UC SANTA CRUZ

Universal Solutions in Temporal Data Exchange

Zehui Cheng
University of California Santa Cruz

Phokion G. Kolaitis
University of California Santa Cruz
IBM Research - Almaden

24 September 2020

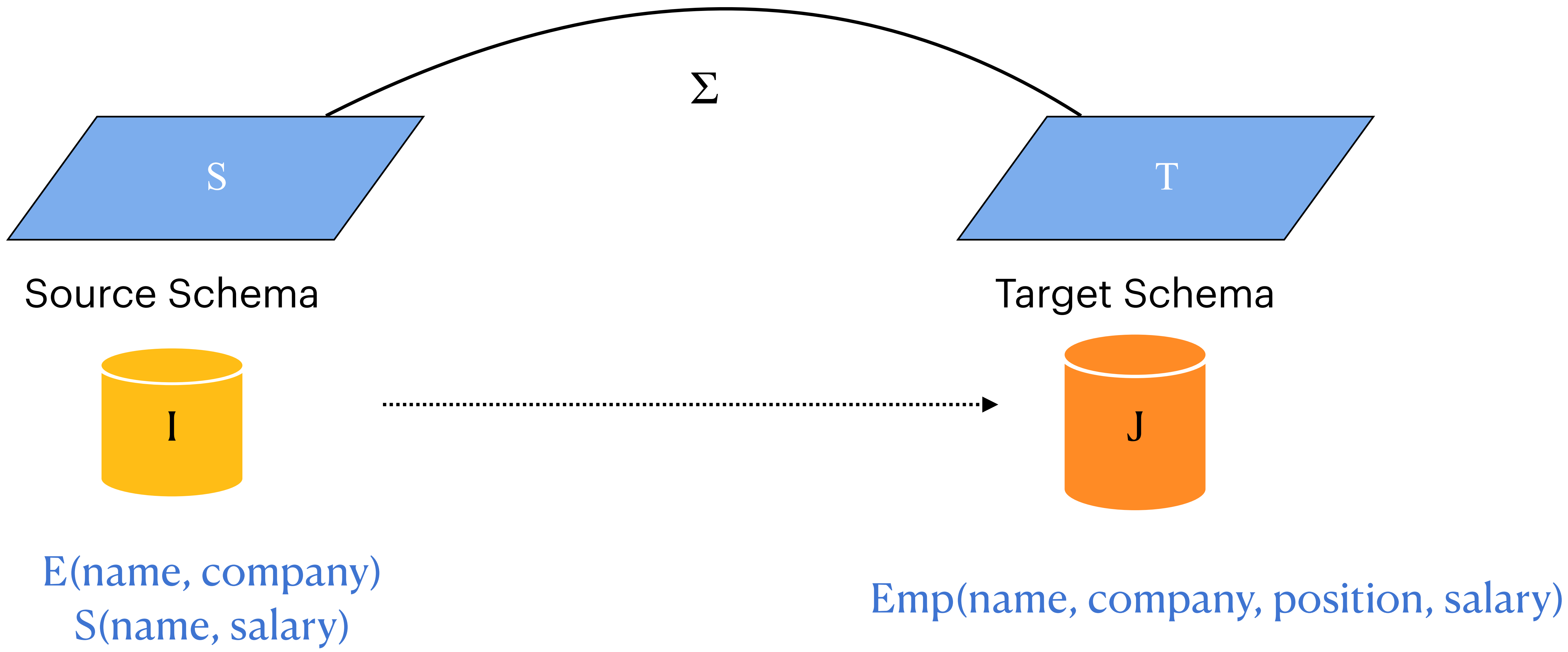
Data Exchange

Fagin, R., Kolaitis, P. G., Miller, R. J., and Popa, L. : Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1): 89-124 (2005)

Transform data structured under a source schema into data structured under a different target schema.

Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$

- Relational source schema $\mathbf{S}$, Relational target schema $\mathbf{T}$
- A set Σ of constraints between $\mathbf{S}$ and $\mathbf{T}$



Data Exchange

Fagin, R., Kolaitis, P. G., Miller, R. J., and Popa, L. : Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1): 89-124 (2005)

Transform data structured under a source schema into data structured under a different target schema.

Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$

- Relational source schema $\mathbf{S}$, Relational target schema $\mathbf{T}$
- A set Σ of constraints between $\mathbf{S}$ and $\mathbf{T}$

