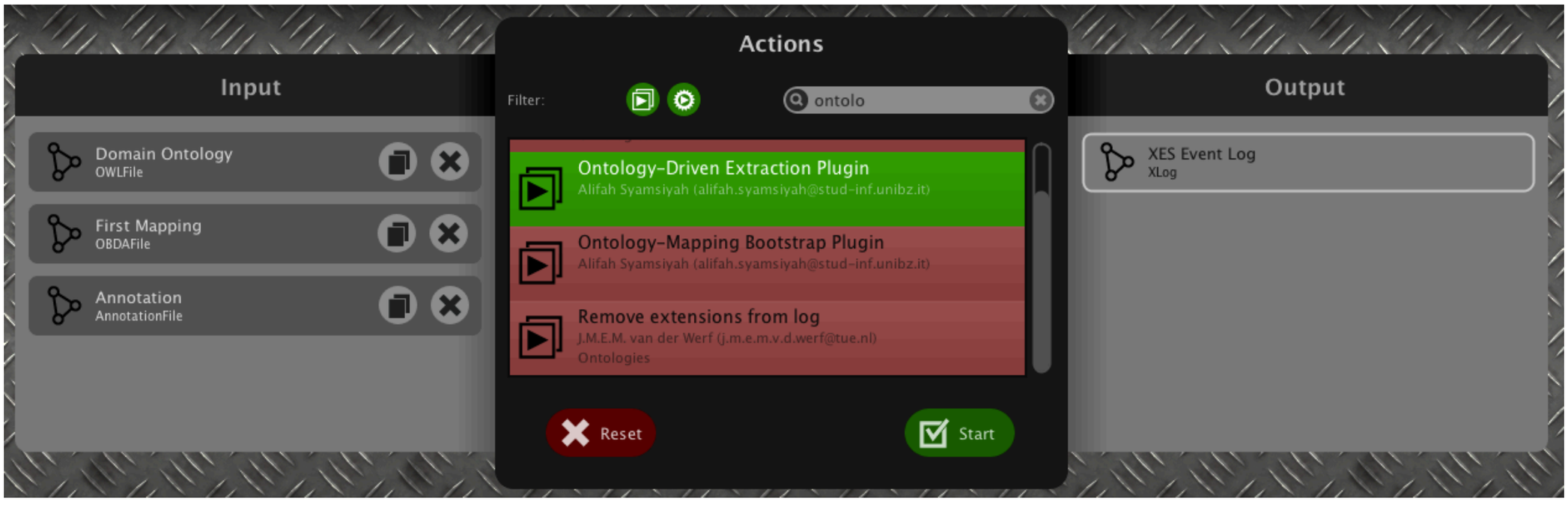

```
SELECT DISTINCT ?t ?v ?e
WHERE {?t :TcontainsA ?ta . ?ta :valueA ?v.
       ?t :TcontainsE ?e.}

SELECT DISTINCT ?e ?t
WHERE {?e :EcontainsA ?a . ?a :typeA ?t.}

SELECT DISTINCT ?e ?t
WHERE {?e :EcontainsA ?a . ?a :keyA ?t.}

SELECT DISTINCT ?e ?t
WHERE {?e :EcontainsA ?a . ?a :valueA ?t.}
```

