



Semantic-Aware Business Intelligence

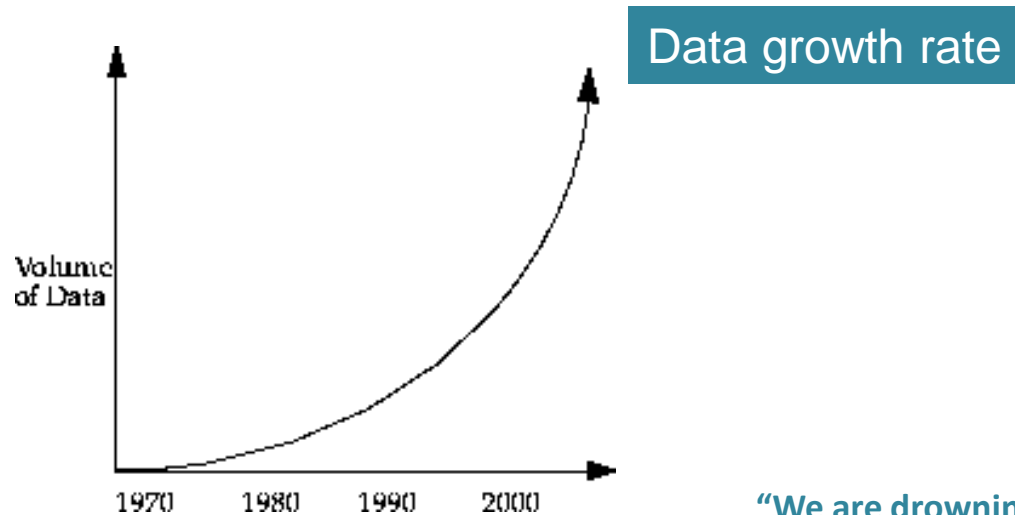
Speaker: Oscar Romero
(oromero@essi.upc.edu)



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

- **Motivation**
 - **Traditional BI Vs. Exploratory BI**
 - **On the Need of Semantic-Aware BI Systems**
- **Semantic-Aware Formalisms**
- **A Characterization of BI Systems**
 - **Current State of the Art**
 - **Discussion and Future Research Trends**
- **An Open-Access Semantic-Aware BI System**
 - **Architecture**
 - **Modules**
 - **AMDO, GEM, ORE and COAL**

“The amount of information in the world doubles every 20 months and the size and number of databases are increasing even faster”



“We are drowning in information but starved for knowledge”
John Naisbitt, “Megatrends” (1982)

Small amounts of data will be ever considered by a human
Automatic Knowledge Discovery is badly **NEEDED**

- Business intelligence is *“the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal”*

H. P. Luhn

A Business Intelligence System

IBM Journal of Research and Development. Vol. 2(4). 1958

- Business intelligence is *“an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance”*

Gartner Reports, IT Glossary, 2013

- **[Business]** Business intelligence is *“the process of collecting business data and turning it into information that is meaningful and actionable towards a strategic goal”*
- **[IT]** Business Intelligence is *“aimed at gathering, transforming and summarizing available business data from available sources to generate analytical information suitable for decision-making tasks”*

Data Analytics: Reporting, OLAP, data mining, prediction (what-if analysis), etc.

Data Modeling: Data warehousing and the multidimensional model

- **[IT] Business Intelligence is “*aimed at gathering, transforming and summarizing available business data from available sources to generate analytical information suitable for decision-making tasks*”**

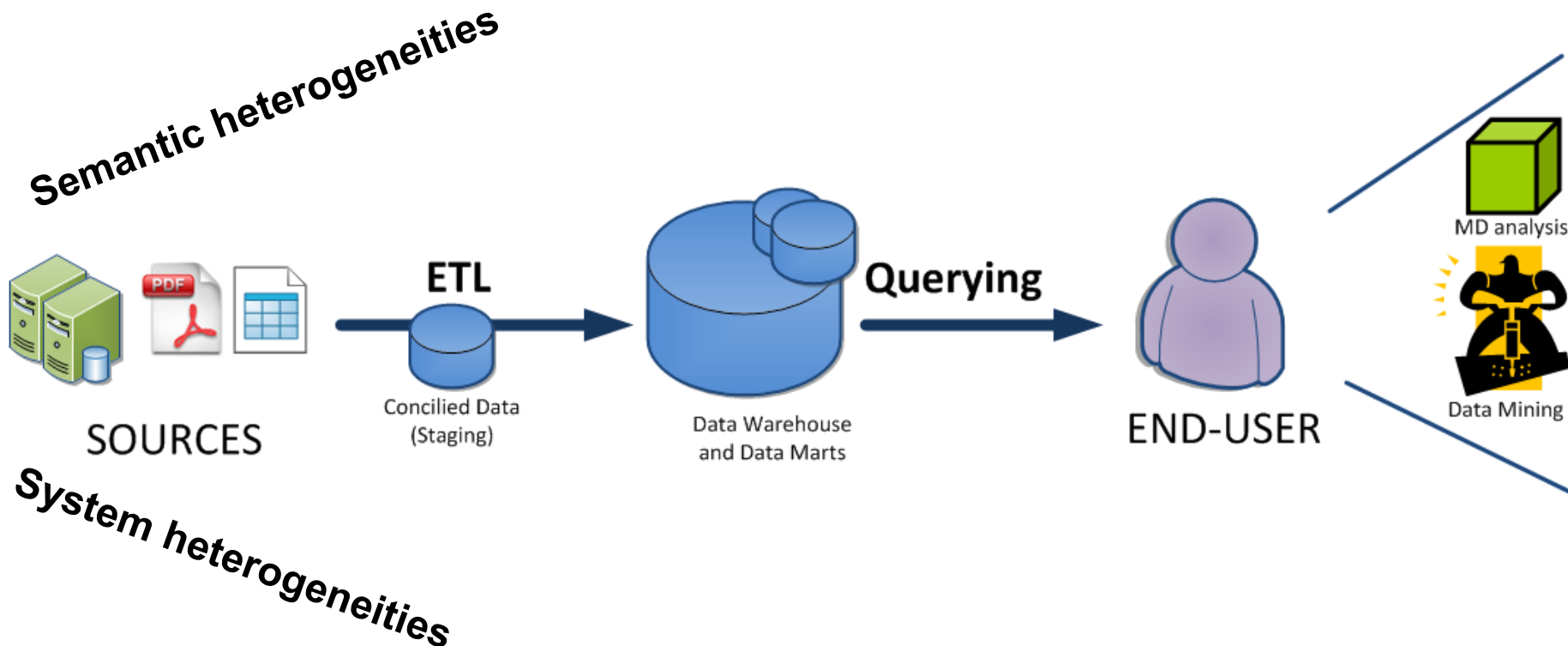
Data Modeling: Data warehousing and the multidimensional model

- **BI has become a huge industrial and scientific domain and a major economic driver**
- **Some numbers:**
 - In 2014, the overall BI market value will be \$100 billion (the Economist)
 - It is expected to have the largest increment in data management and analytics, at almost 10% each year, roughly twice as fast as the software business as a whole.
 - The market growth is estimated in \$14 billion in 2014, up from \$8.8 billion in 2008 (Forrester and the Economist)
 - BI remains unaffected by the economic crisis and it is mentioned among the top priorities of Chief Information Officers worldwide (ranging from 1st to 5th according to the source)

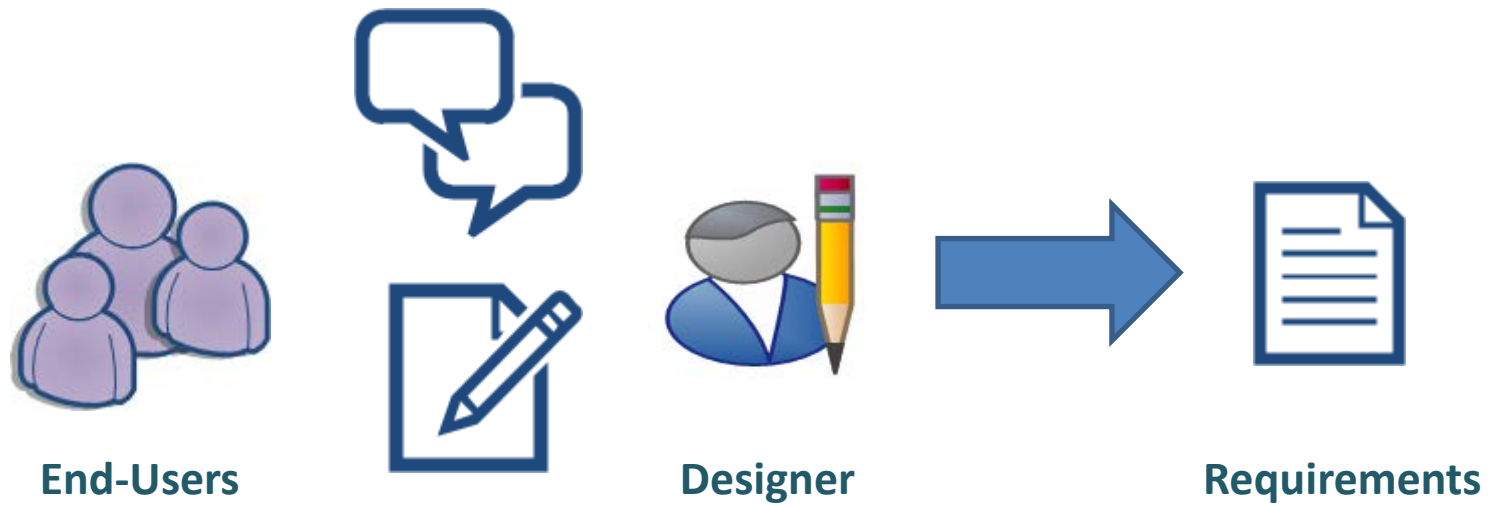
Traditional Vs. Exploratory BI

Joint work with:

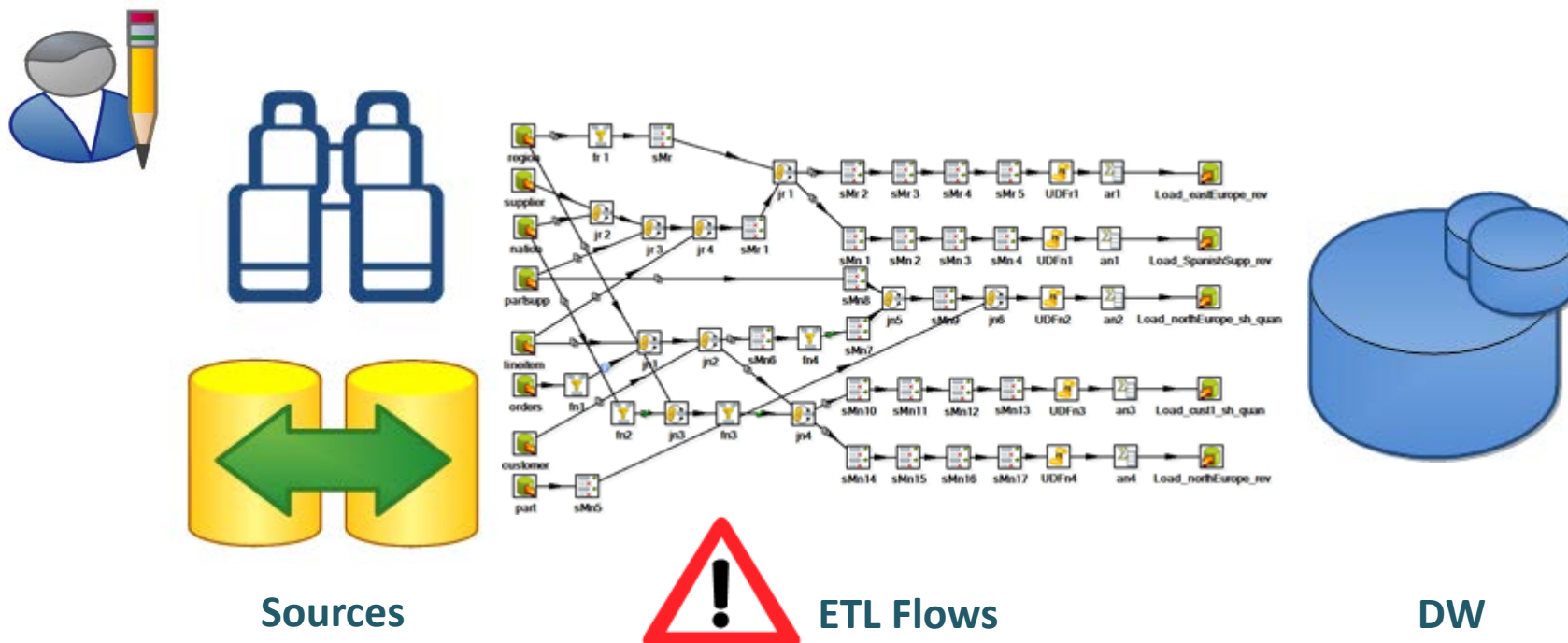
Alberto Abelló (UPC)	Torben Bach Pedersen (AAU)
Rafael Berlanga (UJI)	Victoria Nebot (UJI)
M.J. Aramburu (UJI)	Alkis Simitsis (HP Labs)



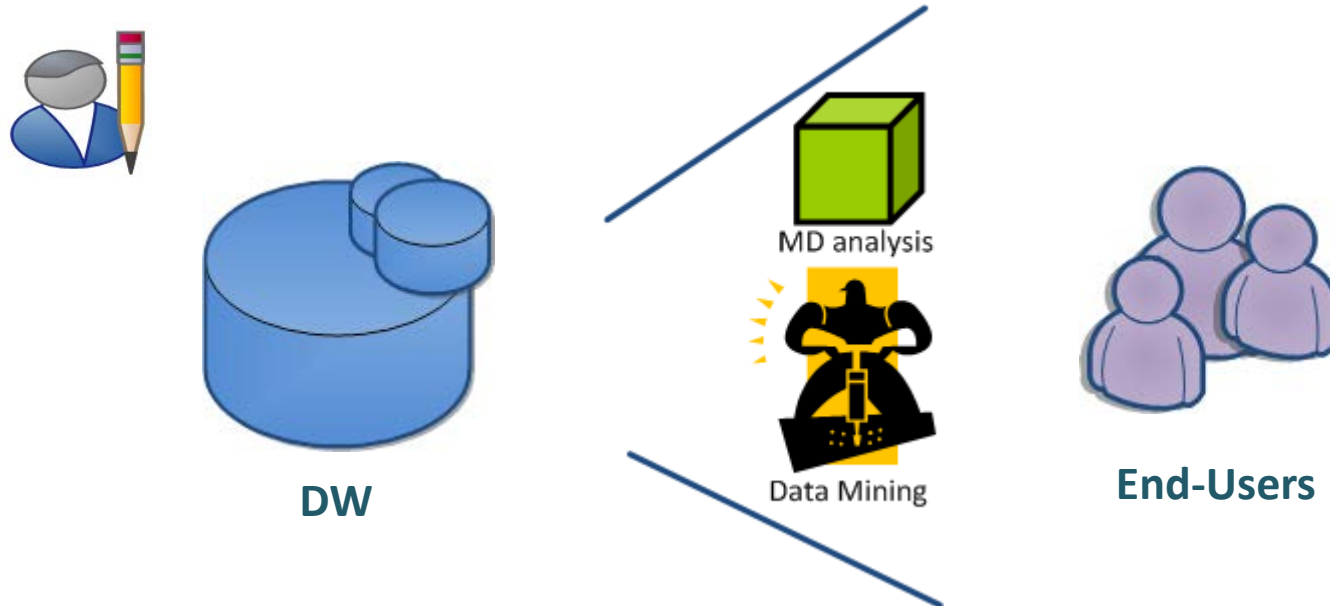
1. Requirement Elicitation



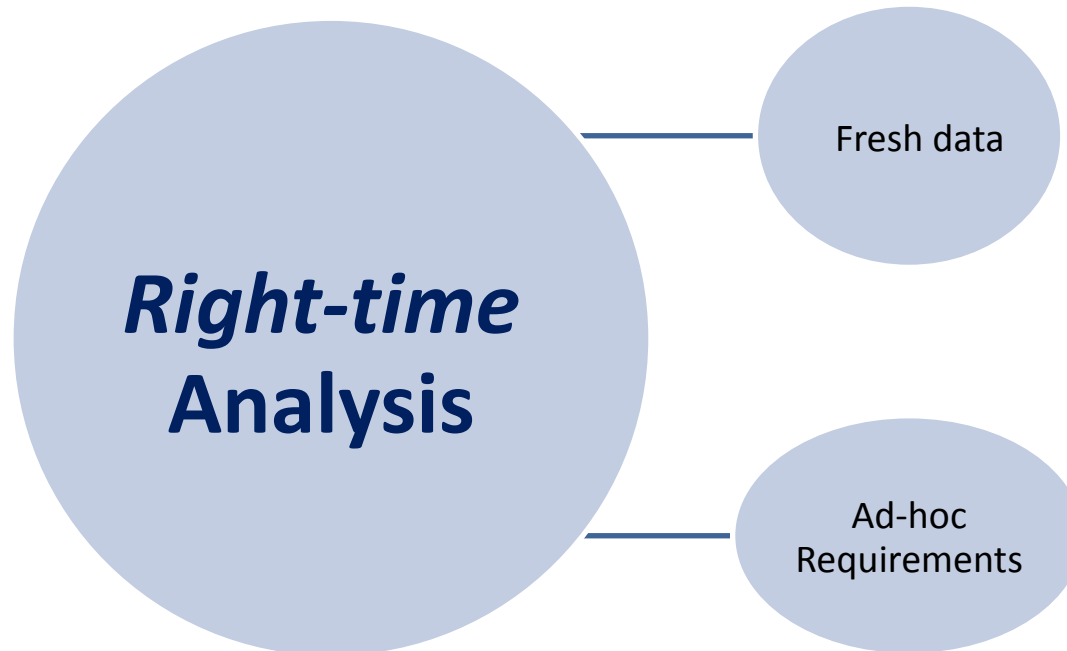
2. Design and Deployment of the DW and ETL flows



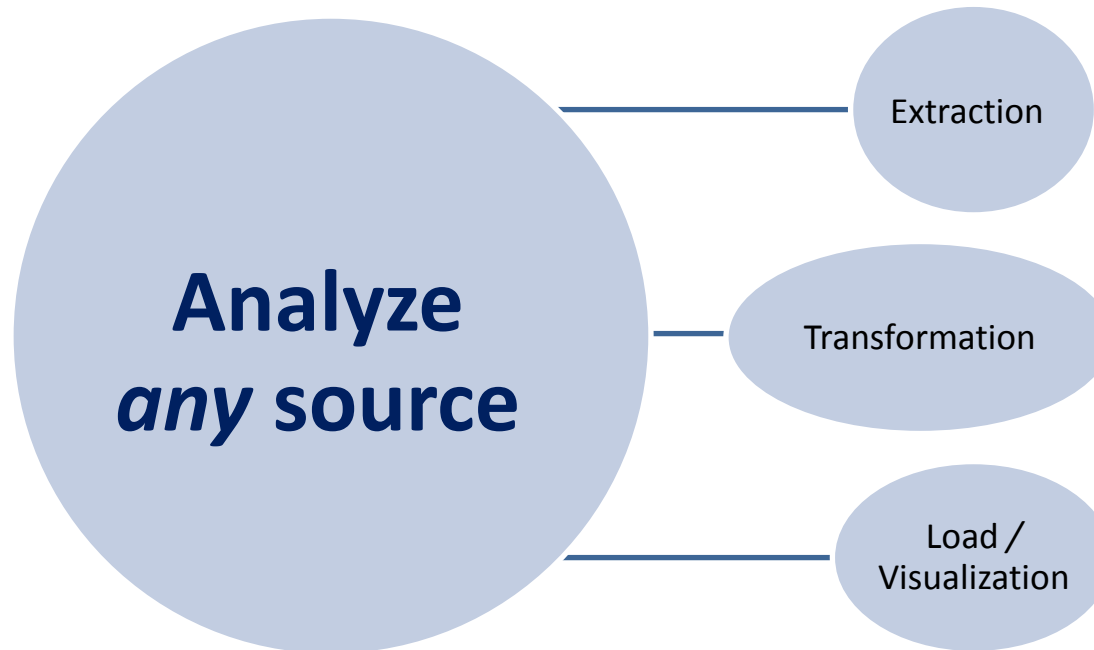
3. Flexible and Dynamic Query Tools

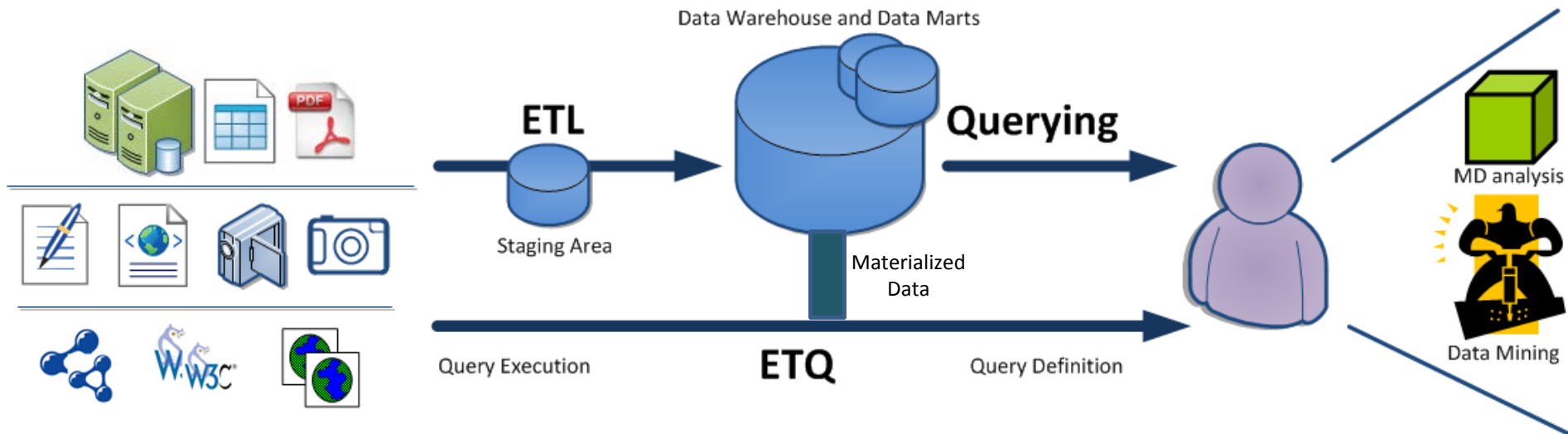


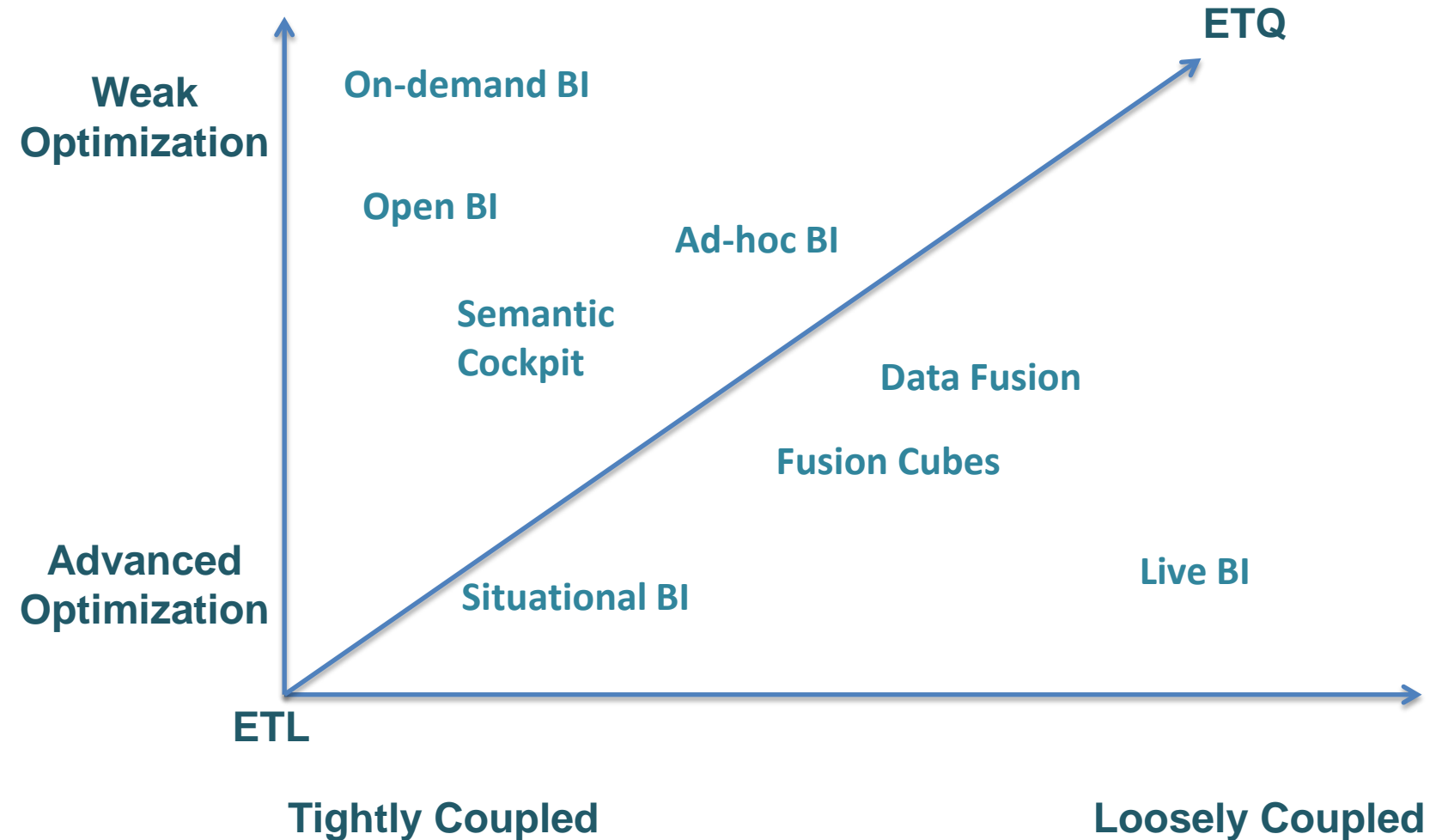
- From the point of view of the end-user:



- From the point of view of the end-user:







- From the IT point of view:



MACHINE PROCESSABLE

End-user Requirement

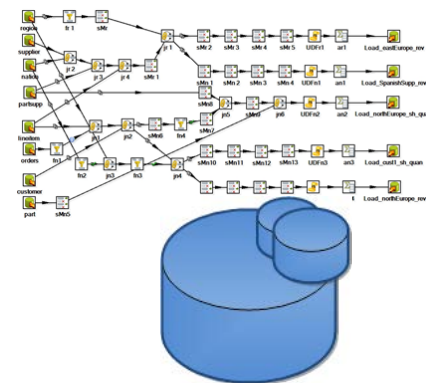
MACHINE PROCESSABLE



Sources

MD Schema

ETL Flows



SUPERVISION



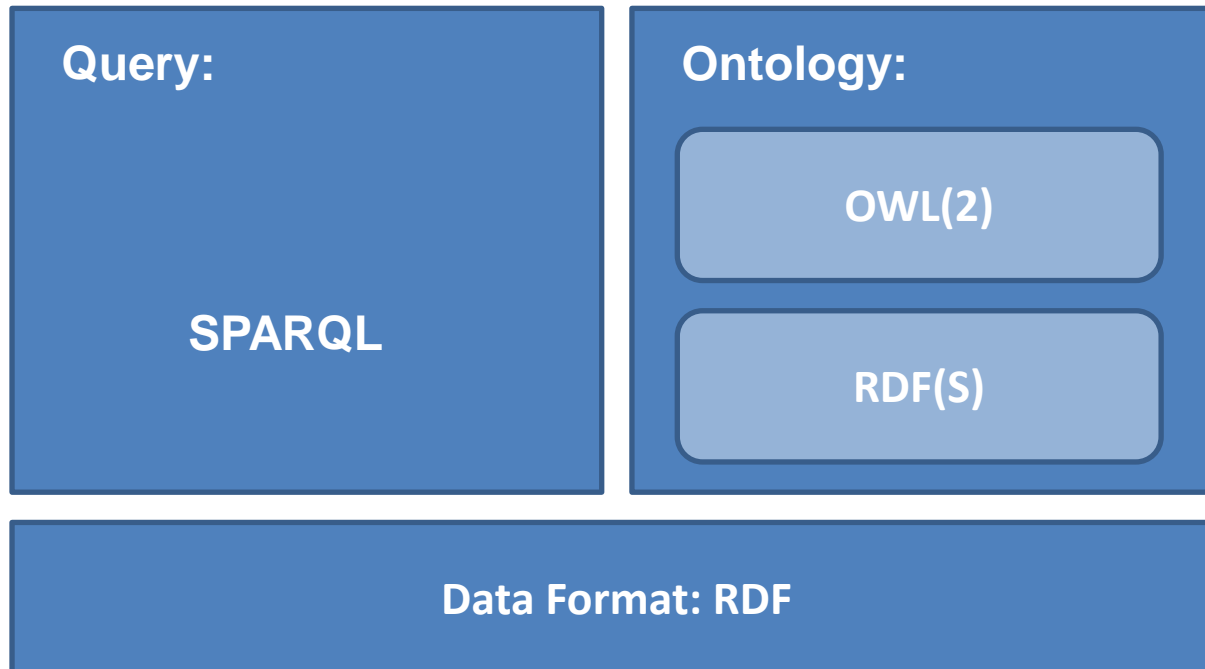
Semantic-Aware Formalisms

Joint work with:

Alberto Abelló (UPC) Torben Bach Pedersen (AAU)

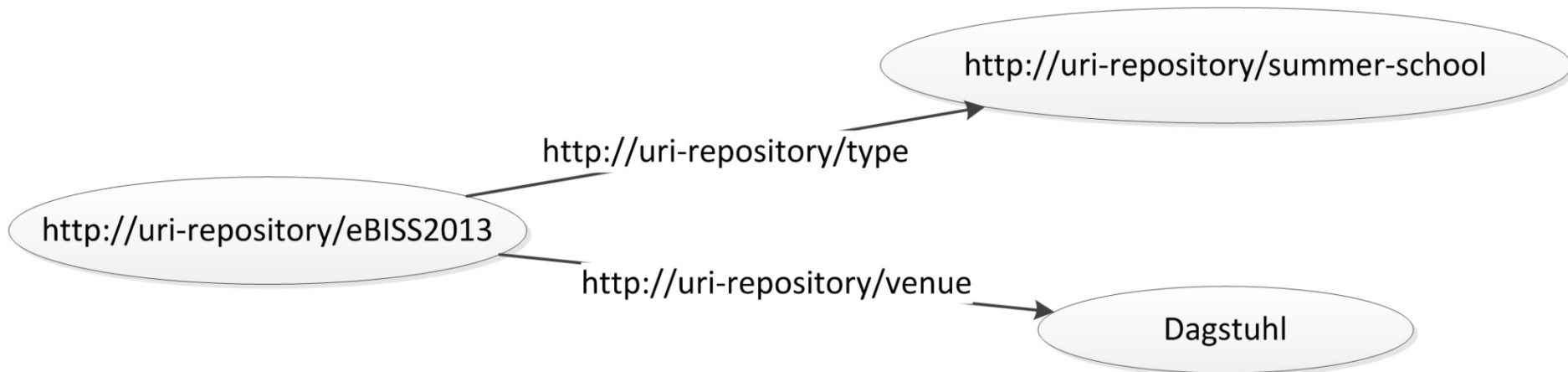
Rafael Berlanga (UJI) Victoria Nebot (UJI)

M.J. Aramburu (UJI) Alkis Simitsis (HP Labs)



Source: Sven Groppe. Data Management and Query Processing in Semantic Web Databases

- **RDF model**



- **Several serialization formats**
 - N3, Turtle, RDF/XML, etc.