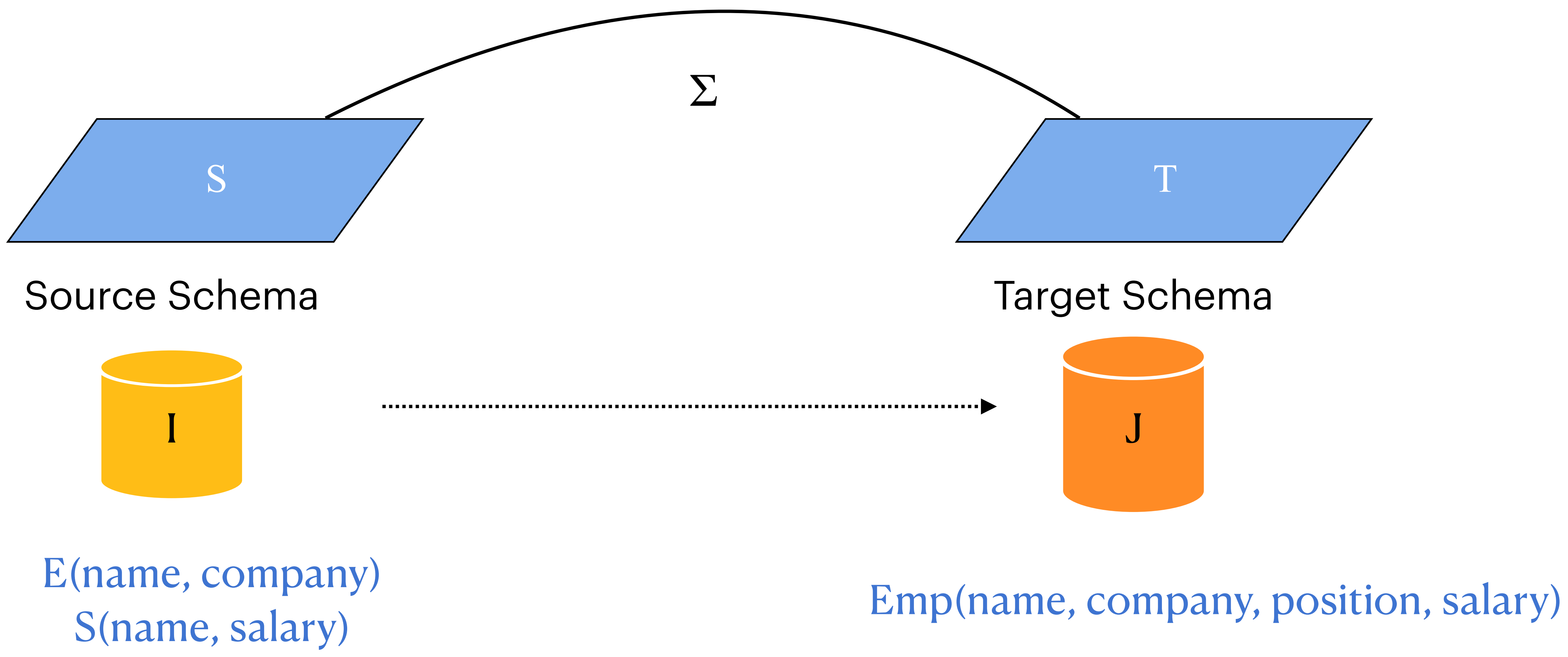
Constraints:

- source-to-target tuple-generating dependencies (s-t tgds)

$$\forall \mathbf{x}(\phi(\mathbf{x}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}))$$

- target equality-generating dependencies (target egds)

$$\forall \mathbf{x}(\theta(\mathbf{x}) \rightarrow x_k = x_l)$$



Example:

(s-t tgd) $\forall n, c, p(E(n, c) \wedge P(n, p) \rightarrow \exists s \; Emp(n, c, p, s))$

(target egd) $\forall n, c, p, s_1, s_2(Emp(n, c, p, s_1) \wedge Emp(n, c, p, s_2) \rightarrow s_1 = s_2)$

Data Exchange

Fagin, R., Kolaitis, P. G., Miller, R. J., and Popa, L. : Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1): 89-124 (2005)

Transform data structured under a source schema into data structured under a different target schema.

Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$

- Relational source schema $\mathbf{S}$, Relational target schema $\mathbf{T}$
- A set Σ of constraints between $\mathbf{S}$ and $\mathbf{T}$

Constraints:

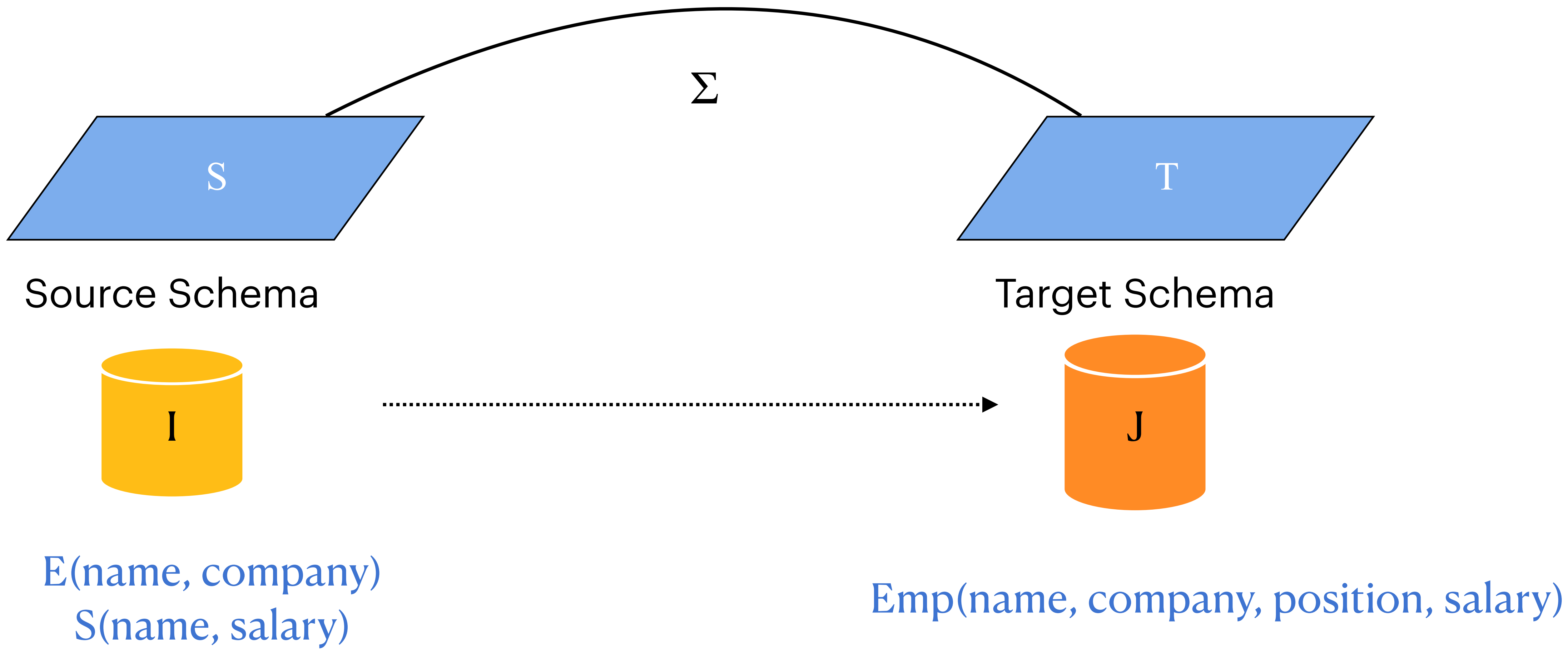
- source-to-target tuple-generating dependencies (s-t tgds)