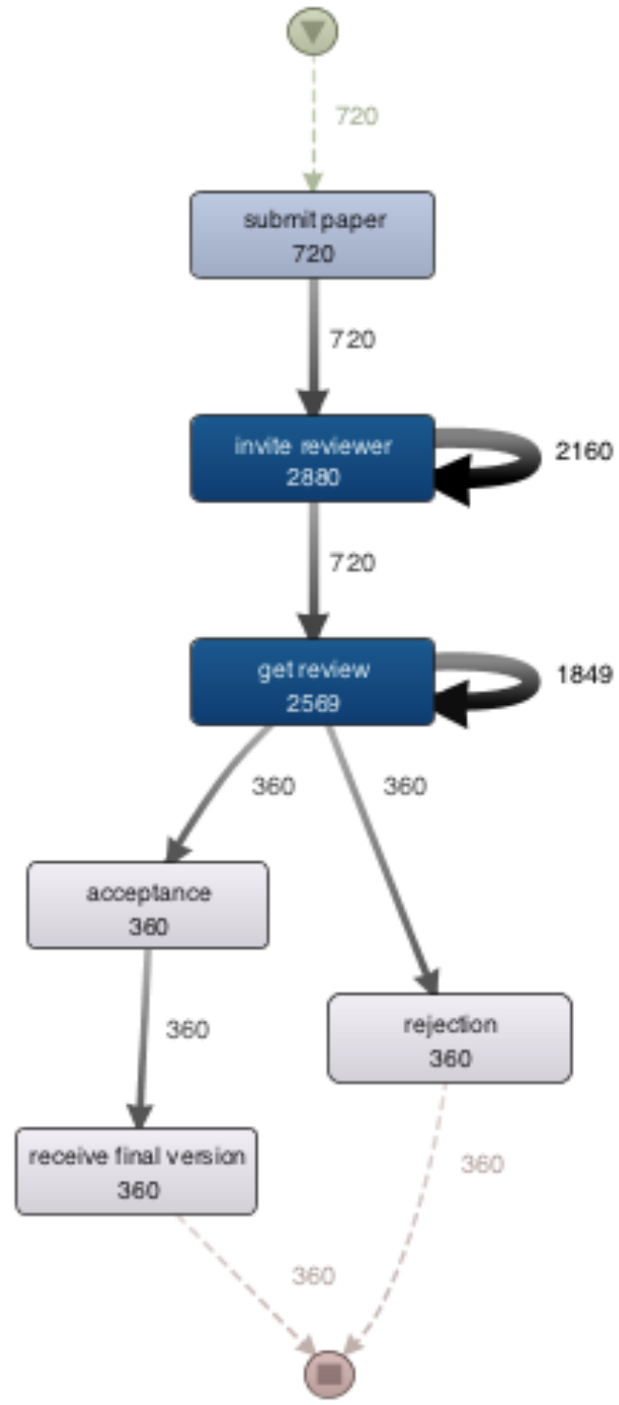
Virtual Log

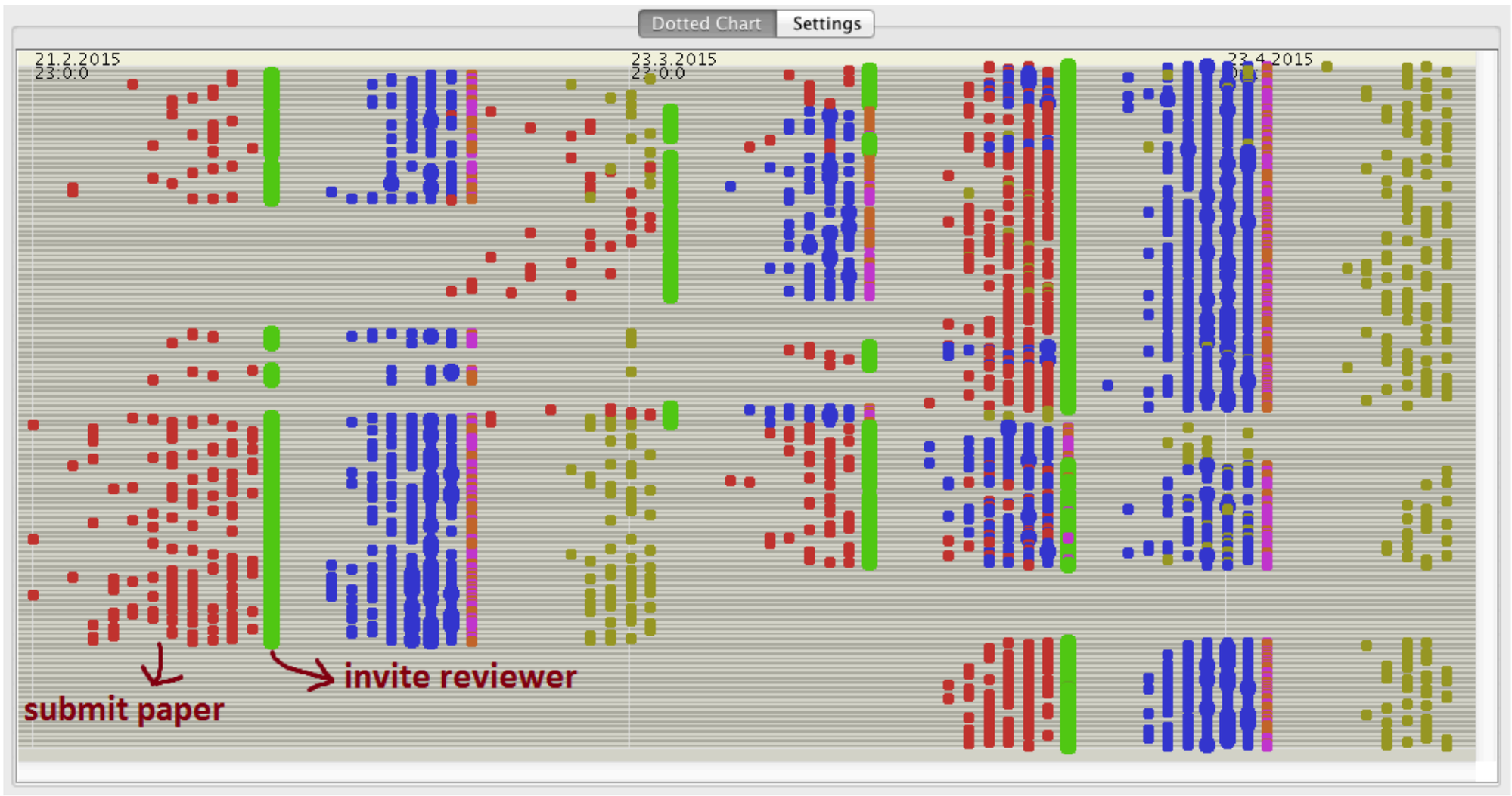


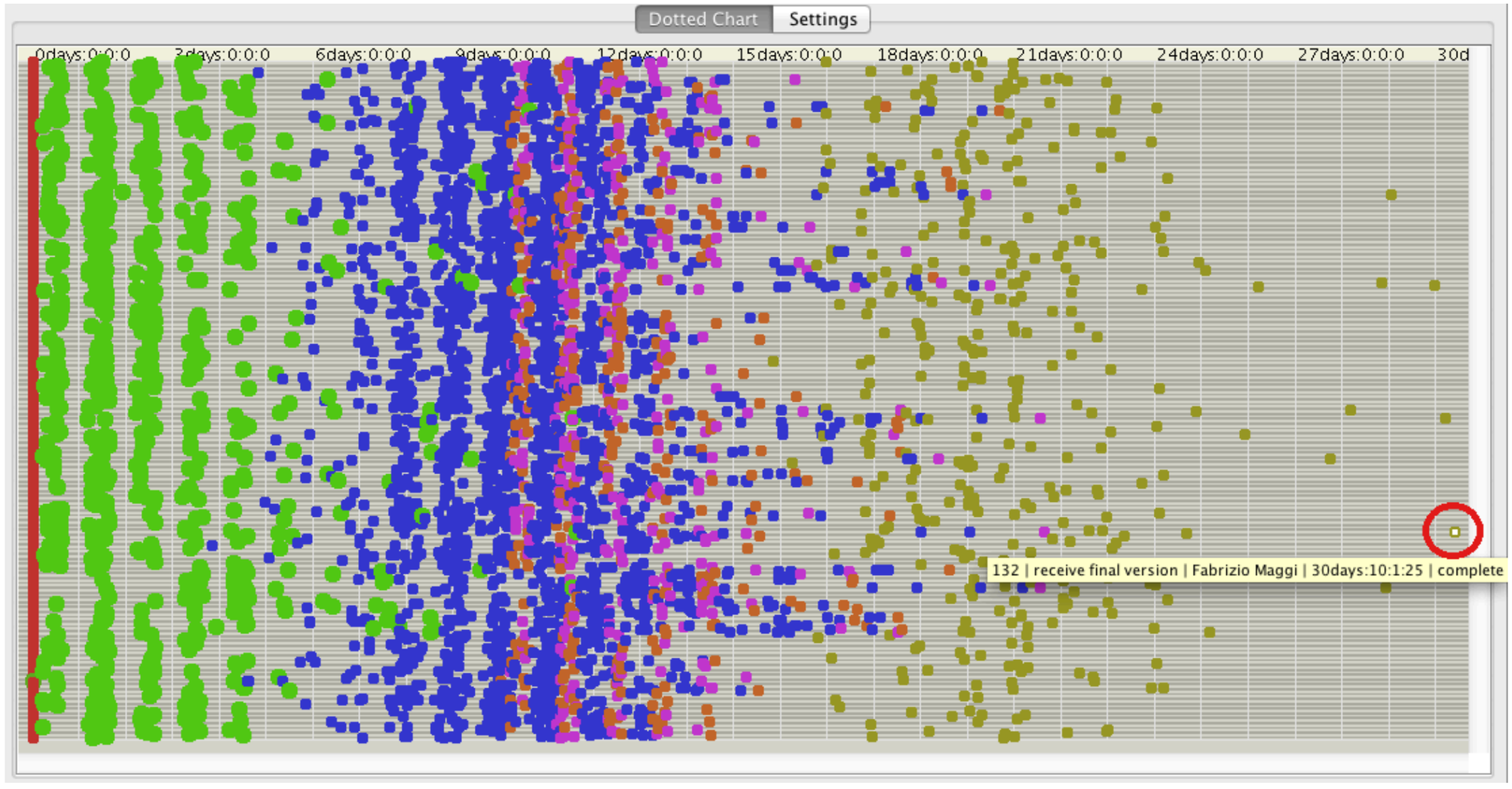

```
XFactoryOnDemandImpl
XLogOnDemandImpl
XTraceOnDemandImpl
XEventOnDemandImpl
XLogOnDemandIterator
XTraceOnDemandIterator
```

