On-demand Relational Concept Analysis

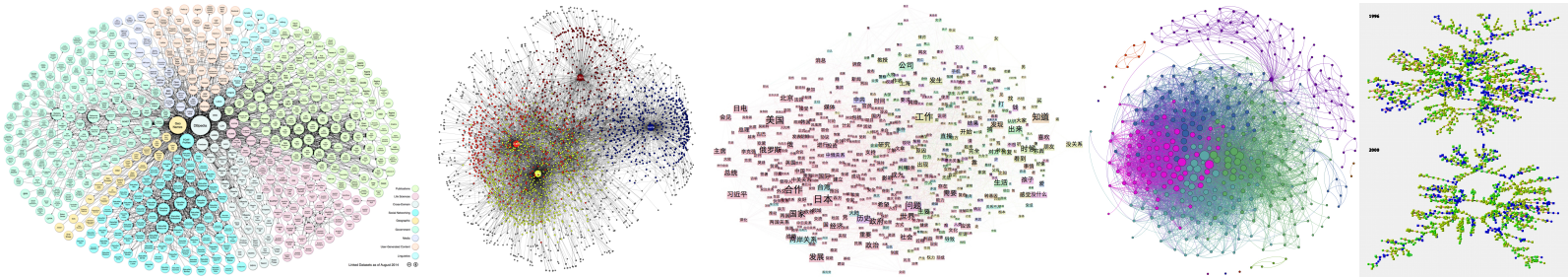
Alexandre Bazin Jessie Carbonnel Marianne Huchard Giacomo Kahn
Priscilla Keip Amirouche Ouzerdine

LORIA, Université de Lorraine, CNRS, Inria
LIRMM, CNRS and Université de Montpellier
LIFO, Université d'Orléans, INSA Centre Val de Loire, LIFO
UPR AIDA, CIRAD, Montpellier

15th International Conference on Formal Concept Analysis
June, 25 - 28, Frankfurt, Germany

Context

Large multi-relational datasets



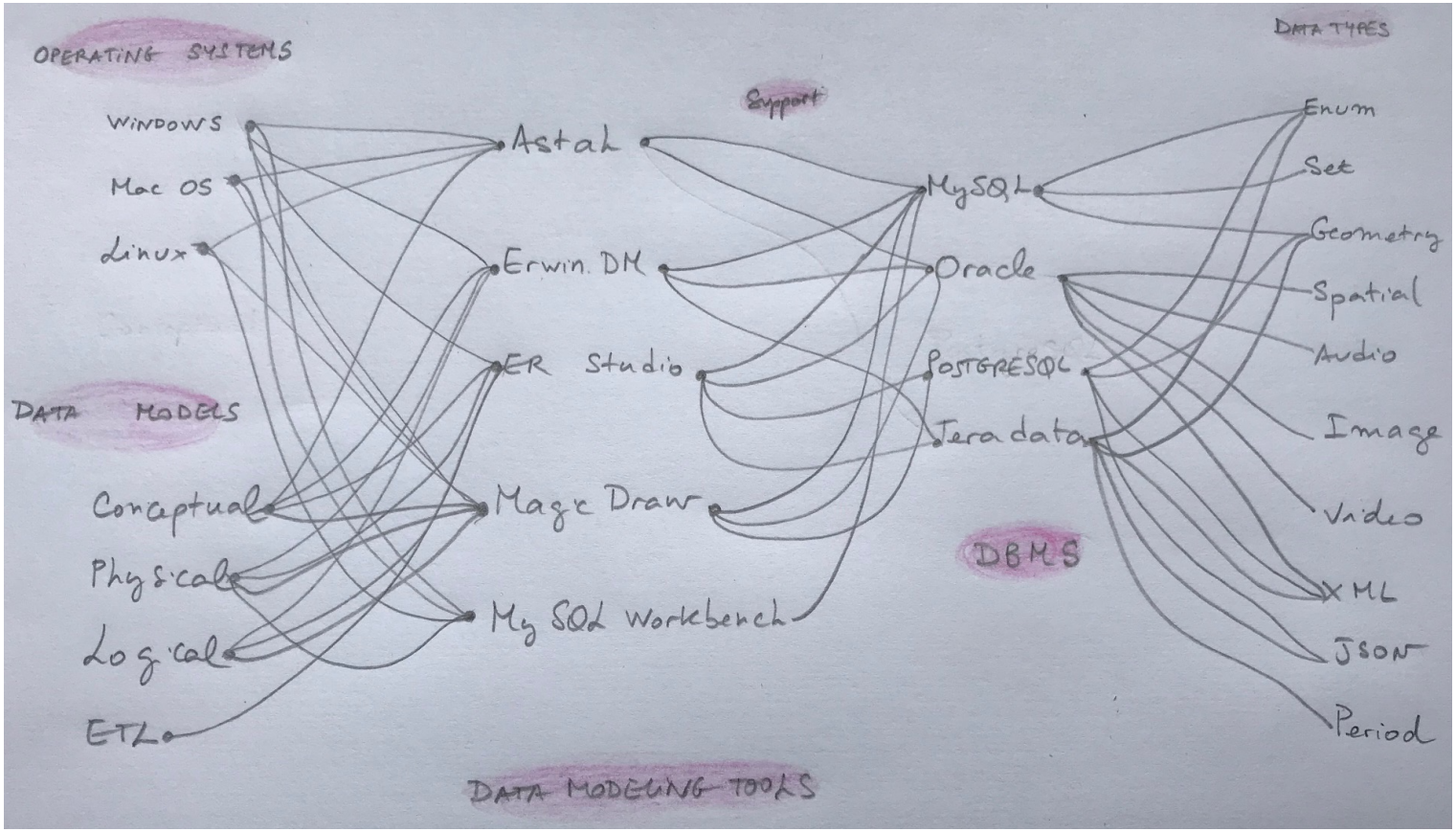
http://www.edeation.fr/wp-content/uploads/2015/01/Graphe-Global-Géant.png - https://gisellezeno.com/category/academic-work.html - https://www.frontiersin.org/articles/10.3389/fdigh.2017.00011/full - http://social-dynamics.org/tag/clustering-algorithm/ - https://www.bmj.com/content/337/bmj.a2338

Data exploitation tasks

- Querying
- Extraction of knowledge patterns
- Classification
- Browsing, Exploring

Illustrative exploration example

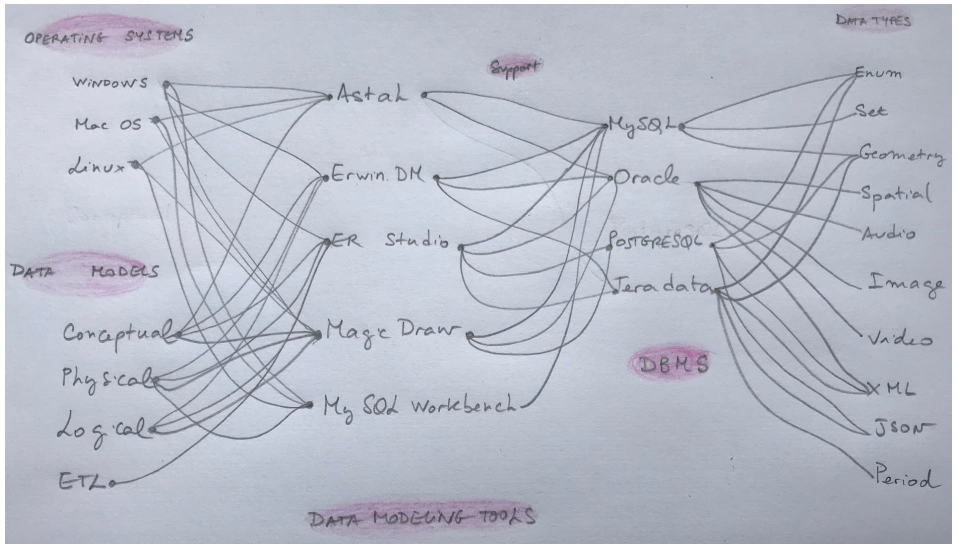
Select a Data Modeling Tool for conceptual model, running on Windows



