

Determining Source Contribution in Integration Systems

A presentation by

Yannis Katsis

on a joint work with

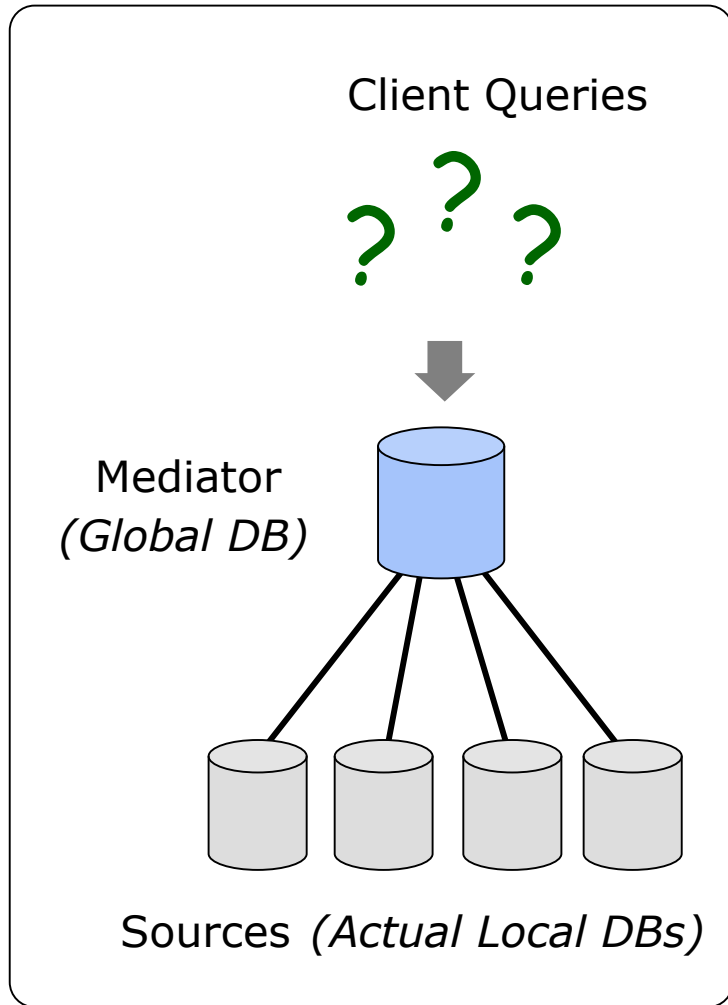
Alin Deutsch

Yannis Papakonstantinou

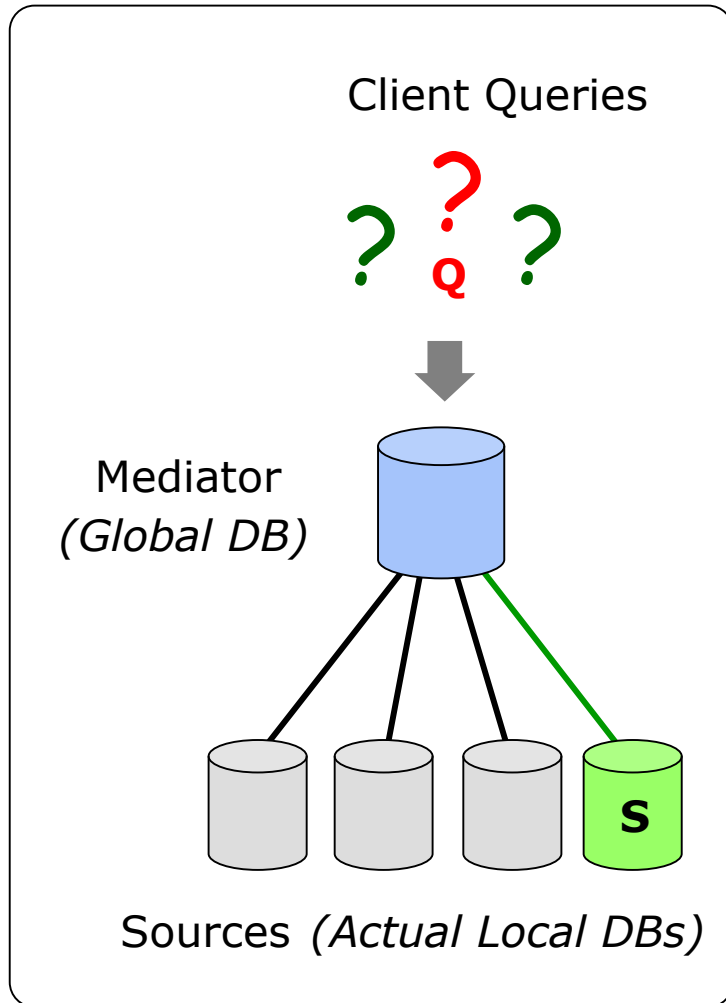


SDSC

The Problem

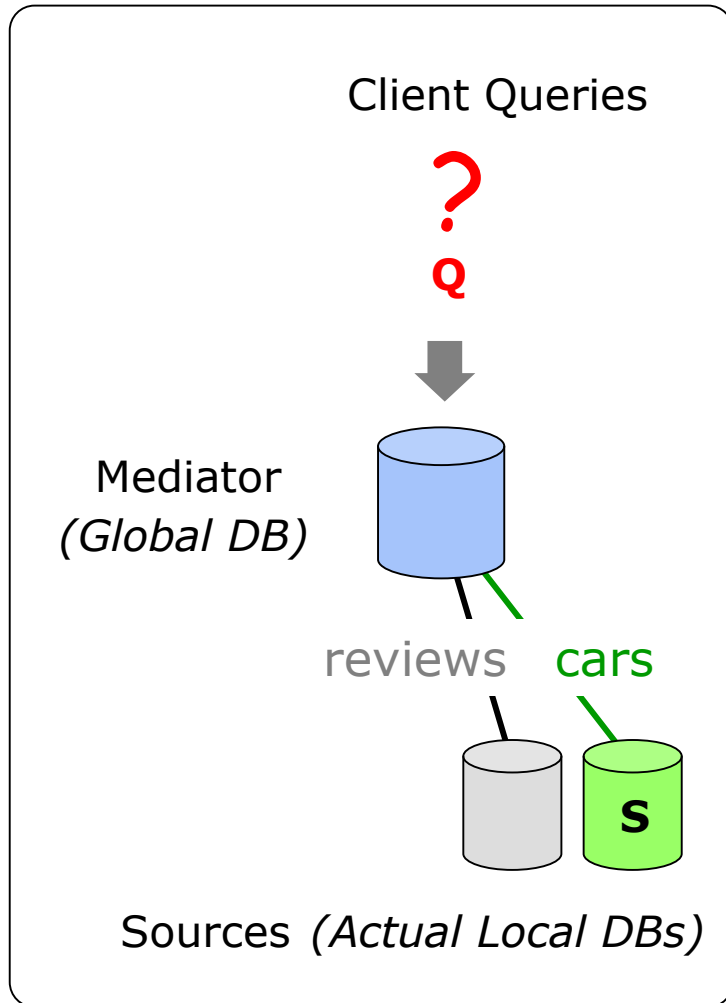


The Problem



- What is the **contribution** of source S to the result of the query Q?

The Problem



- What is the **contribution** of source **S** to the result of the query **Q**?