- Describe the schema within the same framework
 - Extends RDF to allow triples to be defined over classes and properties
- Semantics based on type systems
 - Classes (`rdfs:Class` / `rdf:type`)
 - Subclasses (`rdfs:subClassOf`)
 - Typed properties (`rdfs:domain` / `rdfs:range`)

	Datalog	Description Logics
Focus	Instances	Knowledge
Approach	Centralized	Decentralized
Reasoning	Closed-world assumption	Open-world assumption
Unique name	Unique name assumption	Non-unique name assumption

TBox (*Terminology*)

$$X \sqsubseteq Y \Leftrightarrow \forall \mathcal{I} : X^{\mathcal{I}} \subseteq Y^{\mathcal{I}}$$

$$X \equiv Y \Leftrightarrow X \sqsubseteq Y \wedge Y \sqsubseteq X$$

ABox (*Assertions*)

$C(a)$

$R(a,b)$

Source: DL Handbook

Woman	\equiv	Person \sqcap Female
Man	\equiv	Person $\sqcap \neg$ Woman
Mother	\equiv	Woman $\sqcap \exists$ hasChild.Person
Father	\equiv	Man $\sqcap \exists$ hasChild.Person
Parent	\equiv	Father \sqcup Mother
Grandmother	\equiv	Mother $\sqcap \exists$ hasChild.Parent
MotherWithManyChildren	\equiv	Mother $\sqcap \geq 3$ hasChild
MotherWithoutDaughter	\equiv	Mother $\sqcap \forall$ hasChild. \neg Woman
Wife	\equiv	Woman $\sqcap \exists$ hasHusband.Man

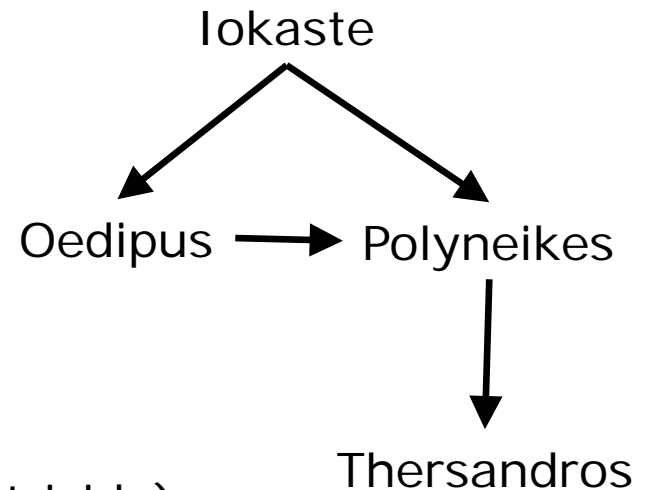
Source: DL Handbook

- **Traditional Reasoning Inferences**
 - **Concept satisfiability:** Is a concept non-contradictory?
 - Example: Concept correctness
 - **Subsumption:** Is an ontology concept *C* more general than another concept *D*?
 - Example: Taxonomies and equivalences
 - **Query Answering:** All asserted instances that satisfy a concept description
 - Example: Arbitrary queries

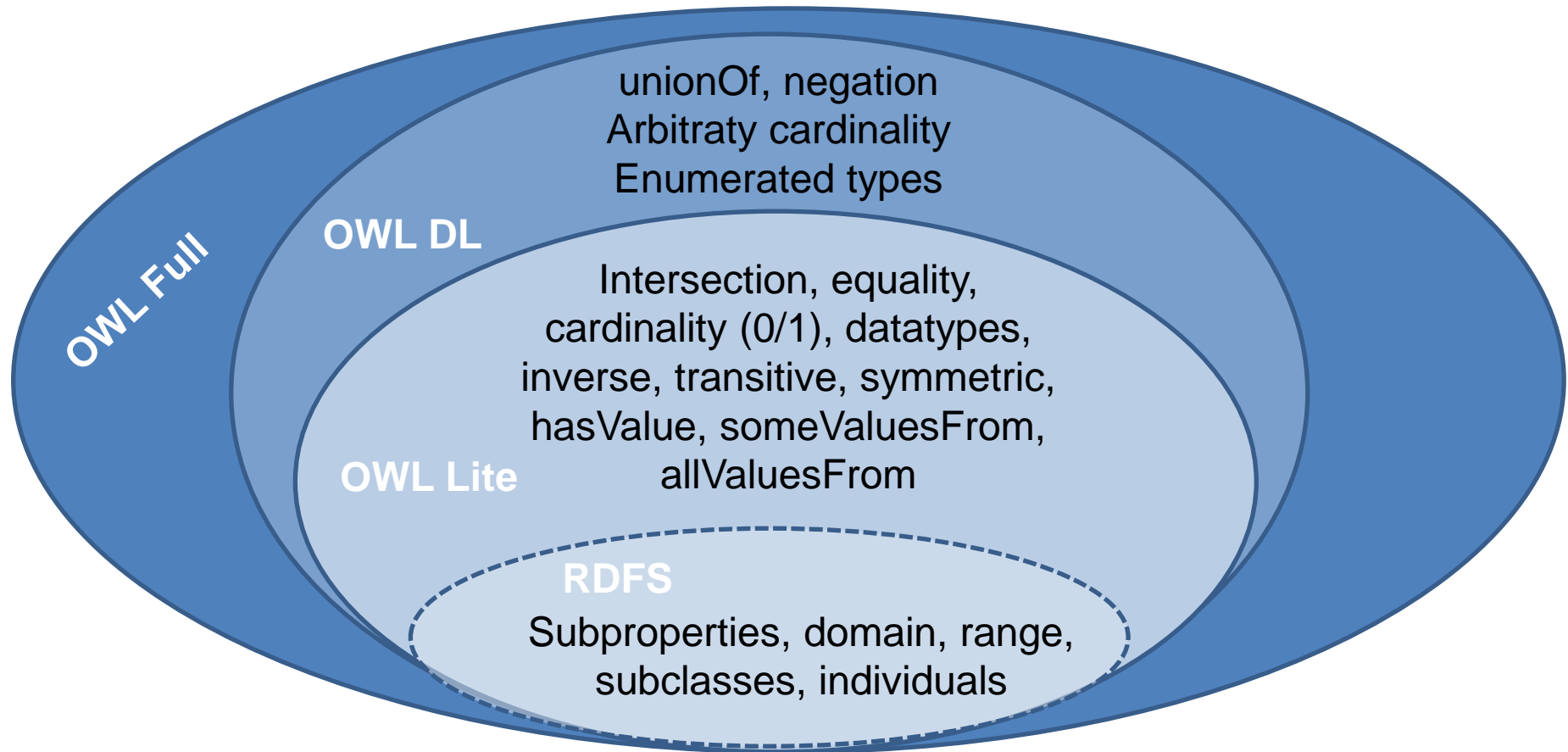
- **Something evaluates false only if it contradicts other information in the ontology**

hasSon(Iokaste, Oedipus)
hasSon(Iokaste, Polyneikes)
hasSon(Oedipus, Polyneikes)
hasSon(Polyneikes, Thersandros)
patricide(Oedipus)
 \neg patricide(Thersandros)

Query $\equiv \exists \text{hasSon}. (\text{patricide} \sqcap \exists \text{hasSon}. \neg \text{patricide})$
ABox \models Query(Iokaste)?



Source: DL Handbook



Source: Sven Groppe. Data Management and Query Processing in Semantic Web Databases

OWL 2 EL:

Based on *EL*++

Large number of
properties / classes

Reasoning:
Polynomial with
regard to the
ontology TBOX

OWL 2 QL:

Based on DL-Lite

Captures (most of)
ER and UML
expressive power

Reasoning:
Reducible to
LOGSPACE (i.e., DBs)

OWL 2 RL:

Based on Description
Logic programs

Scalable reasoning
without sacrificing
much expressivity

Reasoning:
Polynomial with
regard to the size of
the ontology

A Characterization of BI Systems

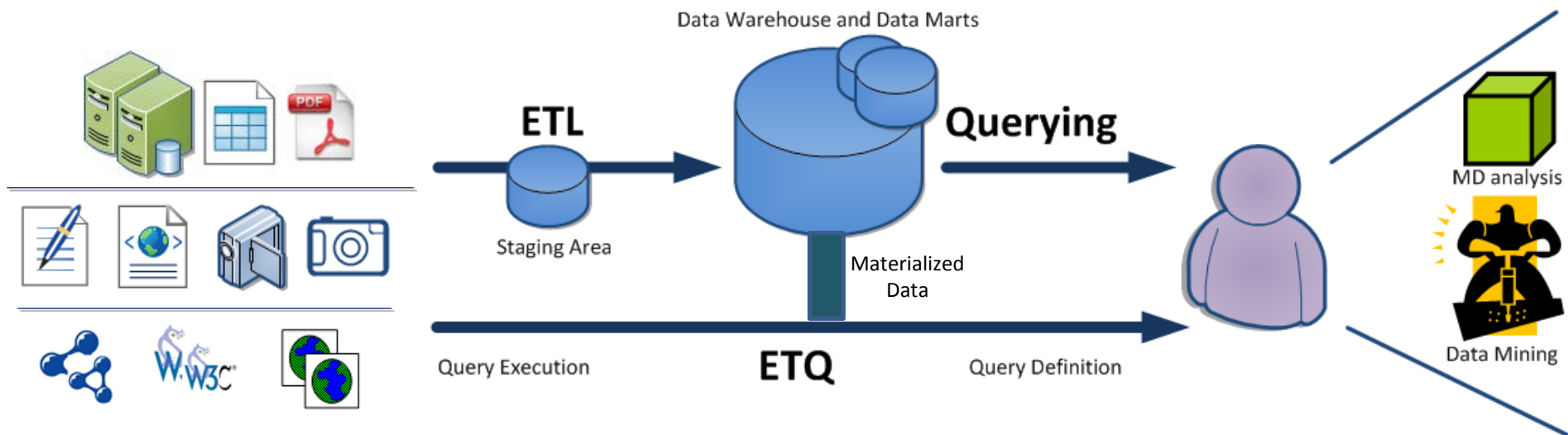
Joint work with:

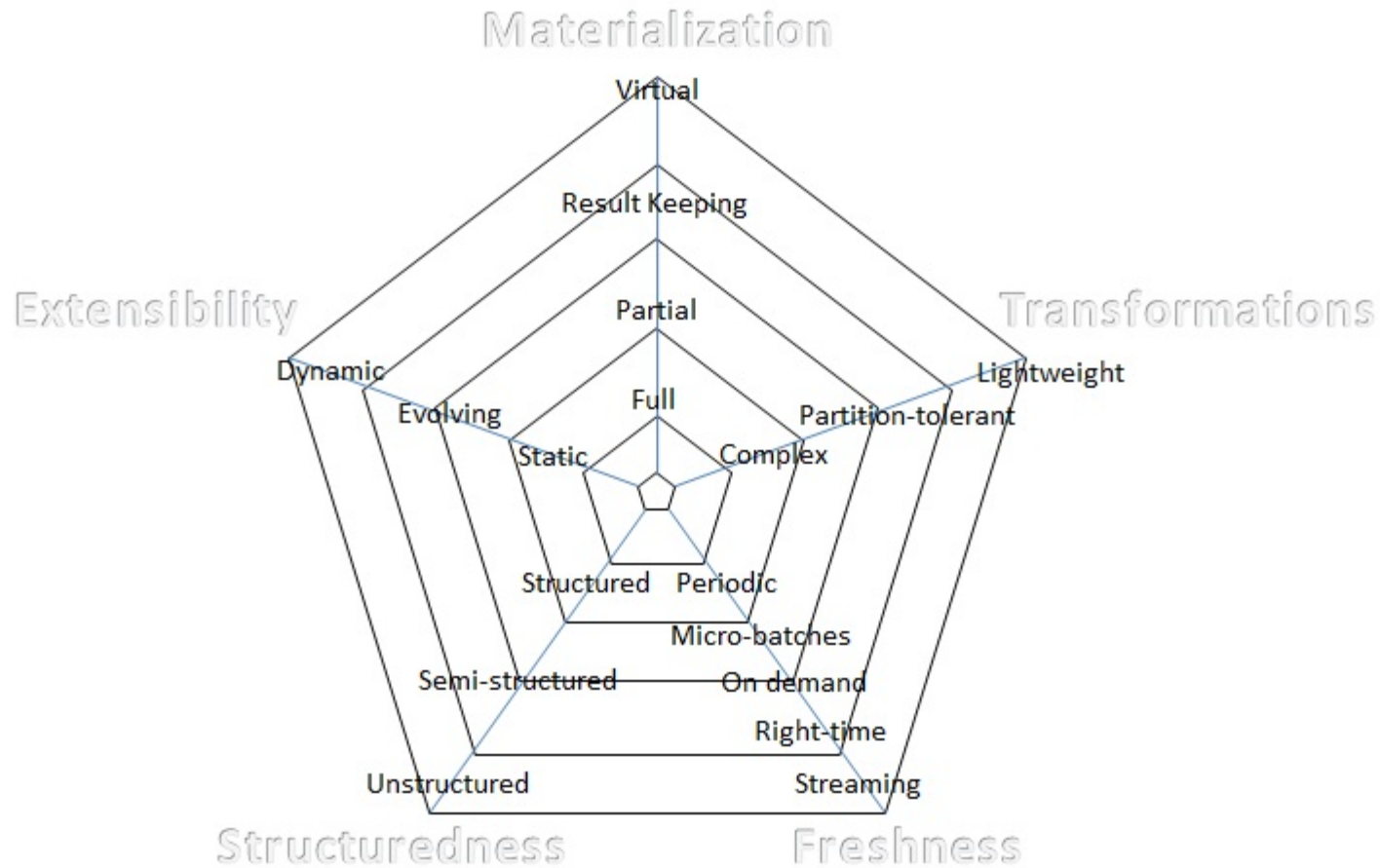
Alberto Abelló (UPC)	Torben Bach Pedersen (AAU)
Rafael Berlanga (UJI)	Victoria Nebot (UJI)
M.J. Aramburu (UJI)	Alkis Simitsis (HP Labs)

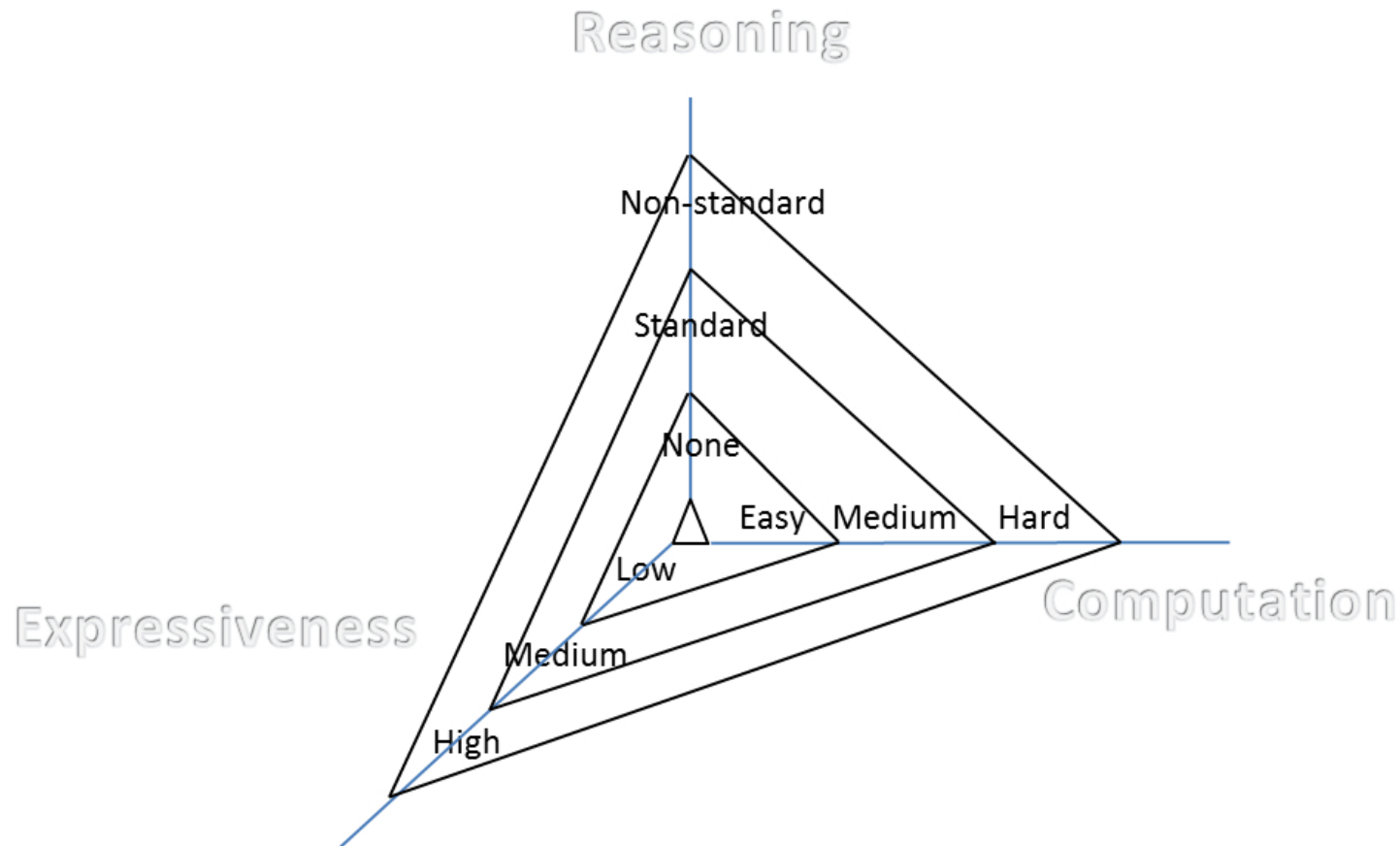
A Characterization of BI Systems



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

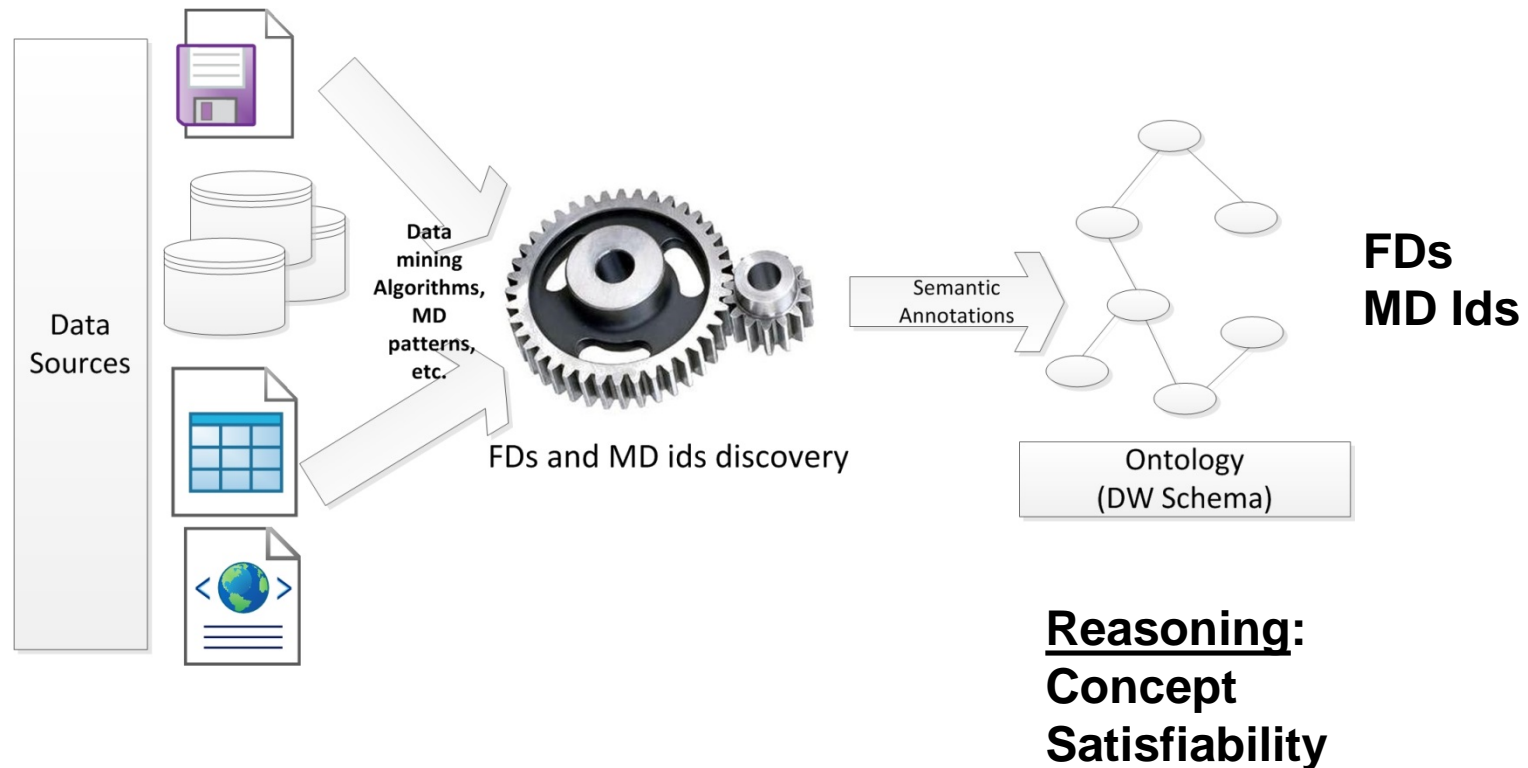




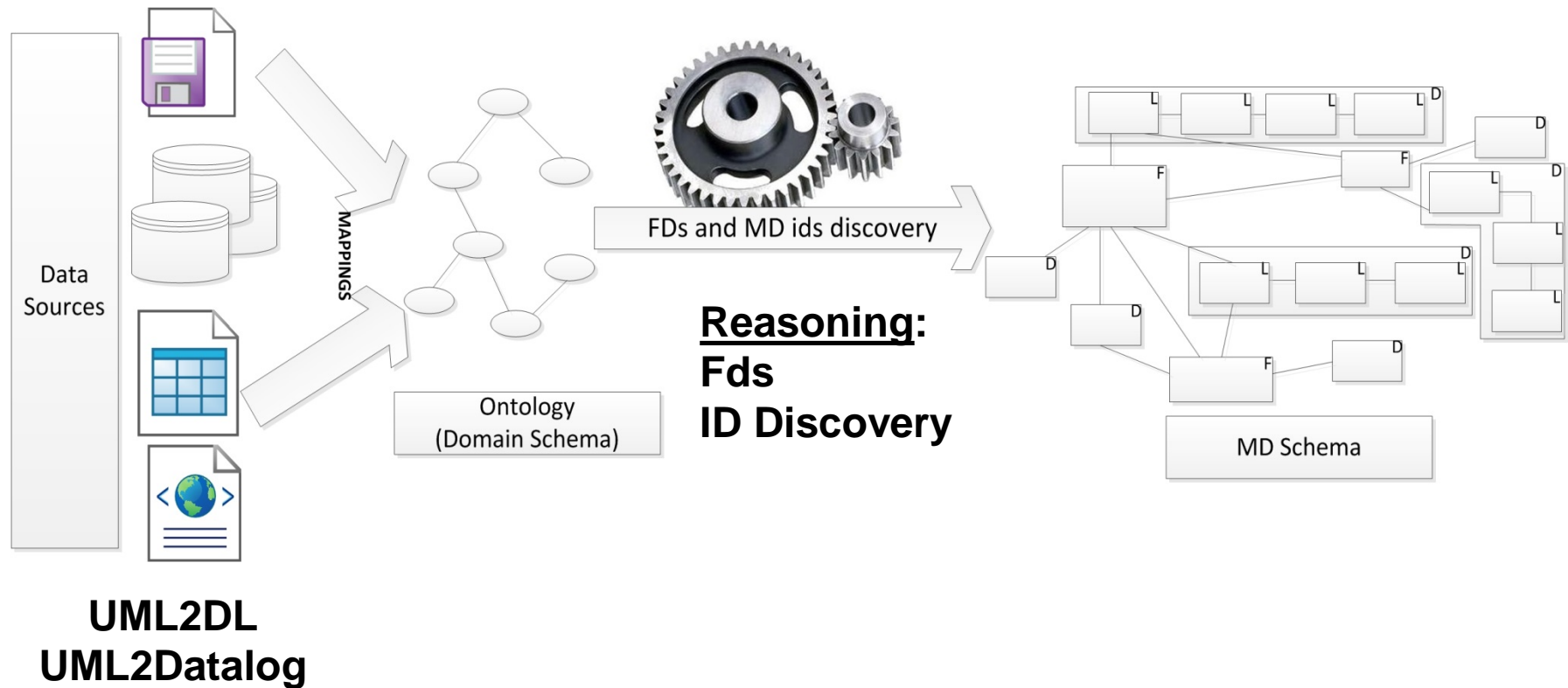


	Reference	Struct.	Mater.	Trans.	Fresh.	Exten.	Reasoning	Expressiveness	Computation
Data Schema Design	Sem.-aware	[33], [61]	○	·	×	×	·	☹	⌚
		[34]	○	·	×	×	·	☹	⌚
		[35]	·	·	×	×	·	☹	☹
		[36]	·	·	×	×	·	☹	☹
		[38], [58]	·	·	×	×	○	⌚	⌚
		[41]	○	·	×	×	○	⌚	☹
	Exploratory	[42]	○	·	×	×	○	⌚	☹
		[47]	⊙	⊙	○	⊙	⊙	☹	☹
		[51]	⊙	·	○	○	⊙	☹	☹
		[52]	⊙	·	○	·	⊙	☹	☹
		[53]	⊙	·	○	·	○	☹	☹
		[55]	⊙	·	○	·	☹	☹	⌚
Data Provisioning	Sem.-aw.	[33], [61]	○	·	⊙	·	○	☹	⌚
		[43], [63]	○	·	○	×	○	☹	☹
		[51]	⊙	·	○	○	⊙	☹	☹
		[62]	○	·	○	×	○	☹	☹
	Expl.	[47]	⊙	⊙	○	⊙	⊙	☹	☹
		[56]	⊙	⊙	⊙	⊙	☹	☹	⌚
		[64]	○	○	○	○	☹	×	☹
		[65]	⊙	⊙	⊙	⊙	☹	☹	⌚

- Data Schema Design (Trend I)



- Data Schema Design (Trend II)



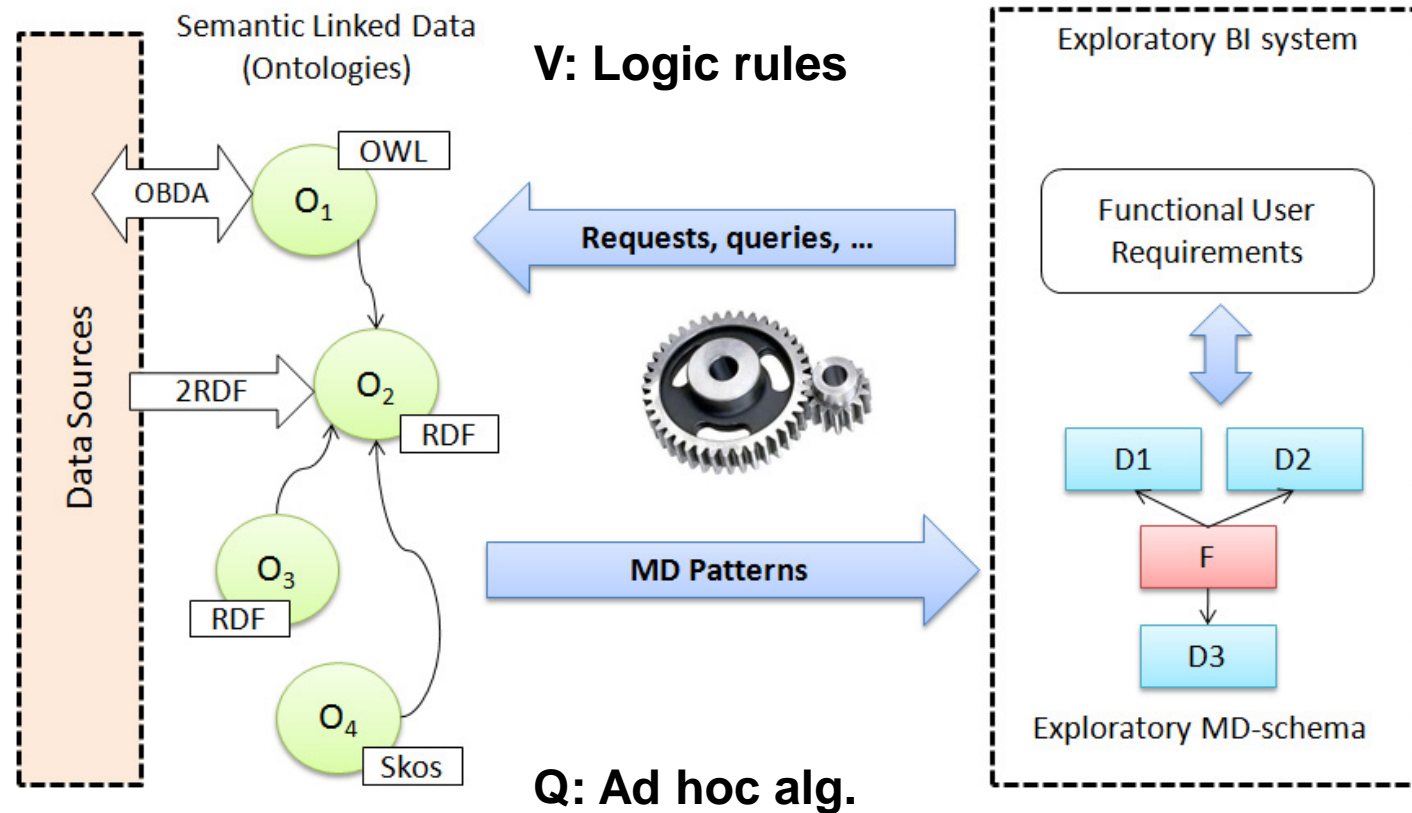
• Data Schema Design

	Reference	Struct.	Mater.	Trans.	Fresh.	Exten.	Reasoning	Expressiveness	Computation
Sem.-aware	[33], [61]	⊙	•	×	×	•	⌚	😄	⌚
	[34]	⊙	•	×	×	•	⌚	😄	⌚
	[35]	•	•	×	×	•	⌚	😞	😞
	[36]	•	•	×	×	•	⌚	😞	😞
	[38], [58]	•	•	×	×	⊙	⌚	⌚	⌚
	[41]	⊙	•	×	×	⊙	⌚	⌚	😞
	[42]	⊙	•	×	×	⊙	⌚	⌚	😞

Legend:

Criterion/Category	Easy (😄)	Feasible (⌚)	Difficult (😞)
Reasoning	None	Standard	Non-standard
Expressiveness	RDF/Taxonomies	Datalog, OWL2 profiles, DL-Lite	OWL-DL
Computation	Less than Polynomial	Polynomial	Exponential

- Data Schema Design (Trend III)