$\rightsquigarrow$ Astah, Erwin DM, Magic Draw, ER/Studio

Illustrative exploration example

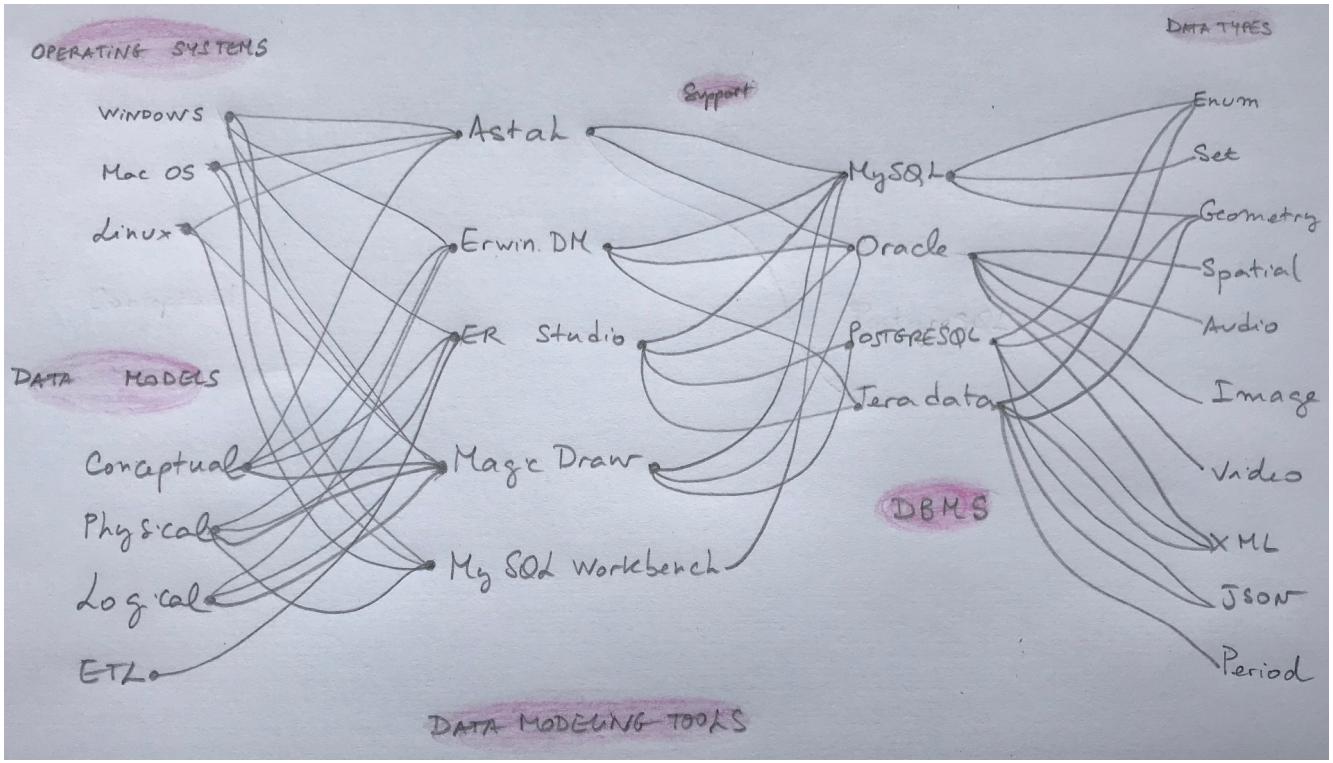
Are Astah, Erwin DM, Magic Draw, ER/Studio equivalent? What are their commonalities and variability in terms of properties and properties of their supported DBMS?



$\rightsquigarrow$ Astah, Magic Draw are available on Mac OS and Linux. Erwin DM, Magic Draw, ER/Studio allow physical and logical modeling (...) Magic Draw is the more complete (except it has no ETL modeling)

Illustrative exploration example

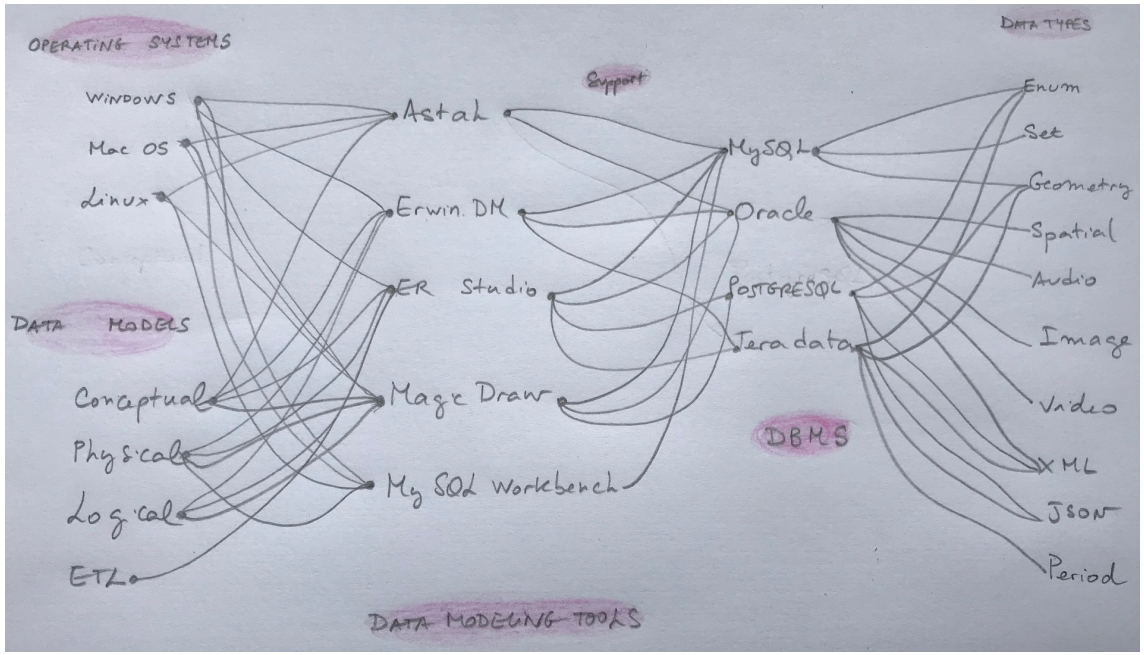
What if I relax the constraint "conceptual modeling"? This is not so important, having conceptual or physical modeling would be great.



$\rightsquigarrow$ All DM tools become relevant - Increases the choice number

Illustrative exploration example

All DM tools share MySQL. My boss is against. Which other similar DBMS could I find? With which additional/lost properties?



$\rightsquigarrow$ Compared with MySQL: PostgreSQL has in common Enum and Geometry data types; PostgreSQL also has JSON; Teradata is even more complete, proposing Period data type; Oracle does not share anything.