Q: cars

S is **Self Sufficient** w.r.t. Q

Q: cars JOIN reviews

S is **Now Complementary** w.r.t. Q

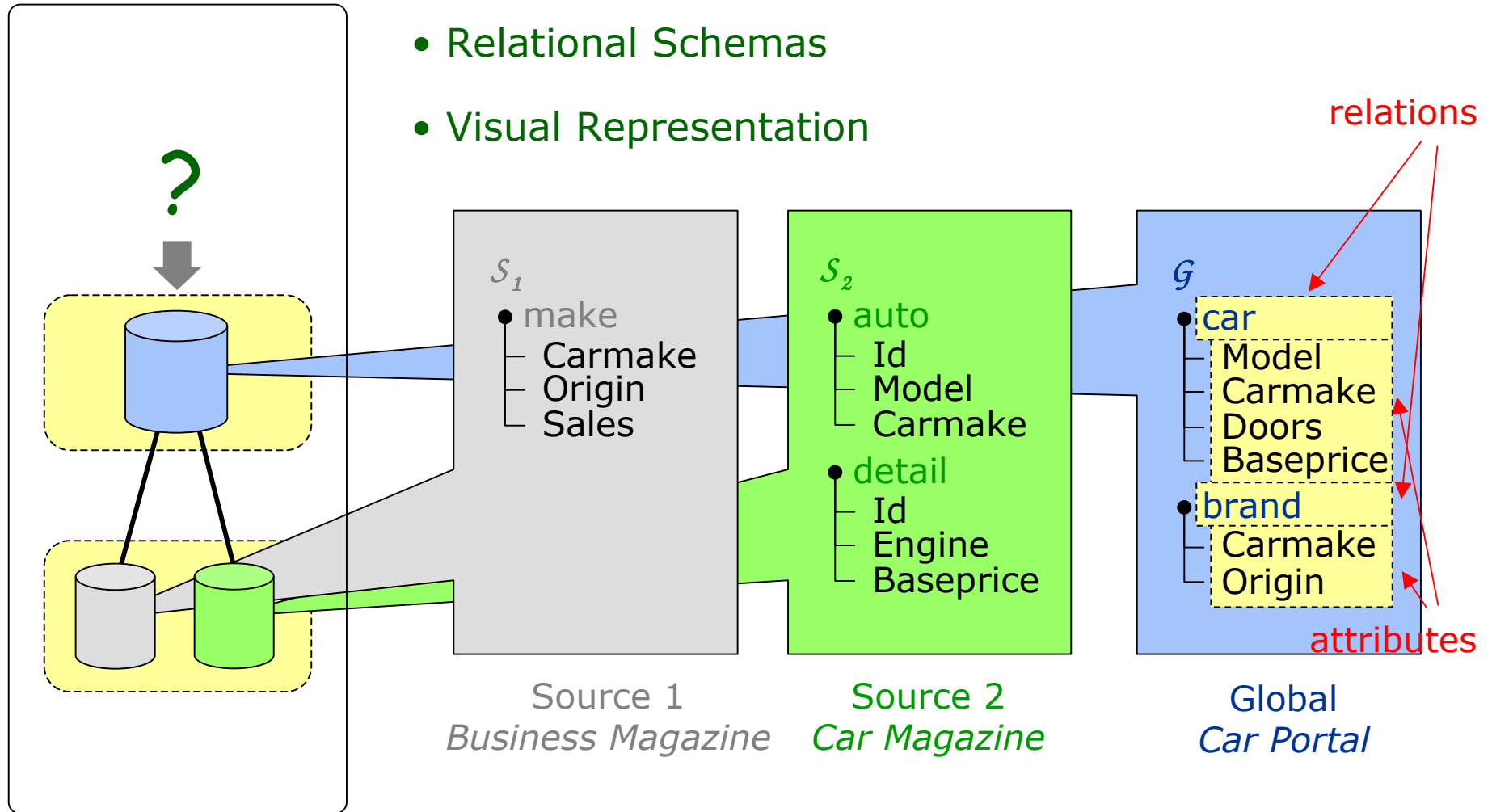
Roadmap

Framework

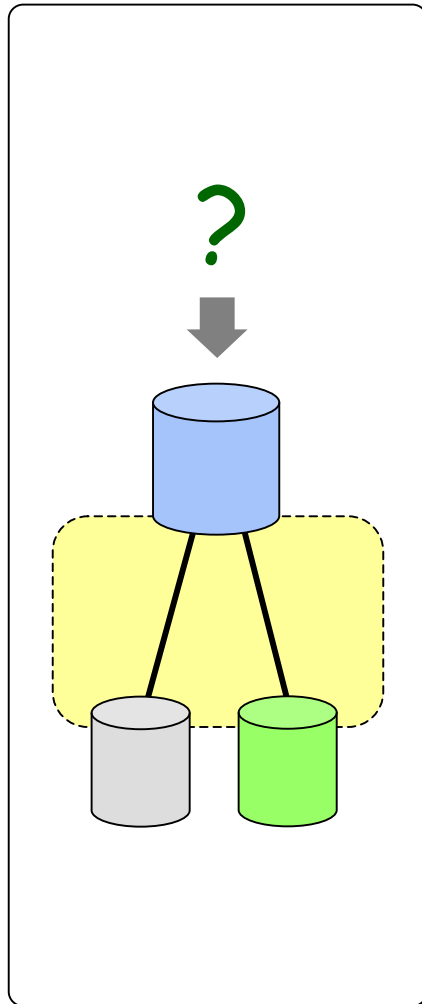
Degrees of Contribution

Decidability Results

Relational Schemas: Local and Global

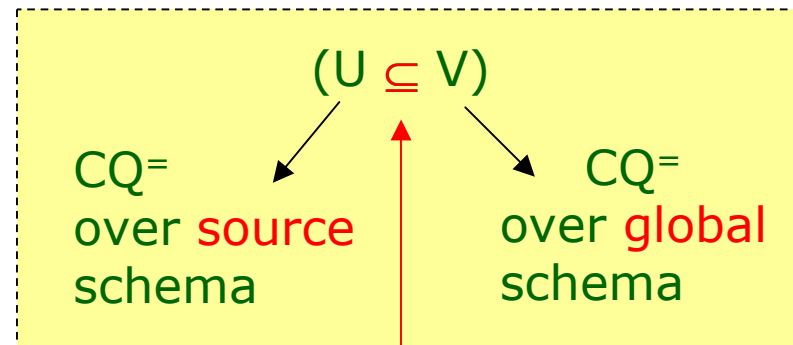


Source Registration using GLAV Mappings



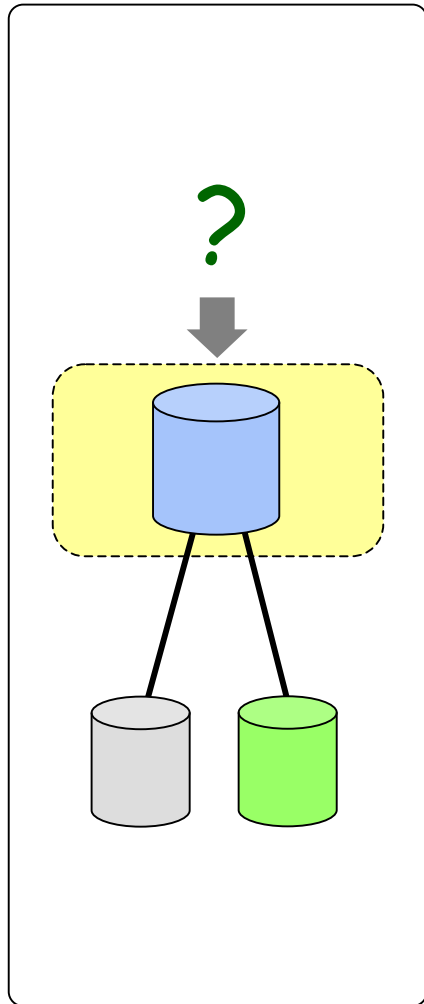
- Source Registration: Correspondence between a source schema and the global schema

=
Set of Mapping Constraints of the form

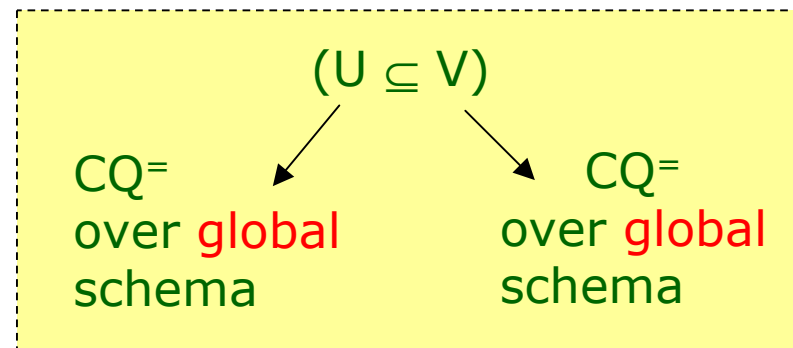


- Open World
- Global and Local As View (GLAV)