`xlog.get(7).get(90)` to retrieve te event in index 7th inside the 90th trace in a log


Input
Domain Ontology
OWLFile
First Mapping
OBDAFile
Annotation
AnnotationFile
Actions
Filter:
ontolo
Ontology-Driven Extraction Plugin
Alifah Syamsiyah (alifah.syamsiyah@stud-inf.unibz.it)
Ontology-Mapping Bootstrap Plugin
Alifah Syamsiyah (alifah.syamsiyah@stud-inf.unibz.it)
Remove extensions from log
J.M.E.M. van der Werf (j.m.e.m.v.d.werf@tue.nl)
Ontologies
Reset
Start
Output
XES Event Log
XLog


720
submit paper
720
720
invite reviewer
2880
2160
720
get review
2569
1849
360
360
acceptance
360
360
rejection
360
receive final version
360
360
360


Dotted Chart
Settings
21.2.2015
23:0:0
23.3.2015
23:0:0
23.4.2015
submit paper
invite reviewer


Dotted Chart
Settings
0days:0:0:0
3days:0:0:0
6days:0:0:0
9days:0:0:0
12days:0:0:0
15days:0:0:0
18days:0:0:0
21days:0:0:0
24days:0:0:0
27days:0:0:0
30d
132 | receive final version | Fabrizio Maggi | 30days:10:1:25 | complete

Social Commitments

Semantics for agent interaction that abstracts away from the internal agent implementation

- [Castelfranchi 1995]: social commitments as a *mediator between an individual and its "normative" relation with other agents*
- Extensively adopted for flexible specification of multiagent interaction protocols, business contracts, interorganizational business processes (cf. work by Singh et al)

Conditional Commitments

CC(debtor,creditor,Φ,Ψ)

- When condition **ϕ** holds, the debtor agent *becomes committed* towards the creditor agent to make condition **Ψ** true
- Agents change the state of affairs implicitly causing conditions to become true/false
- Commitments are consequently progressed reflecting the normative state of the interaction

Literature Example

- Contract between Bob (seller) and Alice (customer):

CC(bob,alice,item_paid,item_owned)

- Actions available to agents:

pay_with_cc **causes** item_paid
send_by_courier **causes** item_owned
deliver_manually **causes** item_owned

Literature Example

- Contract between Bob (seller) and Alice (customer):

CC(bob,alice,item_paid,item_owned)

- Actions available to agents:

```
pay_with_cc causes item_paid
send_by_courier causes item_owned
deliver_manually causes item_owned
```

Is this satisfactory???

Reality

- *Multiple* customers, sellers, items
  —> *Many-to-many* business relations established as *instances* of the same contractual commitment
- Need of **co-referencing commitment instances** through agents and the exchanged data
  - If **Bob** gets paid by **Alice** for **a laptop**, then **Bob** is commitment to ensure that **Alice** owns **that laptop**
- More in general, see work by Ferrario and Guarino on service foundations

From the Literature to Reality

(At least) two fixes required [Montali et al, 2014]:

1. Agent actions/messages must carry an **explicit data payload** (Alice pays *an item* with cc)
2. **Commitments** and dynamics have to become **data-aware**

```
forall Seller S, Customer C, Item I.
   CC(S,C,Paid(C,I,S),Owned(C,I))
```

Relational Commitments

- Ongoing research with Matteo Baldoni, Cristina Baroglio, Diego Calvanese
- Lifting all 4 commitment components
  - Debtor
  - Creditor
  - Guard
  - Condition
    To relational structures!
- Combination of direct operations on commitments (cancel, delegate) with indirect operations obtained as a result of update on data (binding with creditor, discharge)
- The same “commitment schema” now provides the basis for many-to-many business interactions

Example

- Commitment schema on “deliver-on-payment”
- Debtors: all agents that are sellers and that are registered to the marketplace
- Creditors of debtor **x**: all agents that are buyers and that sign a contract with **x**
- Guard: creditor **y** pays to debtor **x** for a given item **i**
- Condition: **y** commits towards **x** to deliver **i**

Commitment Trees