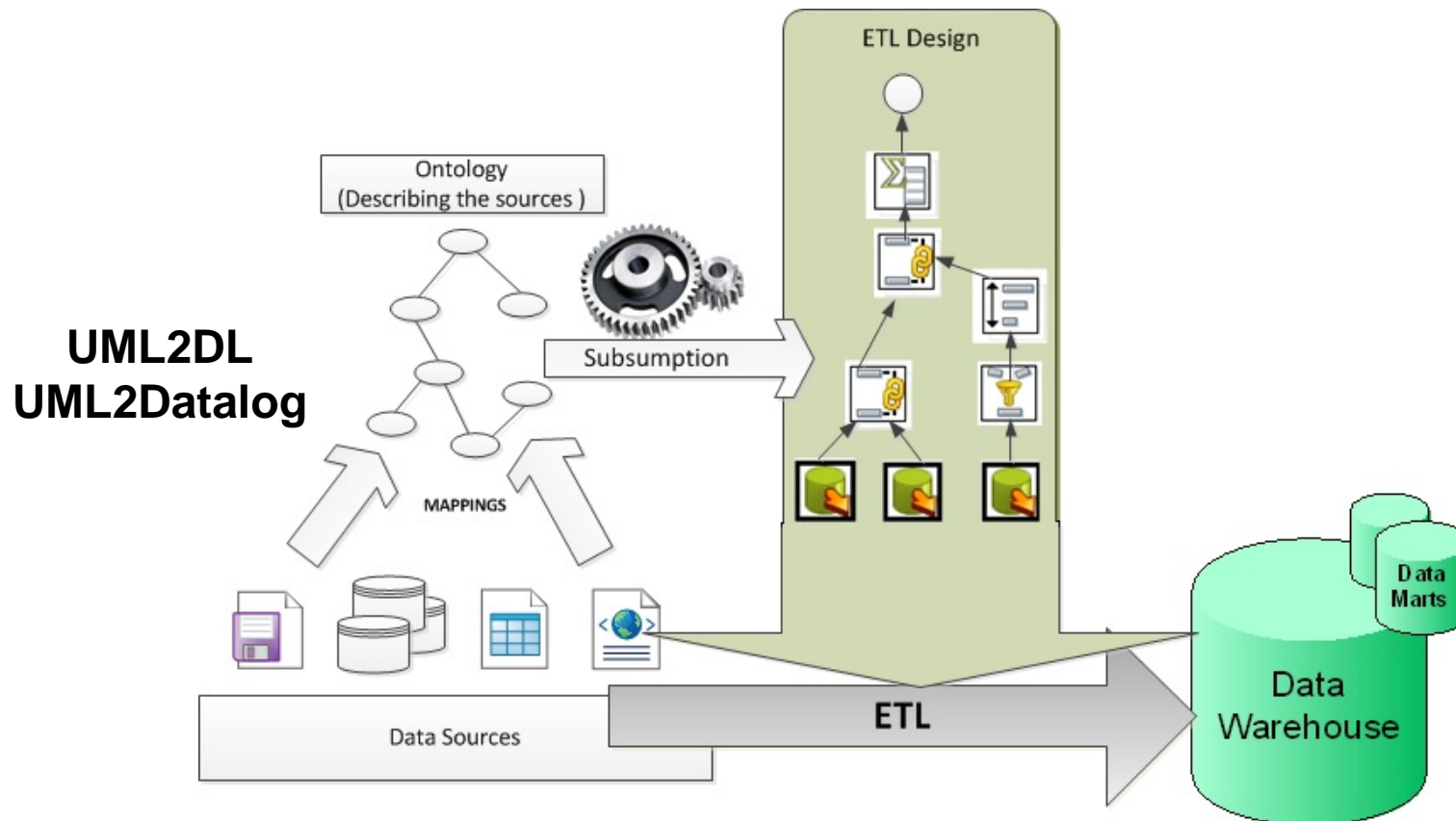


	Reference	Struct.	Mater.	Trans.	Fresh.	Exten.	Reasoning	Expressiveness	Computation
Exploratory	[47]	⊙	⊙	⊙	⊙	⊙	⌚	☹	☹
	[51]	⊙	.	⊙	⊙	⊙	⌚	☹	☹
	[52]	⊙	.	⊙	.	⊙	⌚	☹	☹
	[53]	⊙	.	⊙	.	⊙	⌚	☹	☹
	[55]	⊙	.	⊙	.	⊙	😊	😊	⌚
	[56]	⊙	⊙	⊙	⊙	⊙	😊	😊	⌚

Legend:

Criterion/Category	Easy (😊)	Feasible (⌚)	Difficult (☹)
Reasoning	None	Standard	Non-standard
Expressiveness	RDF/Taxonomies	Datalog, OWL2 profiles, DL-Lite	OWL-DL
Computation	Less than Polynomial	Polynomial	Exponential

- Data Provisioning



Data Provisioning	Sem.-aw.	Reference	Struct.	Mater.	Trans.	Fresh.	Exten.	Reasoning	Expressiveness	Computation
		[33], [61]	⊙	·	⊙	·	⊙	⌚	😊	⌚
		[43], [63]	⊙	·	⊙	×	⊙	⌚	😞	😞
		[51]	⊙	·	⊙	⊙	⊙	⌚	😞	😞
		[62]	⊙	·	⊙	×	⊙	⌚	😊	😊

Legend:

Criterion/Category	Easy (😊)	Feasible (⌚)	Difficult (😞)
Reasoning	None	Standard	Non-standard
Expressiveness	RDF/Taxonomies	Datalog, OWL2 profiles, DL-Lite	OWL-DL
Computation	Less than Polynomial	Polynomial	Exponential

		Reference	Struct.	Mater.	Trans.	Fresh.	Exten.	Reasoning	Expressiveness	Computation
Data Provisioning	Expl.	[47]	○	○	○	○	○	⌚	☹	☹
		[56]	○	○	○	○	○	😊	😊	⌚
		[64]	○	○	○	○	○	😊	×	😊
		[65]	○	○	○	○	.	😊	😊	⌚

Legend:

Criterion/Category	Easy (😊)	Feasible (⌚)	Difficult (☹)
Reasoning	None	Standard	Non-standard
Expressiveness	RDF/Taxonomies	Datalog, OWL2 profiles, DL-Lite	OWL-DL
Computation	Less than Polynomial	Polynomial	Exponential

- **Small steps towards exploratory BI**
 - Far from being mature!
- **Expressivity and Computation is correlated (expected!)**
 - Non-standard reasoning services not considered
- **Reasoning at the instance level is prohibitive**
 - Size matters for data complexity
 - Current solutions for ETL yield combined complexity
 - Datalog and Ontology-Based Data Access
- **Reasoning at the schema level is feasible...**
 - Combined complexity

... but exploratory systems demand further advances

 - Aggregations have been completely overlooked
 - LaV Vs. GaV

- **The multidimensional model as a mature modeling paradigm for exploratory BI**
 - Specific ontology languages (RDF Data Cube Vocabulary)
 - Detecting MD expressions
 - MD optimization
 - Expressiveness Vs. Reasoning Vs. Kind of Queries
 - Cloud and parallel computing (TROWL)
- **The Web and Data Silos / Data Niches**
 - Autonomy Vs. Consistency
 - Registries (e.g., Freebase, DBPedia)

A Semantic-Aware Business Intelligence System

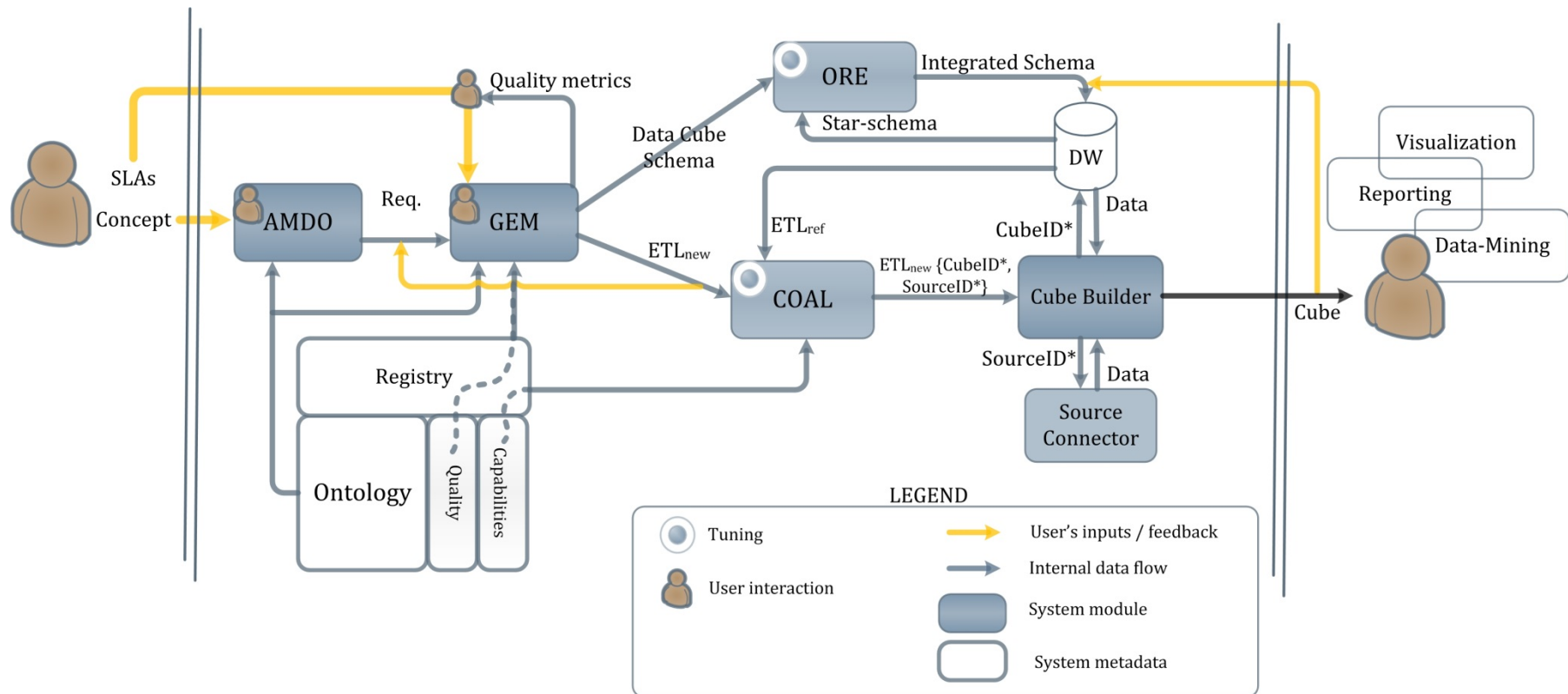
Joint work with:

Alberto Abelló (UPC)

Petar Jovanovic (UPC)

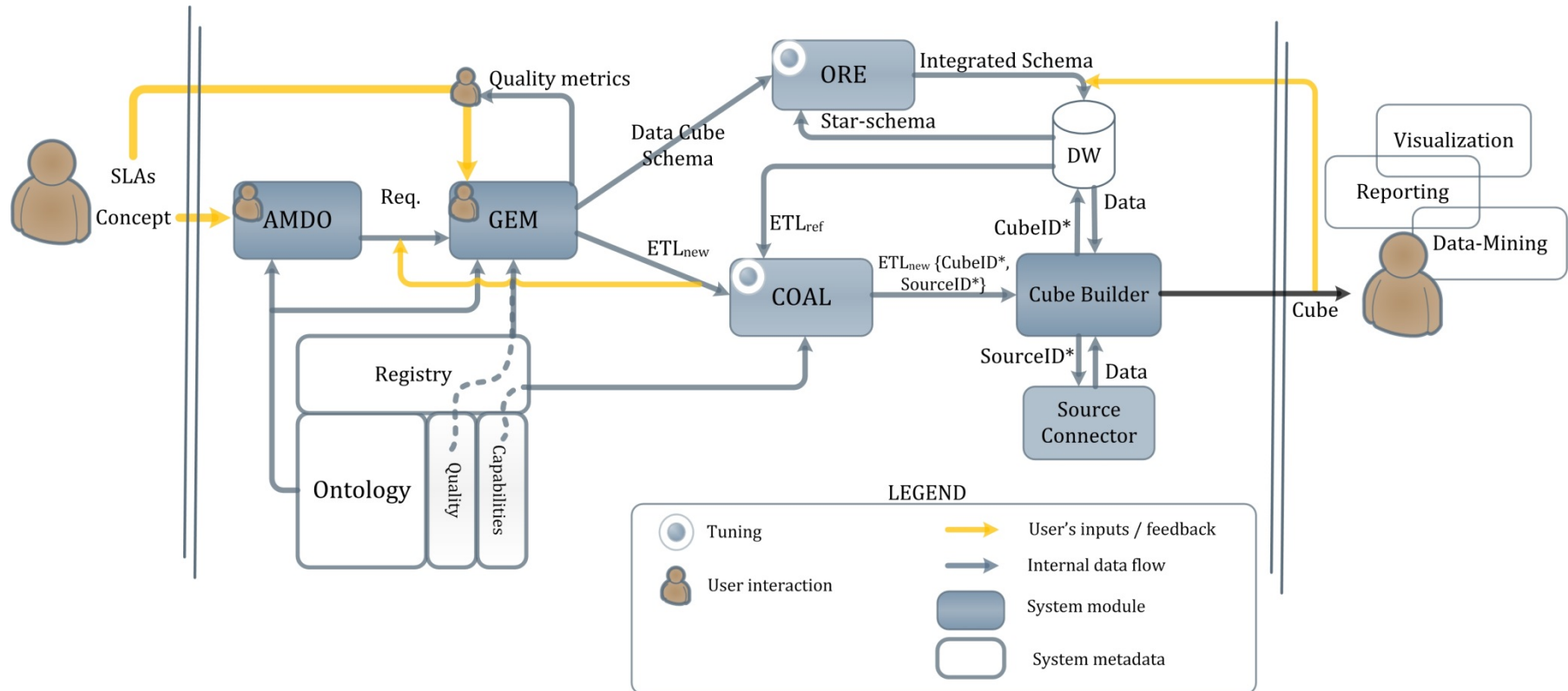
Alkis Simitsis (HP Labs)

• Towards Exploratory BI:

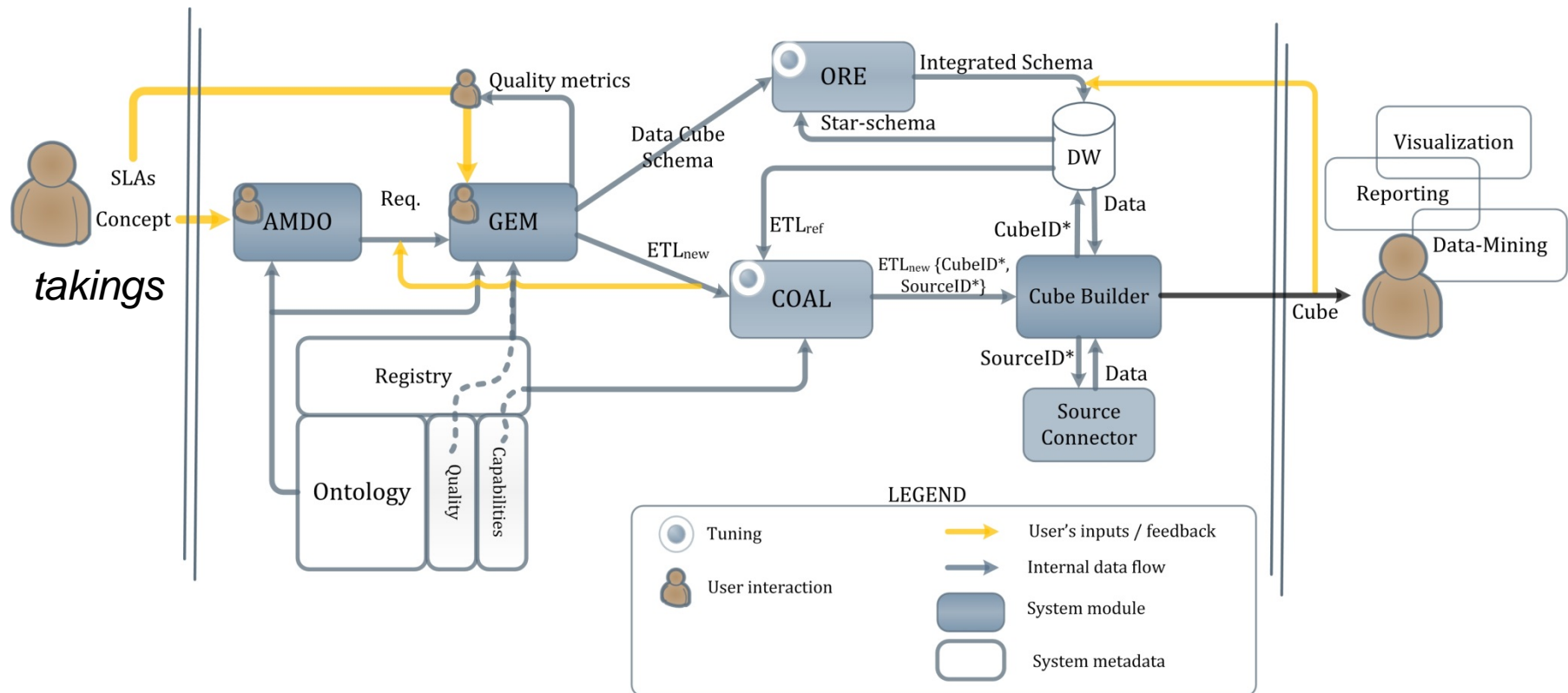


MULTIDIMENSIONAL MODEL AND BPMN

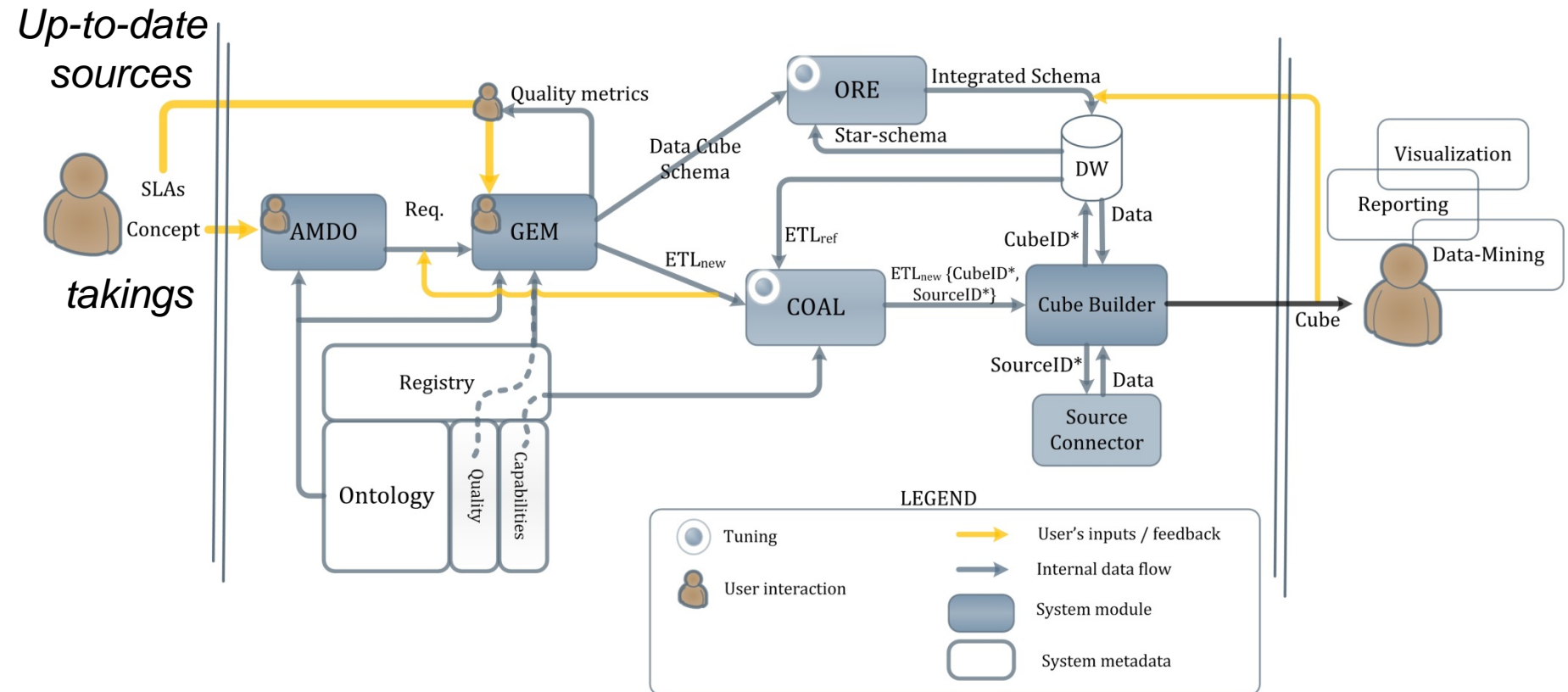
A Toy Example



A Toy Example

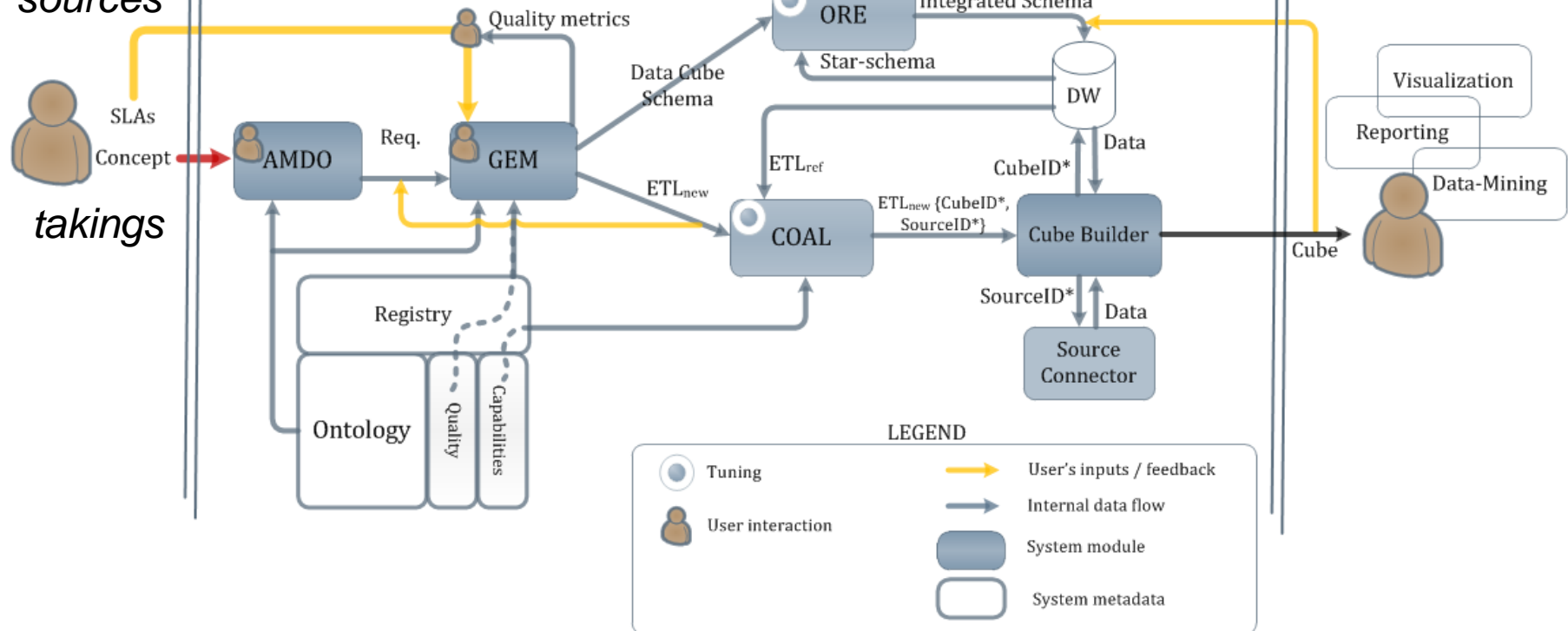


A Toy Example



A Toy Example

Up-to-date
sources



A Toy Example

Up-to-date
sources



SLAs
Concept

takings



Re

Proposed fact: *takings* (100%)

Measures	Dimensions
income (100%)	producer (93%)
expenses (95%)	movie (90%)
directorWage (80%)	director (87%)
...	year(released) (83%)
	...

Registry

Ontology

Quality

Capabilities

ata

der

Data
Source
Connector

Cube

Visualization

Reporting

Data-Mining

LEGEND



Tuning



User interaction



User's inputs / feedback



Internal data flow



System module



System metadata

A Toy Example

Up-to-date
sources



SLAs
Concept

takings



Re

Proposed fact: *takings* (100%)

Measures	Dimensions
income (100%)	producer (93%)
expenses (95%)	movie (90%)
directorWage (80%)	director (87%)
...	year(released) (83%)
	...

Registry

Ontology

Quality

Capabilities

ata

der

Cube

Data
Source
Connector

Visualization

Reporting

Data-Mining

LEGEND



Tuning



User interaction



User's inputs / feedback



Internal data flow

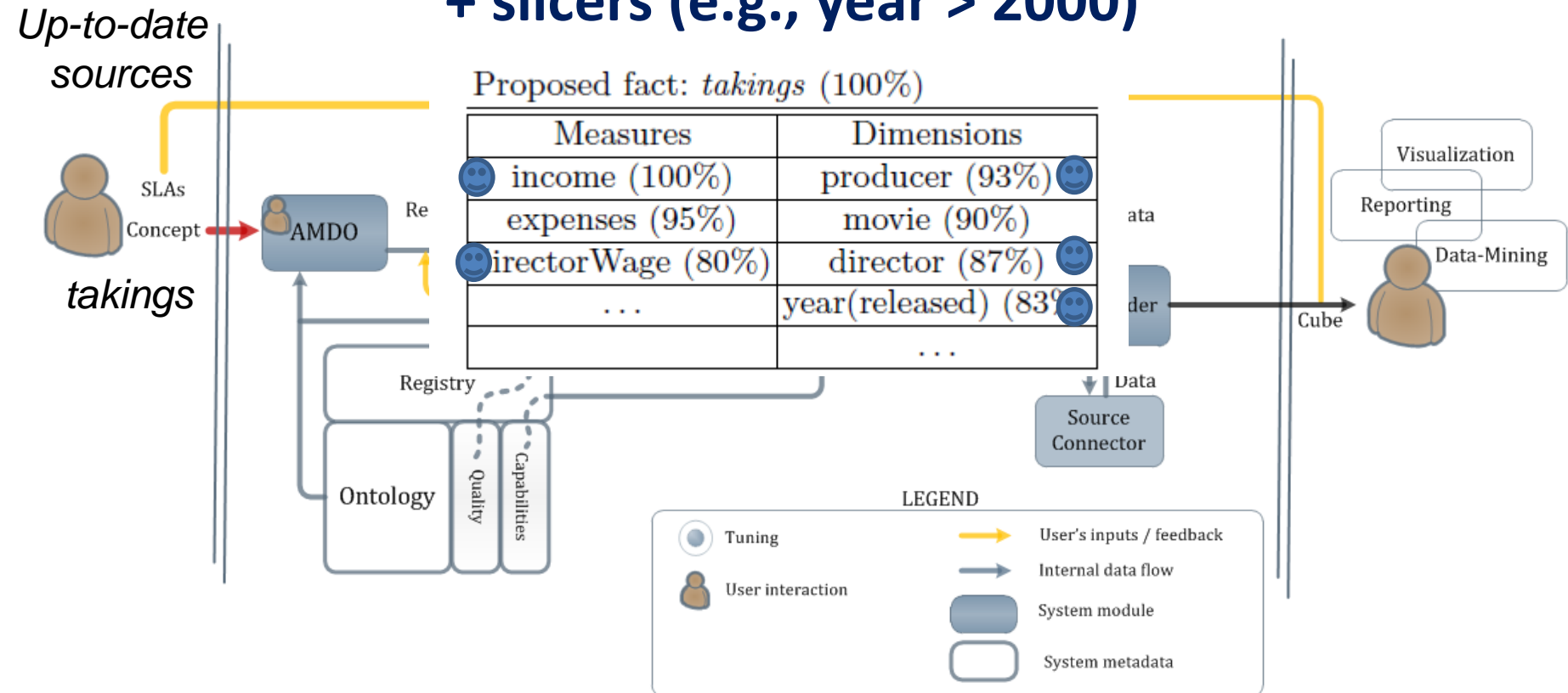


System module



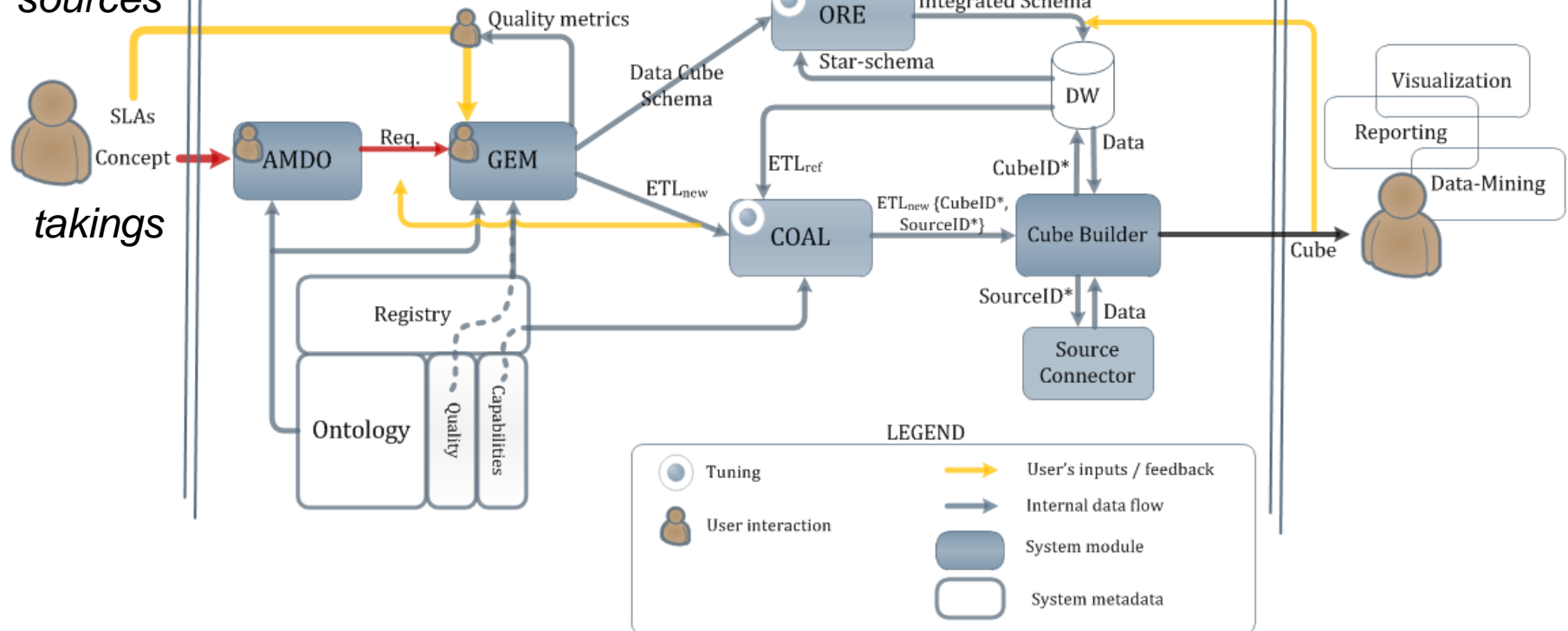
System metadata

+ slicers (e.g., year > 2000)