Target Constraints



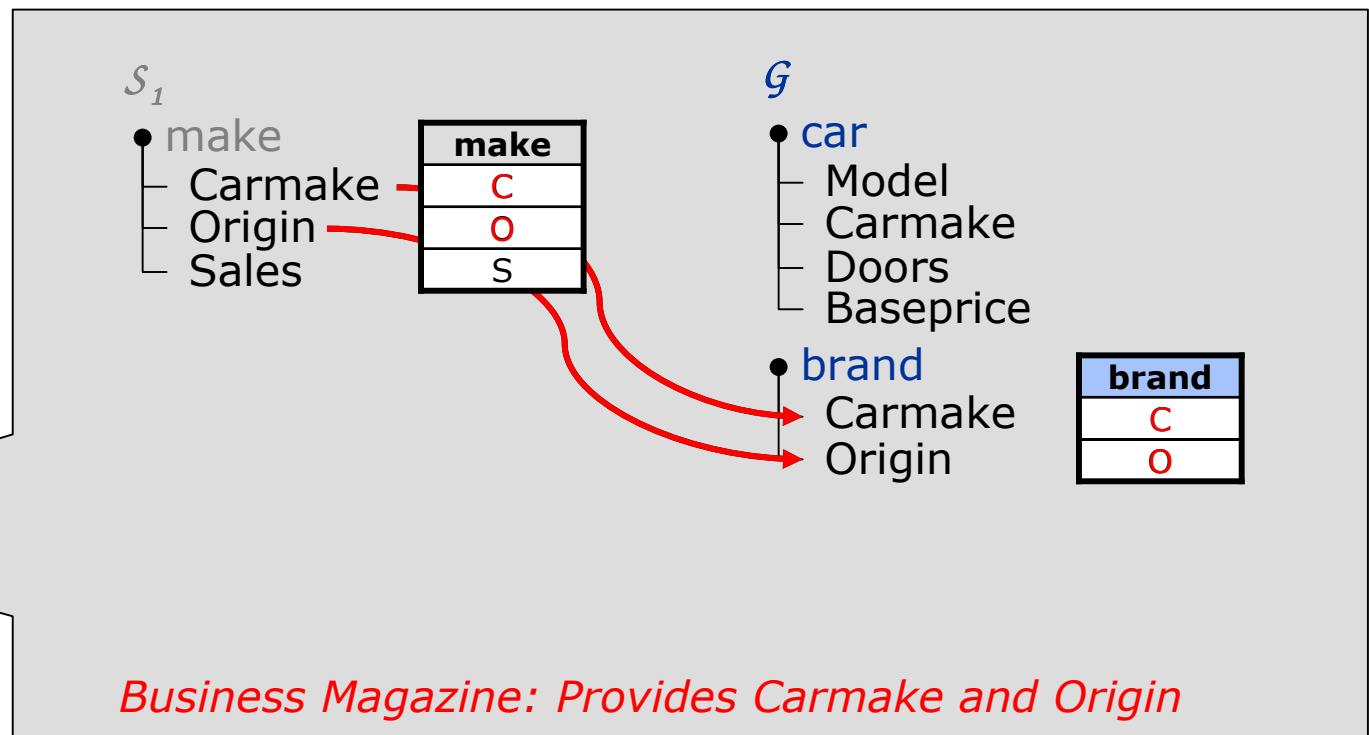
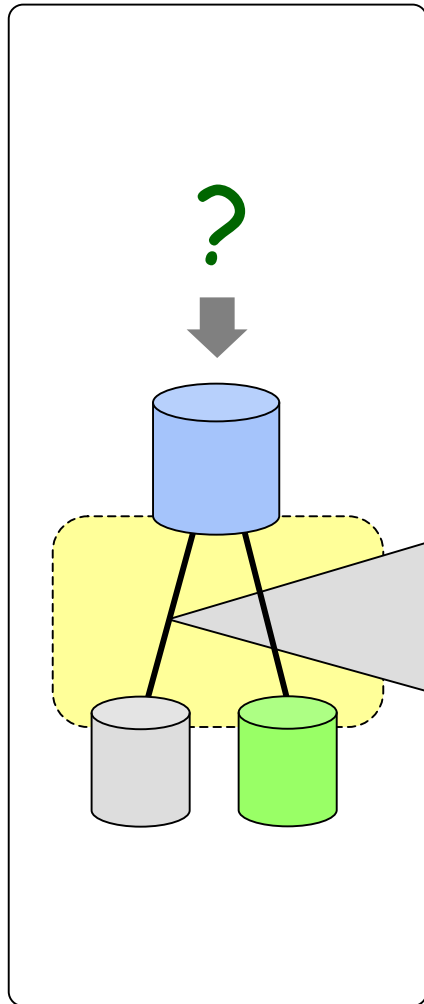
- Constraints on the global schema
=
Set of Constraints of the form



- Expresses Embedded Dependencies (PKs, RICs, ...)

Visual Representation of Mappings (1)

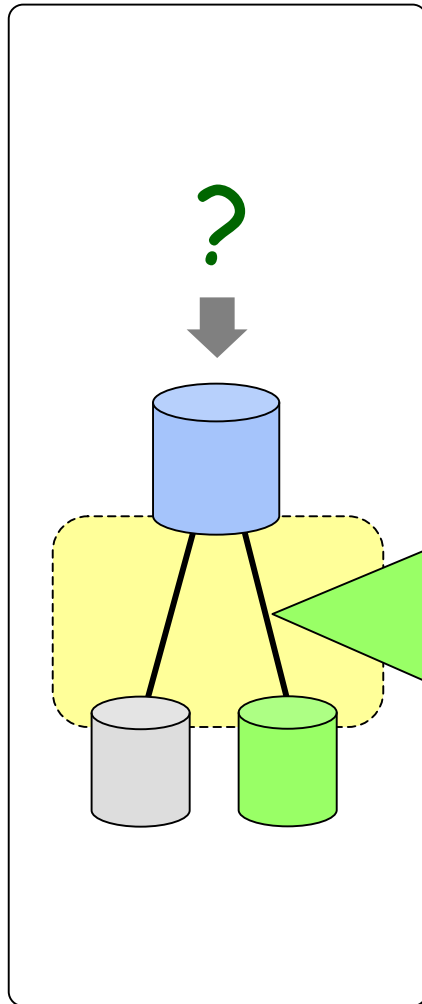
- Visual Representation (IBM Clio)



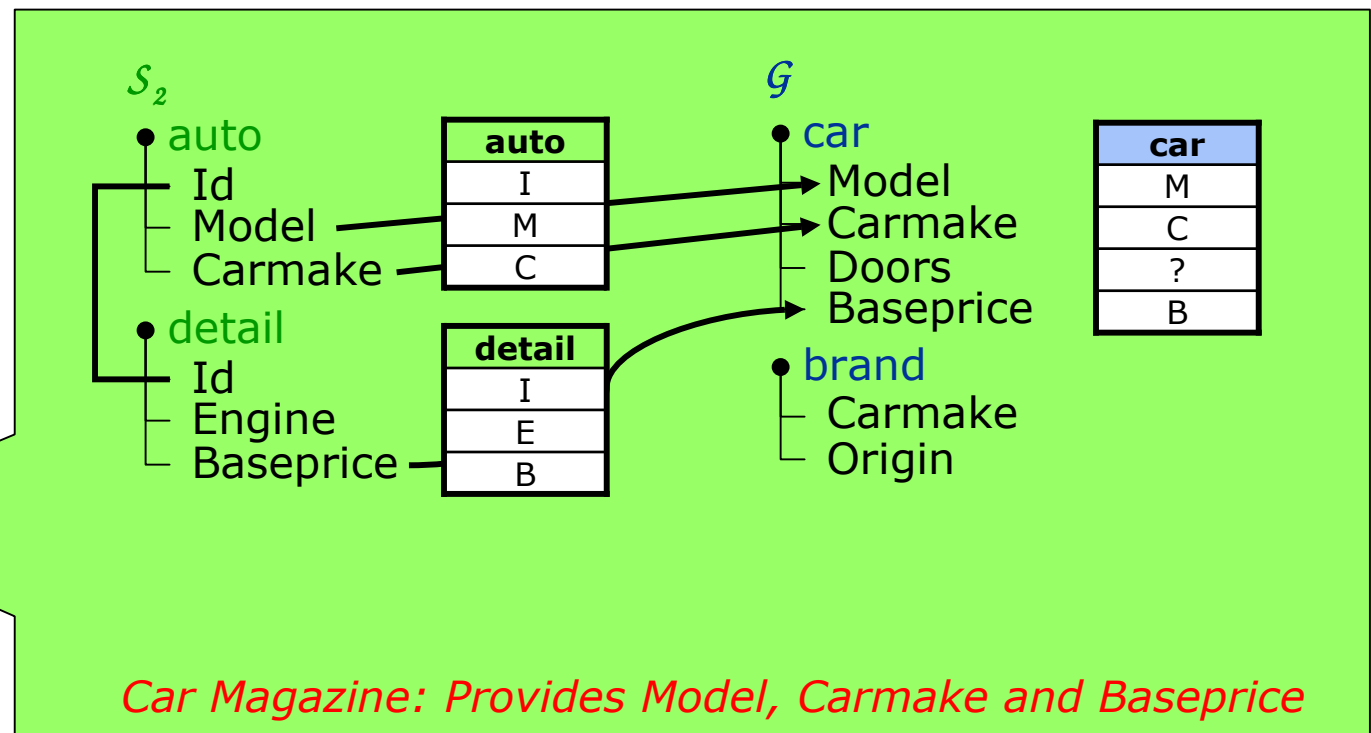
$(U_1 \subseteq V_1)$

$U_1(C, O) :- \text{make}(C, O, S)$
 $V_1(C, O) :- \text{brand}(C, O)$

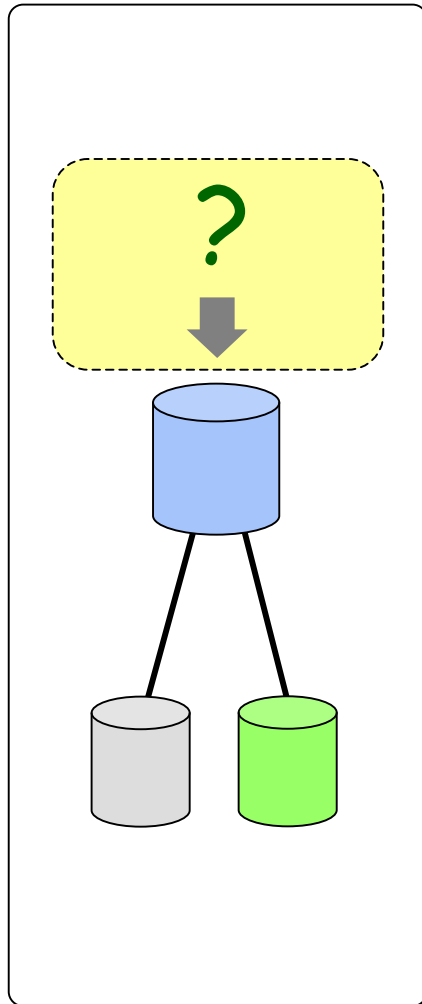
Visual Representation of Mappings (2)



