WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

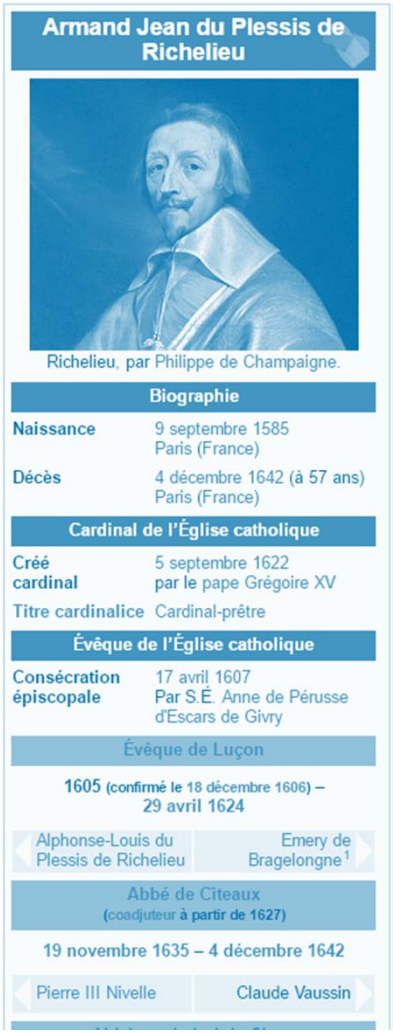
Updating Wikipedia via DBpedia Mappings and SPARQL

Vadim.Savenkov@wu.ac.at

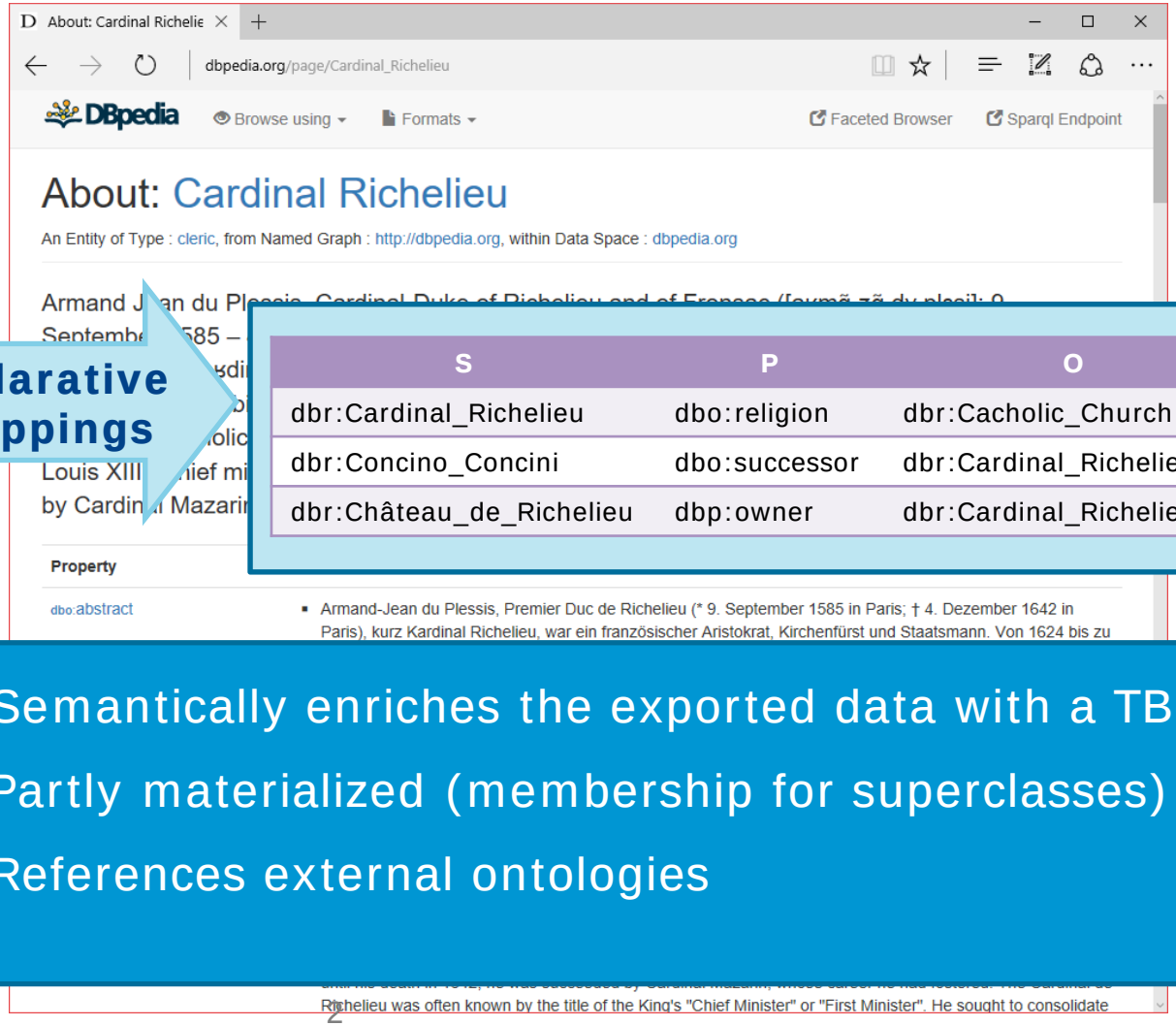
joint work with **Albin Ahmeti**, **Javier Fernández** and **Axel Polleres**