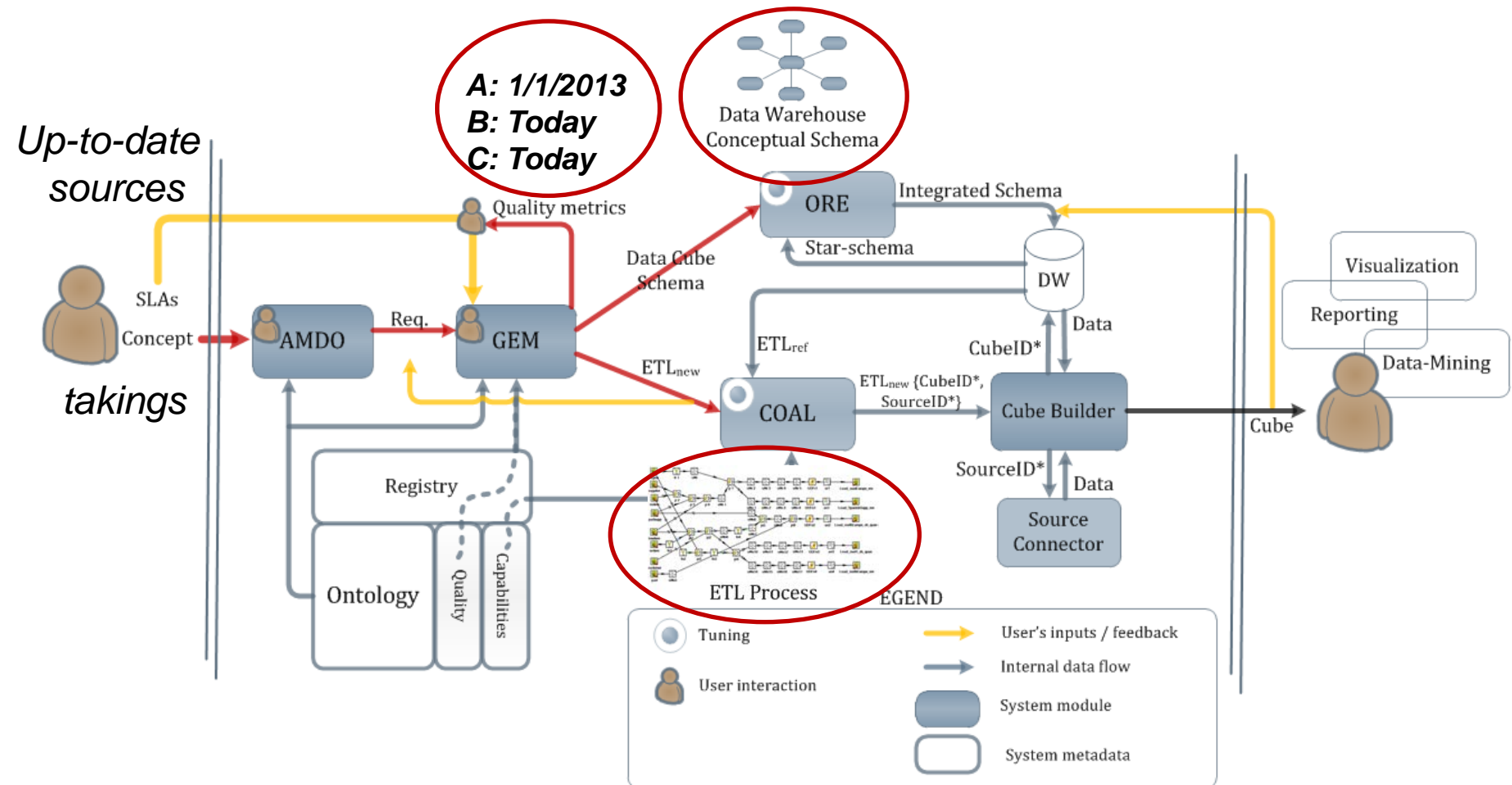


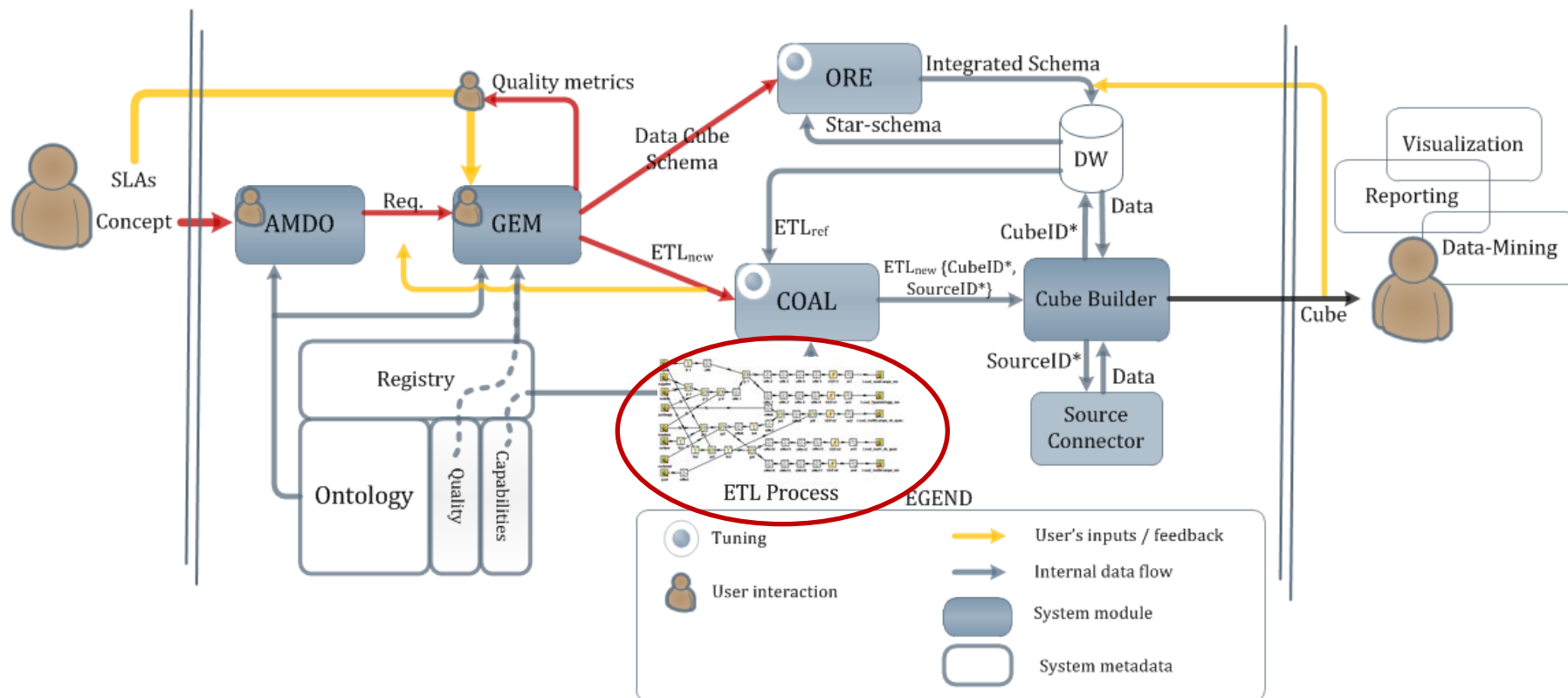
A Toy Example

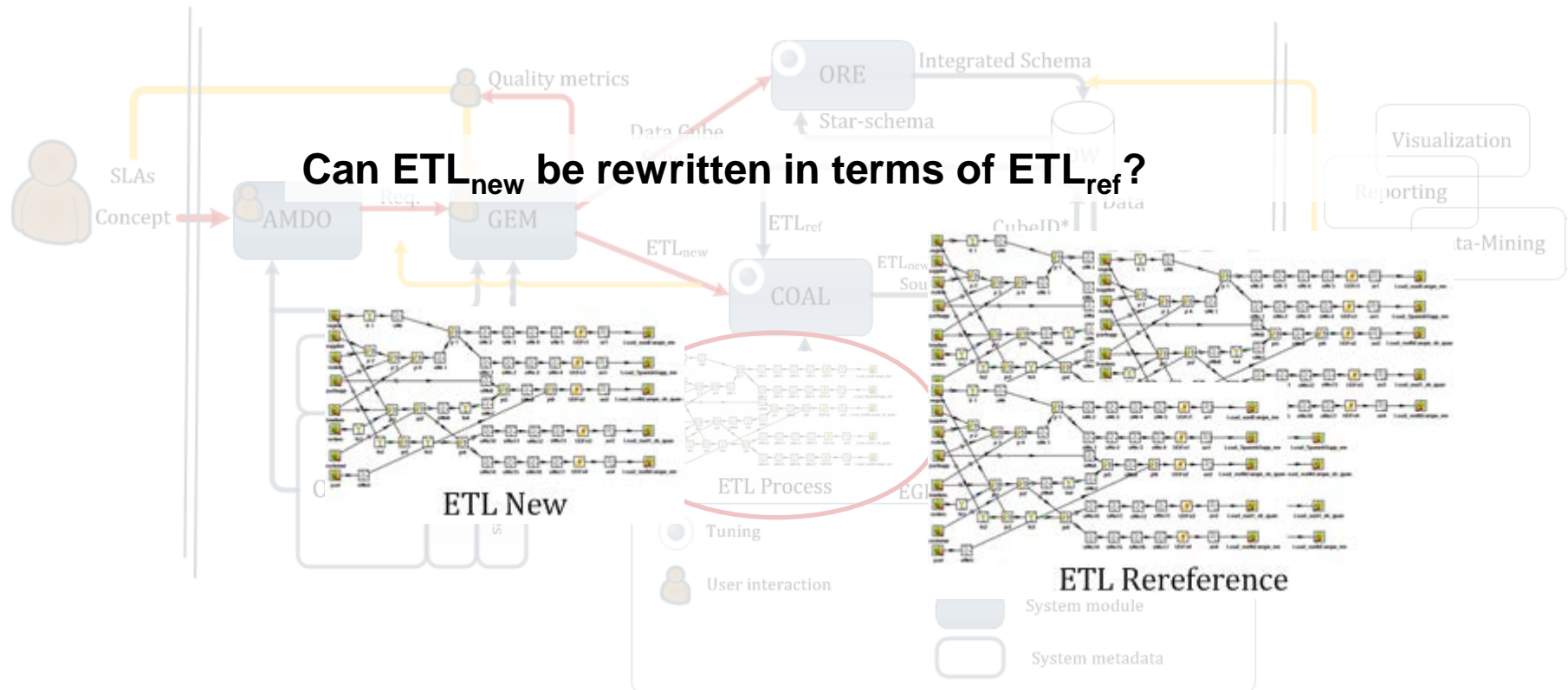
Up-to-date
sources



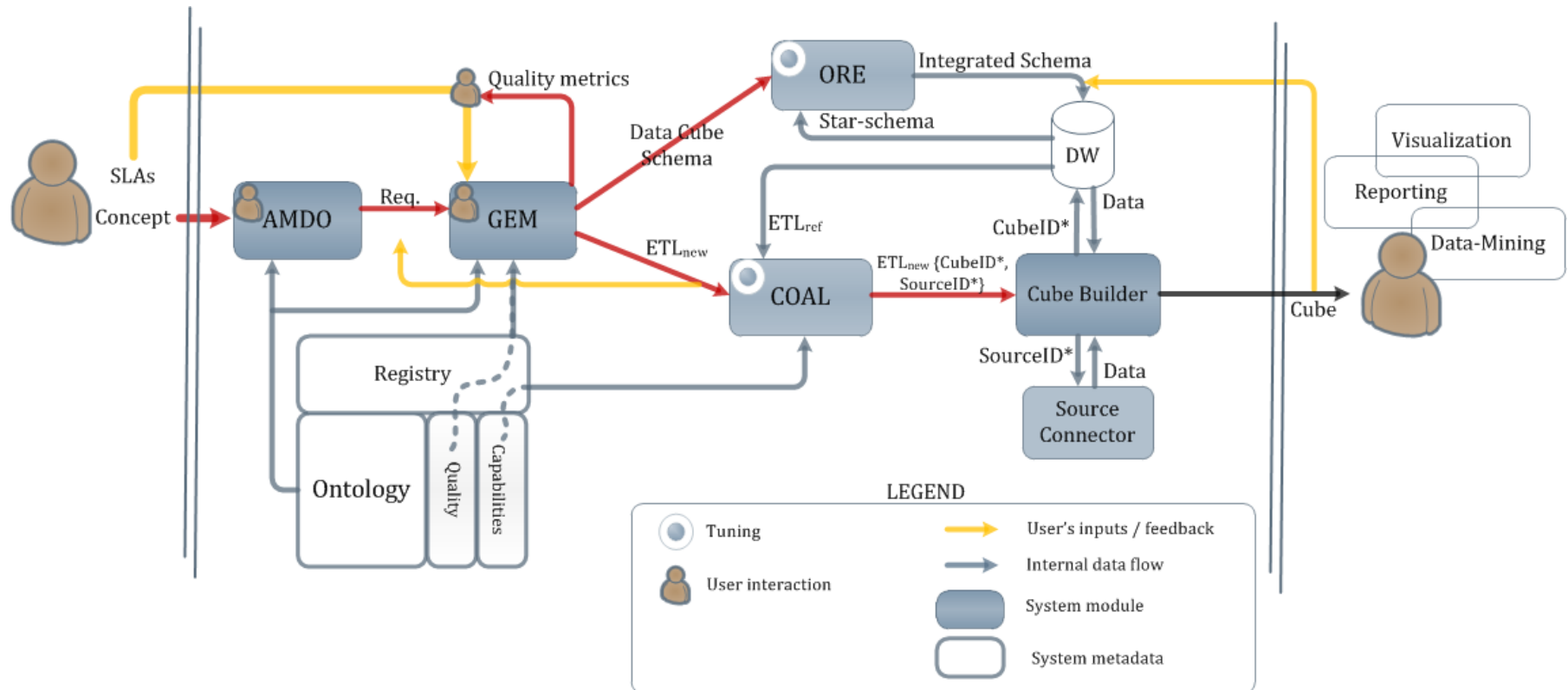
A Toy Example



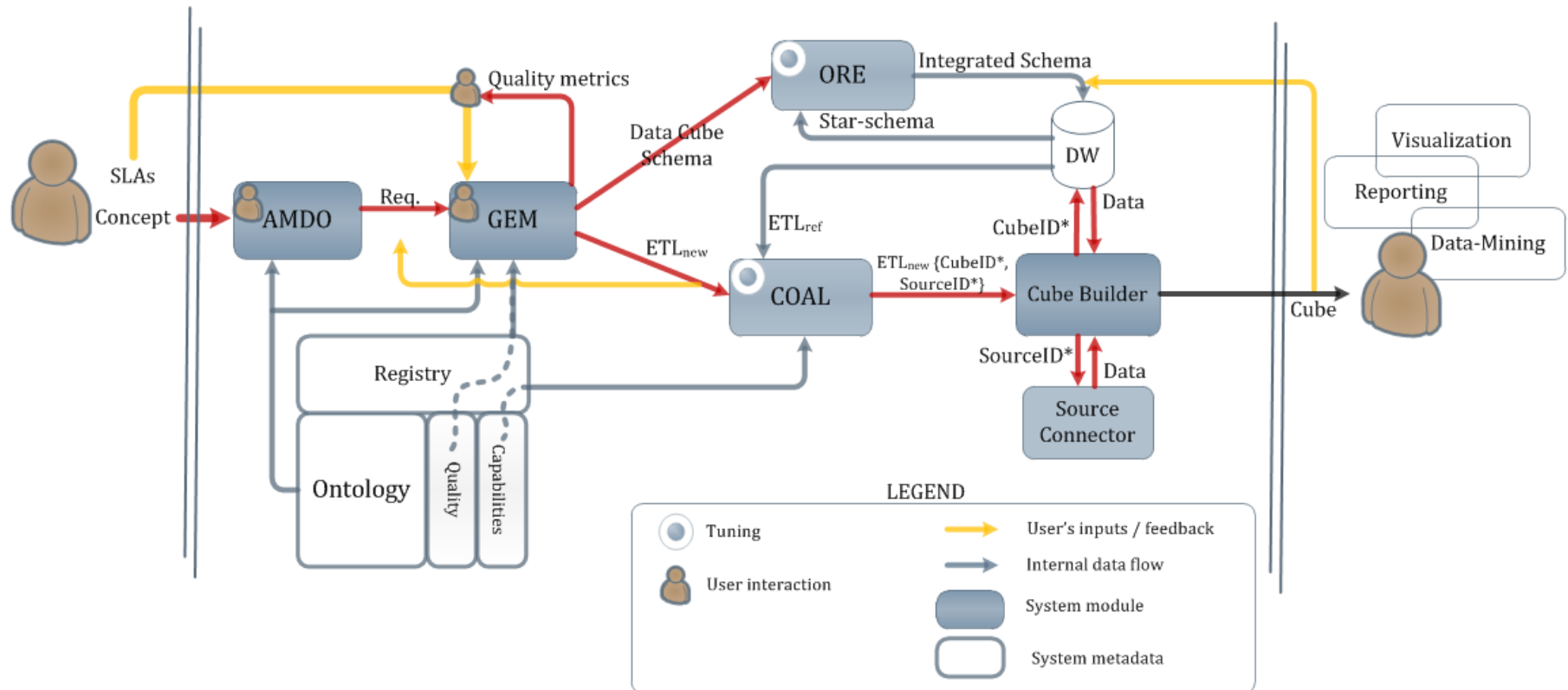




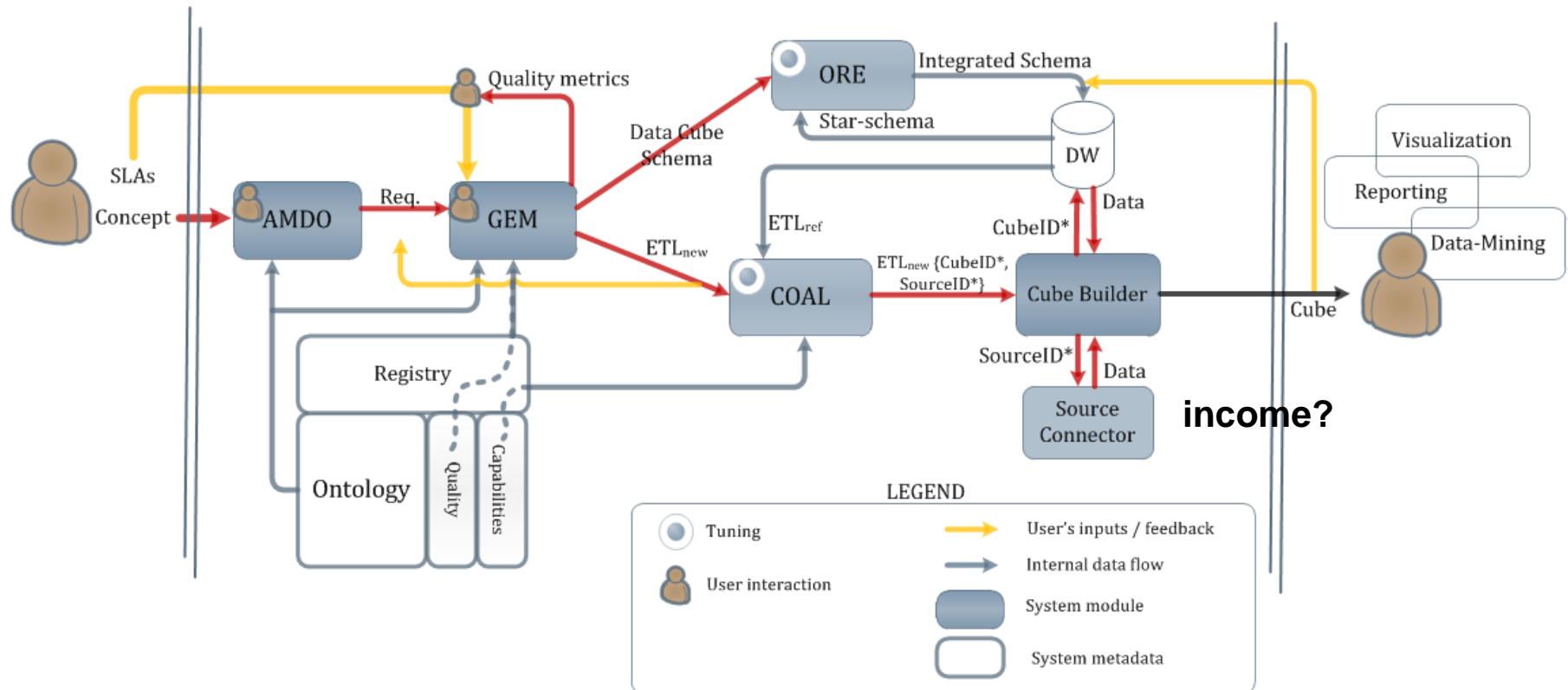
A Toy Example



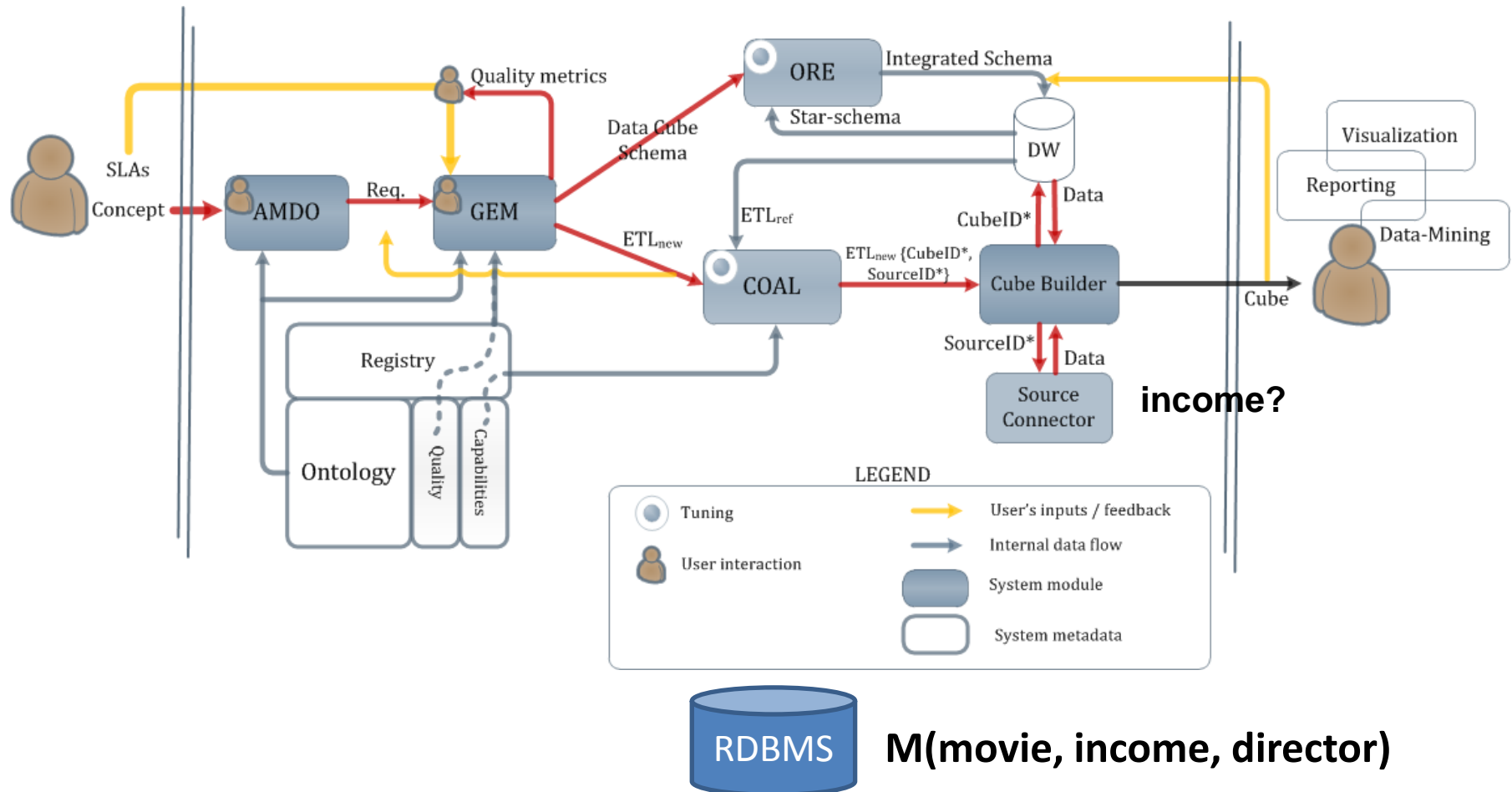
A Toy Example



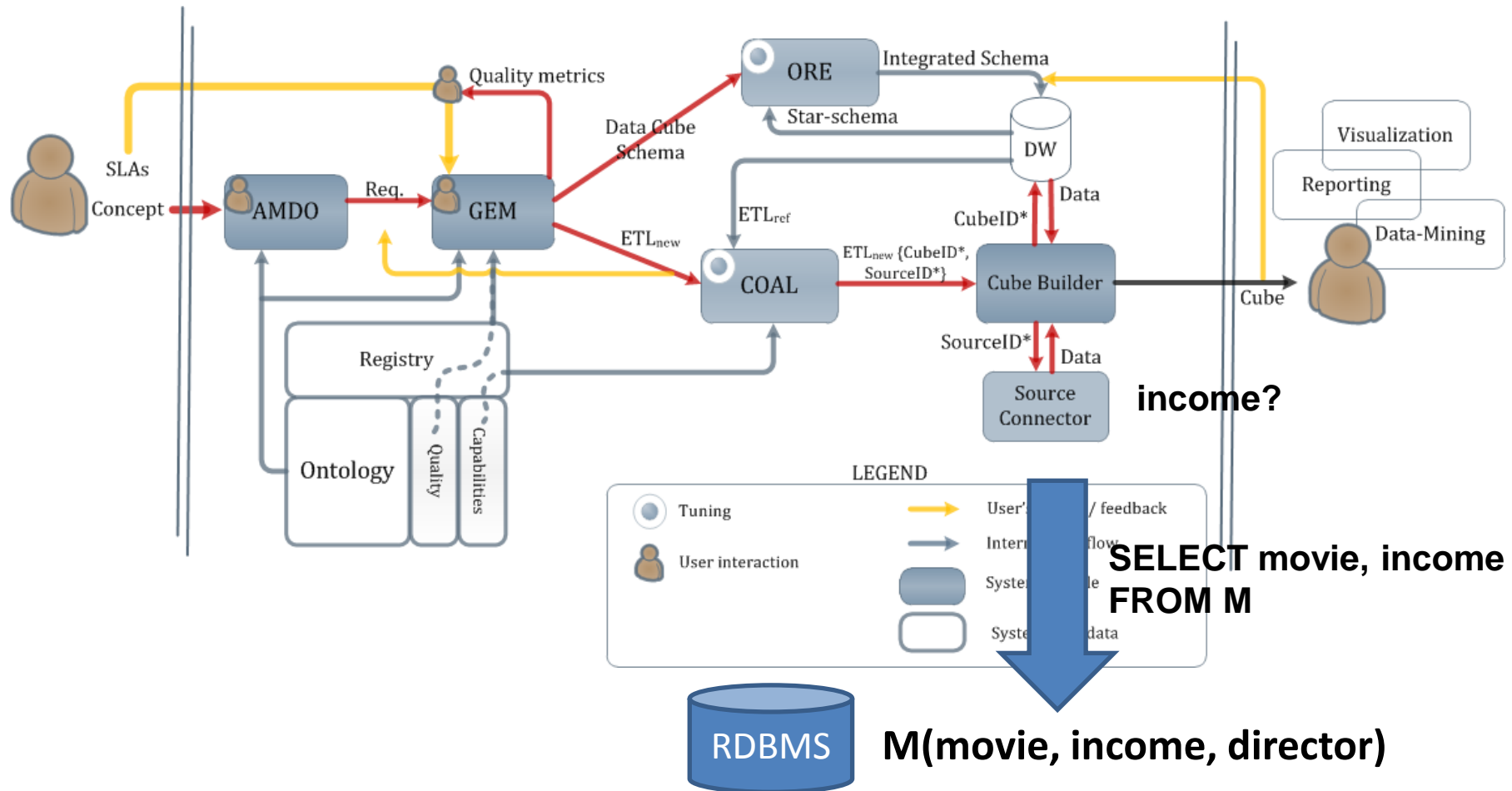
A Toy Example



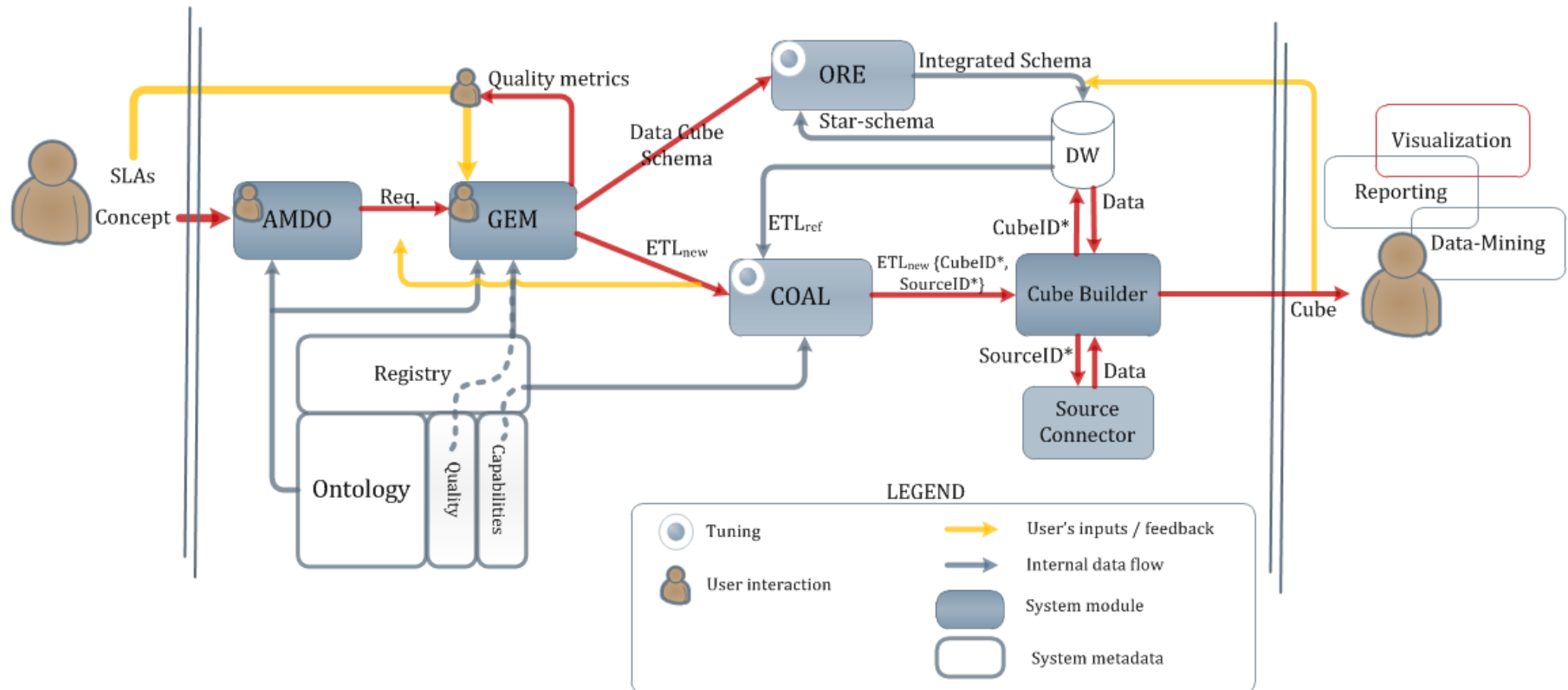
A Toy Example

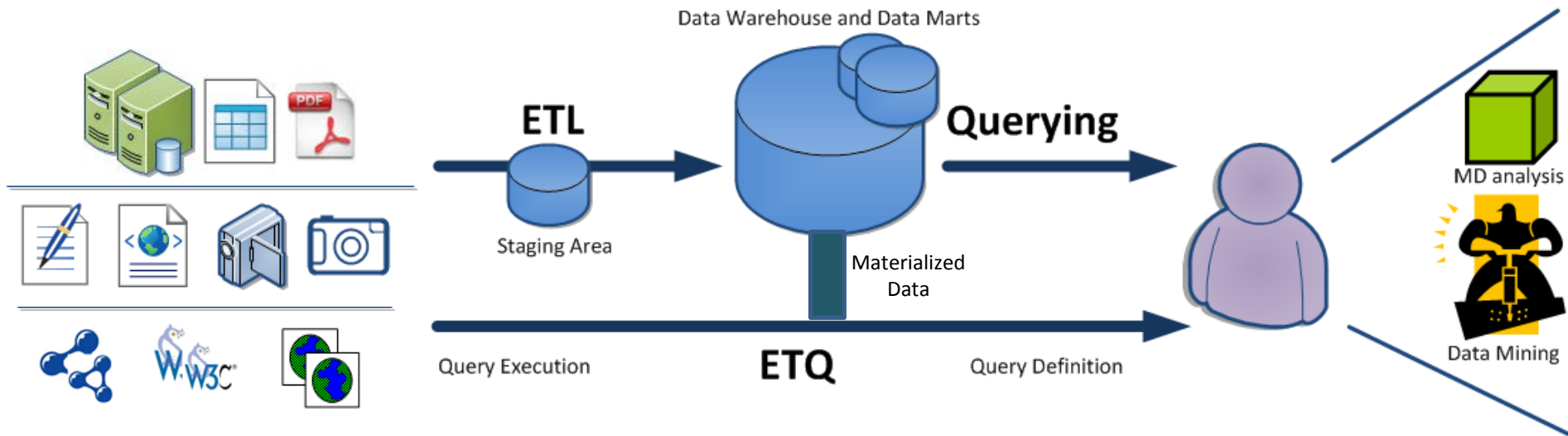


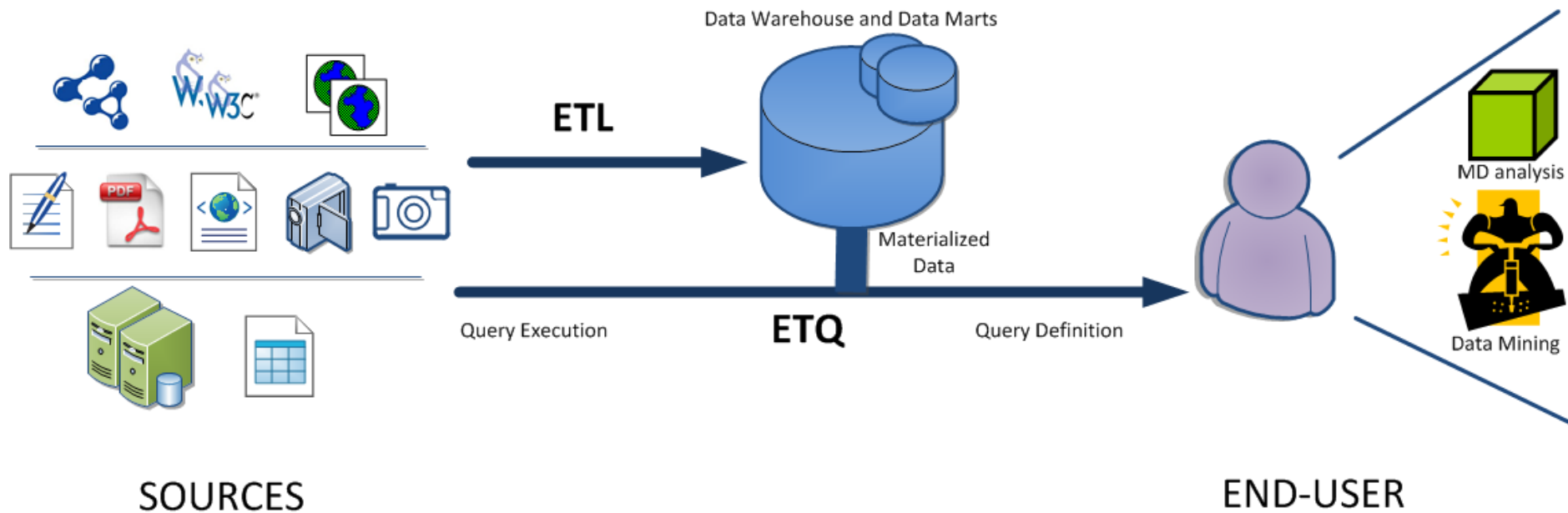
A Toy Example



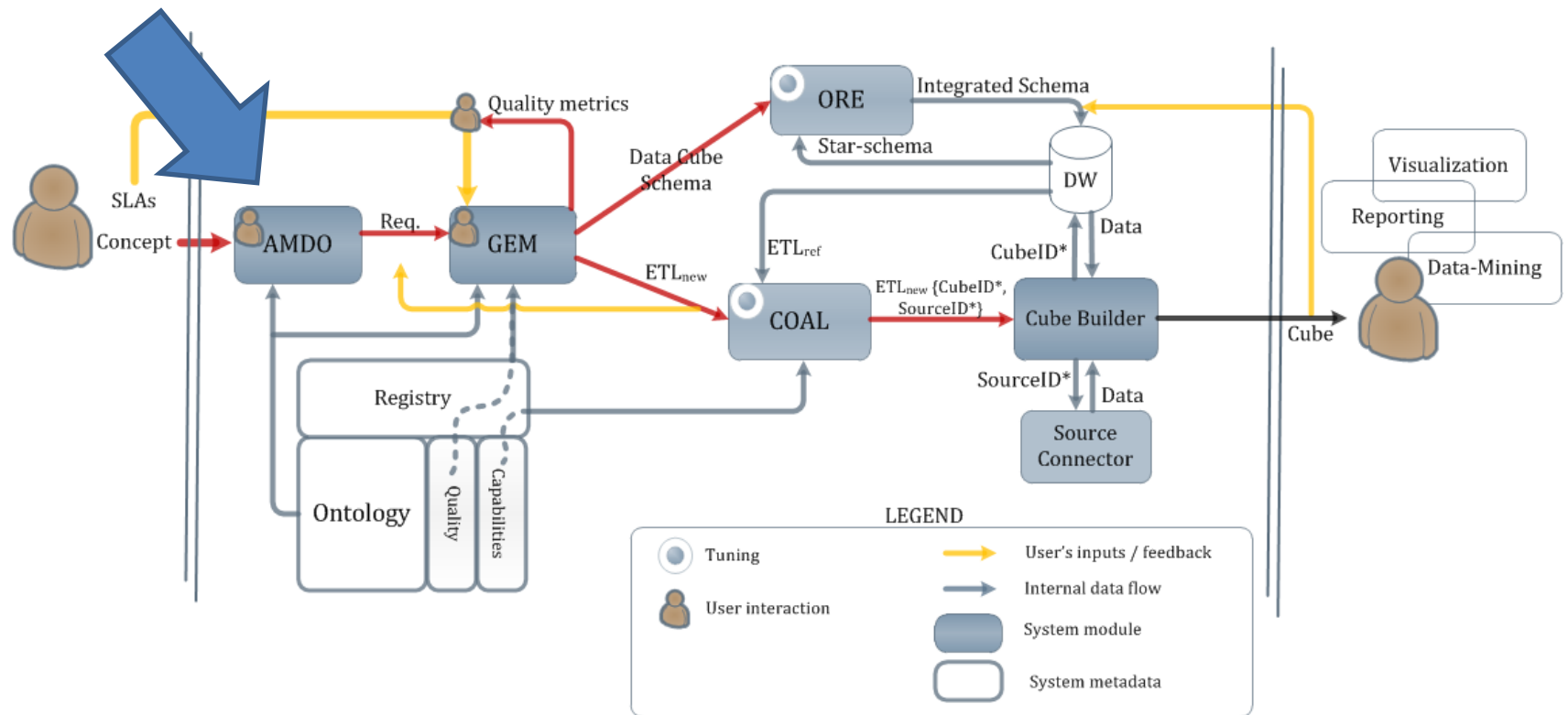
A Toy Example

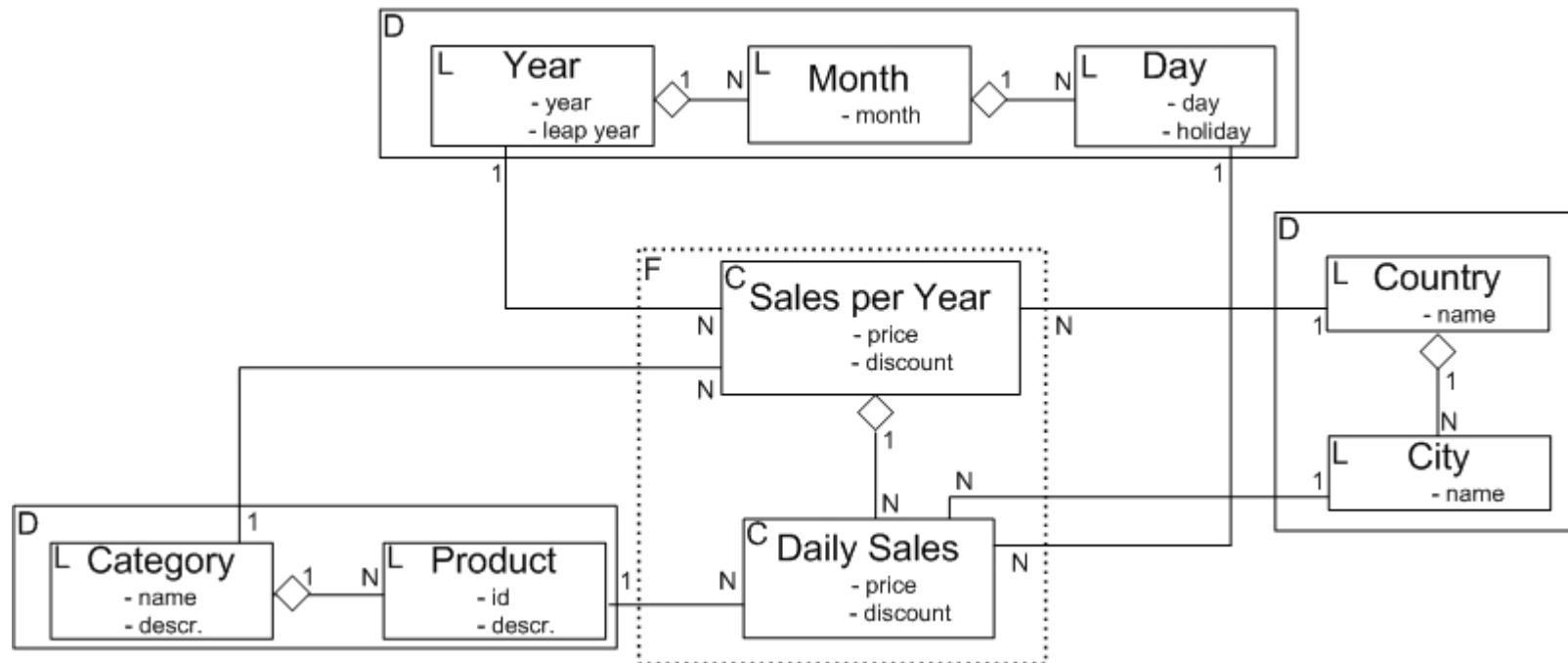






The AMDO Module

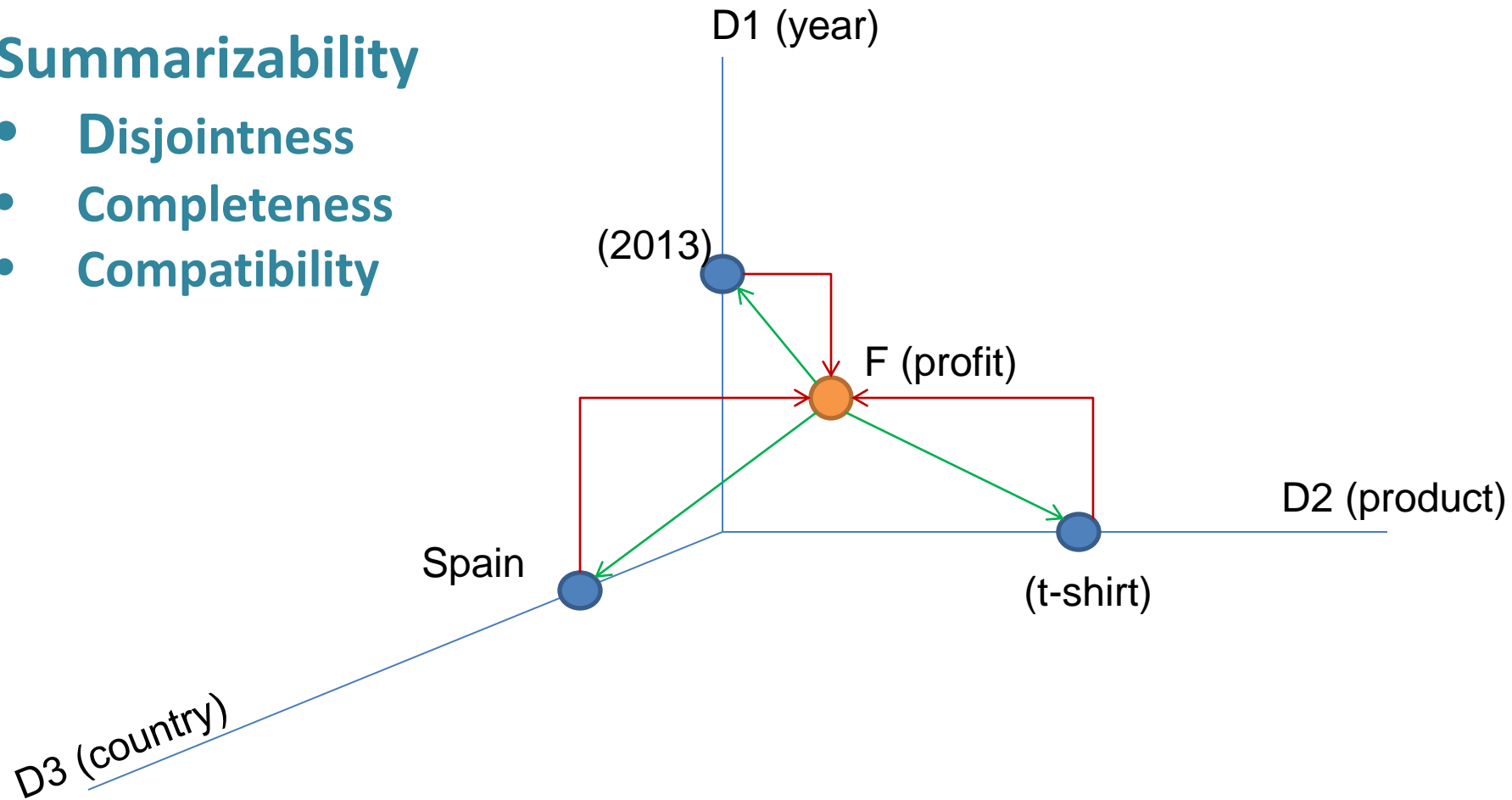


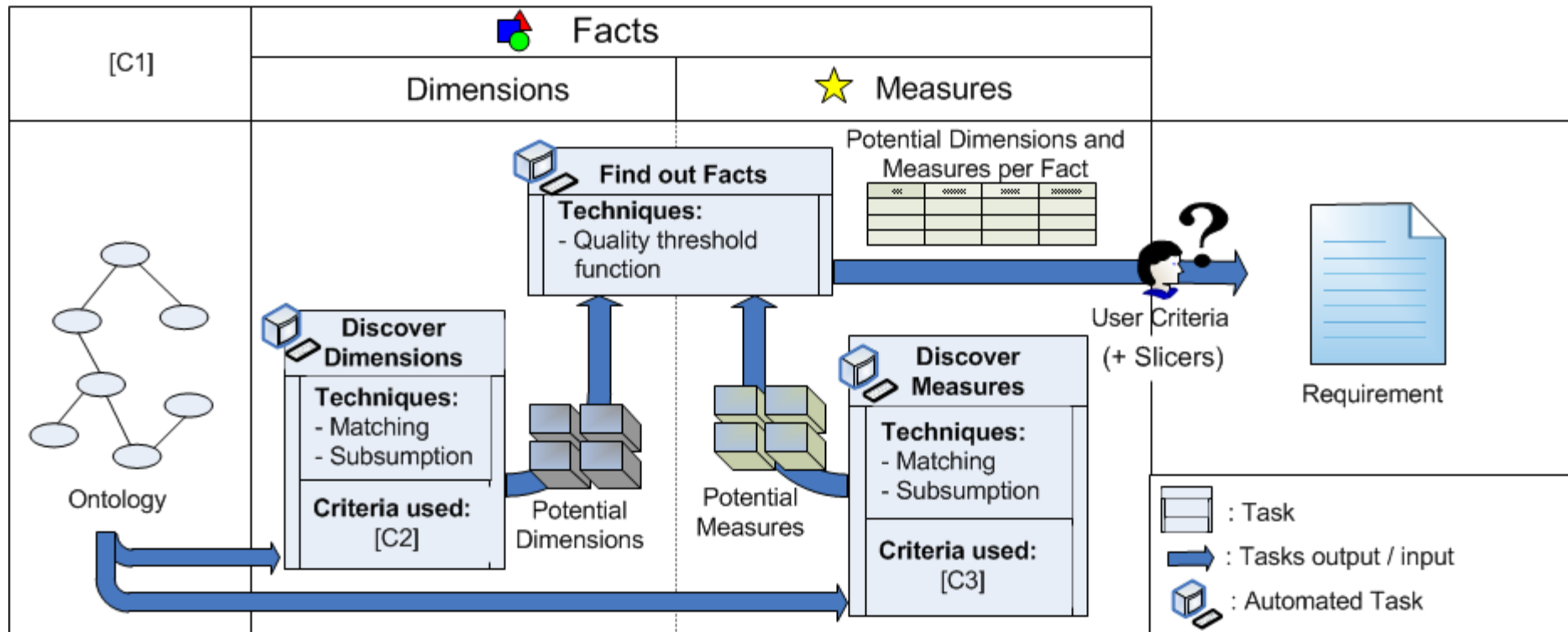


The MD Space

Summarizability

- **Disjointness**
- **Completeness**
- **Compatibility**

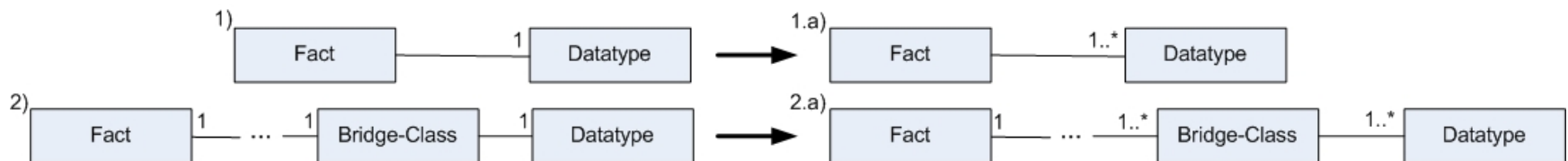




- **IC: Facts are related to concepts by means of to-one relationships**