Illustrative exploration example

Lessons learned

- Using tables, or following links between individual objects is tricky
- Receiving a flat set of answers does not help so much

⇝Need for data structuring to foster exploration

- object aggregation by categories (groups)
  - *DBMS with Enum and Geometry*
- link aggregation (between-group links)
  - *DM tools* that support one *DBMS with Enum and Geometry*
- informed ranking, observed patterns
  - quantitative similarity may help (how much similar)
  - qualitative similarity is necessary (why/how similar, specialization)
  - *Teradata is more complete than PostgreSQL, proposing Period data type*
  - *Json* $\rightarrow$ *XML*

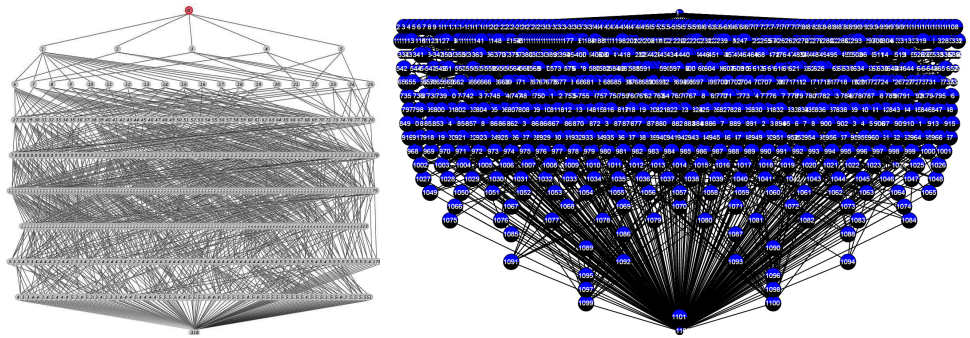
Solution tracks

Concepts in multi-relational datasets

- Power contexts families [Wille, 2002]
- Triadic [Lehmann and Wille, 1995], Polyadic Concept Analysis [Voutsadakis, 2002]
- Cubes of Concepts [Ferré et al., 2012]
- Relational, windowed structures [Kötters, 2013]
- Graph-FCA [Ferré, 2015]

→ Relational Concept Analysis [Hacene et al., 2013]

Concept lattices may be large (Villerd et al., Alam et al. (Latviz))



→Adopt exploratory approaches, cf. Abstract Conceptual Navigation [Ferré, 2014]

Relational Concept Analysis [Hacene et al., 2013]

- Extends the purpose of FCA for taking into account object categories and links between objects
- Main principles:
  - a relational model based on an entity-relationship model
  - integrate relations between objects as *relational* attributes
  - a variety of quantifiers for creating the *relational* attributes: $\exists$, $\exists\forall$, $\exists \supseteq$, ...
  - iterative process
- RCA provides a set of interconnected lattices
- Produced structures can be represented as ontology concepts within a knowledge representation formalism such as description logics (DLs)

Relational Context Family

DM_tools	OS:Windows	OS:Mac OS	OS:Linux	DM:Conceptual	DM:Physical	DM:Logical	DM:ETL
Astah	x	x	x	x			
Erwin DM	x			x	x	x	
ER/Studio	x			x	x	x	x
Magic Draw	x	x	x	x	x	x	
MySQL Workb.	x	x	x		x		

Formal context *DM_tools*

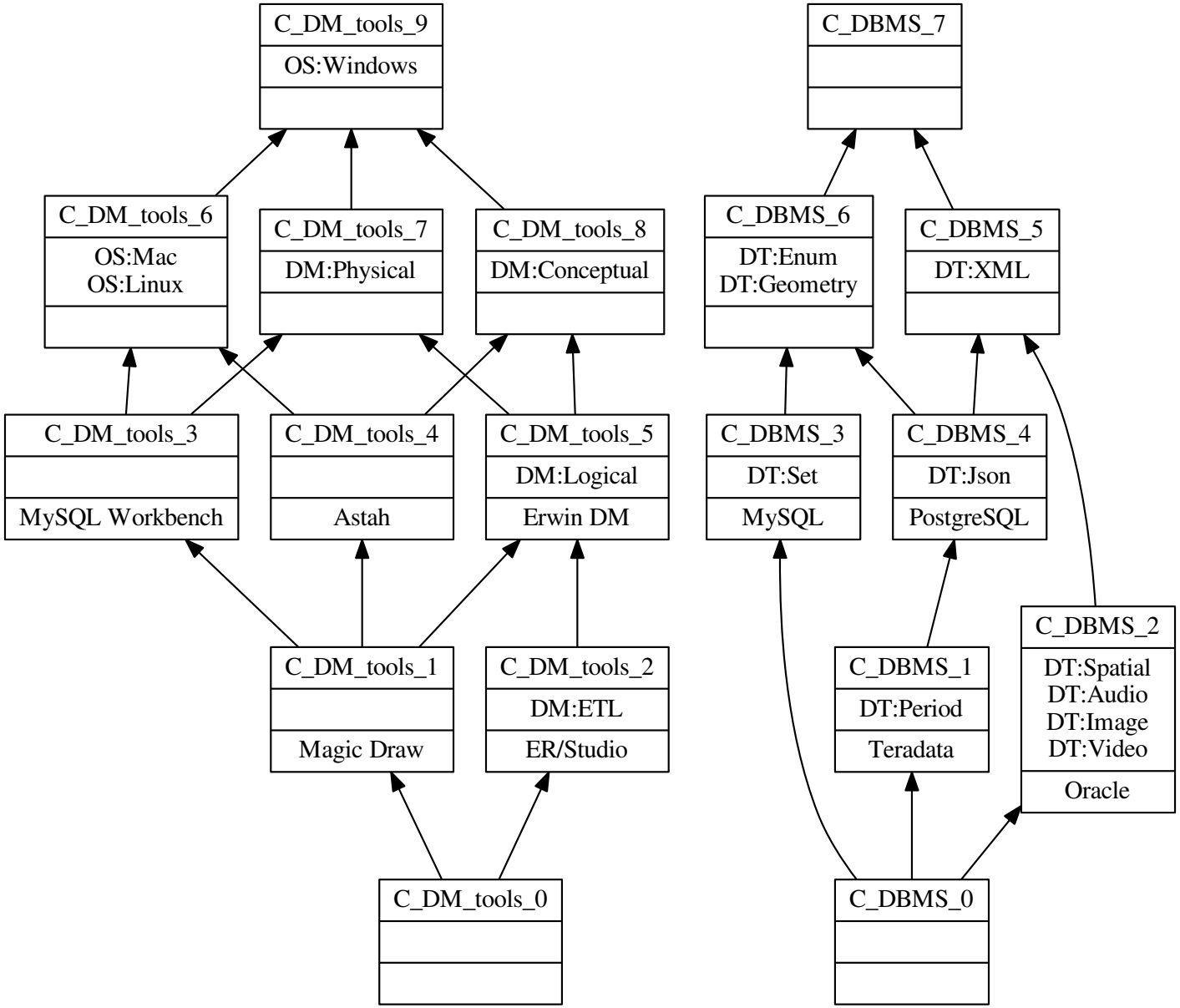
DBMS	DT:Enum	DT:Set	DT:Geometry	DT:Spatial	DT:Audio	DT:Image	DT:Video	DT:XML	DT:JSON	DT:Period
MySQL	x	x	x							
Oracle				x	x	x	x	x		
PostgreSQL	x		x					x	x	
Teradata	x		x					x	x	x

Formal context *DBMS*

support	MySQL	Oracle	PostgreSQL	Teradata
Astah	x	x		
Erwin DM	x	x		x
ER/Studio	x	x	x	x
Magic Draw	x	x	x	
MySQL Workb.	x			