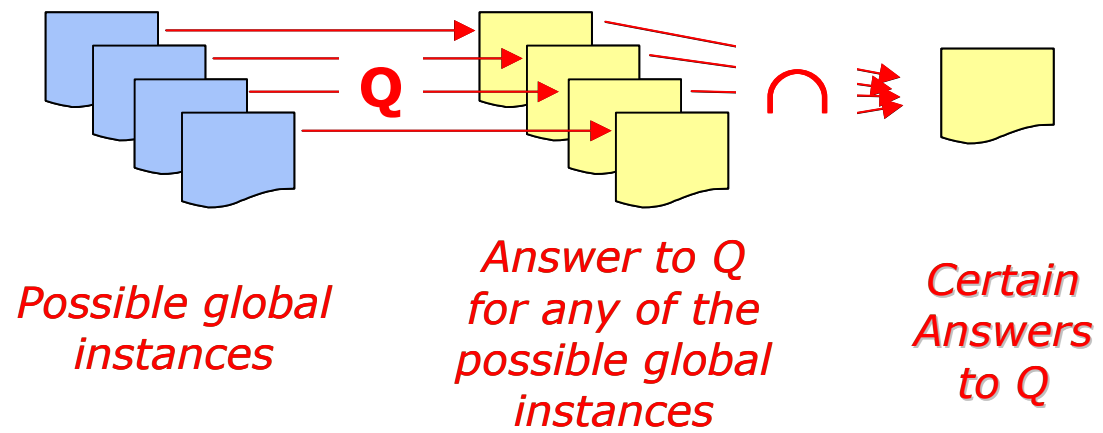
- Visual Representation (IBM Clio)



Query Semantics



- Queries in UCQ=
 - Set of Possible Global Instances
Set of global instances that satisfy all constraints
 - Query Answers = Set of Certain Answers
The tuples appearing in the answer to Q for any possible global instance



Roadmap

Framework

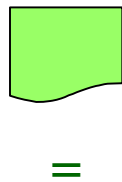
Degrees of Contribution

Decidability Results

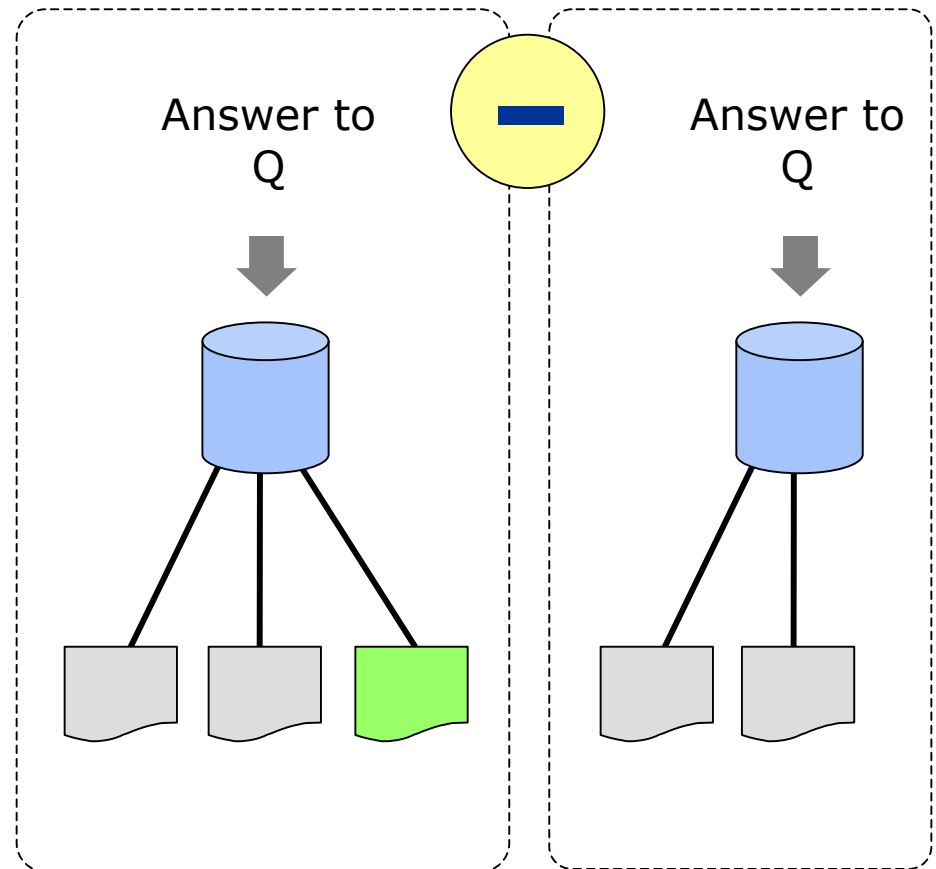
Source Instance's Contribution

- For given instances of the sources

Contribution to Q of Source Instance



The tuples in answer of Q not provided by the other sources



Source Registration's Contribution

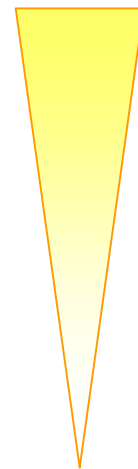
- Source Registration: Source Mappings
- Degrees of Source Registration's Contribution