- **Definition: *A dimensional concept is defined by an ending concept and a path of properties. The path must be considered because it adds relevant semantics.***

$$F \sqsubseteq = 1r.D, \text{ where } r \equiv (r_1 \circ \dots \circ r_n)$$

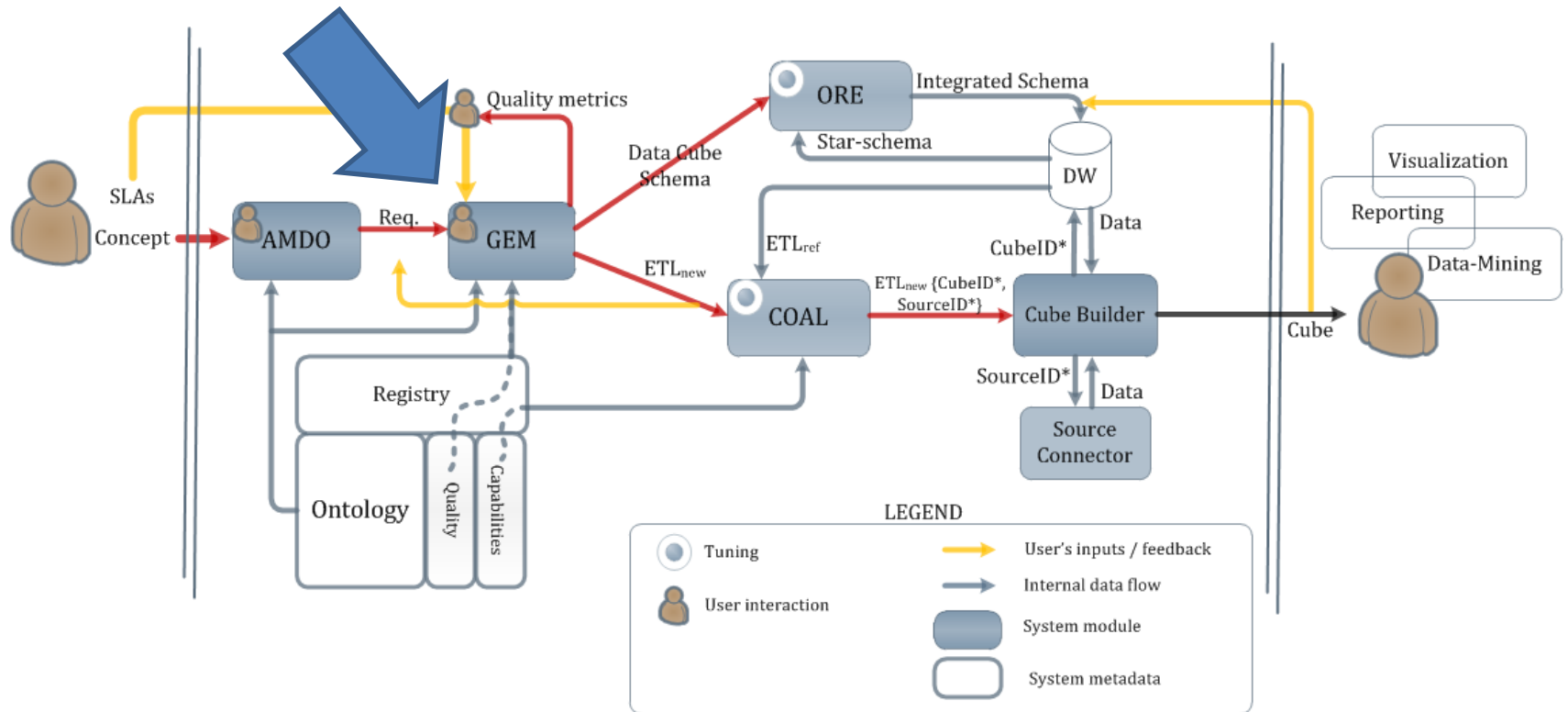
- **IC: Measures are numerical attributes enabling correct data aggregation (datatype)**
- **Definition: A measure is defined by a datatype and a path of properties (i.e., a composite property)**



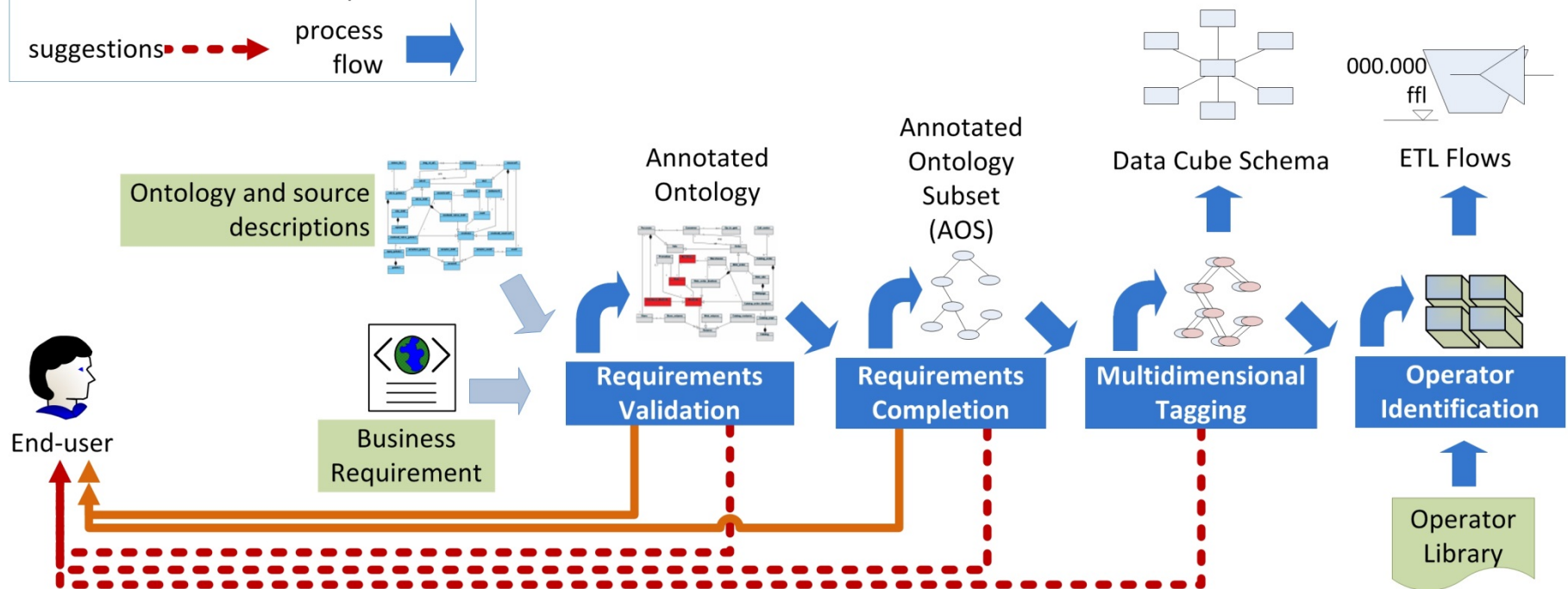
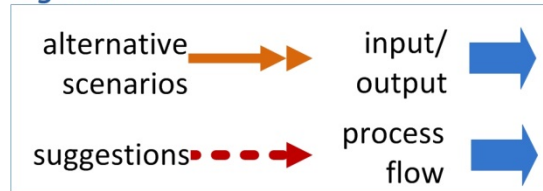
- Fact Estimation
 - $f(\#D, \#M)$
 - Other parameters could be considered

Concept	#Dimensional Concepts	#Potential Measures	FactEstimation
LateReturn	78	5	88
DamageCost	81	3	87
Prepared	81	3	87
AssignedCar	80	3	86
PaidWithPointsRental	74	4	82
ClosedRental	74	4	82
EarlyReturn	74	4	82

The GEM Module

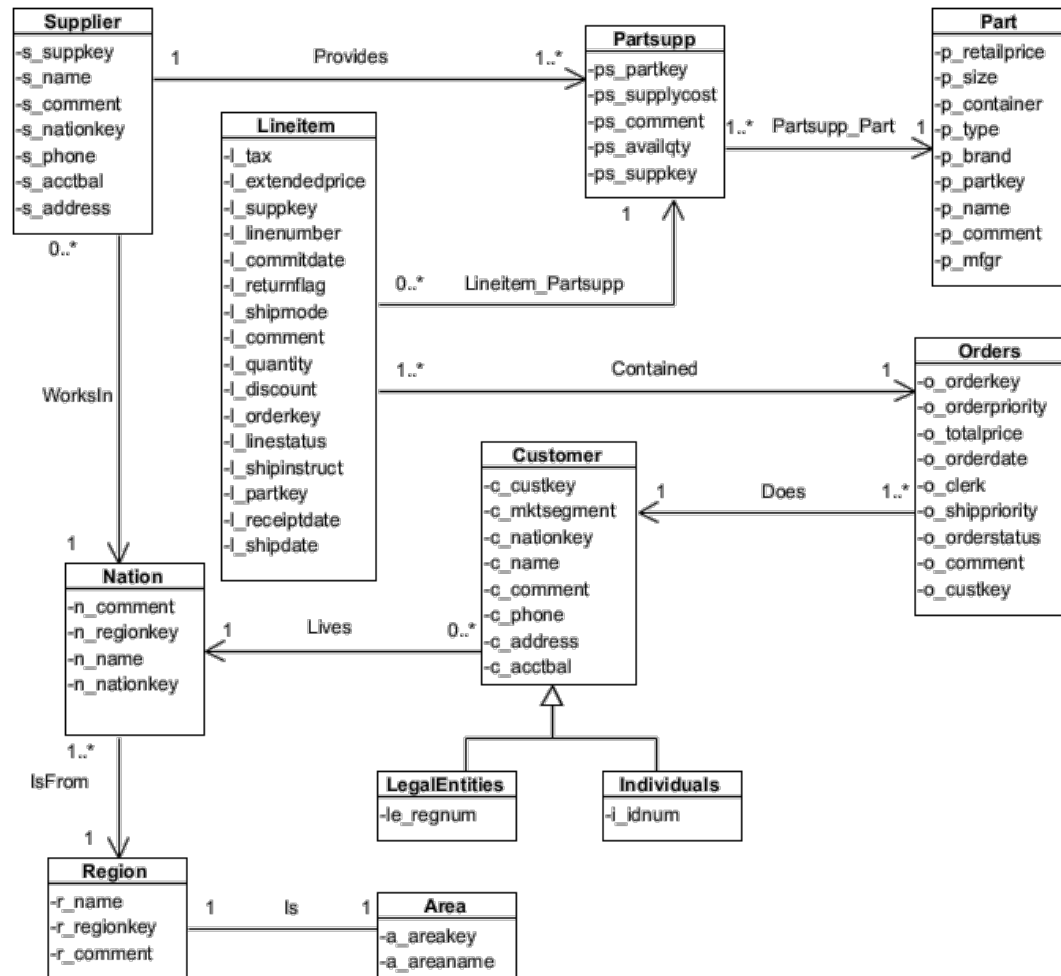


legend:



Requirement: *Revenue* related to **Serbian customers**.

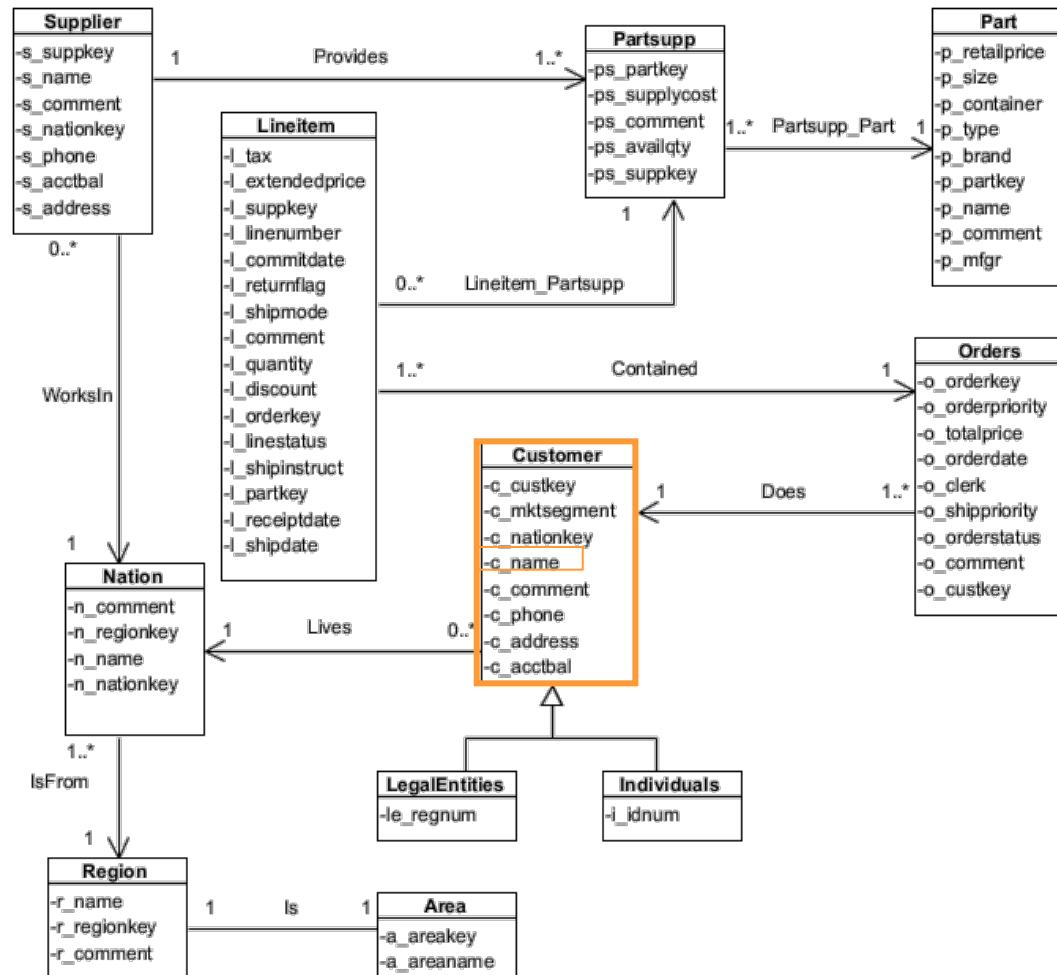
Requirement: *Revenue* related to *Serbian customers*.



Requirement: *Revenue* related to *Serbian customers*.

Concept's tagging...