European Semantic Web Conference @ Portoroz, Slovenia, Mai 30, 2017

Wikipedia as a Data Service



declarative mappings



S	P	O
dbr:Cardinal_Richelieu	dbo:religion	dbr:Cacholic_Church
dbr:Concino_Concini	dbo:successor	dbr:Cardinal_Richelieu
dbr:Château_de_Richelieu	dbp:owner	dbr:Cardinal_Richelieu

- ✓ Semantically enriches the exported data with a TBox
- ✓ Partly materialized (membership for superclasses)
- ✓ References external ontologies

Motivation for batch updates

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS


```
DELETE {
 ?X dbo:region dbr:Upper_Normandy .
 ?Y dbo:region dbr:Lower_Normandy .}
INSERT {
 ?X dbo:region dbr:Normandy .
 ?Y dbo:region dbr:Normandy}
WHERE {
 {?X dbo:region dbr:Upper_Normandy}
 UNION
 {?Y dbo:region dbr:Lower_Normandy}
}
```

Ontology Based Data Management (OBDM)

Ontology as a unified interface for querying and updating the data in external sources.

Updates+Entailment: First Steps

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Without mappings: updates can be unambiguously resolved for some simple ontology languages.

Example: RDFS¬ = RDFS core + Class disjointness

[ESWC 2016: Ahmeti, Calvanese, Polleres, S.]

- **Subclass** $A \sqsubseteq B$, **subproperty** $p \sqsubseteq q$
- **Domain and range restrictions**

  domain of p is restricted to A: $\exists p \sqsubseteq A$

  range of p is restricted to A: $\exists p^{-1} \sqsubseteq A$
- **Class disjointness** $A \sqsubseteq \neg B$

DBpedia TBox features

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

RDFS¬ +

- Inverse properties $p \sqsubseteq r^-$
- Property disjointness $p \sqsubseteq \neg r$
- Functionality (**data properties only**)
- DifferentFrom $c_1 \neq c_2$
- sameAs $c_1 = c_2$ (to link **external ontologies**)
- redirectsTo: equality with the **canonical representative** of each equivalence class

French Catholic Clergy mapping

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

```
Mapping fr:Infobox Prélat catholique{{ConditionalMapping
| cases =
 {{Condition  | templateProperty = titre  | operator = equals  | value = Pape
          | mapping = {{TemplateMapping | mapToClass = Pope | mappings =
             {{IntermediateNodeMapping  |  nodeClass = PersonFunction  | correspondingProperty = occupation
             | mappings =
               {{ConstantMapping | ontologyProperty = title | value = Pape }}
               {{PropertyMapping | templateProperty = début pontificat | ontologyProperty = activeYearsStartYear }}
               {{PropertyMapping | templateProperty = fin pontificat | ontologyProperty = activeYearsEndYear }}
               ...
               {{PropertyMapping | templateProperty = prédécesseur pape | ontologyProperty = predecessor }}
               {{PropertyMapping | templateProperty = successeur pape | ontologyProperty = successor }}
             }}
           }}
 }}
 {{Condition | templateProperty = titre | operator = equals  | value = Évêque
          | mapping = {{TemplateMapping | mapToClass = ChristianBishop }}
 }}
 ...
{{Condition | operator = otherwise | mapping = {{TemplateMapping |mapToClass = Cleric }} }}

| defaultMappings =
  {{PropertyMapping | templateProperty = nom | ontologyProperty = foaf:name }}
  ...
  {{PropertyMapping | templateProperty = nom naissance   | ontologyProperty = birthName }}
  {{PropertyMapping | templateProperty = devise | ontologyProperty = motto }}
  ...
  {{PropertyMapping | templateProperty = bénédiction abbatiale | ontologyProperty = abbeychurchBlessing  }}
}}
```

TemplateMapping and PropertyMapping

Infobox type

```
Mapping fr:Infobox Prélat catholique{{ConditionalMapping
| cases =
 {{Condition  | templateProperty = titre  | operator = equals  | value = Pape
    ...
 }}
 {{Condition | templateProperty = titre | operator = equals  | value = Évêque
        | mapping = {{TemplateMapping | mapToClass = ChristianBishop }}
}}
{{Condition | templateProperty = titre | operator = equals  | value = Cardinal
        | mapping = {{TemplateMapping | mapToClass = Cardinal }}
}}
...
{{Condition | operator = otherwise
        | mapping = {{TemplateMapping |mapToClass = Cleric }}
}}
| defaultMappings =
 {{PropertyMapping | templateProperty = nom | ontologyProperty = foaf:name }}
 ...
 {{PropertyMapping | templateProperty = devise | ontologyProperty = motto }}
 ...
}}
```

<Page URI> a dbr:Pope.