$$\forall \mathbf{x}(\phi(\mathbf{x}) \rightarrow \exists \mathbf{y}\psi(\mathbf{x}, \mathbf{y}))$$

- target equality-generating dependencies (target egds)

$$\forall \mathbf{x}(\theta(\mathbf{x}) \rightarrow x_k = x_l)$$

The data exchange problem: given a source instance I, find a solution for I, that is, a target instance J such that the pair (I, J) satisfies the constraints Σ of $\mathscr{M}$.



Example:

(s-t tgd) $\forall n, c, p(E(n, c) \wedge P(n, p) \rightarrow \exists s\ Emp(n, c, p, s))$

(target egd) $\forall n, c, p, s_1, s_2(Emp(n, c, p, s_1) \wedge Emp(n, c, p, s_2) \rightarrow s_1 = s_2)$

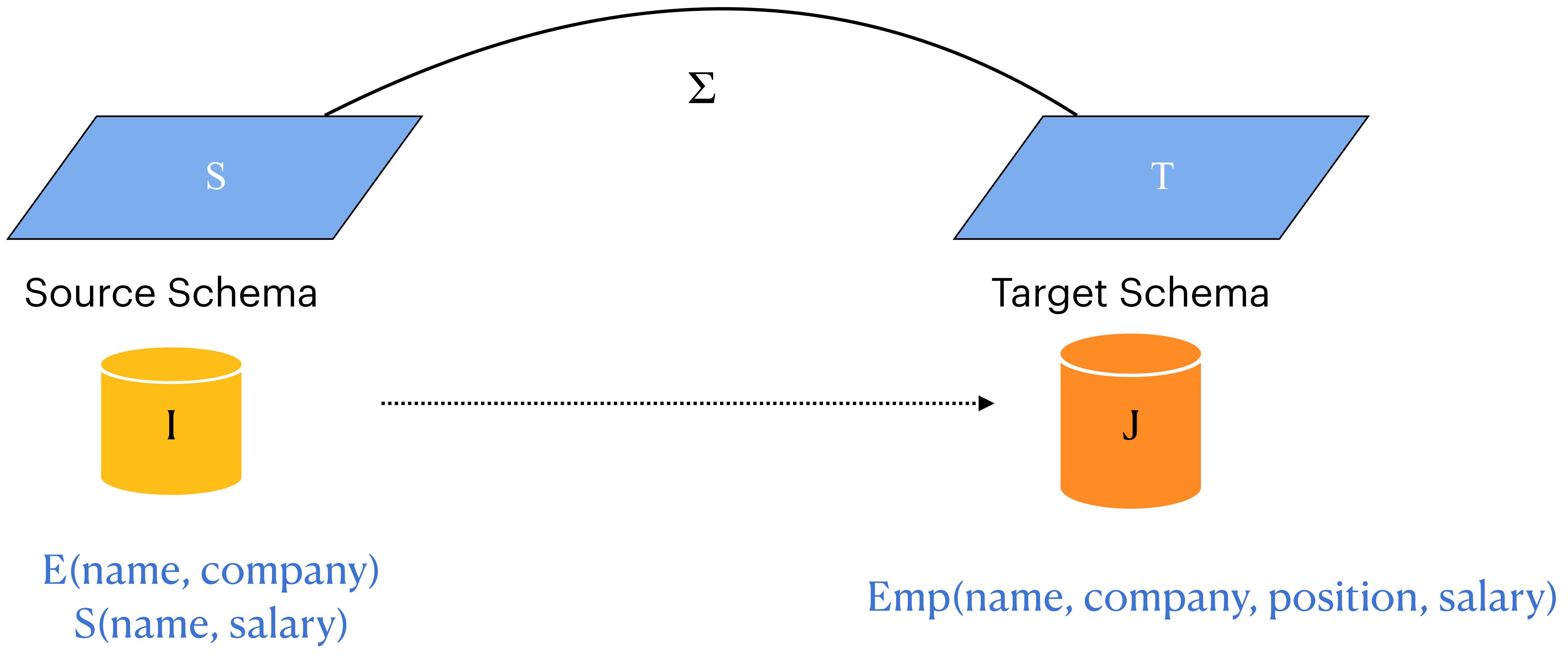
Data Exchange

Fagin, R., Kolaitis, P. G., Miller, R. J., and Popa, L. : Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1): 89-124 (2005)

Transform data structured under a source schema into data structured under a different target schema.

Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma)$

- Relational source schema $\mathbf{S}$, Relational target schema $\mathbf{T}$
- A set Σ of constraints between $\mathbf{S}$ and $\mathbf{T}$



Constraints:

- source-to-target tuple-generating dependencies (s-t tgds)

$$\forall \mathbf{x}(\phi(\mathbf{x}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}))$$

- target equality-generating dependencies (target egds)

$$\forall \mathbf{x}(\theta(\mathbf{x}) \rightarrow x_k = x_l)$$

Example:

(s-t tgd) $\forall n, c, p(E(n,c) \wedge P(n,p) \rightarrow \exists s\ Emp(n,c,p,s))$

(target egd) $\forall n, c, p, s_1, s_2(Emp(n,c,p,s_1) \wedge Emp(n,c,p,s_2) \rightarrow s_1 = s_2)$

The data exchange problem: given a source instance I, find a solution for I, that is, a target instance J such that the pair (I, J) satisfies the constraints Σ of $\mathscr{M}$.

Universal Solutions: The "most general" solutions (notion formalized using homomorphisms)

Temporal Databases

Abstract Model of Time: Time points (natural numbers)

Concrete Model of Time: Time intervals (intervals [s, e))

Temporal Schema: A schema with temporal relations, such as $R(Name, Company, Time)$
Abstract Instance: temporal attributes range over time points.
Concrete Instance: temporal attributes range over time intervals.

Semantic Function [[.]]: convert all tuples of the form $\mu = (c_1, \ldots, c_m, [s, e))$ into $[[\mu]] = \{(c_1, \ldots, c_m, t) : s \leq t < e\}$.

Name	Company	Time
Ada	IBM	[2012,2015)