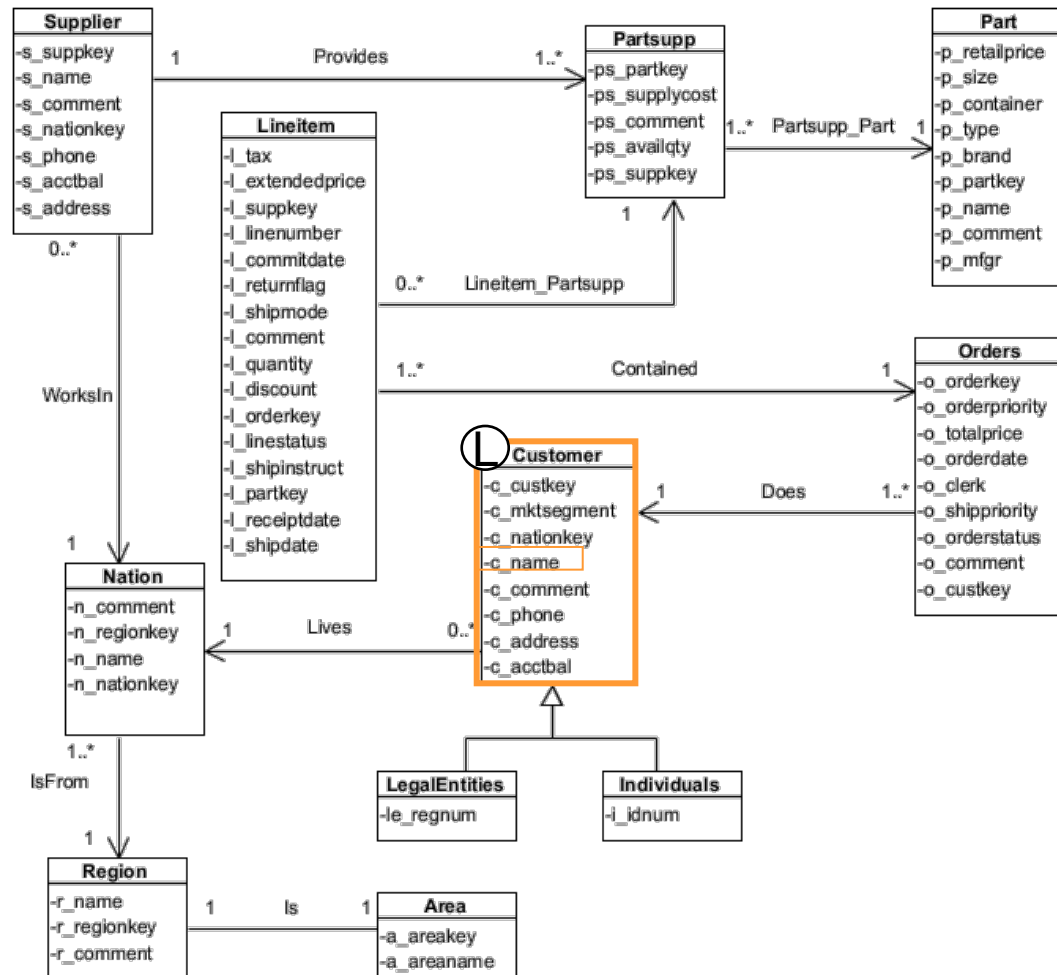
- Identifying concepts in the ontology



Requirement: *Revenue* related to *Serbian customers*.

Concept's tagging...

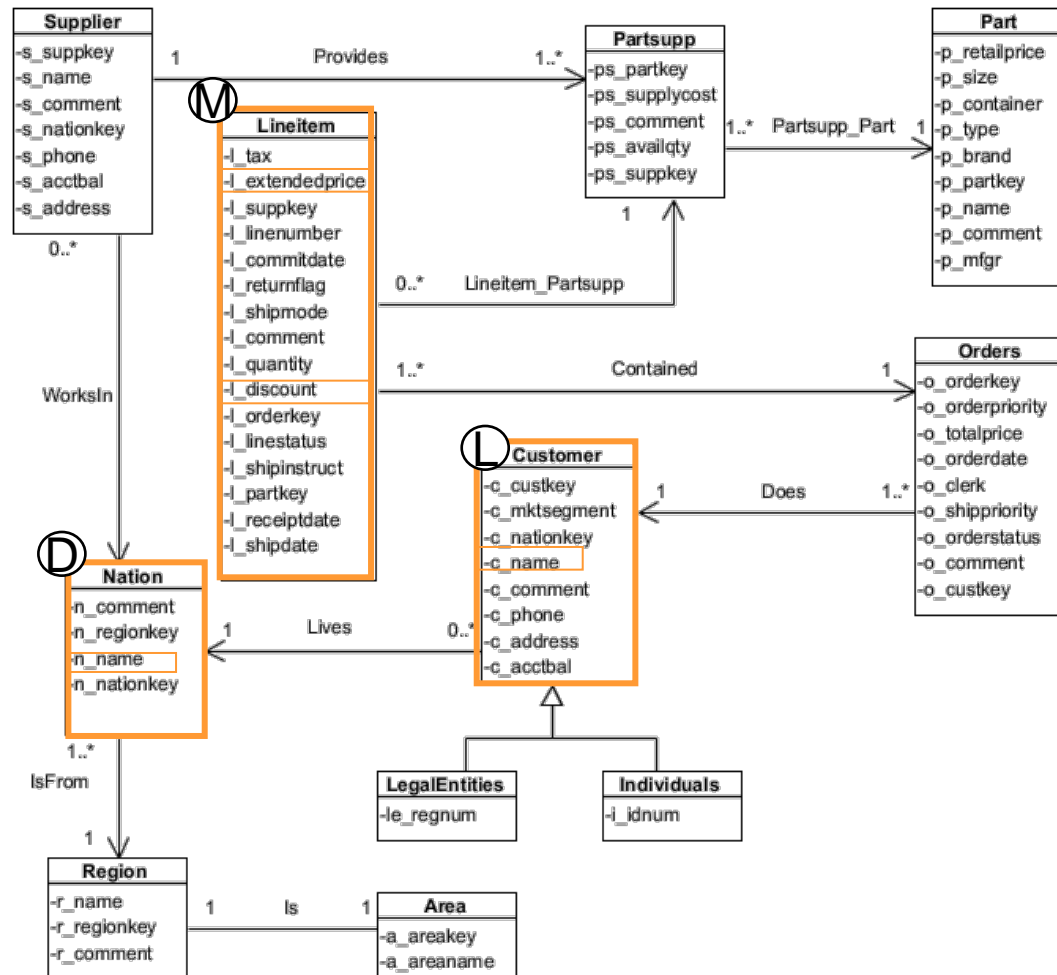
- Identifying concepts in the ontology
- Tagging the concepts with appropriate MD roles



Requirement: *Revenue* related to *Serbian customers*.

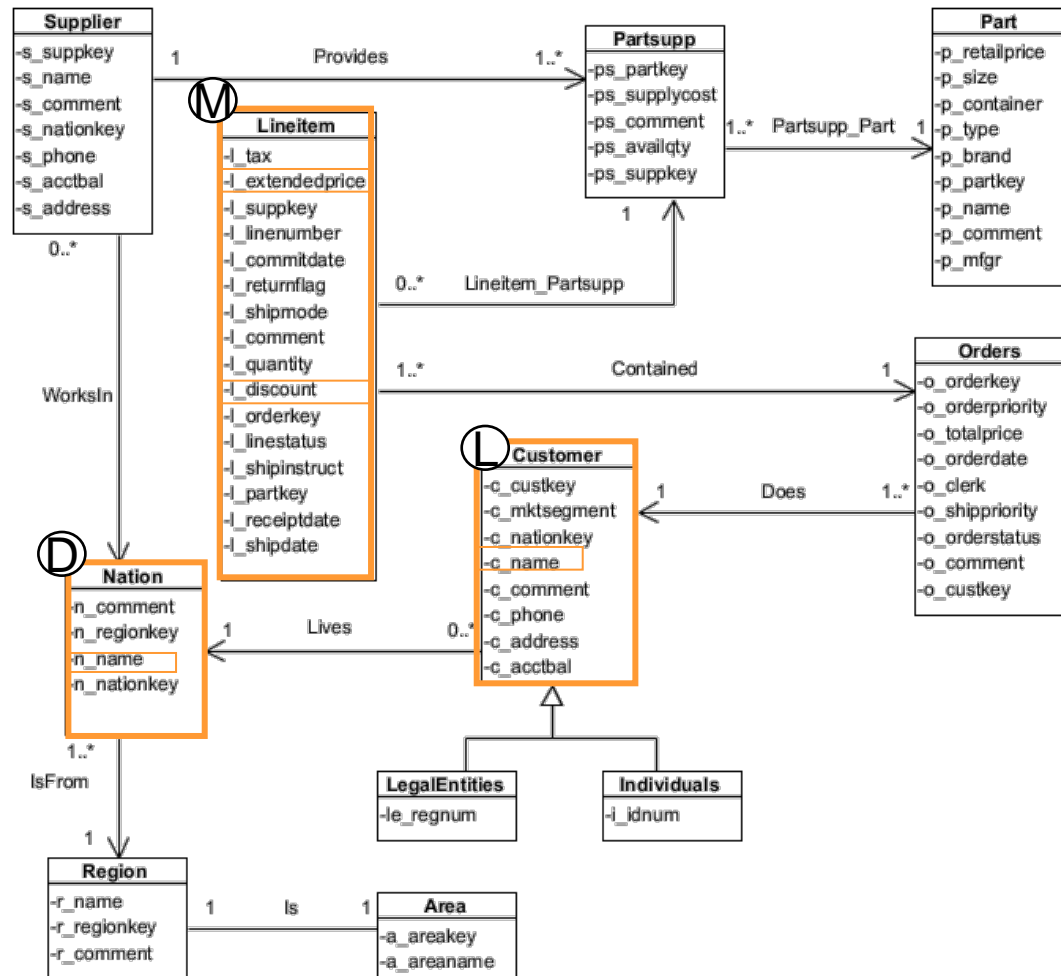
Concept's tagging...

- Identifying concepts in the ontology
- Tagging the concepts with appropriate MD roles



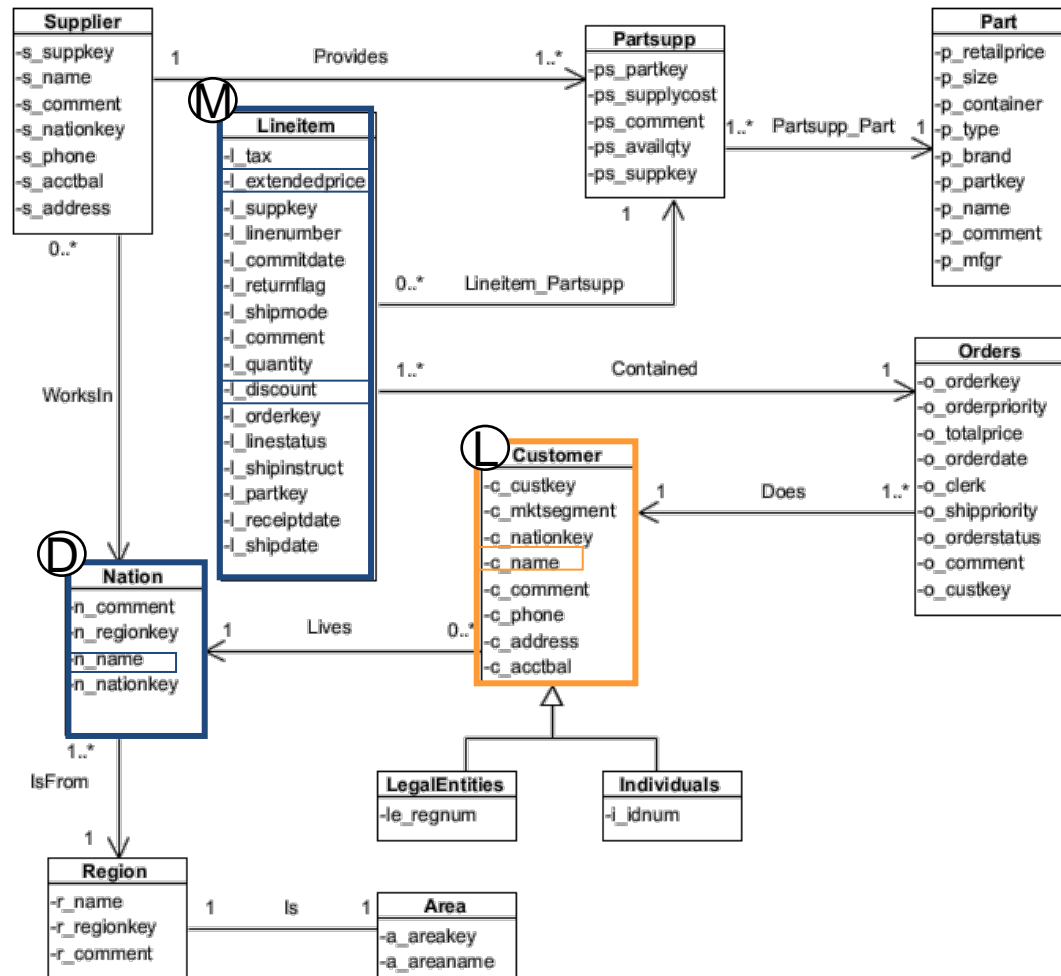
Requirement: *Revenue* related to *Serbian customers*.

Concept available?



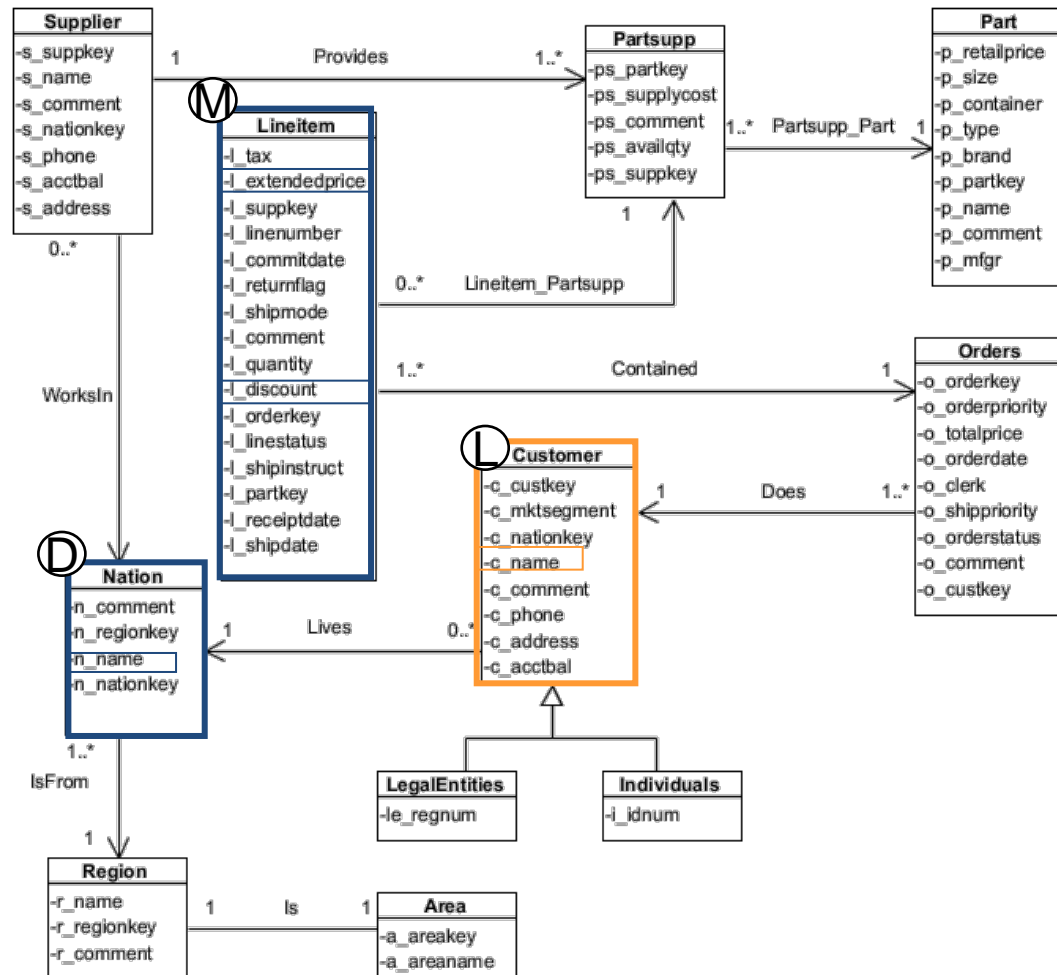
Requirement: *Revenue* related to *Serbian customers*.

Concept available?



Requirement: *Revenue* related to *Serbian customers*.

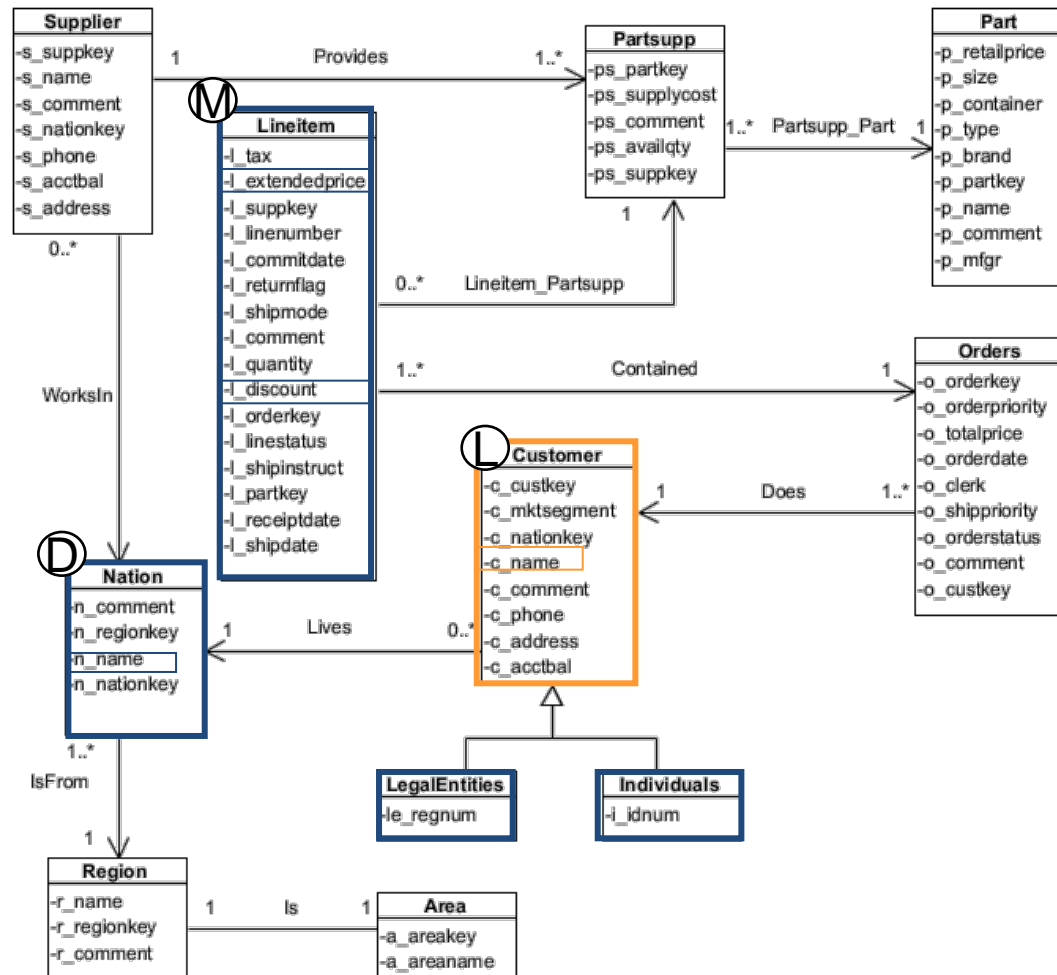
Concept available?
If no source can supply
that concept:



Requirement: *Revenue* related to *Serbian customers*.

Concept available?
If no source can supply
that concept:

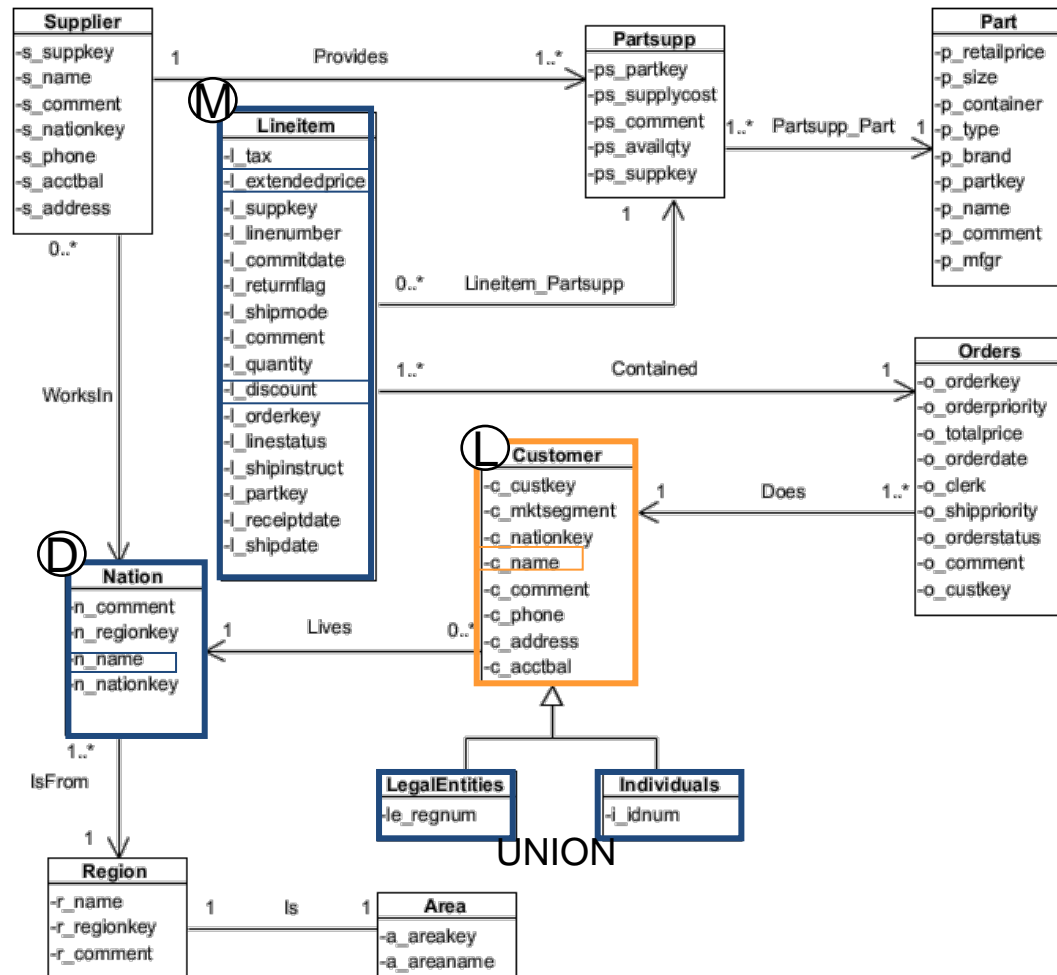
- Search for taxonomies or synonyms
- Operations



Requirement: *Revenue* related to *Serbian customers*.

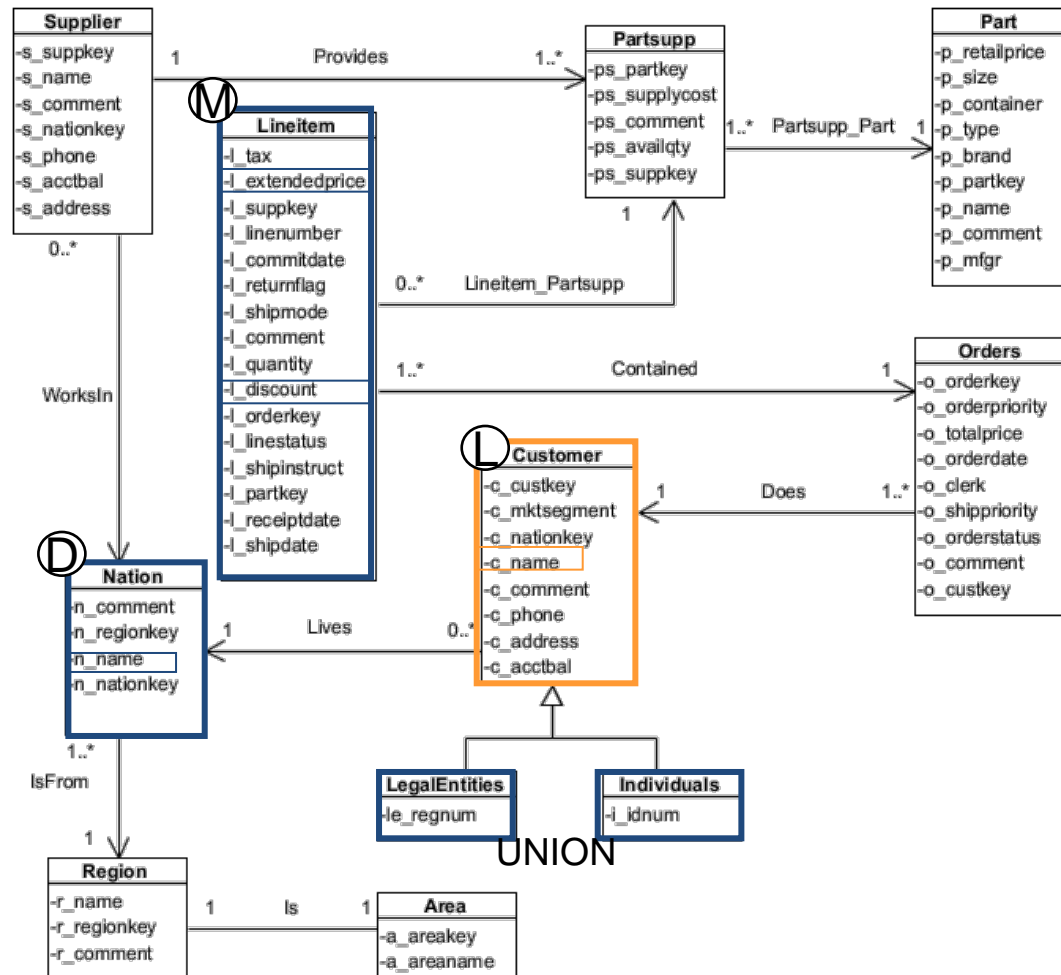
Concept available?
If no source can supply
that concept:

- Search for taxonomies or synonyms
- Operations



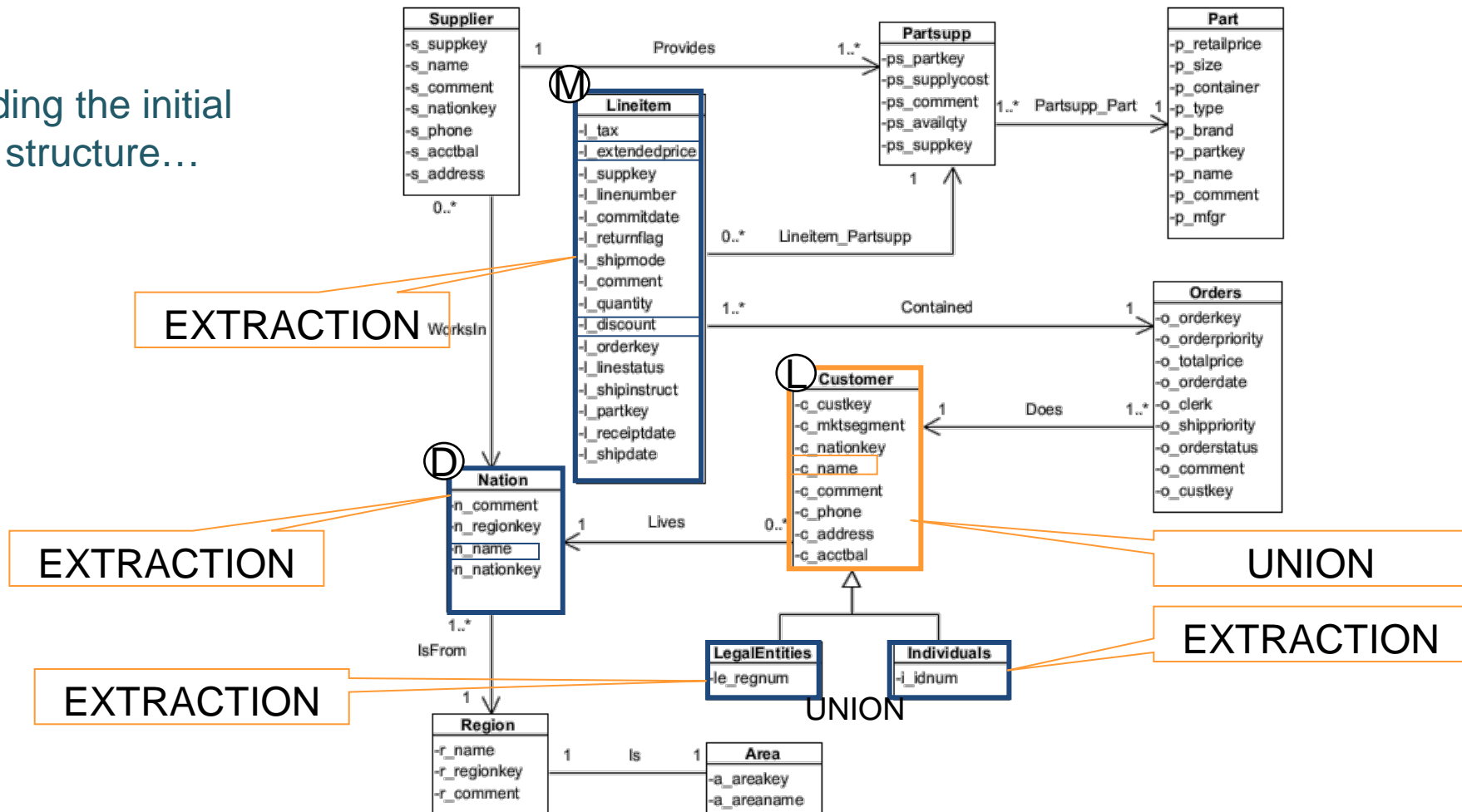
Requirement: *Revenue* related to *Serbian customers*.

Building the initial
ETL structure...

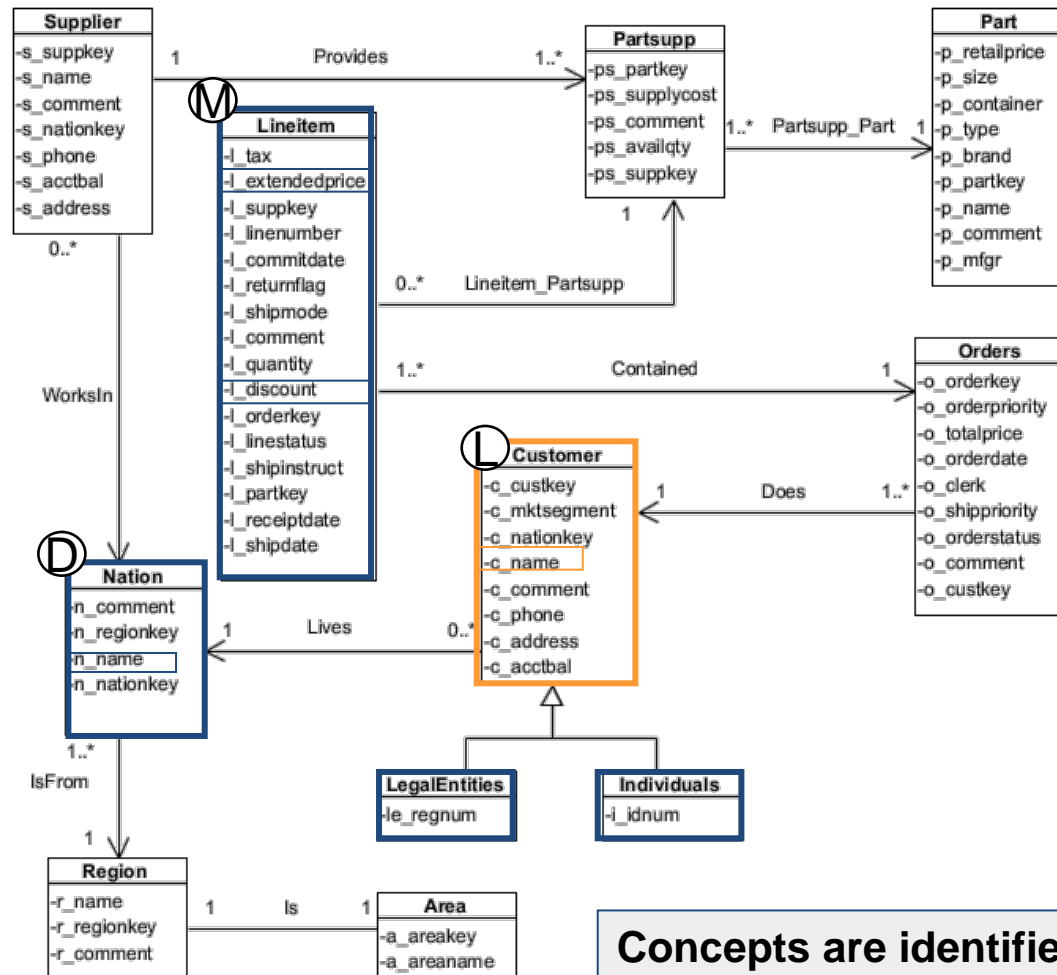


Requirement: *Revenue* related to *Serbian customers*.

Building the initial
ETL structure...