Relational context *support*

Initial Lattices



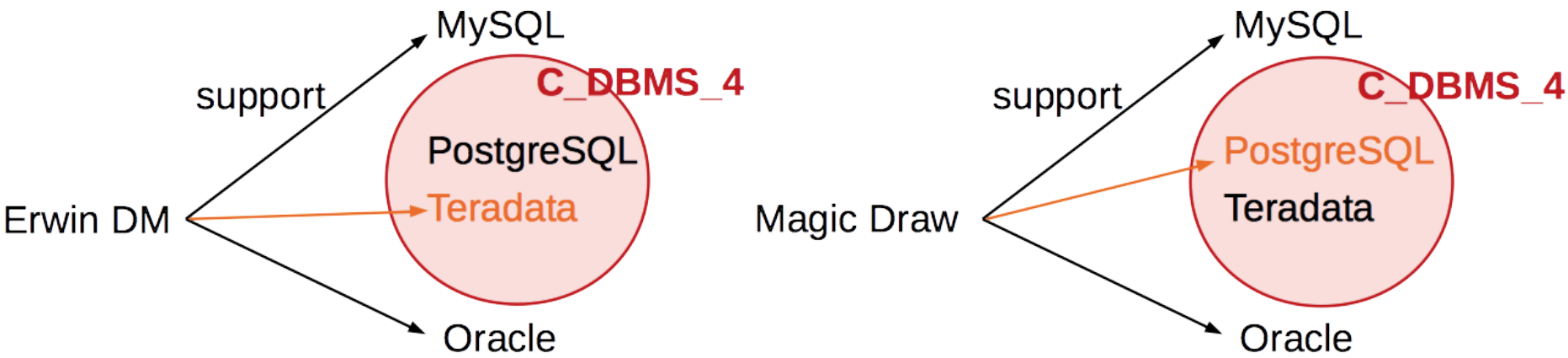
Enriching with *support*

DM_tools	OS:Windows	OS:Mac OS	OS:Linux	DM:Conceptual	DM:Physical	DM:Logical	DM:ETL	support:MySQL	support:Oracle	support:PostgreSQL	support:Teradata
Astah	x	x	x	x				x	x		
Erwin DM	x			x	x	x		x	x		x
ER/Studio	x			x	x	x	x	x	x	x	x
Magic Draw	x	x	x	x	x	x		x	x	x	
MySQL Workb.	x	x	x		x			x			

- Extracted knowledge: ErwinDM and Magic_Draw share MySQL and Oracle
- Knowledge not extracted:
  - Magic_Draw supports PostgreSQL; ErwinDM supports Teradata
  - PostgreSQL and Teradata admit DT:JSON
  - ErwinDM and Magic_Draw both support a DBMS admitting DT:Json

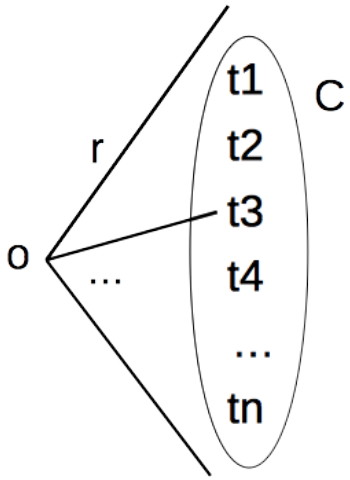
Consider quantifiers

Relational attribute ∃r(C)

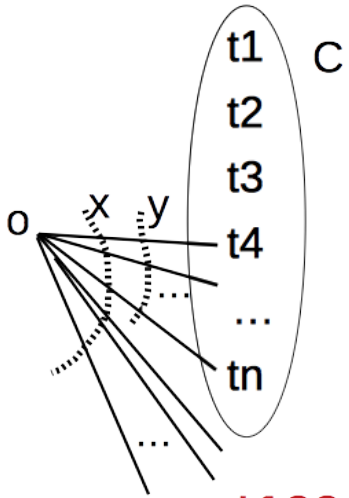


∃ support(C_DBMS_4)

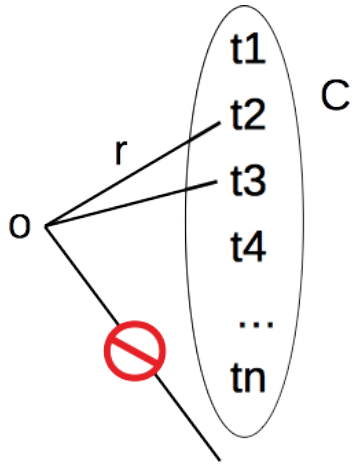
Consider quantifiers



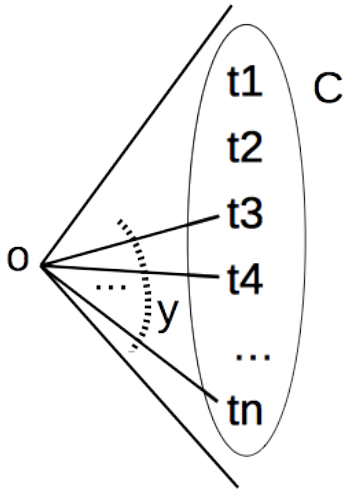
$\exists$ r (C)



$\exists\forall \geq y*100/x\%$ r (C)

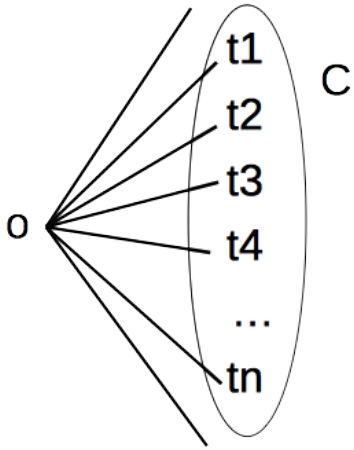


$\exists\forall$ r (C)



|C| = x

$\exists\supseteq \geq y*100/x\%$ r (C)