Name	Company	Time
Ada	IBM	2012
Ada	IBM	2013
Ada	IBM	2014

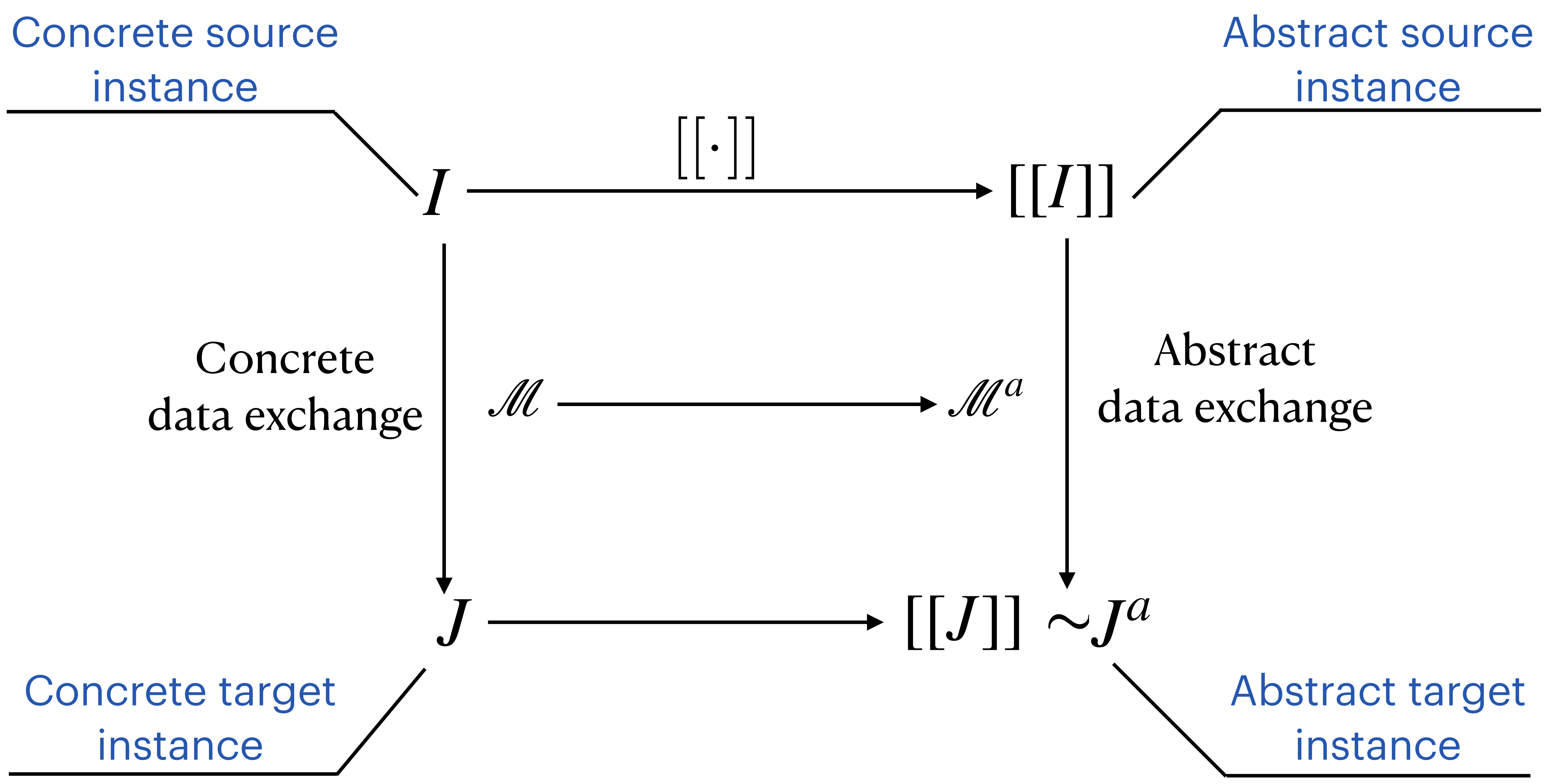
Temporal Data Exchange

Temporal Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$

- Source temporal schema $\mathbf{S}$, Target temporal schema $\mathbf{T}$
- Temporal s-t tgds Σ_{st}, Temporal target egds Σ_t

Temporal s-t tgds
$\forall \mathbf{x}, \mathbf{t}(\phi(\mathbf{x}, \mathbf{t}) \rightarrow \exists \mathbf{y}\psi(\mathbf{x}, \mathbf{y}, \mathbf{t}))$

Temporal target egds
$\forall \mathbf{x}, \mathbf{t}(\theta(\mathbf{x}, \mathbf{t}) \rightarrow x_k = x_l)$



Semantic Adequacy: Let I be a concrete source instance. We say that a concrete target instance J is ***semantically adequate*** for I if the abstract target instance $[[J]]$ is a universal solution for $[[I]]$ w.r.t. $\mathscr{M}^a$.

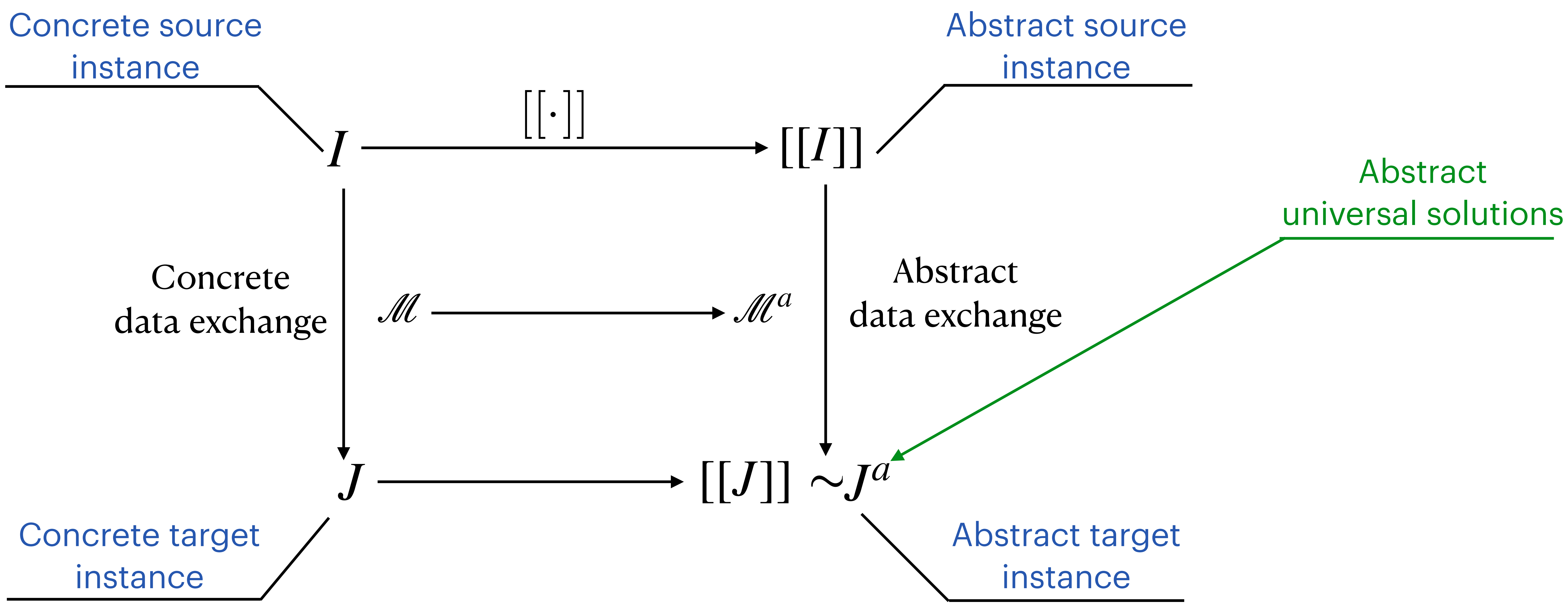
Temporal Data Exchange

Temporal Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$

- Source temporal schema $\mathbf{S}$, Target temporal schema $\mathbf{T}$
- Temporal s-t tgds Σ_{st}, Temporal target egds Σ_t

Temporal s-t tgds
$\forall \mathbf{x}, \mathbf{t}(\phi(\mathbf{x}, \mathbf{t}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}, \mathbf{t}))$

Temporal target egds
$\forall \mathbf{x}, \mathbf{t}(\theta(\mathbf{x}, \mathbf{t}) \rightarrow x_k = x_l)$



Semantic Adequacy: Let I be a concrete source instance. We say that a concrete target instance J is ***semantically adequate*** for I if the abstract target instance $[[J]]$ is a universal solution for $[[I]]$ w.r.t. $\mathscr{M}^a$.