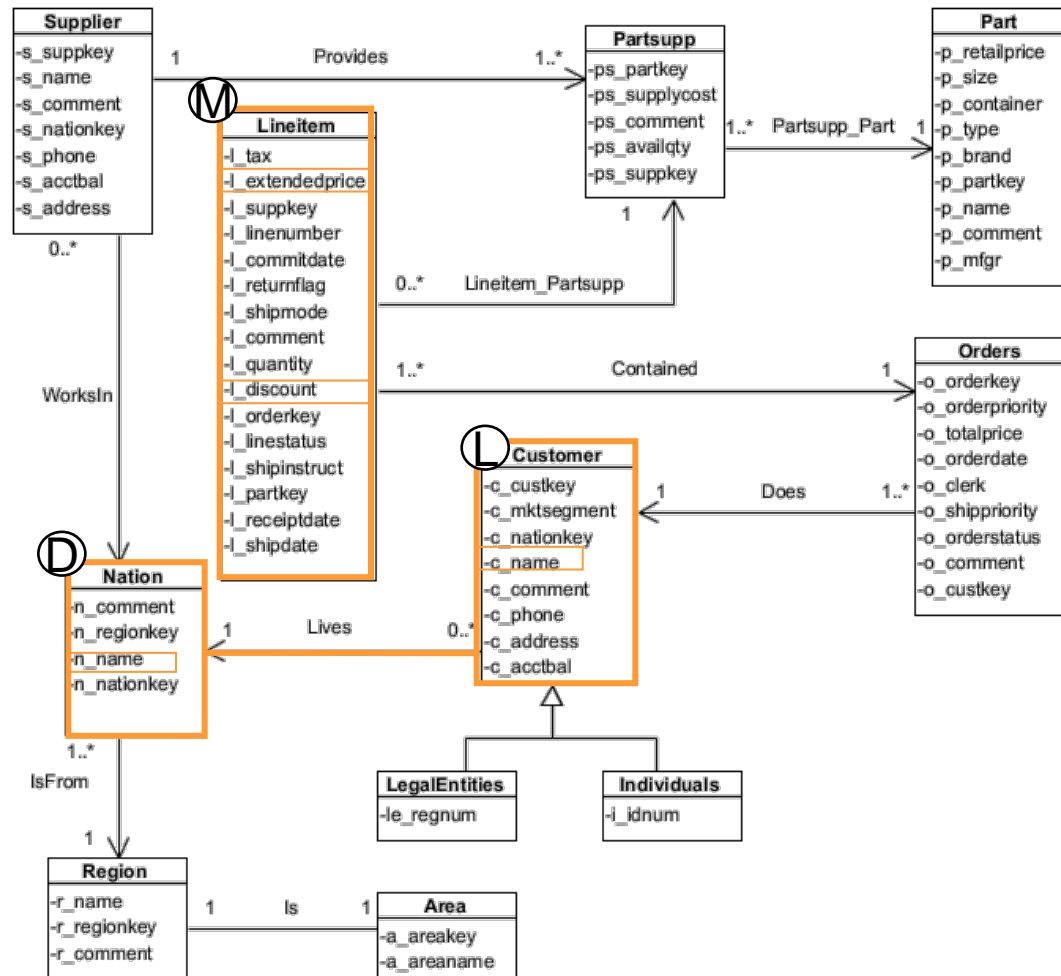
Requirement: *Revenue* related to *Serbian customers*.

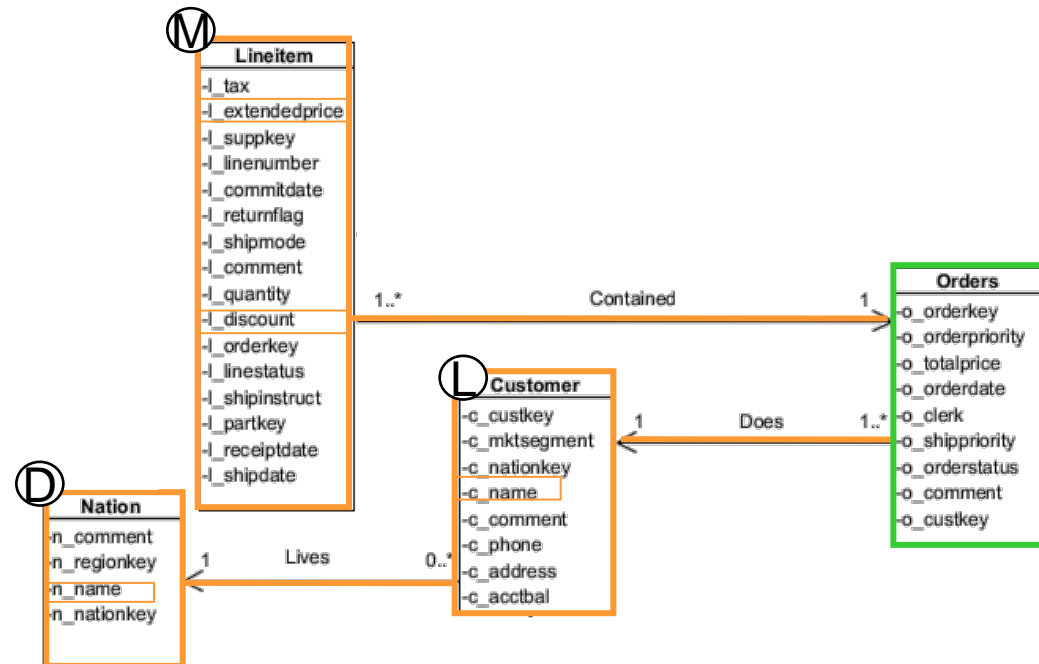


Concepts are identified in the ontology, tagged with a MD role and it is guaranteed at least one source can supply them

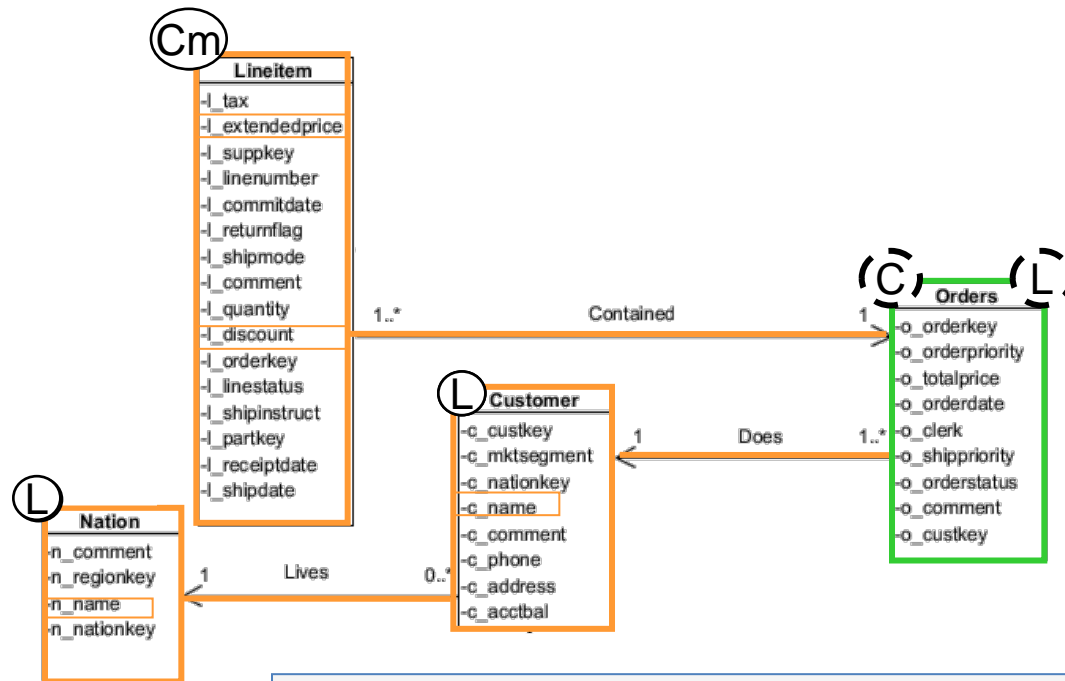
Identifying paths...

- Paths between tagged concepts

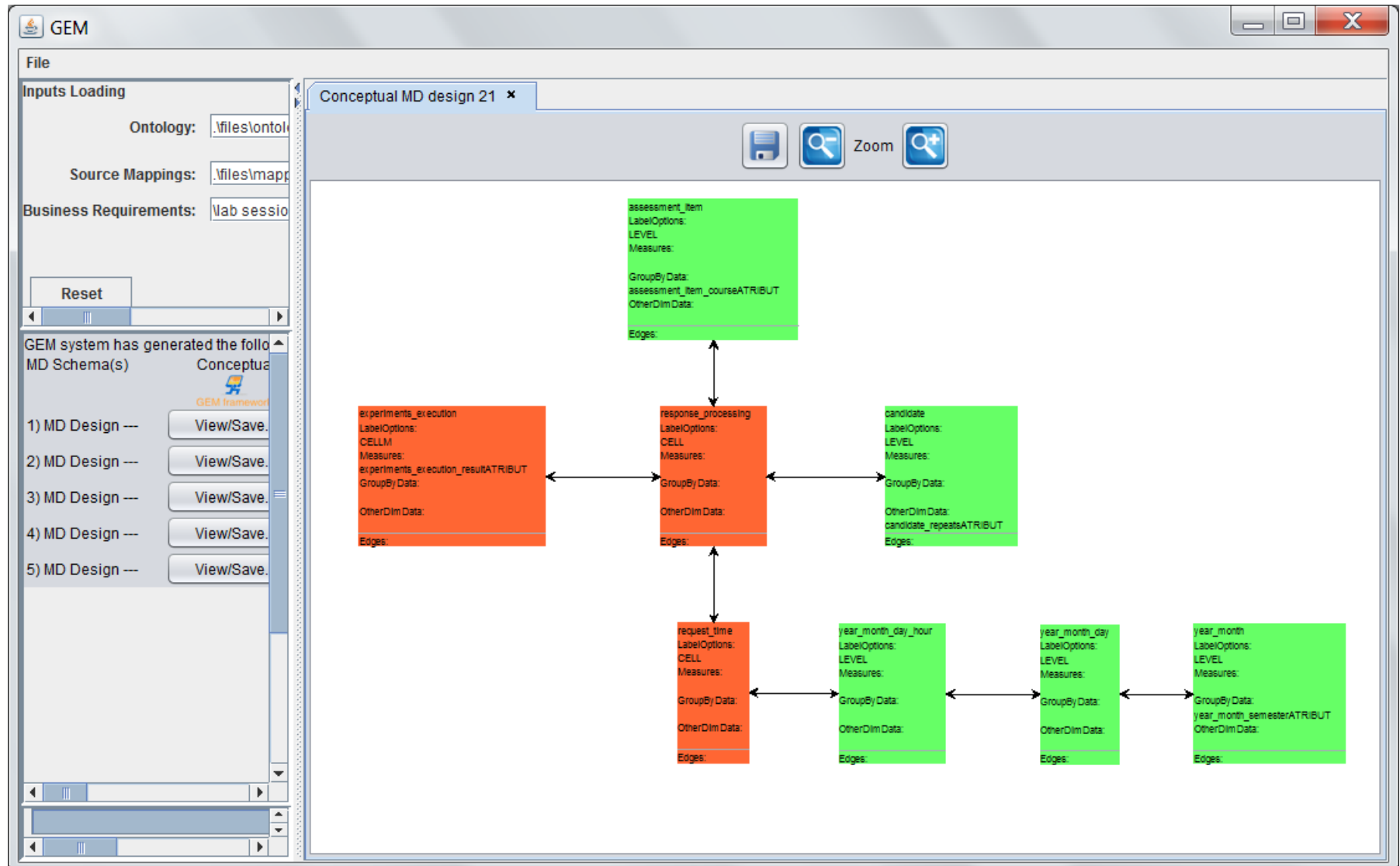


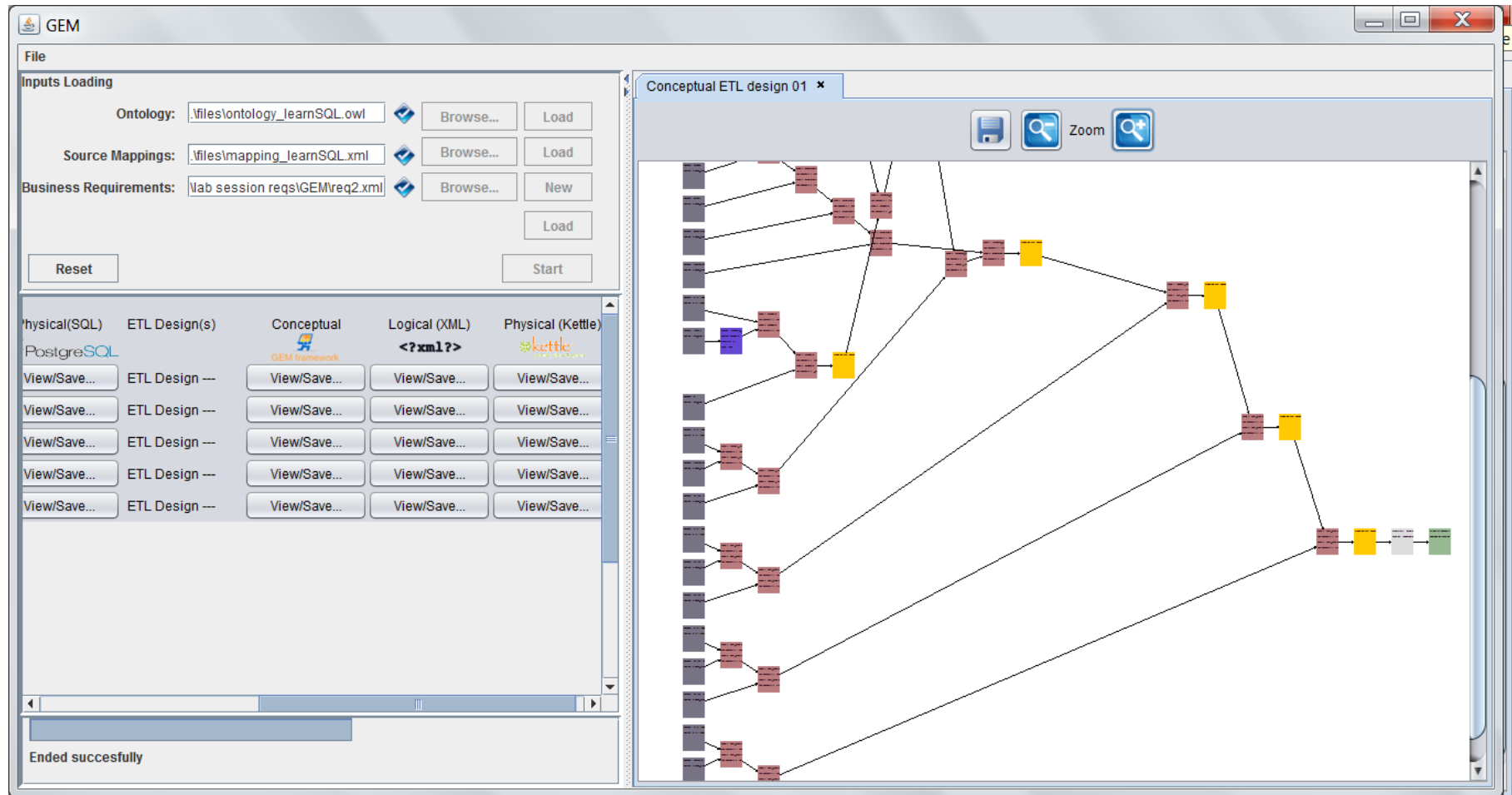


Paths between the required concepts are identified, concepts are partially tagged with MD roles

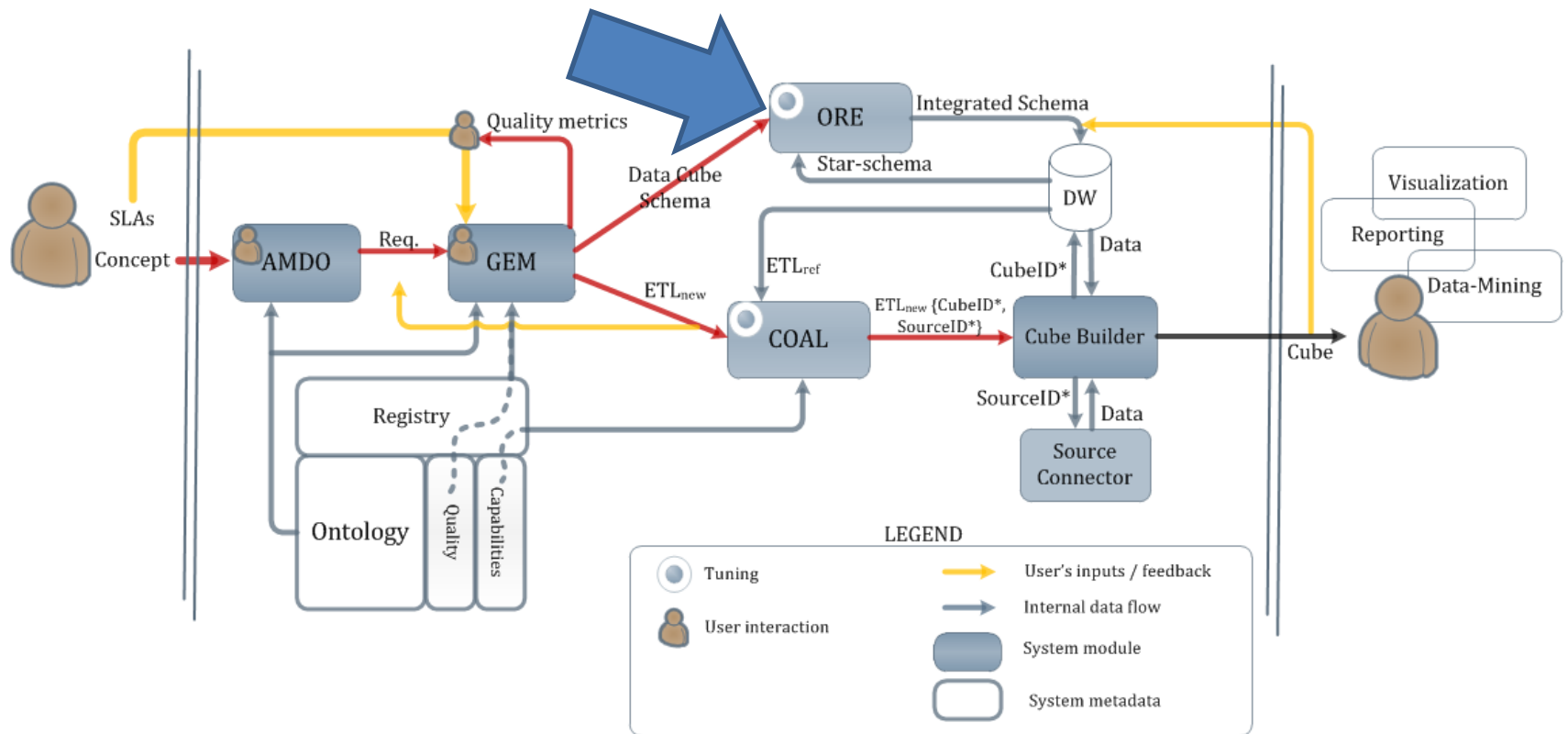


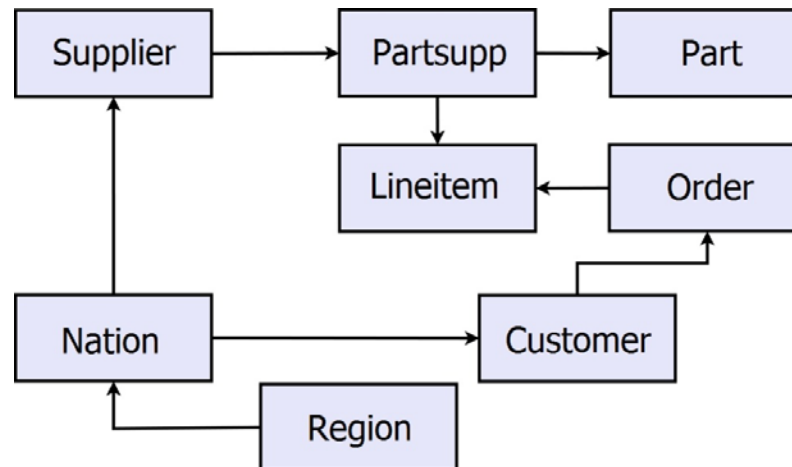
All concepts are appropriately tagged and generated schemas are validated according to MD integrity constraints. Then, the schema transformations are lowered to the instance level by means of ETL operators (Operator Library)





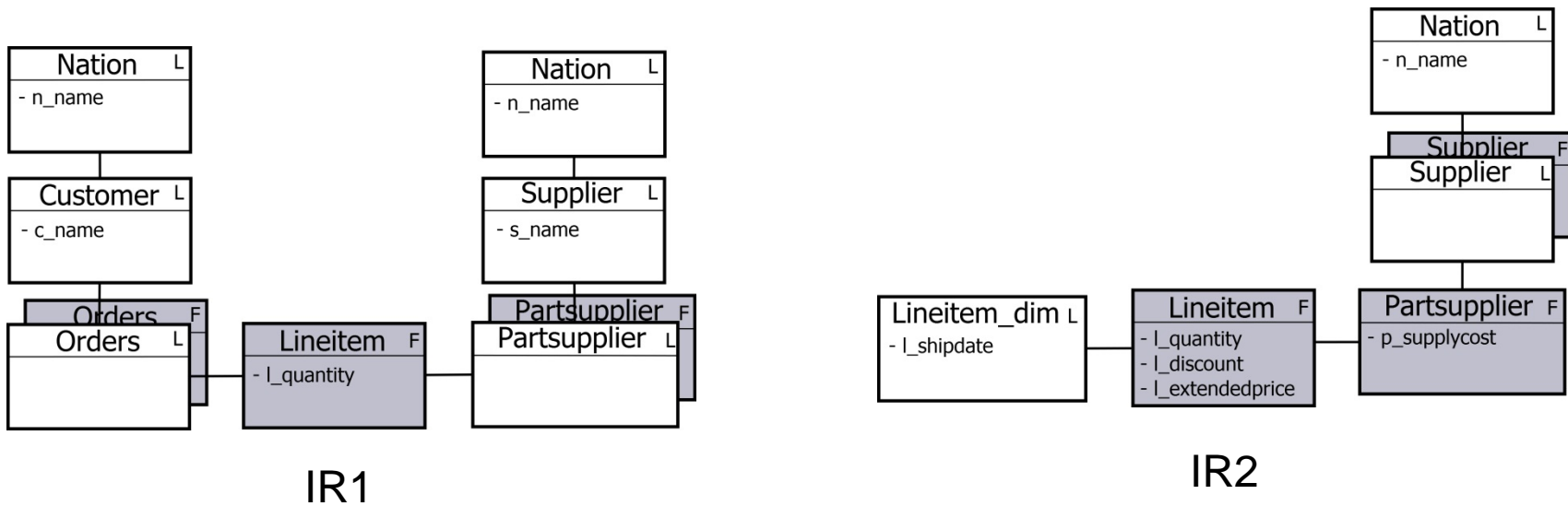
The ORE Module



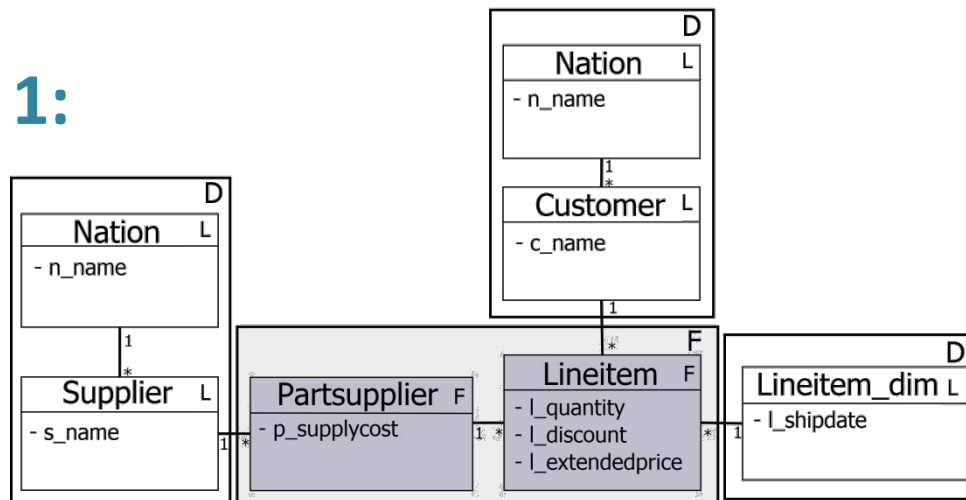


► Example information requirements:

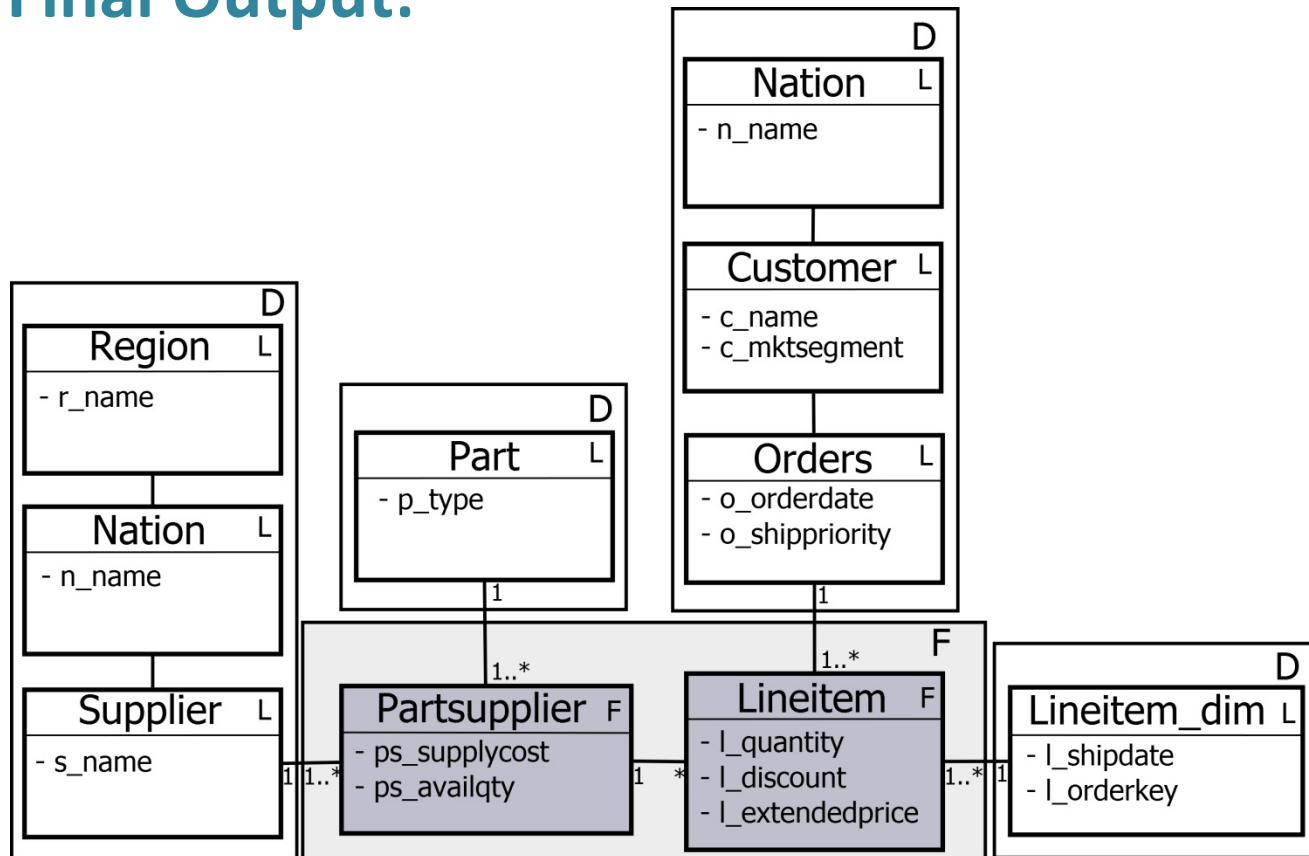
- **IR1:** The total quantity of the parts shipped from Spanish suppliers to French customers.
- **IR2:** For each nation, the profit for all supplied parts, shipped after 01/01/2011.
- **IR3:** The total revenue of the parts supplied from East Europe.
- **IR4:** For German suppliers, the total available stock value of supplied parts.
- **IR5:** Shipping priority and total potential revenue of the parts ordered before certain date and shipped after certain date to a customer of a given segment.