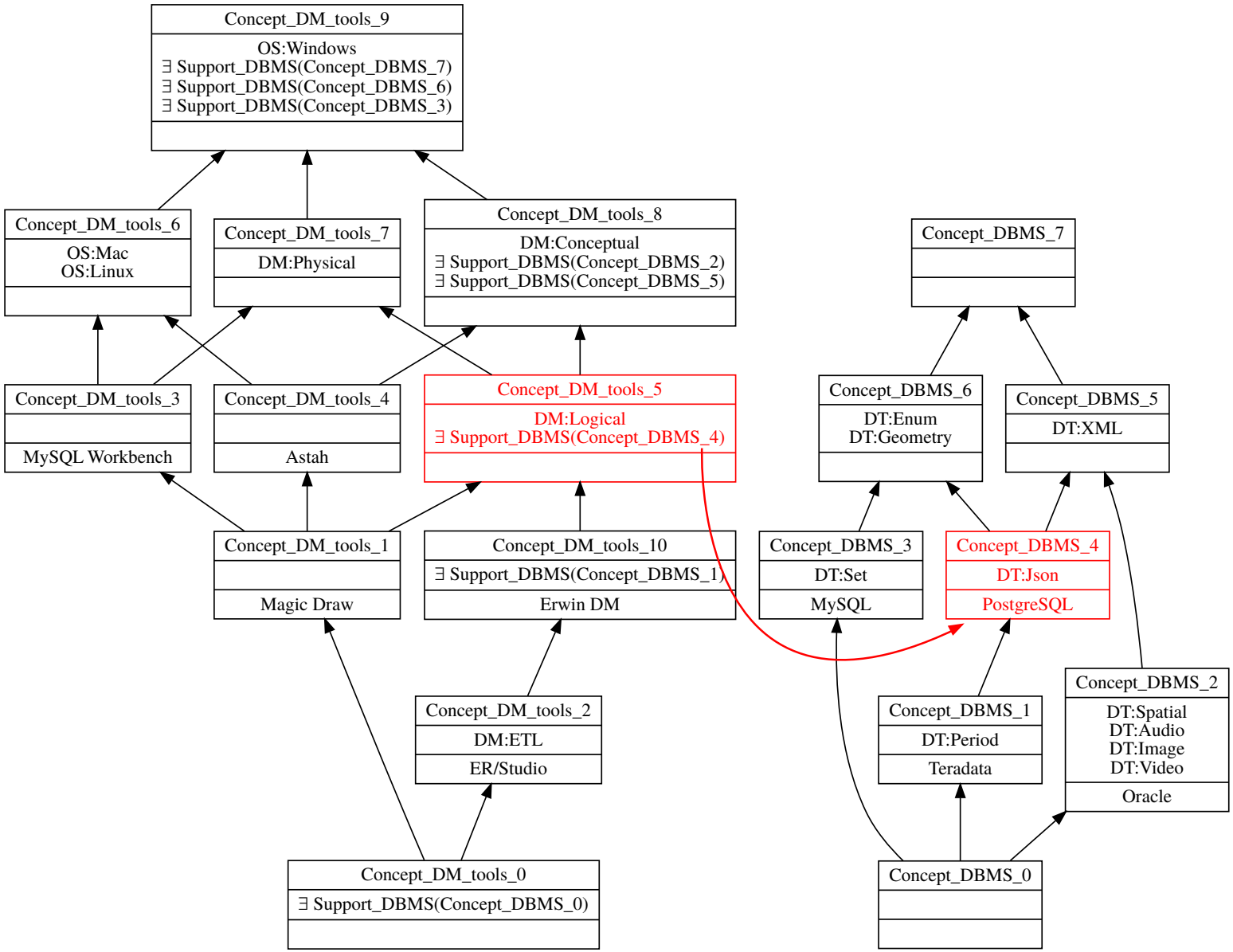
$\exists\supseteq$ r (C)

Enriching with ∃*support*

DM_tools	OS:Windows	OS:Mac	OS:Linux	DM:Conceptual	DM:Physical	DM:Logical	DM:ETL	∃ Support_DBMS(Concept_DBMS_7)	∃ Support_DBMS(Concept_DBMS_6)	∃ Support_DBMS(Concept_DBMS_3)	∃ Support_DBMS(Concept_DBMS_0)	∃ Support_DBMS(Concept_DBMS_2)	∃ Support_DBMS(Concept_DBMS_4)	∃ Support_DBMS(Concept_DBMS_5)	∃ Support_DBMS(Concept_DBMS_1)
Astah	×	×	×	×				×	×	×		×		×	
Erwin DM	×			×	×	×		×	×	×		×	×	×	×
ER/Studio	×			×	×	×	×	×	×	×		×	×	×	×
Magic Draw	×	×	×	×	×	×		×	×	×		×	×	×	
MySQL Workbench	×	×	×		×			×	×	×					

Now ErwinDM and Magic_Draw share "support a DBMS with DT:Json" (Concept_DBMS_4)

Connected concept lattices

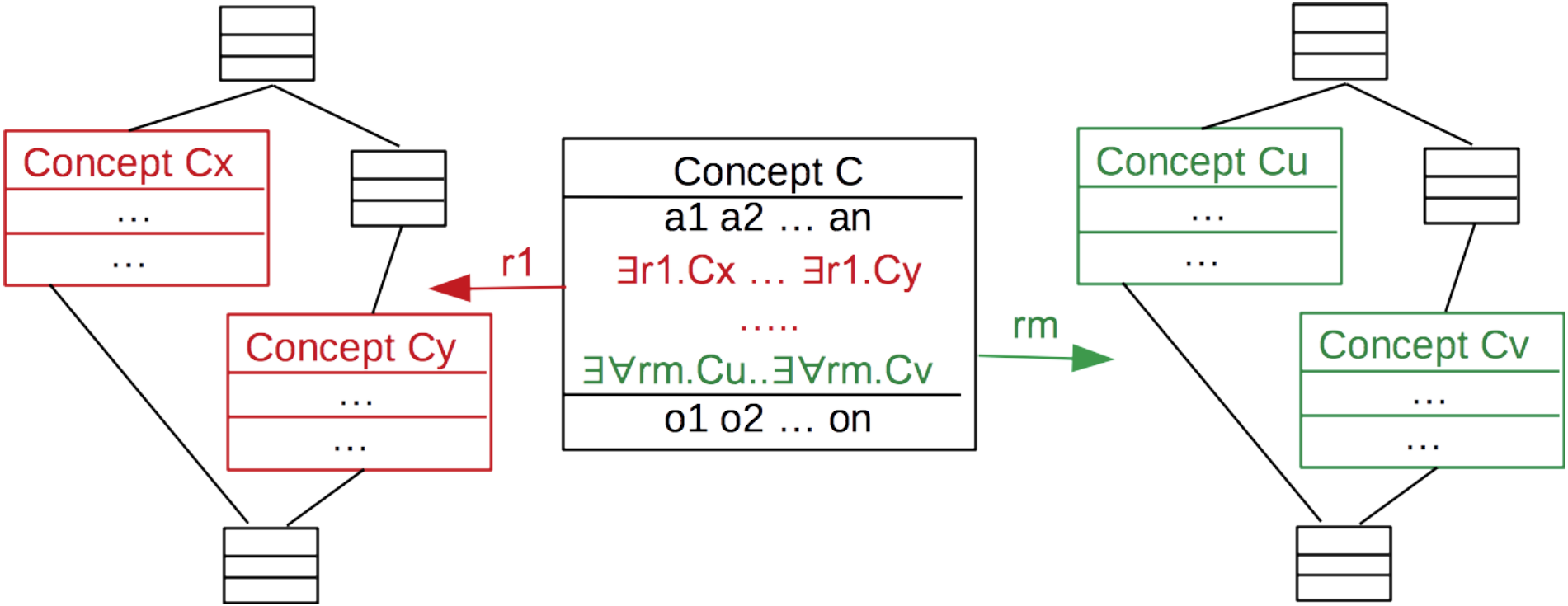


On-Demand: input

Concept C
a1 a2 ... an
o1 o2 ... on

A strategy: pairs of (r_i, ρ_j)
r_i a relation
ρ_j a quantifier ($\exists$, $\exists\forall$)

On-Demand: (1) relational covers



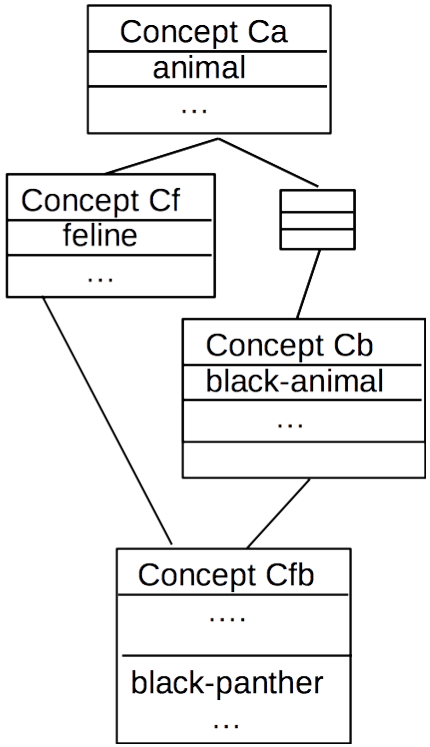
Limiting the number of relational attributes