① Self Sufficient

② Now Complementary

③ Later Complementary

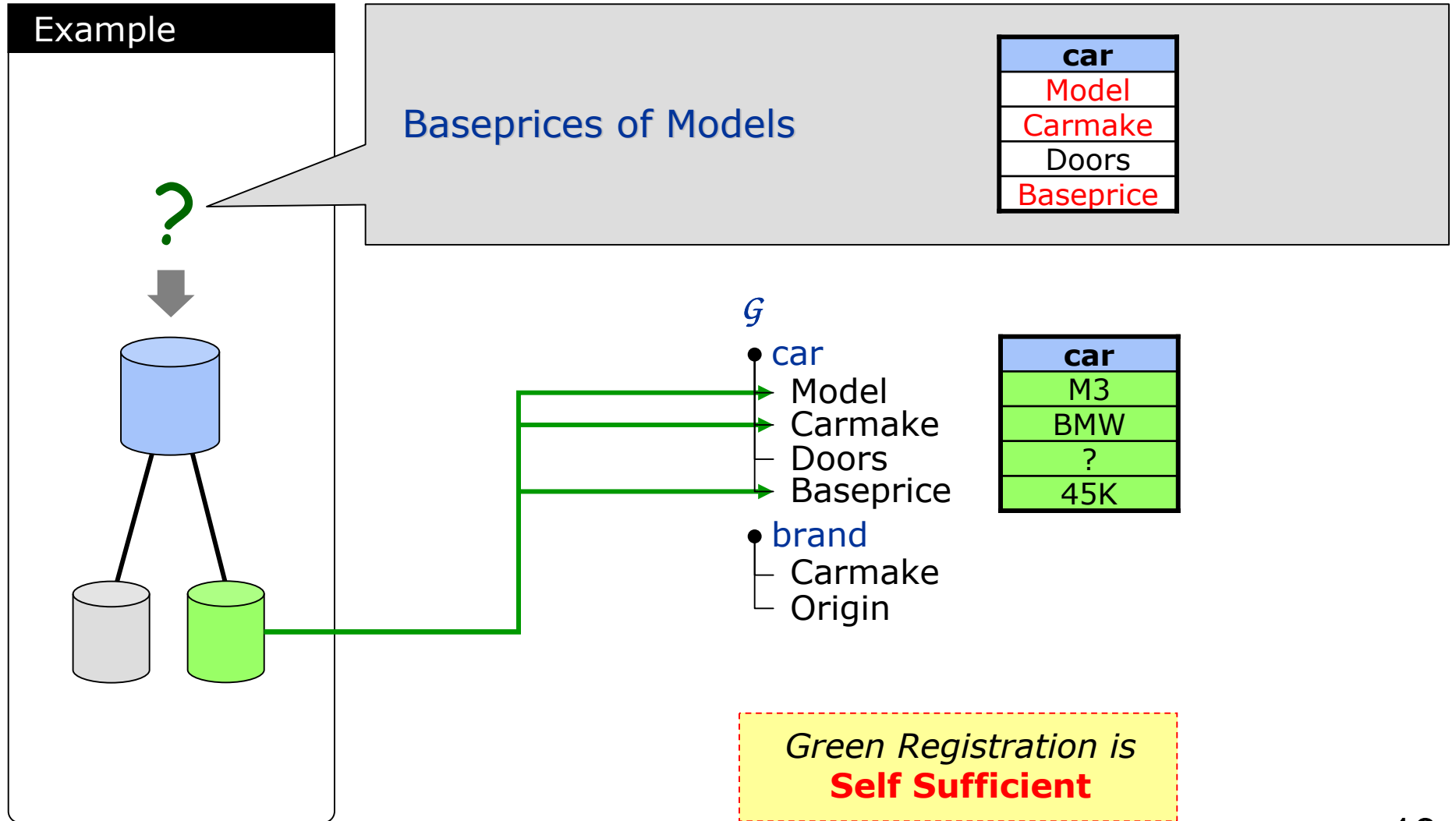
④ Unusable



More contribution

Less contribution

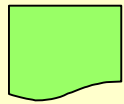
Self Sufficient Registration: Example



Self Sufficient Registration: Definition

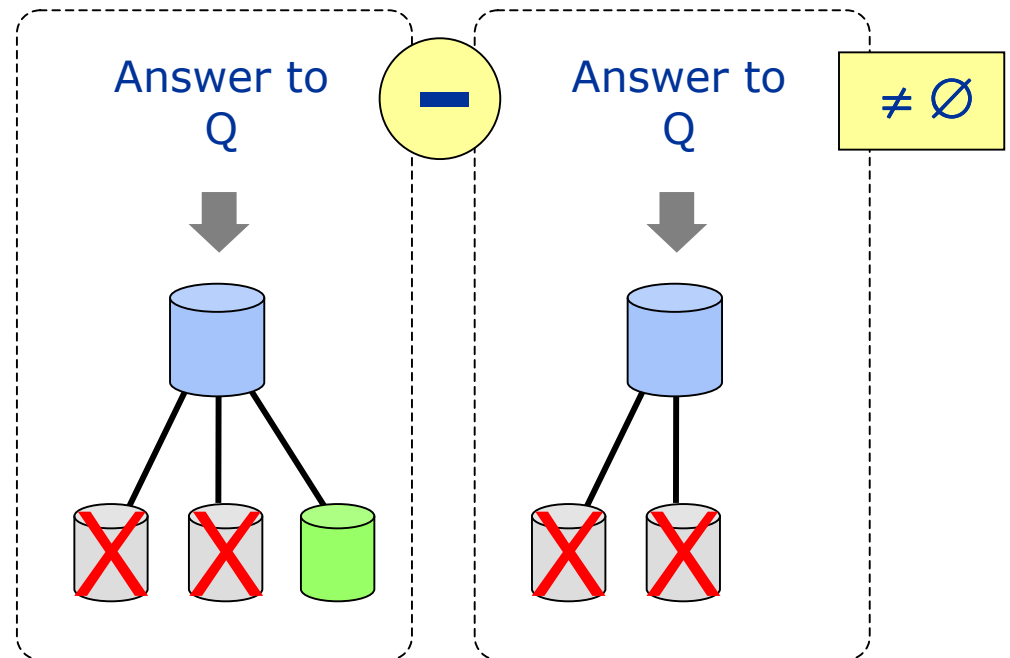
1 Self Sufficient

\exists Source instance

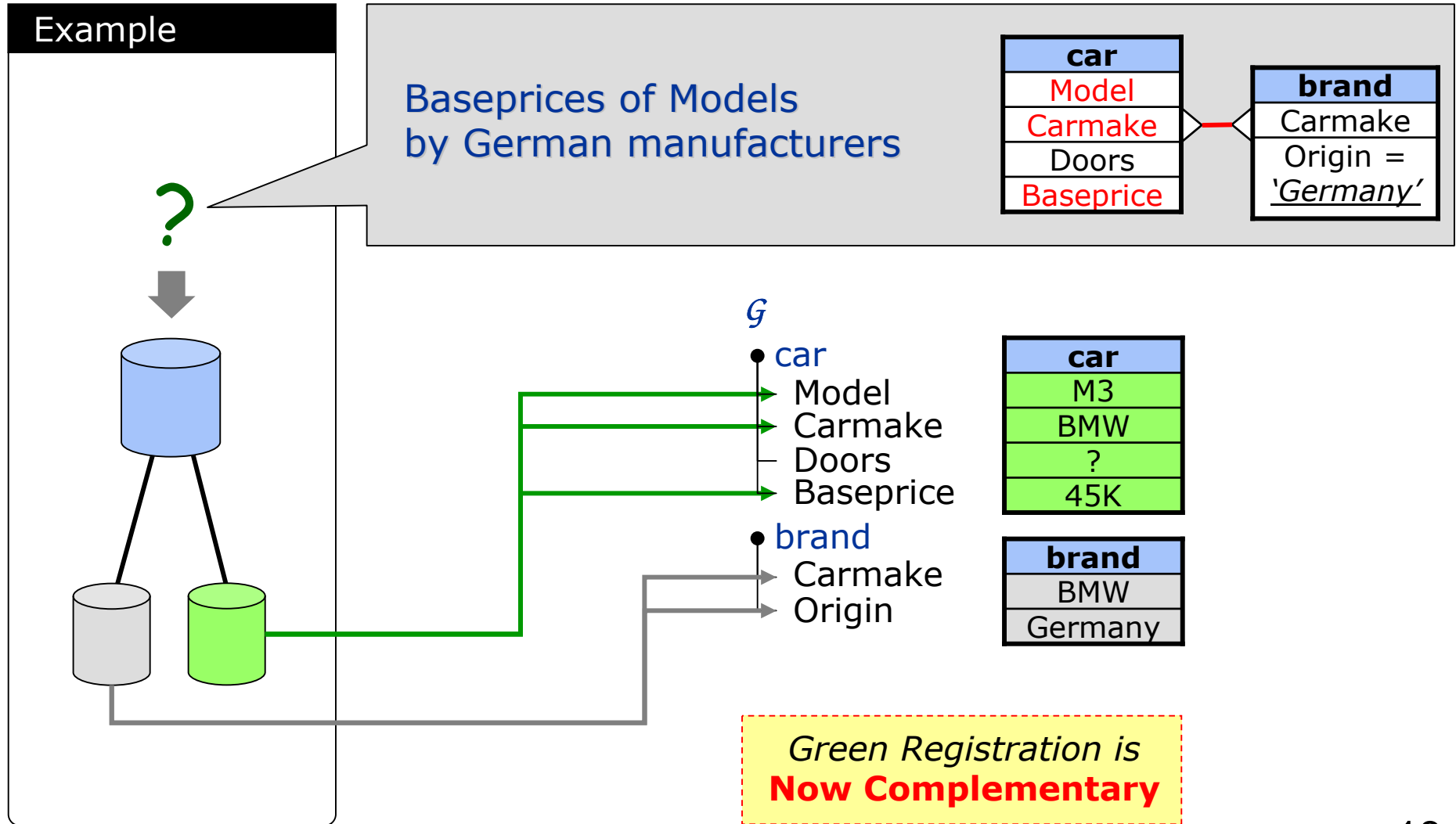


s.t.

The source has a non empty contribution in the **absence** of the other sources



Now Complementary Registration: Example



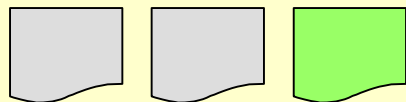
Now Complementary Registration: Definition