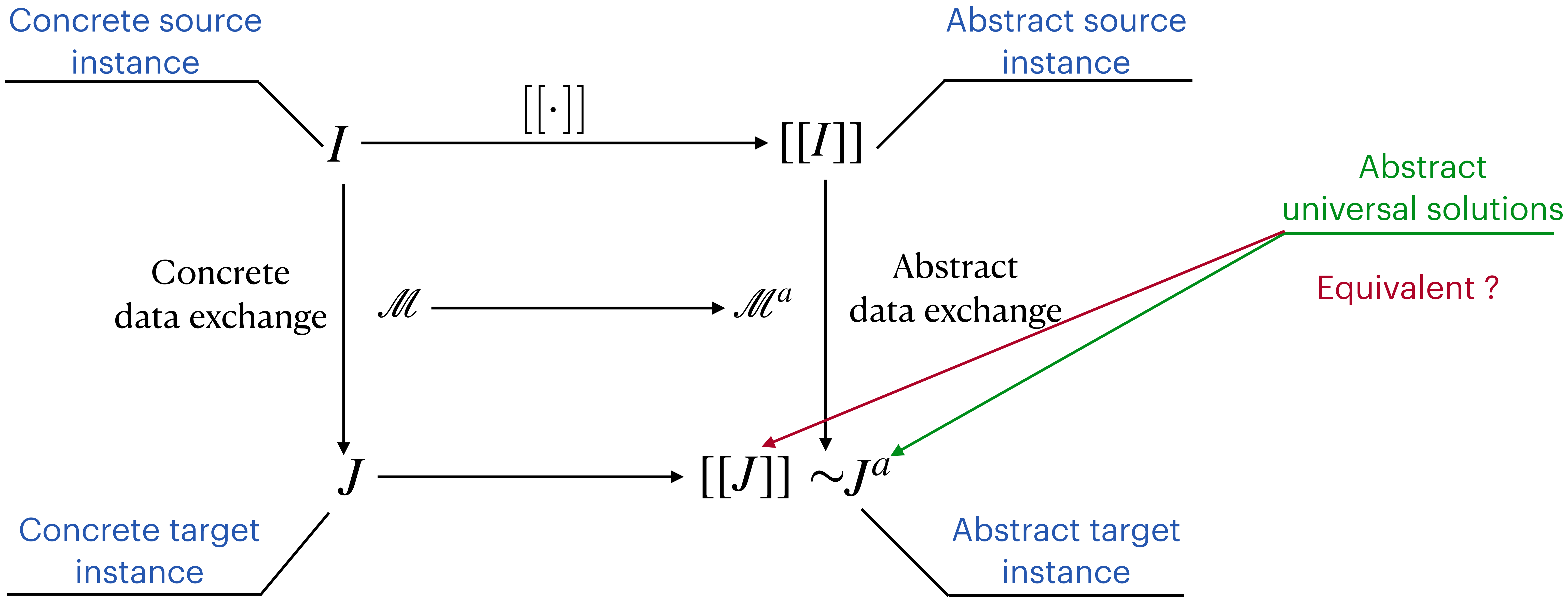
Temporal Data Exchange

Temporal Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$

- Source temporal schema $\mathbf{S}$, Target temporal schema $\mathbf{T}$
- Temporal s-t tgds Σ_{st}, Temporal target egds Σ_t

Temporal s-t tgds
$\forall \mathbf{x}, \mathbf{t}(\phi(\mathbf{x}, \mathbf{t}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}, \mathbf{t}))$

Temporal target egds
$\forall \mathbf{x}, \mathbf{t}(\theta(\mathbf{x}, \mathbf{t}) \rightarrow x_k = x_l)$



Semantic Adequacy: Let I be a concrete source instance. We say that a concrete target instance J is ***semantically adequate*** for I if the abstract target instance $[[J]]$ is a universal solution for $[[I]]$ w.r.t. $\mathscr{M}^a$.