- Using only Object-Concepts (introducing at least an object) to avoid computing whole lattices
- Limited to relations in the strategy
- Specific attribute set intersection

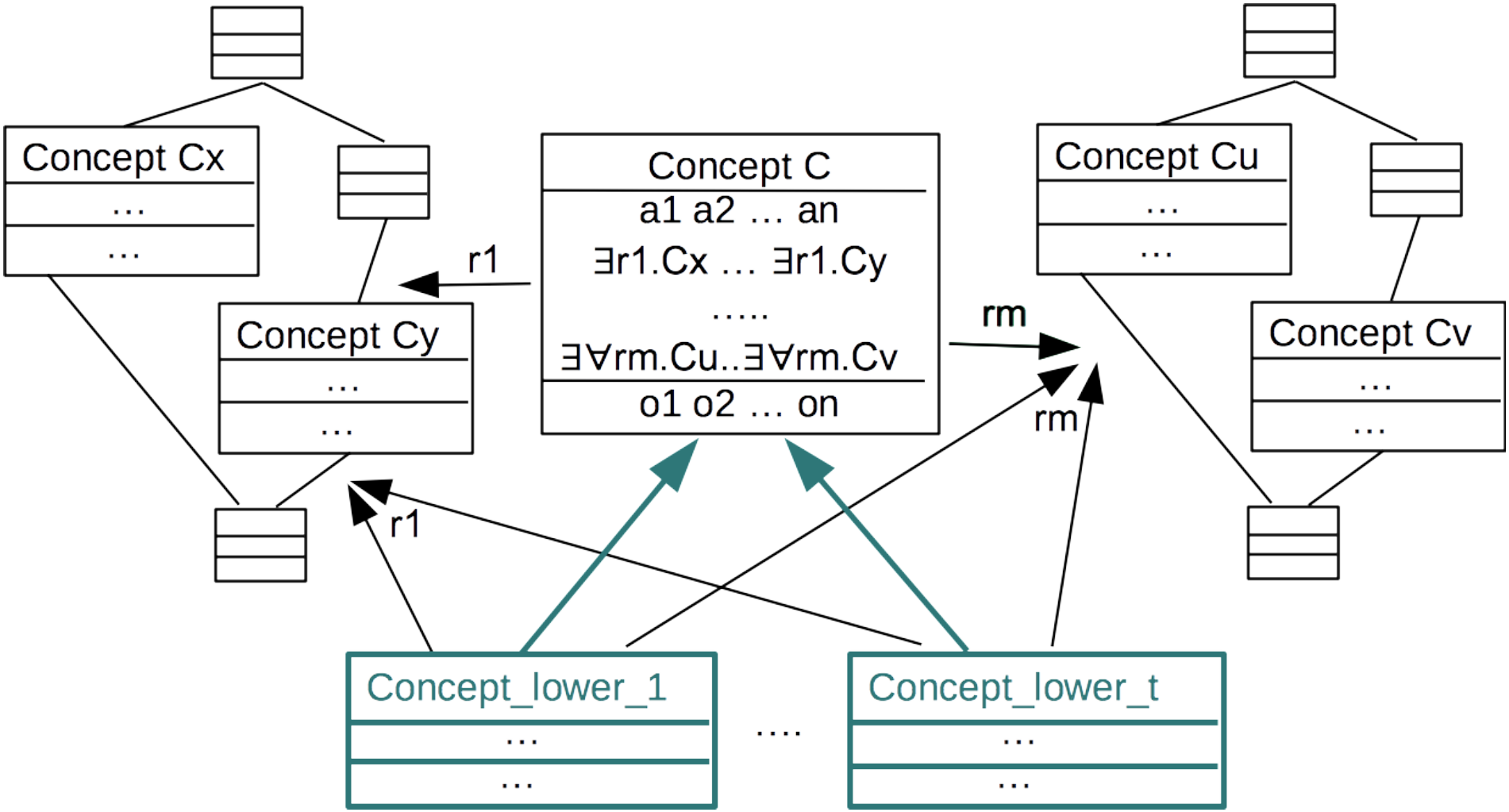
A={∃raises.Cf, ...}

B={∃raises.Cb, ...}

A ∩ B = {∃raises.Ca, ...}

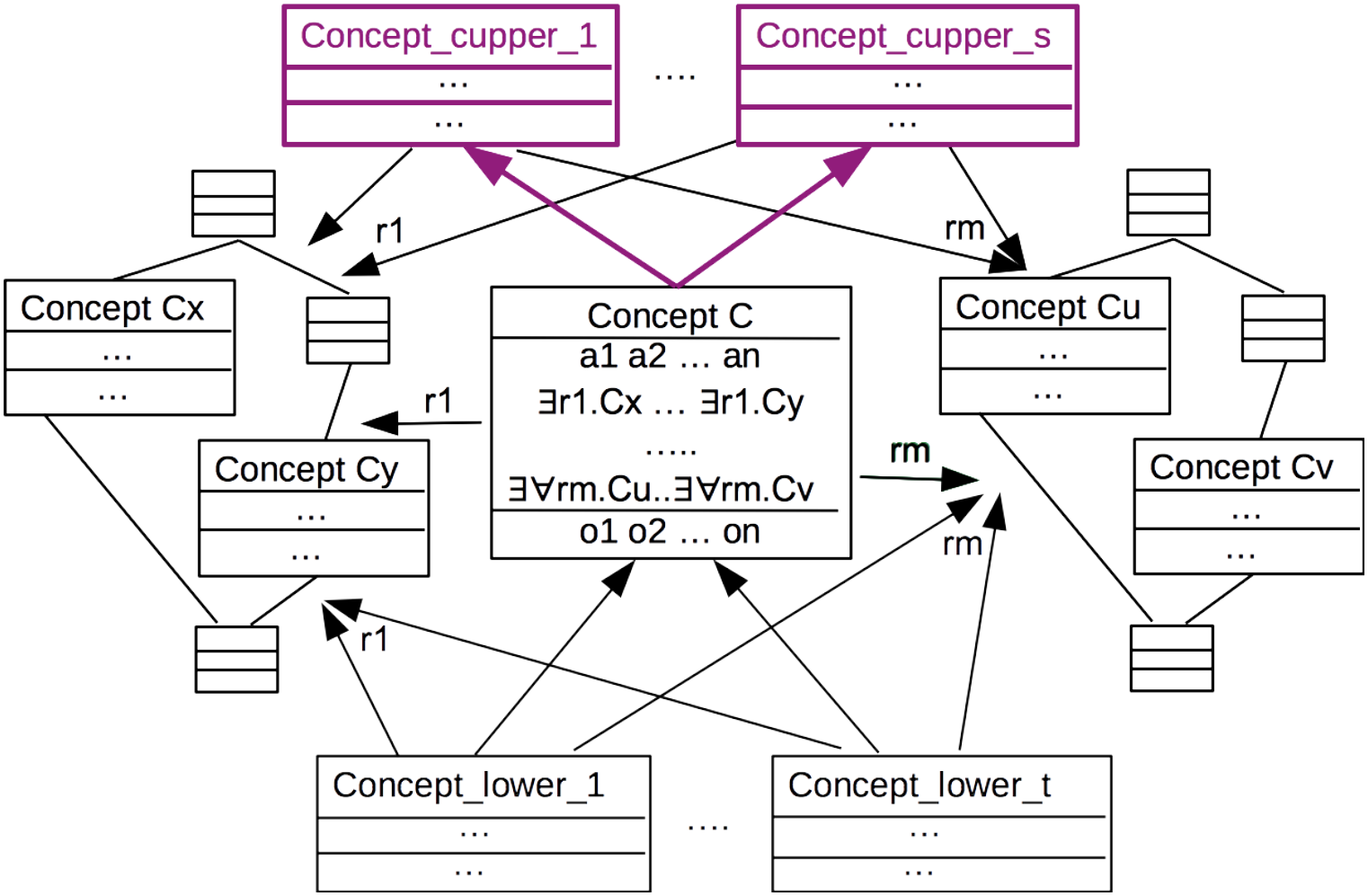


On-Demand: (2) lower covers



→ Compute concepts with the maximal extents that are contained in C'extent and do not contain any of its minimal generators = removing from the extent of C a minimal transversal of the set of minimal generators of C's extent