② Now Complementary

Not Self Sufficient

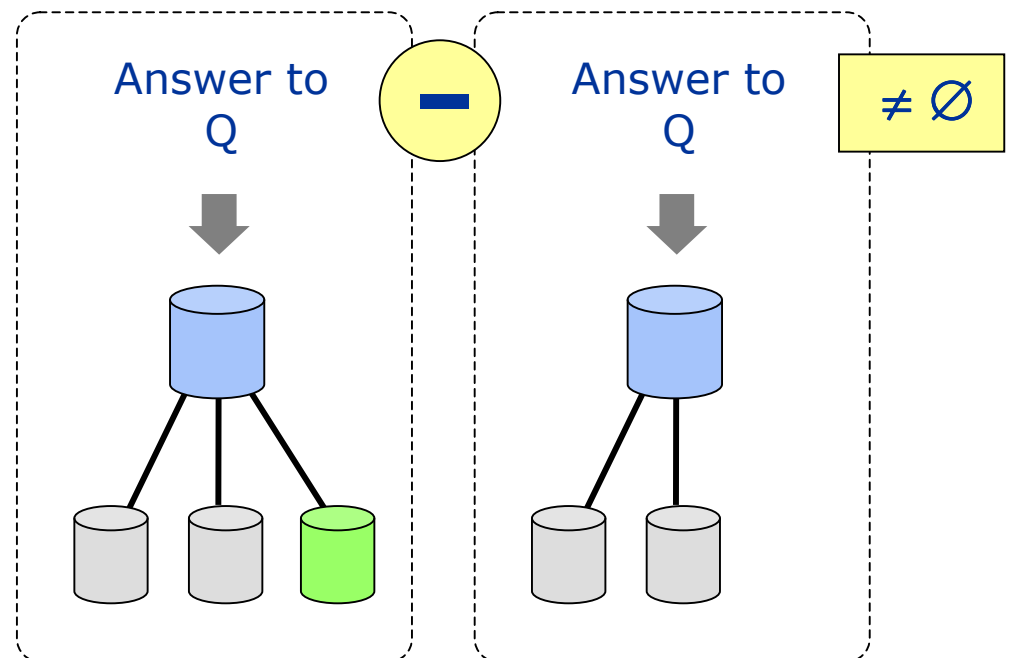
&

∃ Source instances

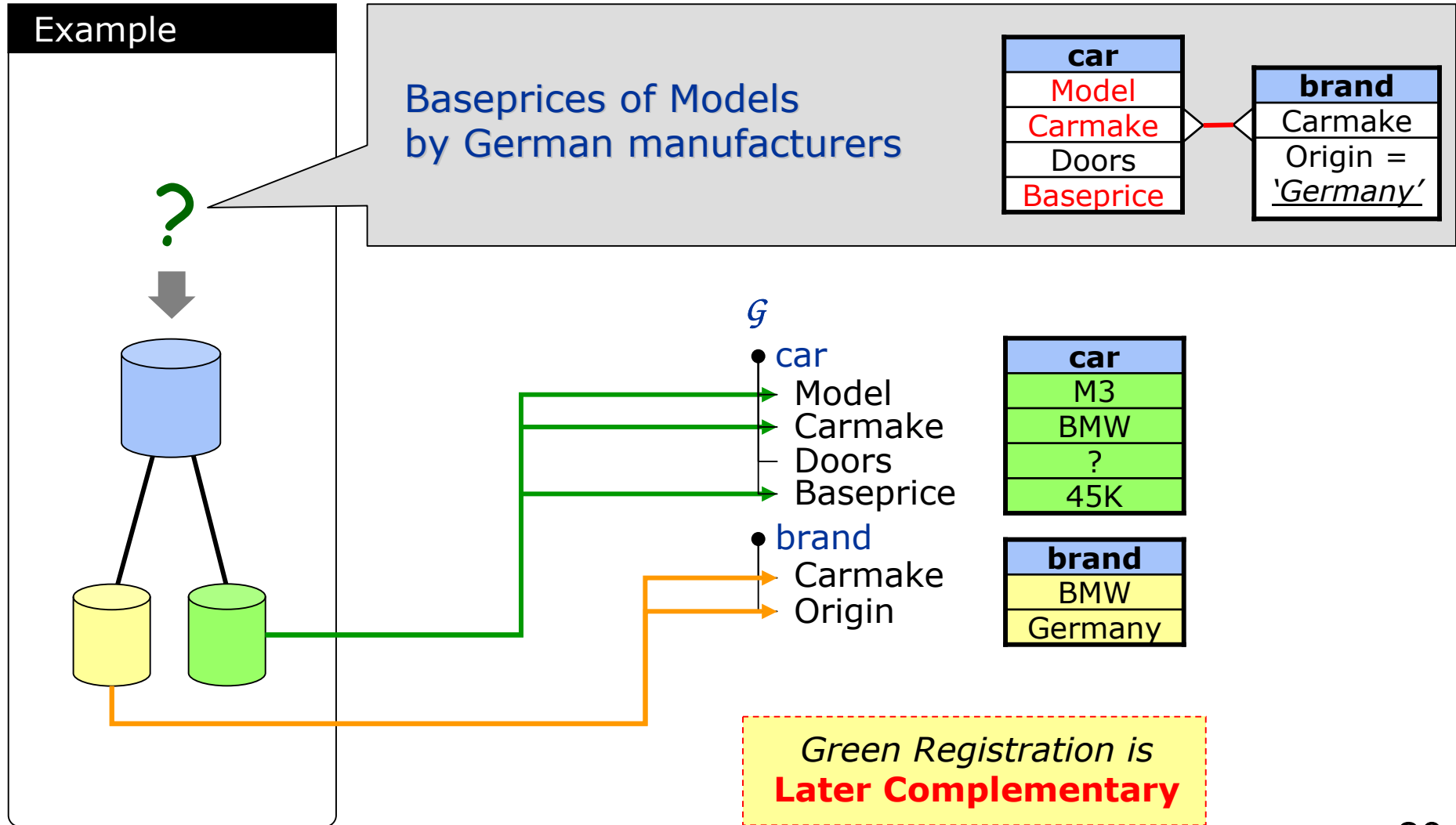


s.t.

The source has a non empty contribution in combination with the other **existing** sources



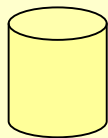
Later Complementary Registration: Example



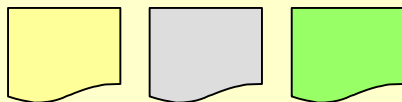
Later Complementary Registration: Definition

③ Later Complementary

*Not Self Sufficient &
Not Now Complementary
&
 \exists Potential future sources*

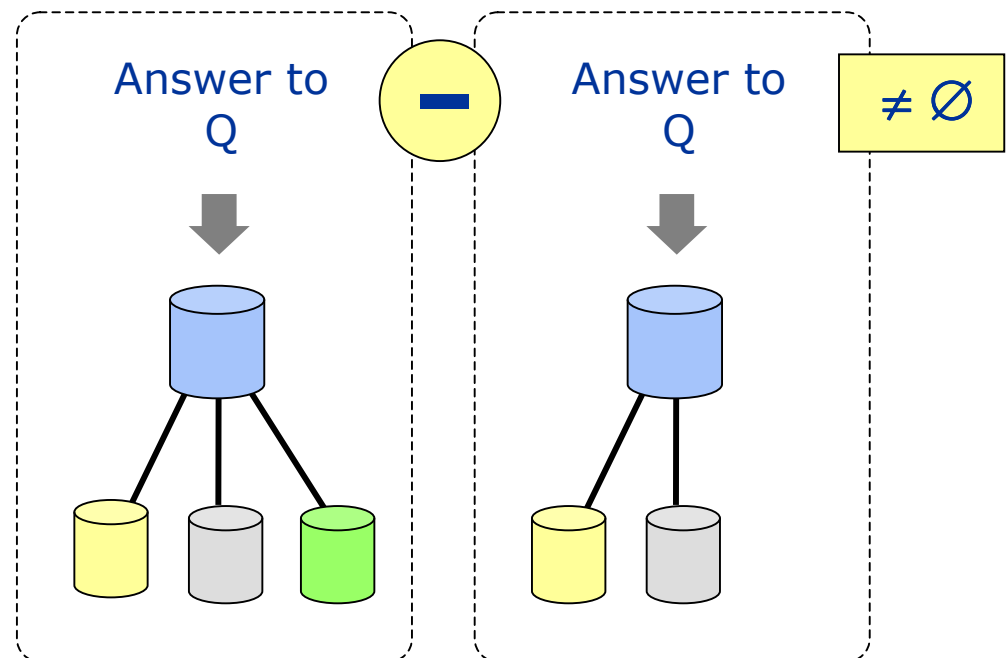


& Source instances

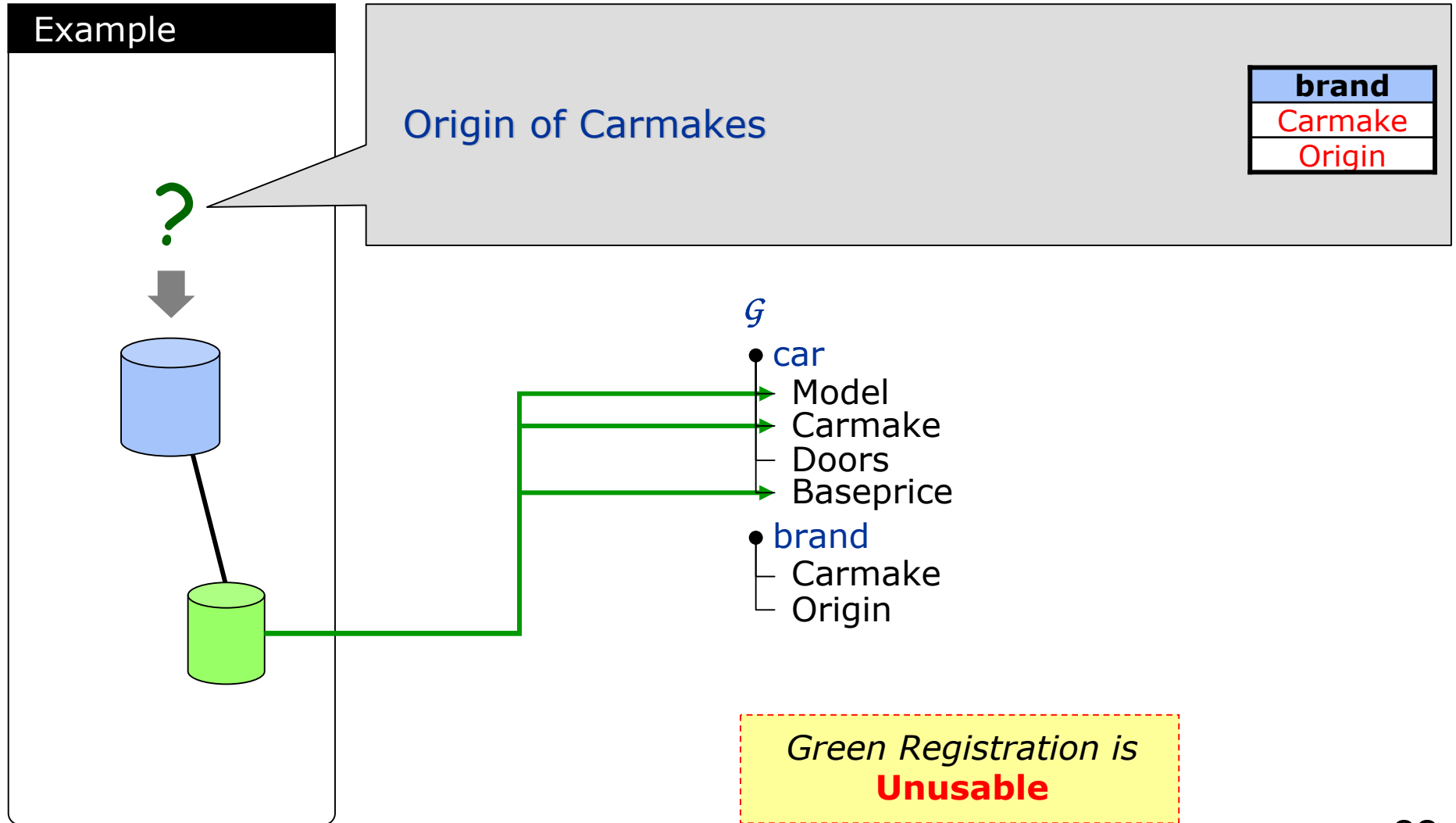


s.t.