Temporal Data Exchange

Temporal Schema Mapping $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$

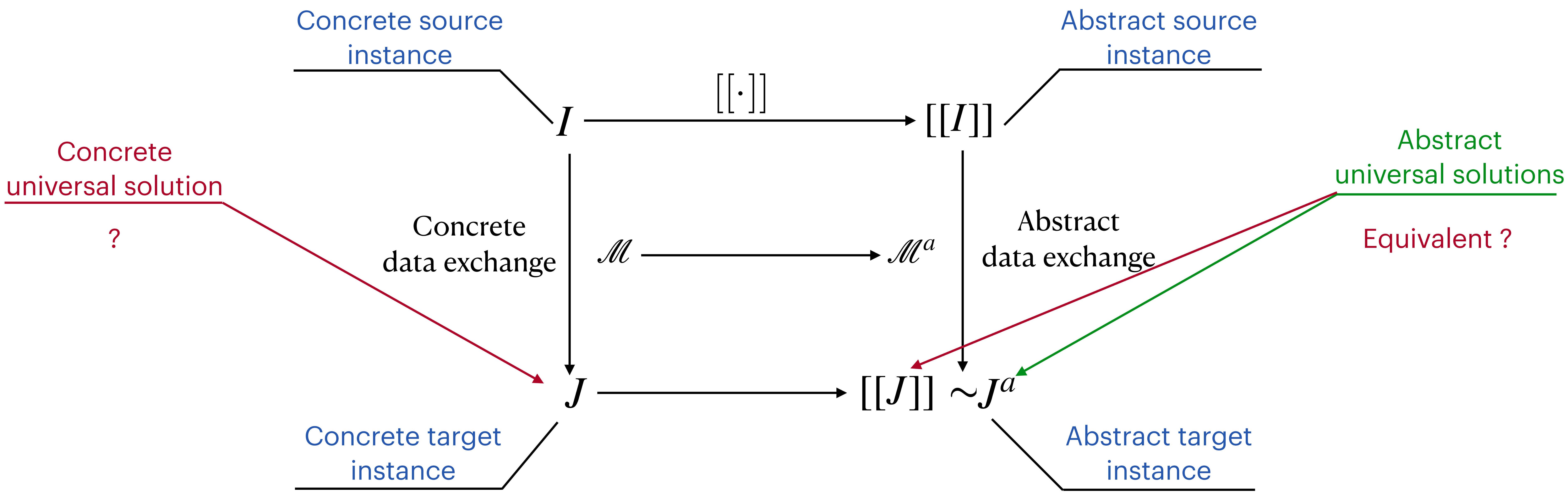
- Source temporal schema $\mathbf{S}$, Target temporal schema $\mathbf{T}$
- Temporal s-t tgds Σ_{st}, Temporal target egds Σ_t

Temporal s-t tgds

$$\forall \mathbf{x}, \mathbf{t}(\phi(\mathbf{x}, \mathbf{t}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}, \mathbf{t}))$$

Temporal target egds

$$\forall \mathbf{x}, \mathbf{t}(\theta(\mathbf{x}, \mathbf{t}) \rightarrow x_k = x_l)$$



Semantic Adequacy: Let I be a concrete source instance. We say that a concrete target instance J is ***semantically adequate*** for I if the abstract target instance $[[J]]$ is a universal solution for $[[I]]$ w.r.t. $\mathscr{M}^a$.

Challenges in Temporal Data Exchange

- In data exchange, universal solutions are produced using the chase algorithm which introduce labelled nulls N_j to witness $\exists \mathbf{y}$ in $\forall \mathbf{x}(\phi(\mathbf{x}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}))$.

- In temporal data exchange, time-stamped nulls N_j^t need to be introduced to witness $\exists \mathbf{y}$ in $\forall \mathbf{x}, t(\phi(\mathbf{x}, t) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}, t))$

- Managing time-stamped nulls requires special care when it comes to target constraints $\forall \mathbf{x}, t(\theta(\mathbf{x}, t) \rightarrow x_k = x_l)$

- This becomes a challenging problem when multiple temporal variables are used in tgds.

Temporal Data Exchange: Single Temporal Variable

Golshanara, L. and Chomicki, J.: Temporal data exchange. Inf. Syst. 87 (2020)

Considered **temporal schema mappings** $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$, in which

- Every relation symbol in $\mathbf{S}$ and $\mathbf{T}$ has one temporal attribute
- Each constraint in $\Sigma_{st} \cup \Sigma_t$ contains exactly one temporal variable
- Hence, the only temporal variable occurs in every atom of the consequent of s-t tgd; no temporal variable is existentially quantified.

Temporal s-t tgd
$\forall \mathbf{x}, t(\phi(\mathbf{x}, t) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}, t))$

Temporal target egd
$\forall \mathbf{x}, t(\theta(\mathbf{x}, t) \rightarrow x_k = x_l)$

Example:

S
E(name, company, time)
S(name, salary, time)

T
Emp(name, company, position, salary, time)

(s-t tgd) $\forall n, c, p, t(E(n, c, t) \wedge P(n, p, t) \rightarrow \exists s\ Emp(n, c, p, s, t))$

(target egd) $\forall n, c, p, s_1, s_2, t(Emp(n, c, p, s_1, t) \wedge Emp(n, c, p, s_2, t) \rightarrow s_1 = s_2)$

Theorem. Let $\mathscr{M}$ be a temporal schema mapping as above. There is a polynomial-time algorithm such that, given a concrete source instance I,

a) if a solution for $[[I]]$ exists, then the algorithm produces a semantically adequate concrete target instance J for I w.r.t. $\mathscr{M}$.

b) if there is no solution for $[[I]]$, then the algorithm fails.

Note: They did not address the question of whether there is always a concrete universal solution that is semantically adequate for I.