On-Demand: (3) upper covers



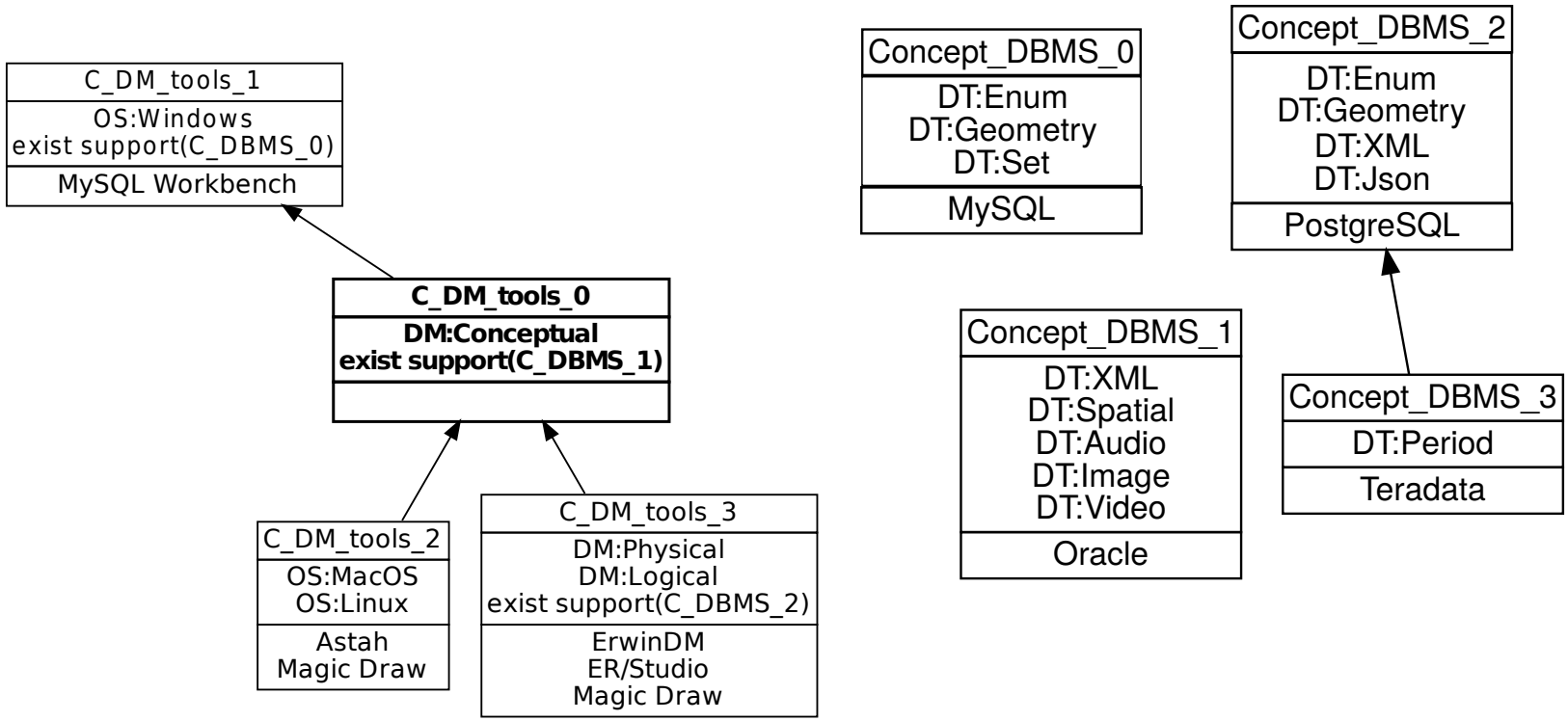
→ Computed by adding an object o to the extent of C

Integration in RCAExplore

- selection of a formal context, an attribute set
- from concept to concept
- minimal transversal computed with MTMiner [Hébert et al., 2007]

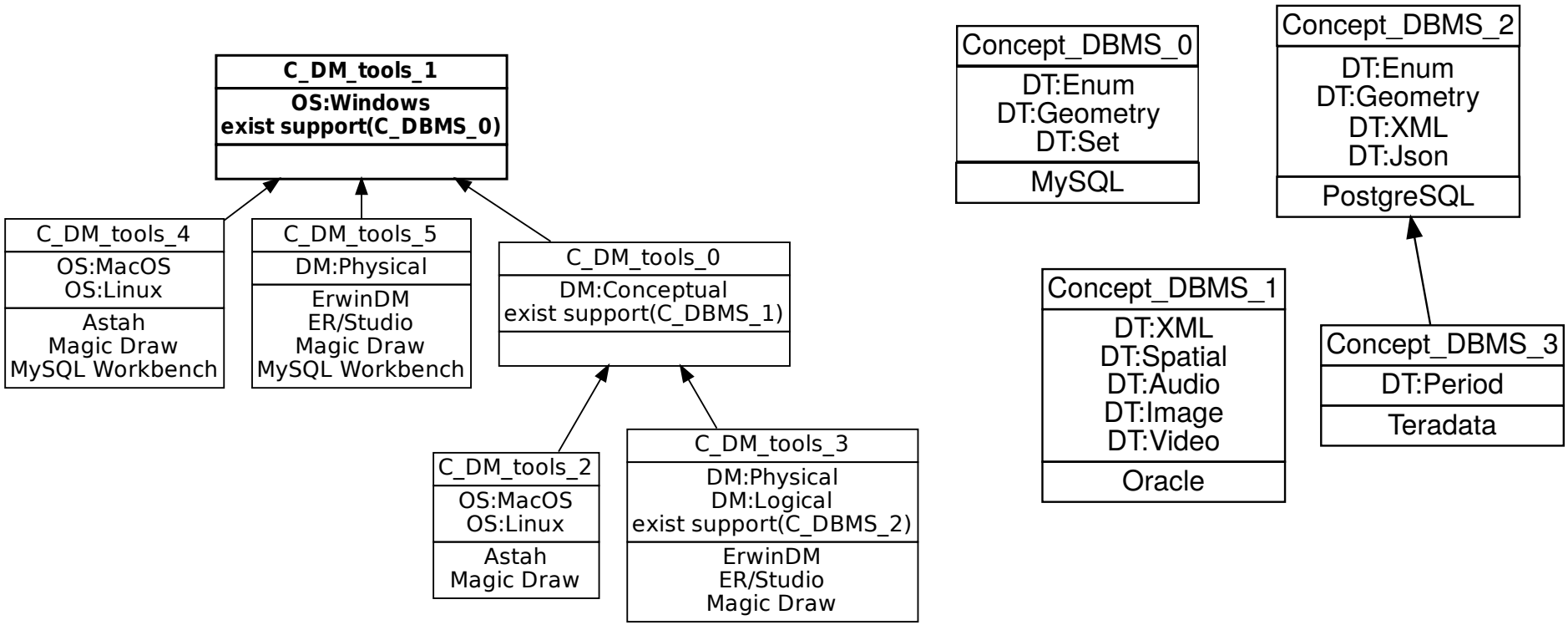
On-demand illustrated (Step 1)