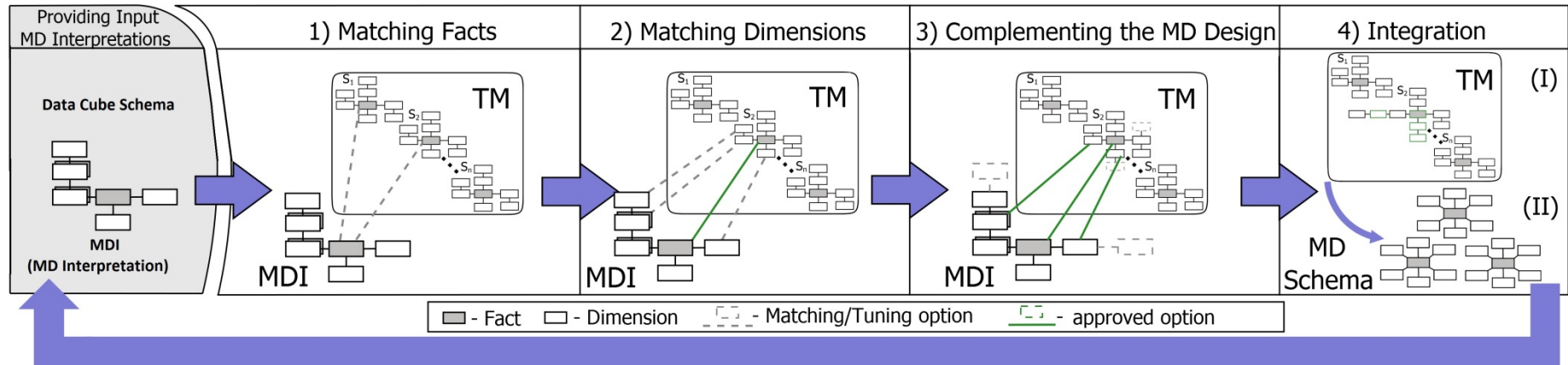
Iteration 1:



Final Output:



MD Schema satisfying IR1- IR5



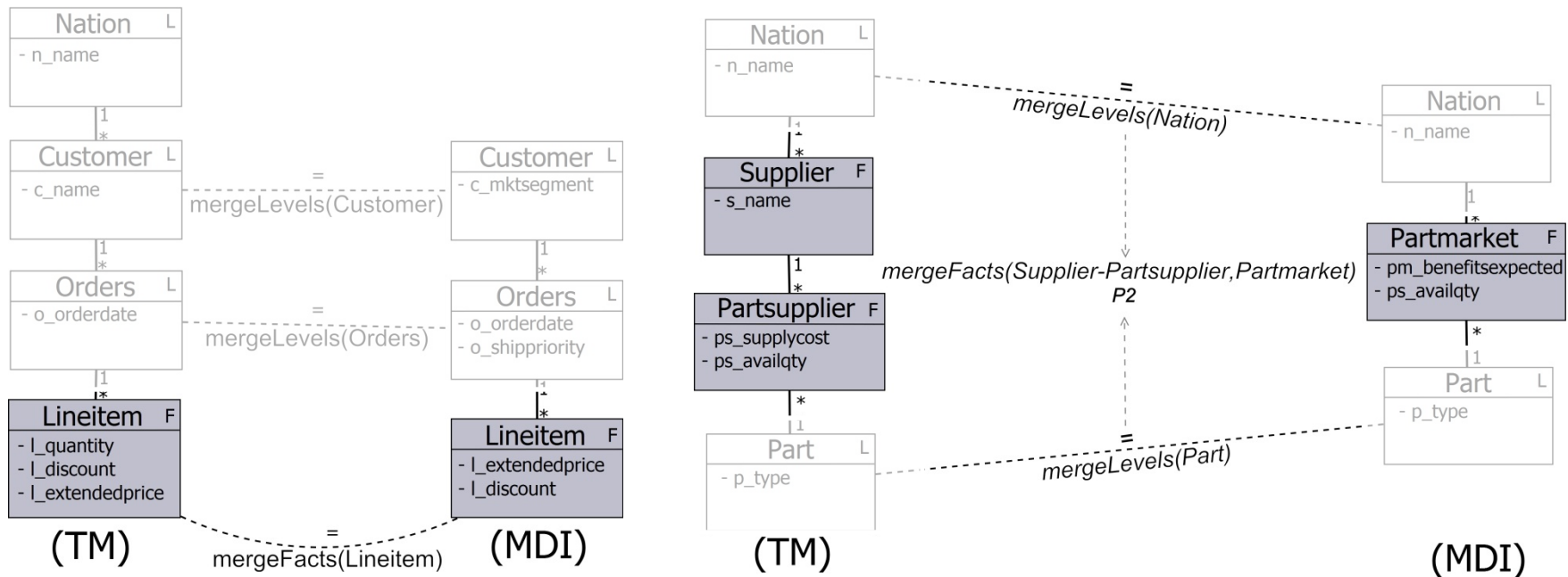
► Inputs

- MD interpretations of requirements (GEM)

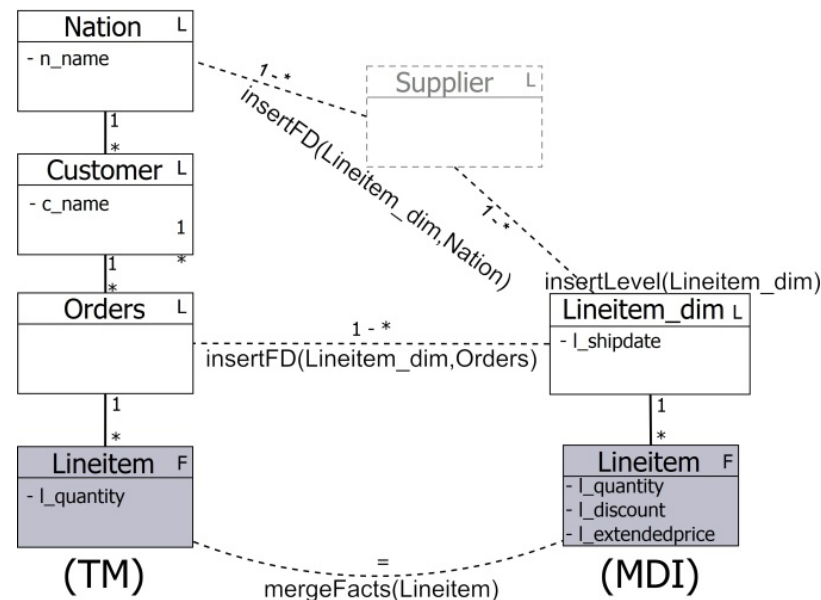
► Stages

1. Matching Facts
2. Matching Dimensions
3. Complementing the MD Design
4. Integration

- Two facts match if they produce an equivalent set of points in the MD space
- Alternative solutions with different costs



- Dimension - partially ordered set of individual levels (DAG)
- We search for possible matchings among the individual levels
- Match levels through the shortest path producing a valid MD relation (=, 1-1, 1-* or *-1) between them
- Alternative solutions with different costs



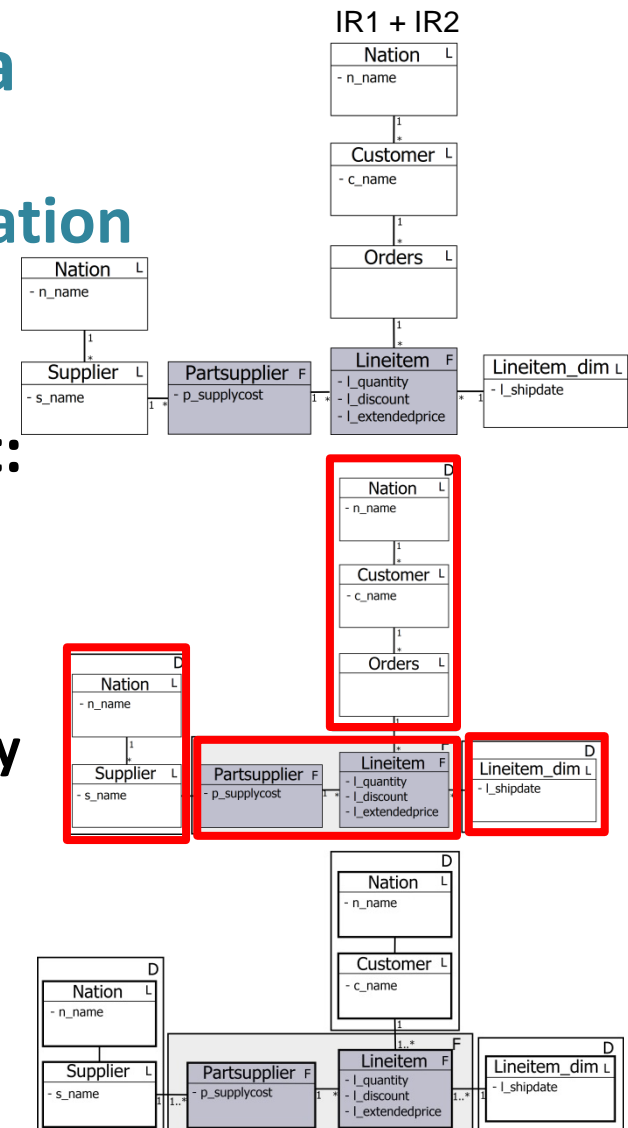
- Producing the final MD schema
 - Relaxing the final schema from currently irrelevant information

• Two phases

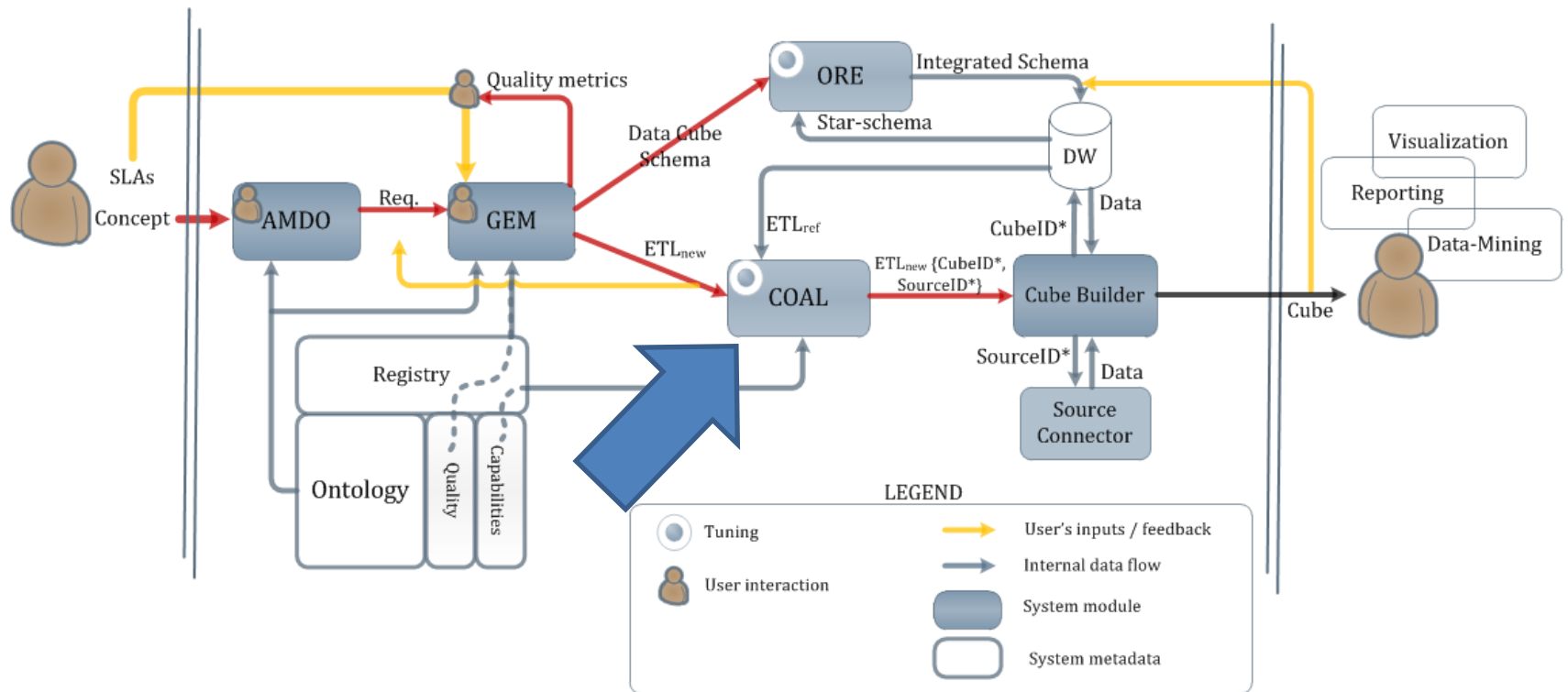
- Partitioning**
grouping different concepts that:
 - Produce a connected subgraph
 - Have the same MD interpretation

- Folding**

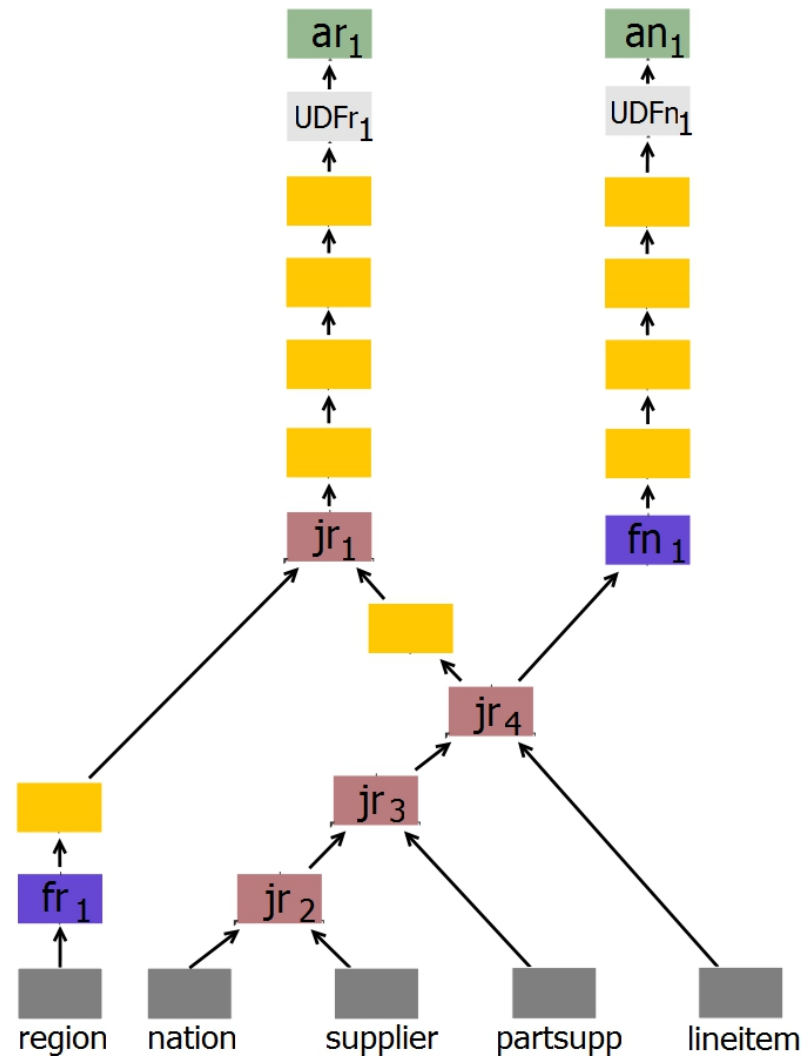
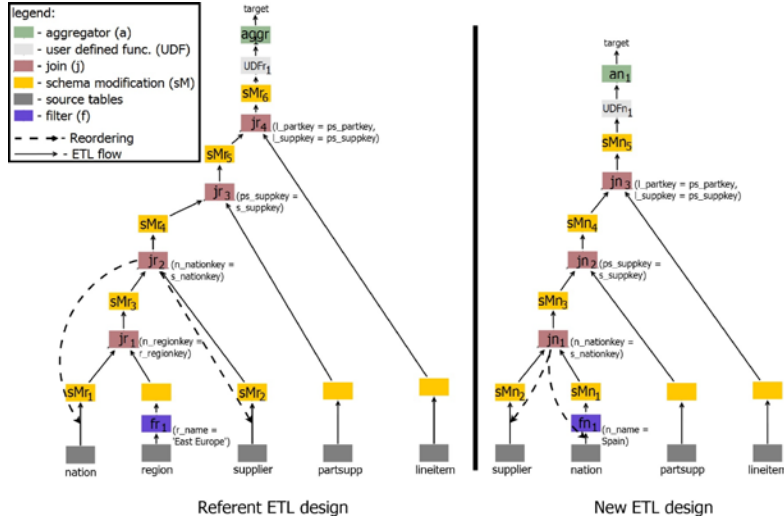
- Consider only the concepts currently required by the user
- All the knowledge still preserved in TM for future integration steps

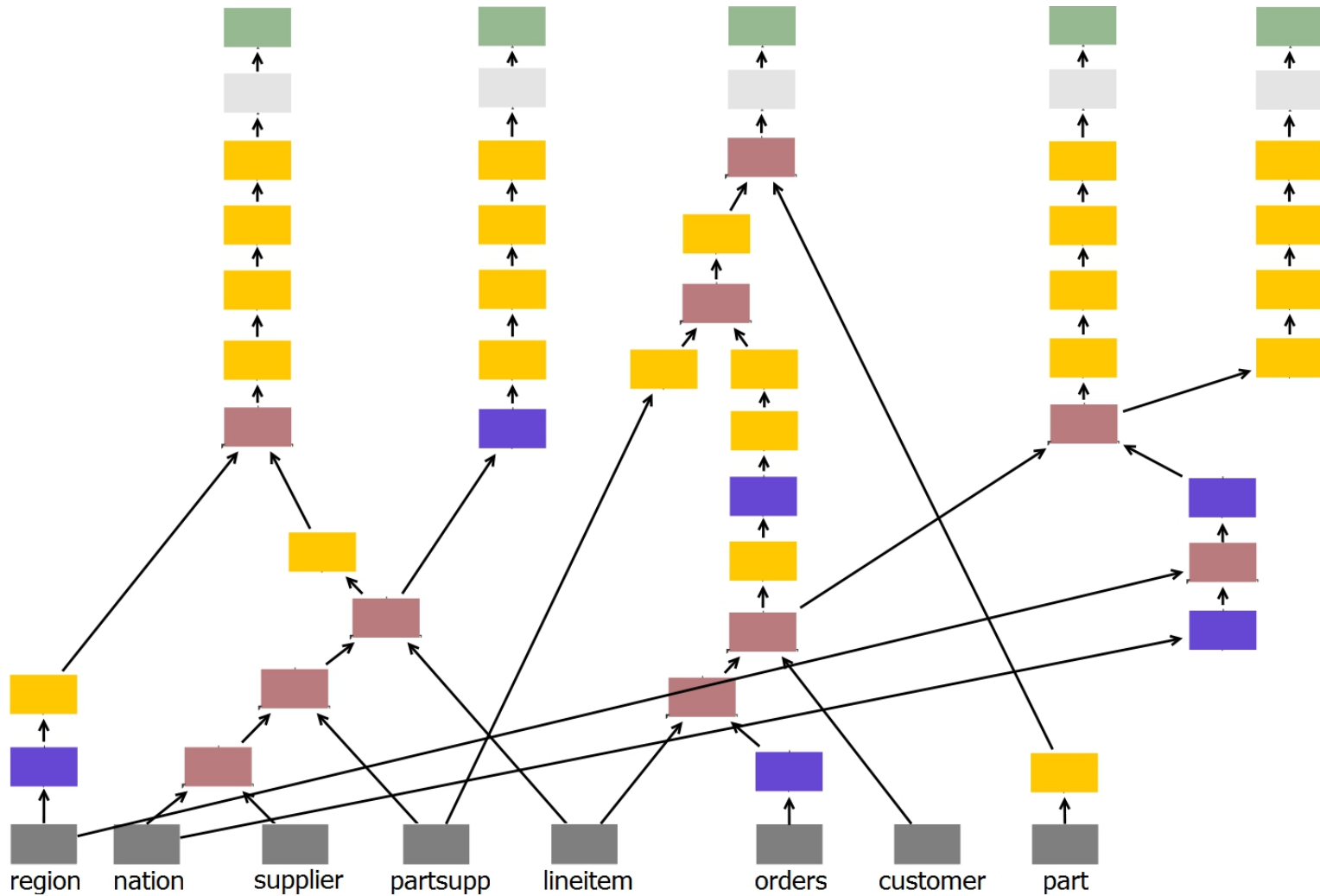


The COAL Module

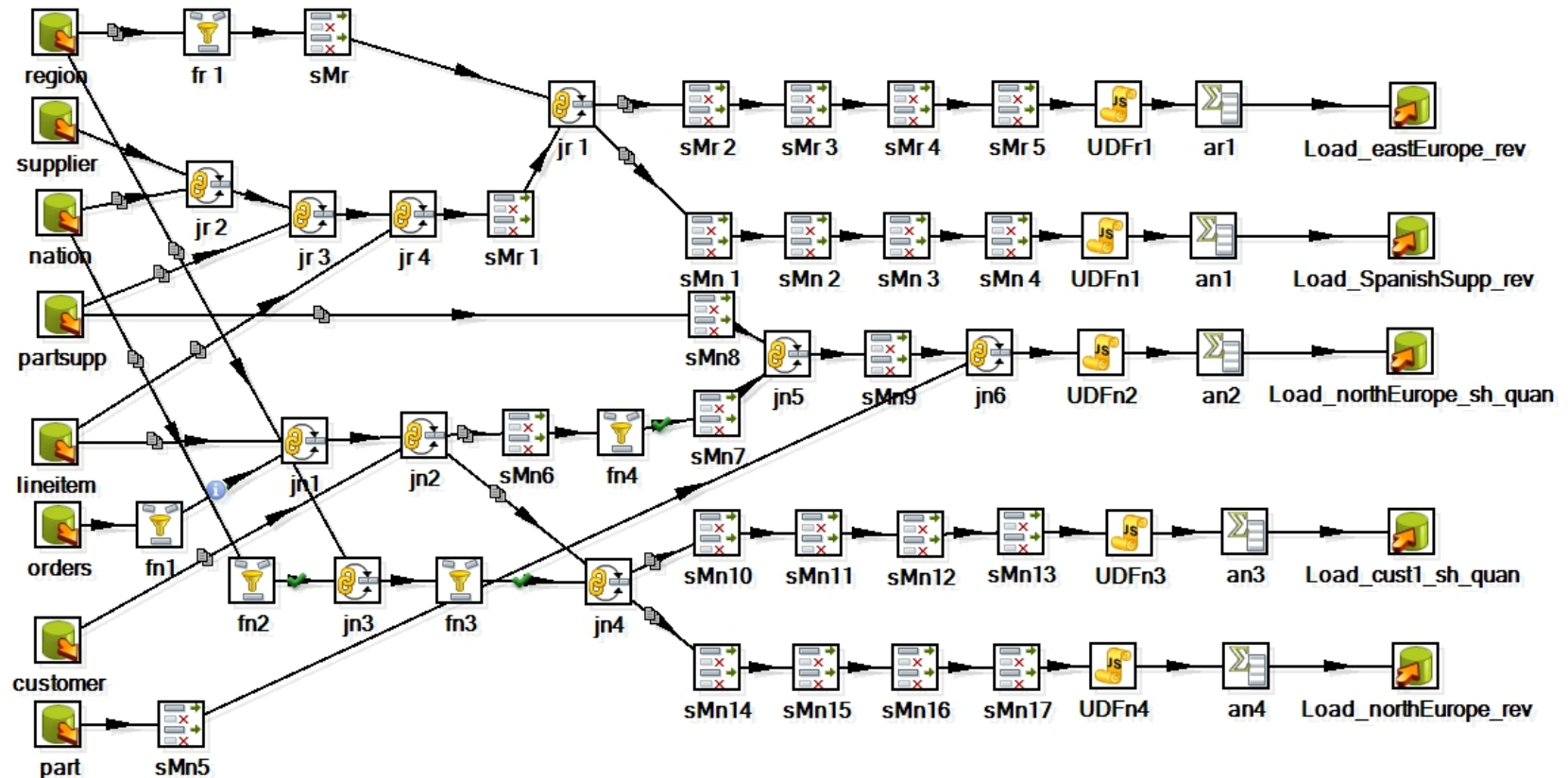


COAL: TPC-H Example





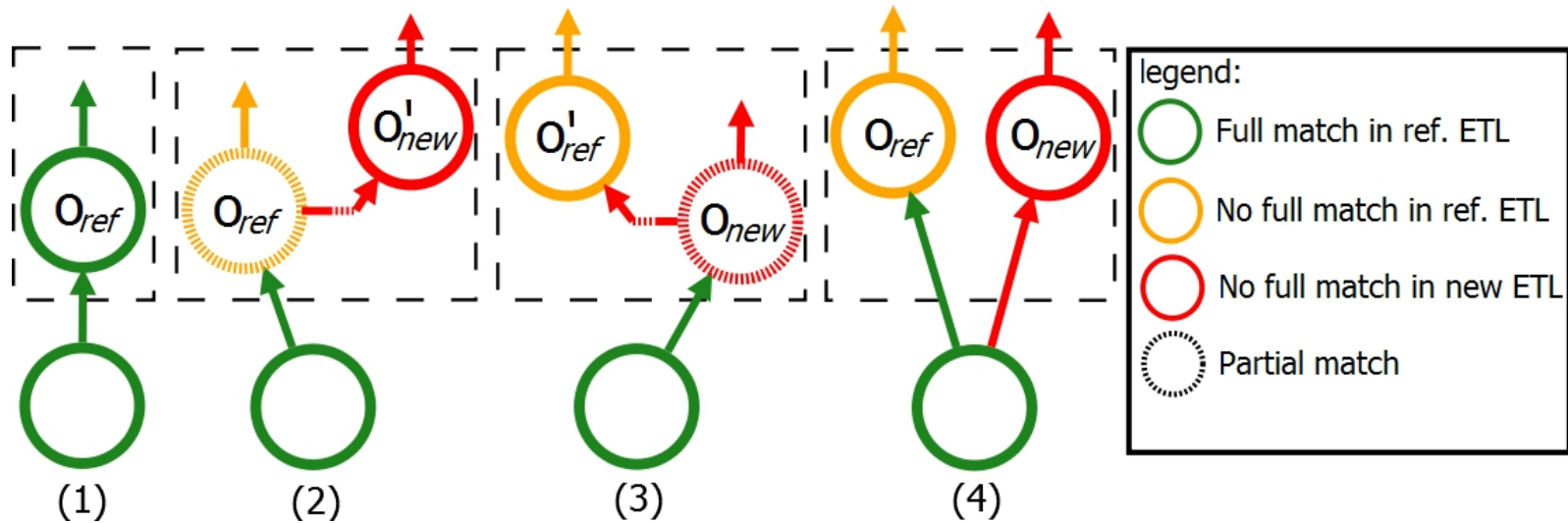
COAL: TPC-H Example



- **WHAT?**
 - Incremental integration of ETLs
 - Consolidation Algorithm
 - Apply equivalence rules
 - Maximize matching area
 - Consider partial matching
 - Consider execution costs
- **HOW?**
 - Consider the ETL as DAG
 1. Match the inputs
 2. Match paths from the same input(s)
 - Consider a pair of operations can match
 - a. Fully
 - b. Partially
 - Two operations may match if their inputs coincide

- **Unary Operations:**
 - Swap
 - Distribute
 - Factorize
- **Binary Operations:**
 - Associate
 - Distribute

oper.	f	sM	j	\cup	a	UDF	SK
f	\sqrt{swp}	\sim_{swp}	$\sqrt{dst/fct}$	$\sqrt{dst/fct}$	\sim_{swp}	\sqrt{swp}	\sqrt{swp}
sM	\sqrt{swp}	\times	$\sqrt{dst/fct}$	$\sqrt{dst/fct}$	\sim_{swp}	\sqrt{swp}	\sqrt{swp}
j	$\sqrt{dst/fct}$	$\sim_{dst/fct}$	\sqrt{asc}	\times	$\sim_{dst/fct}$	$\sim_{dst/fct}$	$\sim_{dst/fct}$
\cup	$\sqrt{dst/fct}$	$\sqrt{dst/fct}$	\sqrt{distr}	\sqrt{asc}	\times	$\sqrt{dst/fct}$	$\sim_{dst/fct}$
a	\sim_{swp}	\sim_{swp}	$\sim_{dst/fct}$	\times	\times	\sim_{swp}	\sim_{swp}
UDF	\sim_{swp}	\sim_{swp}	$\sim_{dst/fct}$	$\sqrt{dst/fct}$	\sim_{swp}	\sim_{swp}	\sim_{swp}
SK	\sim_{swp}	\sim_{swp}	$\sim_{dst/fct}$	$\sqrt{dst/fct}$	\sim_{swp}	\sim_{swp}	\times



• Filter

- $\mathbf{f}_{\psi_1}(R) = \mathbf{f}_{\psi_2}(S) \text{ iff } R=S \wedge \psi_1 \equiv \psi_2;$
- $\mathbf{f}_{\psi_1}(R) \prec \mathbf{f}_{\psi_2}(S) \text{ iff } R=S \wedge \psi_1 \Leftarrow \psi_2;$

• Schema modification

- $\mathbf{sM}_{a_1,\dots,a_n}(R) = \mathbf{sM}_{b_1,\dots,b_m}(S) \text{ iff } R=S \wedge \{a_1,\dots,a_n\} = \{b_1,\dots,b_m\};$
- $\mathbf{sM}_{a_1,\dots,a_n}(R) \prec \mathbf{sM}_{b_1,\dots,b_m}(S) \text{ iff } R=S \wedge \{a_1,\dots,a_n\} \supset \{b_1,\dots,b_m\};$

• Join

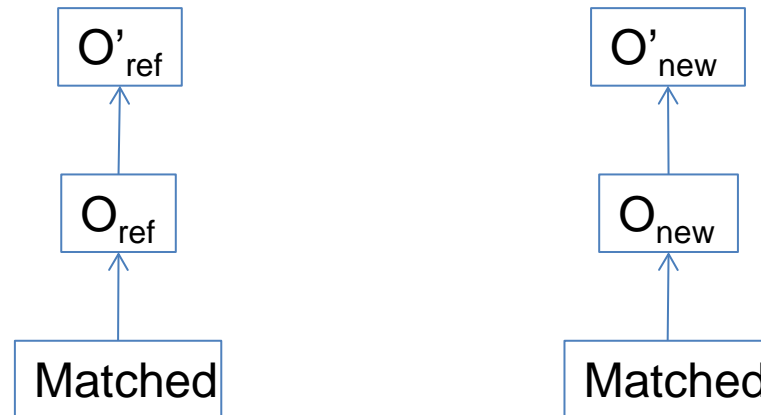
- $P \mathbf{j}_{\psi_1} Q = R \mathbf{j}_{\psi_2} S \text{ iff } ((P=R \wedge Q=S) \vee (P=S \wedge Q=R)) \wedge \psi_1 \equiv \psi_2;$
- $P \mathbf{j}_{\psi_1} Q \prec R \mathbf{j}_{\psi_2} S \text{ iff } ((P=R \wedge Q=S) \vee (P=S \wedge Q=R)) \wedge \psi_1 \Leftarrow \psi_2;$

• Union

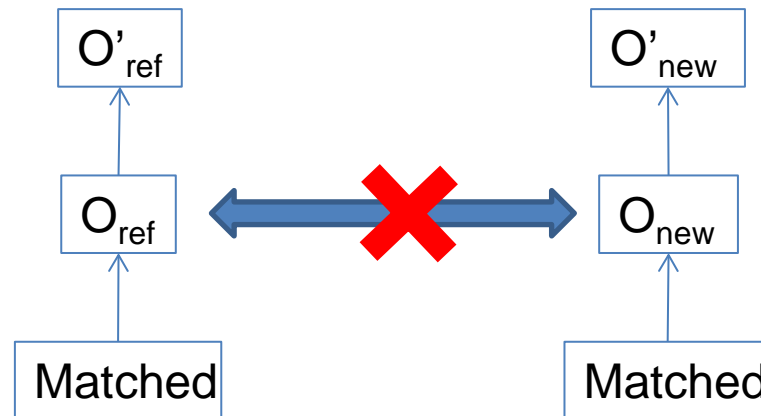
- $P \cup Q = R \cup S \text{ iff } ((P=R \wedge Q=S) \vee (P=S \wedge Q=R));$

1. Match leaves-sources → queue
2. While queue is not empty
 - I. Try to match next (topological) operation
 - II. Else for each pair of operations
 - i. Reorder both graphs
 - ii. If full match then queue
 - iii. Else if partial match then add to output list
 - III. If no match then add to output list
3. Choose the lowest cost in output list

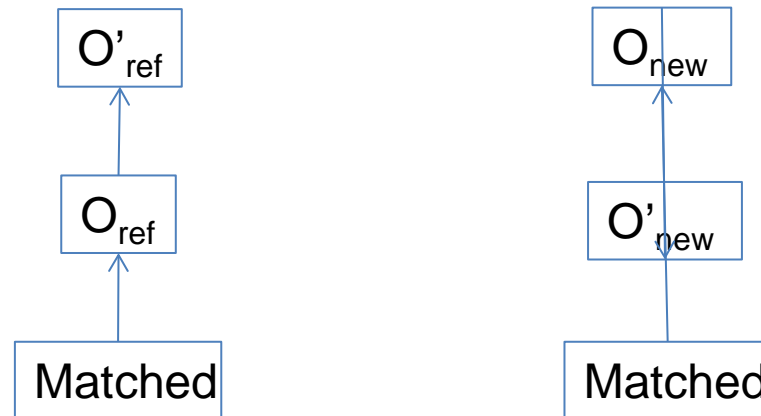
- **Invariants**
 - One pair of operations matches
 - A new match is only considered if inputs match
- **Rule**
 - Compare operations before trying to move



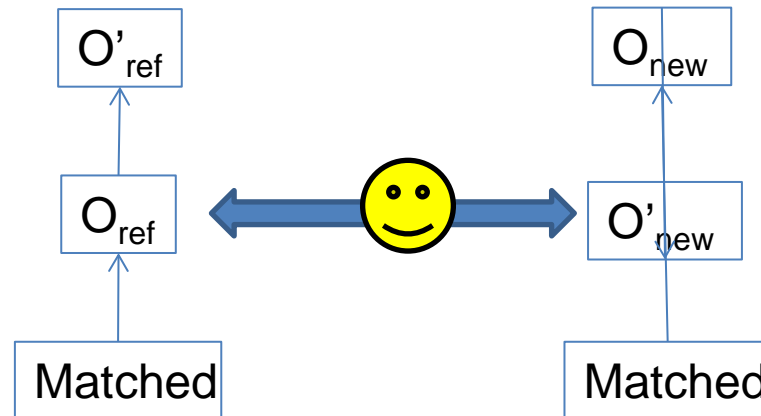
- Invariants
 - One pair of operations matches
 - A new match is only considered if inputs match
- Rule
 - Compare operations before trying to move



- **Invariants**
 - One pair of operations matches
 - A new match is only considered if inputs match
- **Rule**
 - Compare operations before trying to move



- **Invariants**
 - One pair of operations matches
 - A new match is only considered if inputs match
- **Rule**
 - Compare operations before trying to move



- **BI Systems Characterization**
 - Exploratory BI
 - State of the Art: Trends
- **Technical Challenges**
 - Semantic-Aware Systems
 - Autonomy Vs. Consistency
 - Automation
- **Our Vision: A System to Enable Exploratory BI**
 - AMDO
 - GEM
 - ORE
 - COAL
- **Still a lot to do!**

Gràcies per la seva Atenció!
¡Gracias por su Atención!
Thank you for your Attention!



Questions?

Oscar Romero
(oromero@essi.upc.edu)

EM Master IT4BI: <http://it4bi.univ-tours.fr/>
EM Joint Doctorate IT4BI-DC: <http://it4bi-dc.ulb.ac.be/>



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**