Temporal Data Exchange: Single Temporal Variable

Theorem 1. There are

- a temporal schema mapping $\mathcal{M}^* = (\mathbf{S}, \mathbf{T}, \Sigma^*_{st}, \Sigma^*_t)$ with one temporal variable in each constraint in $\Sigma^*_{st} \cup \Sigma^*_t$,

and

- a concrete source instance I^*,

such that

- there is a concrete universal solution for I^* w.r.t. $\mathcal{M}^*$, but no concrete universal solution is semantically adequate for I^*;
- there is a concrete universal solution for $\mathcal{N}(I^*)$ w.r.t. $\mathcal{M}^*$, but no concrete universal solution is semantically adequate for $\mathcal{N}(I^*)$.

  ($\mathcal{N}(I^*)$ is a normalized version of I^*, computed in the first step of Golshanara & Chomicki's algorithm)

Hint of proof.

Schema mapping $\mathcal{M}^*$

$$\text{s-t tgd: } \forall n, s, c, t(E(n,c,t) \wedge S(n,s,t) \rightarrow Emp(n,c,s,t))$$

$$\text{s-t tgd: } \forall n, c, p, t(P(n,p,t) \rightarrow \exists c EmpPos(n,c,p,t))$$

$$\text{target egd: } \forall n, c_1, c_2, s, p, t(\underbrace{Emp(n,c_1,s,t)} \wedge \underbrace{EmpPos(n,c_2,p,t)} \rightarrow c_1 = c_2)$$

Source instance I^*

(a) E

Name	Company	Time
Ada	IBM	[2013, 2018)
Bob	IBM	[2012, 2015)

(b) S

Name	Salary	Time
Ada	18000	[2014, 2018)
Bob	13000	[2013, 2015)

(c) P

Name	Position	Time
Ada	Manager	[2015, 2017)
Bob	Consultant	[2012, 2015)

Temporal Data Exchange: Single Temporal Variable

Theorem 2. Let $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ be a temporal schema mapping such that

(a) each s-t tgd contains at most one temporal variable;

(b) if an s-t tgd contains a temporal variable, then that temporal variable occurs in every atom of its consequent;

(c) each target egd contains at most one temporal atom in its antecedent (it may contain non-temporal atoms).

There is a version of chase algorithm such that, given a concrete source instance I,

- if a solution for $\mathscr{N}(I)$ exists, then the algorithm produces a semantically adequate concrete universal solution for $\mathscr{N}(I)$;
- if the algorithm fails, then there is no solution for $[[I]]$ (note $[[I]] = [[\mathscr{N}(I)]]$).

Temporal Data Exchange: Multiple Temporal Variables

Allen's relations on time intervals: m (meets), o (overlaps), $\prec$(before), $\succ$ (after) , and $=$(equals).

For instance, [2010, 2013) o [2011, 2014)

Concrete s-t tgds $\forall \mathbf{x}, \mathbf{t}(\phi(\mathbf{x}, \mathbf{t}) \wedge \pi(\mathbf{t}) \rightarrow \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}, \mathbf{t}))$

Concrete target egds $\forall \mathbf{x}, \mathbf{t}(\theta(\mathbf{x}, \mathbf{t}) \wedge \rho(\mathbf{t}) \rightarrow x_k = x_l)$

Theorem 3. Let $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ be a concrete schema mapping such that each s-t tgd is full and each target egd contains at most one temporal atom. There is a polynomial-time algorithm such that given a concrete source instance I,

a) if a solution for I exists, then the algorithm produces a semantically adequate concrete universal solution J for I w.r.t. $\mathscr{M}$;

b) if the algorithm fails, then there is no solution for $[[I]]$.

Temporal Data Exchange: Multiple Temporal Variables

Allen's relations on time intervals: m (meets), o (overlaps), $\prec$(before), $\succ$ (after) , and $=$(equals).

For instance, [2010, 2013) o [2011, 2014)

Concrete s-t tgds $\forall \mathbf{x}, \mathbf{t}(\phi(\mathbf{x}, \mathbf{t}) \land \pi(\mathbf{t}) \rightarrow \exists \mathbf{y}\psi(\mathbf{x}, \mathbf{y}, \mathbf{t}))$

Concrete target egds $\forall \mathbf{x}, \mathbf{t}(\theta(\mathbf{x}, \mathbf{t}) \land \rho(\mathbf{t}) \rightarrow x_k = x_l)$

(full s-t tgd) $\forall x_1, x_2, x_3, x_4, t_1, t_2(R_2(x_1, x_2, x_3, t_1) \land R_3(x_1, x_4, t_2) \land (t_2 \text{ m } t_1) \rightarrow T_2(x_1, x_3, t_2))$

Theorem 3. Let $\mathscr{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ be a concrete schema mapping such that each s-t tgd is full and each target egd contains at most one temporal atom. There is a polynomial-time algorithm such that given a concrete source instance I,

a) if a solution for I exists, then the algorithm produces a semantically adequate concrete universal solution J for I w.r.t. $\mathscr{M}$;

b) if the algorithm fails, then there is no solution for $[[I]]$.

Concluding Remarks

- We showed that **semantically adequate universal solutions** may **not exist** even for temporal schema mappings with single temporal variable.

- We identified classes of schema mappings with **single temporal atom** in each target egd, for which semantically adequate universal solutions exist.

- We expanded the original framework of temporal data exchange studied by Golshanara and Chomicki via considering temporal schema mappings **with multiple temporal variables**.

- Important Open Problem: Study temporal schema mappings with existentially quantified temporal variables.

Thank you for your attention