Select a Data Modeling Tool for conceptual model, running on Windows



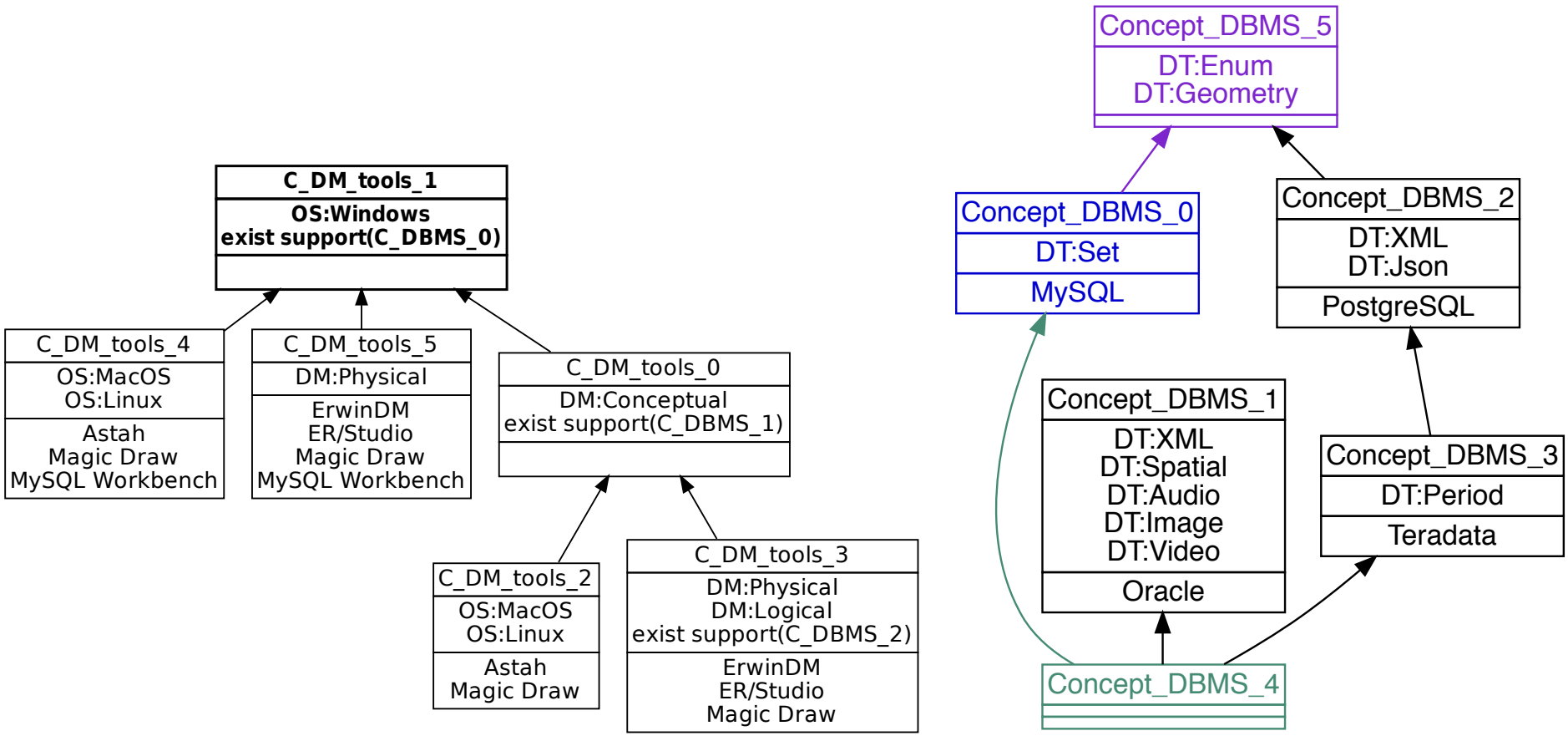
On-demand illustrated (Step 2)

What if I relax constraint "conceptual modeling"? This is not so important, having conceptual or physical modeling would be great.



On-demand illustrated (Step 3)

All DM tools share MySQL. My boss is against. Which other similar DBMS could I find? With which additional/lost properties?



Conclusion

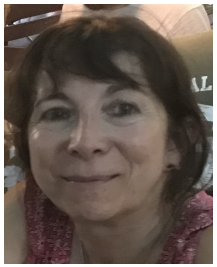
- on-demand computation for RCA
- considers $\exists$ and $\exists\forall$
- relational, upper and lower covers of a concept
- RCAExplore implementation with MTMiner and lattice completion at each step (with-memory approach)

Future work

- graphical user interface, integration within Cogui (conceptual graphs) platform http://www.lirmm.fr/cogui/
- reduce computation time for the minimal transversals, e.g. [Murakami and Uno, 2014]
- propose helpers (such as history, breadcrumb trail)
- propose variations (such as no memory progression)
- consider other quantifiers
- user experiment in Fresqueau (hydro-ecology) and Knomana (pesticide plants) projects

Thank you!







Ferré, S. (2014).
Reconciling Expressivity and Usability in Information Access.
Habilitation à diriger des recherches, Université de Rennes 1.


Ferré, S. (2015).
A proposal for extending formal concept analysis to knowledge graphs.
In *Formal Concept Analysis - 13th International Conference, ICFCA 2015, Nerja, Spain, June 23-26, 2015, Proceedings*, pages 271–286.


Ferré, S., Allard, P., and Ridoux, O. (2012).
Cubes of concepts: Multi-dimensional exploration of multi-valued contexts.
In *Formal Concept Analysis - 10th International Conference, ICFCA 2012, Leuven, Belgium, May 7-10, 2012. Proceedings*, pages 112–127.


Hacene, M. R., Huchard, M., Napoli, A., and Valtchev, P. (2013).
Relational concept analysis: mining concept lattices from multi-relational data.
Ann. Math. Artif. Intell., 67(1):81–108.


Hébert, C., Bretto, A., and Crémilleux, B. (2007).
A Data Mining Formalization to Improve Hypergraph Minimal Transversal Computation.

Fundam. Inform., 80:415–433.


Kötters, J. (2013).
Concept lattices of a relational structure.
In *Conceptual Structures for STEM Research and Education, 20th International Conference on Conceptual Structures, ICCS 2013, Mumbai, India, January 10-12, 2013. Proceedings*, pages 301–310.


Lehmann, F. and Wille, R. (1995).
A triadic approach to formal concept analysis.
In Ellis, G., Levinson, R., Rich, W., and Sowa, J. F., editors, *Conceptual Structures: Applications, Implementation and Theory, Third International Conference on Conceptual Structures, ICCS '95, Santa Cruz, California, USA, August 14-18, 1995, Proceedings*, volume 954 of *Lecture Notes in Computer Science*, pages 32–43. Springer.


Murakami, K. and Uno, T. (2014).
Efficient algorithms for dualizing large-scale hypergraphs.
Discrete Applied Mathematics, 170:83 – 94.


Voutsadakis, G. (2002).
Polyadic concept analysis.
Order, 19(3):295–304.


Wille, R. (2002).
Existential concept graphs of power context families.
In *Conceptual Structures: Integration and Interfaces, 10th International Conference on Conceptual Structures, ICCS 2002, Borovets, Bulgaria, July 15-19, 2002, Proceedings*, pages 382–396.