The source has a non empty contribution in combination with the **future** sources



Unusable Registration: Example



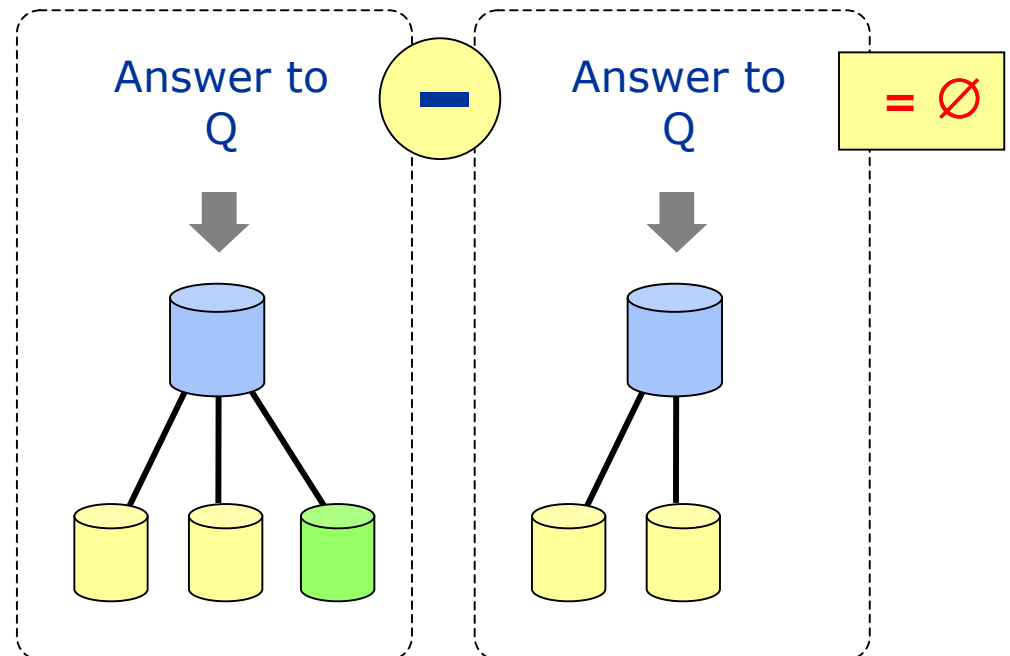
Unusable Registration: Definition

4 Unusable

Not Self Sufficient &
Not Now Complementary &
Not Later Complementary



The source has a empty
contribution regardless of
what sources enter the
system



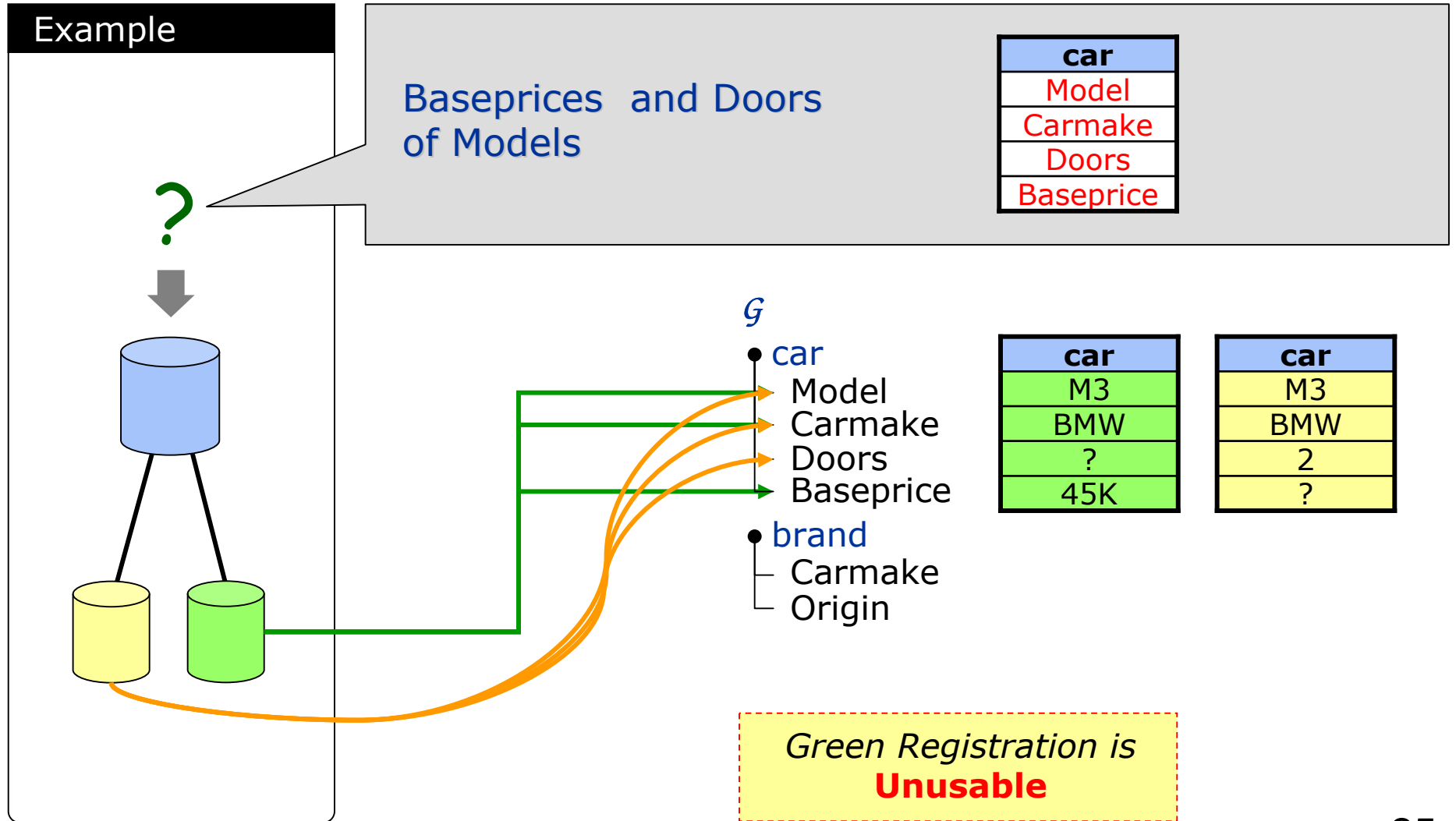
Roadmap

Framework

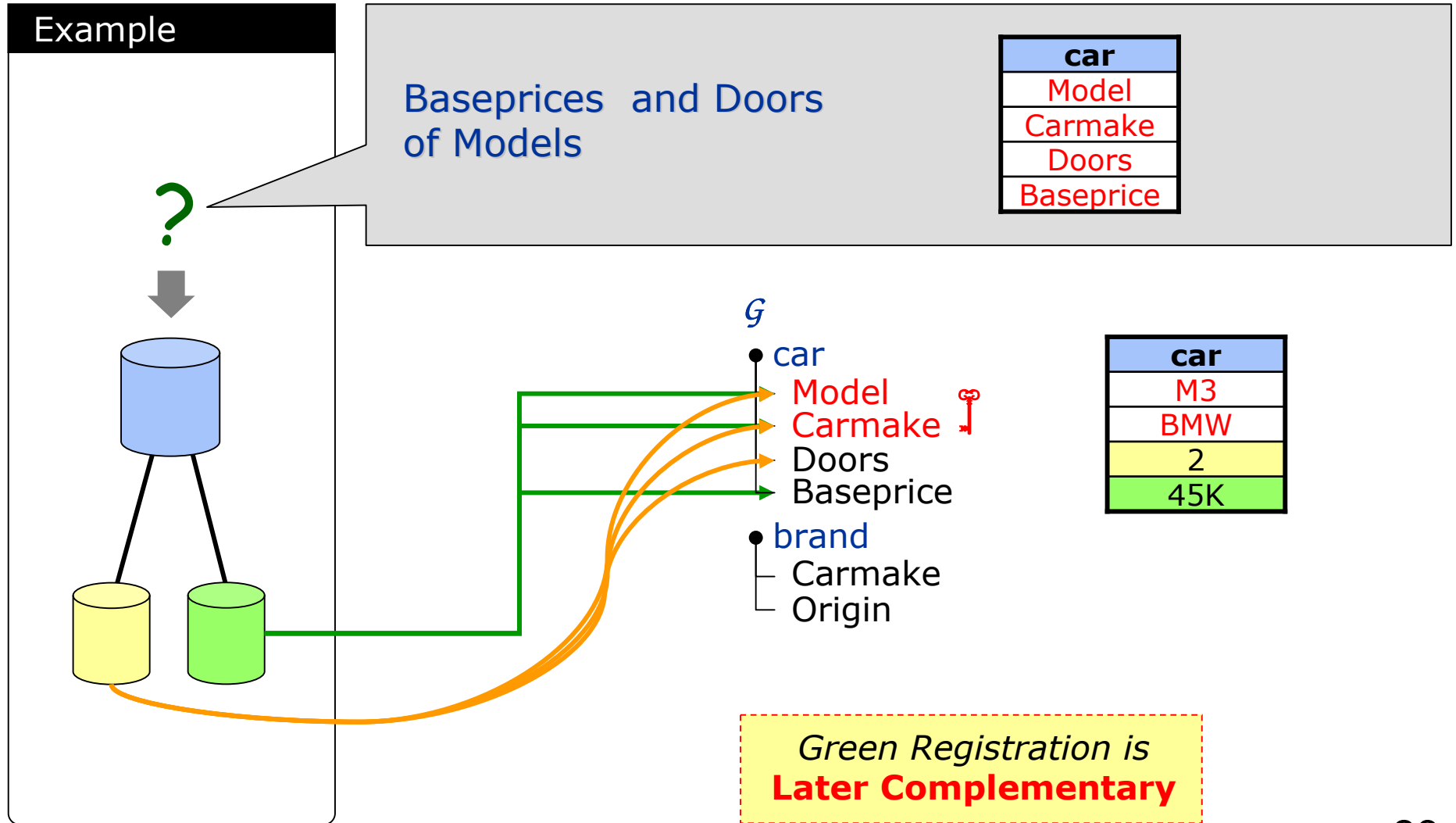
Degrees of Contribution

Decidability Results

Subtleties for Unusable Registrations



In presence of PK Unusable Example becomes Later Complementary



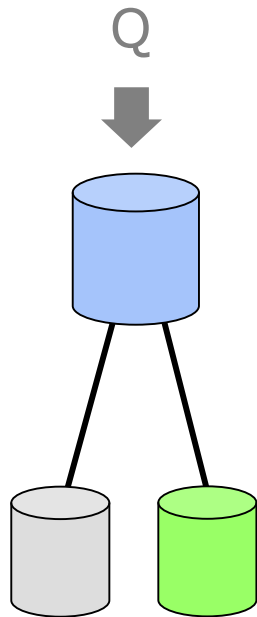
Decidability Results

Overview: What is decidable

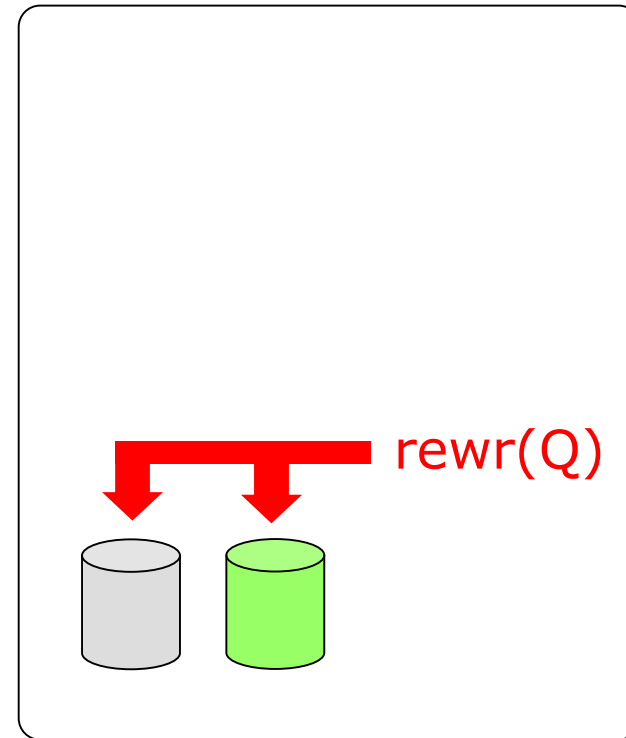
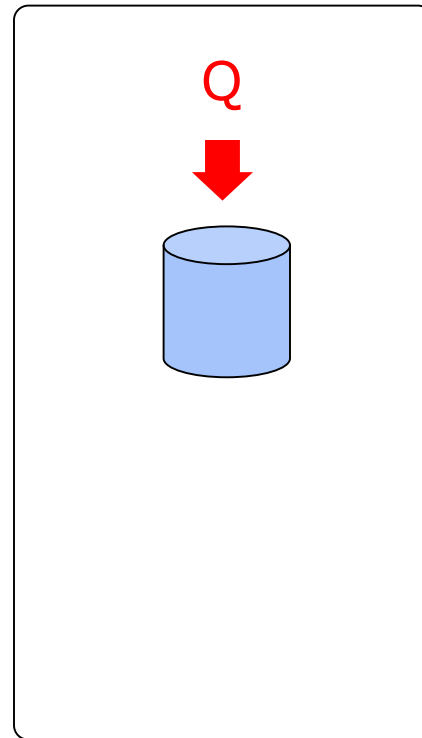
		Target constraints		
		<i>None</i>	<i>Primary keys</i>	<i>Primary keys + Referential Integrity Constraints</i>
D e g r e e s	<i>Self Sufficient</i>	Yes	Yes	No
	<i>Now complementary</i>	Yes	Yes	No
	<i>Later complementary</i>	Yes	Yes	?
	<i>Unusable</i>	Yes	Yes	?

Certain Rewriting

Original Query
Query Q over
global schema



Certain Rewriting
Query $\text{rewr}(Q)$ over
source schemas



Certain Answers to Q = *Answers to $\text{rewr}(Q)$*

Decidability of Self Sufficiency

Self Sufficiency

<i>None</i>	<i>PKs</i>	<i>PKs & RICs</i>
Decidable	Decidable	Undecidable

- *Reduction: Self Sufficiency* → Satisfiability of certain rewriting
 - PKs: Certain rewriting expressible as a potentially recursive datalog program [Duschka, Genesereth]
➔ Self Sufficiency decidable
 - PKs & RICs: No known result for finding certain rewriting
Reduction from the Post Correspondence Problem [new result]
➔ Self Sufficiency undecidable

Decidability of Now Complementarity

Now Complementarity

<i>None</i>	<i>PKs</i>	<i>PKs & RICs</i>
Decidable	Decidable	Undecidable

- *Reduction: Now Complementarity* → Containment of certain rewritings

- PKs: Certain rewriting expressible as a potentially recursive datalog program [Duschka, Genesereth]

Certain rewriting expressible as UCQ⁼ [new result]

➡ Now Complementarity decidable

- PKs & RICs: Reduction from Self Sufficiency

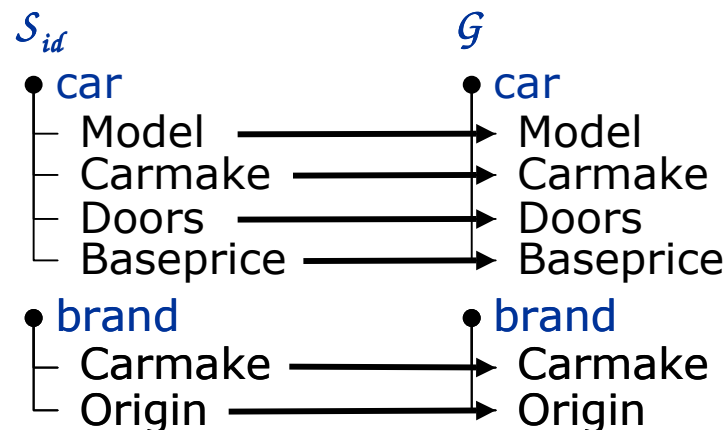
➡ Now Complementarity undecidable

Decidability of Later Complementarity

Later Complementarity

None	PKs	PKs & RICs
Decidable	Decidable	?

- No target constraints: Check whether source registration is now complementary with identity registration



- PKs: This approach does not work. Check if now complementary with more registrations

Conclusion

- Summary:

- *Definition of 4 Degrees of Contribution*
- *Decidability results based on the type of target constraints*
- *Extension of work on Certain Rewritings*

- Future Work:

- *Make definitions instance dependent*
- *Extend classes of supported queries and mappings*
- *Interface for guiding the source owner through source registration*