<Page URI> a dbr:ChristianBishop.

<Page URI> a dbr:Cardinal.

<Page URI> a dbr:Cleric.

PropertyMapping Example

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

```
Mapping fr:Infobox Prélat catholique{{ConditionalMapping
| cases =
  ...
 {{Condition | templateProperty = titre | operator = equals  | value = Évêque
         | mapping = {{TemplateMapping | mapToClass = ChristianBishop }}
 }}
 ...

{{Condition | operator = otherwise
         | mapping = {{TemplateMapping |mapToClass = Cleric }}
 }}

| defaultMappings =
  {{PropertyMapping | templateProperty = nom | ontologyProperty = foaf:name }}
  ...
    {{PropertyMapping | templateProperty = devise | ontologyProperty = motto }}
  ...
}}
```

<Page URI> foaf:name <nom>.

<Page URI> dbo:motto <devise>.

IntermediateNodeMapping

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

<Page URI> dbo:motto <devise>.

```
{{PropertyMapping
  | templateProperty = devise | ontologyProperty = motto }}

{{IntermediateNodeMapping
  | nodeClass = PersonFunction
  | correspondingProperty = occupation
  | mappings =
    {{PropertyMapping
       | templateProperty = ministère 1 | ontologyProperty = title }}

       [...]
    {{PropertyMapping
       | templateProperty = successeur 1 | ontologyProperty = successor
    }}
}}
```

<Fresh IRI> a dbr:PersonFunction.

<Page URI> dbo:occupation <Fresh IRI>.

<Fresh IRI> dbo:title < ministère 1>.

<Fresh IRI> dbo:successor <successeur 1>.

Tuple Generating Dependencies (TGDs)

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Basic (full) tgds
  - $\forall \vec{x}\left(\psi(\vec{x}) \rightarrow CQ(\vec{x})\right)$
  - $\forall x_1 \forall x_2 \left(Cleric(x_1) \wedge WorksFor(x_1, x_2) \rightarrow Church(x_2)\right)$
- $CQ(\vec{x})$ is a conjunctive query with free variables $\vec{x}$.
- ψ is a FO formula (often also just a CQ).
- Enforcing (aka “chasing“) tgds:
  - For each assignment $\vec{a}$ for $\vec{x}$ satisfying ψ on the source instance, yield the facts instantiating $CQ(\vec{a})$.

The Infobox Schema W

- W_i (Page_**U**RI, Infobox_**T**ype, Infobox_**I**D)
- W_d (Infobox_**I**D, **P**roperty_Name, Property_**V**alue)
- Infobox constraints:
  - UT → I
  - IP → V
  - $W_d.I \subseteq W_i.I$

Mapping rules amenable to full TGDs

Mapping fr:Infobox **Prélat catholique**{{ConditionalMapping

$$\forall U \forall I \left(W_i(U, \text{"Prélat catholique"}, I) \wedge W_d(I, \text{"Èvêque"}, X) \rightarrow dbr\text{:}\, ChristianBishop(U) \right)$$

{{Condition | templateProperty = titre | operator = equals | value = Évêque
| mapping = {{TemplateMapping | mapToClass = ChristianBishop }}
}}

$$\forall U \forall I \left(W_i(U, \text{"Prélat catholique"}, I) \wedge W_d(I, \text{"Cardinal"}, X) \rightarrow dbr\text{:}\, Cardinal(U) \right)$$

{{Condition | templateProperty = titre | operator = equals | value = Cardinal
| mapping = {{TemplateMapping | mapToClass = Cardinal }}
}}

...

{{Condition | operator = otherwise
| mapping = {{TemplateMapping |mapToClass = Cleric }}

$$\forall U \forall I \forall X \left(\begin{array}{l} W_i(U, \text{Prélat catholique}, I) \wedge \neg W_d(I, \text{"}Pape\text{"}, X) \\ \wedge \neg W_d(I, \text{"Èvêque"}, X) \wedge \neg W_d(I, \text{"Cardinal"}, X) \wedge \ldots \rightarrow dbr\text{:}\, Cleric(U) \end{array} \right)$$

PropertyMapping Example

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Mapping fr:Infobox **Prélat catholique**{{ConditionalMapping
| cases =
...

{{Condition | operator = otherwise
| mapping = {{TemplateMapping |mapToClass = Cleric }}
}}

| defaultMappings =
{{**PropertyMapping** | templateProperty = nom | ontologyProperty = foaf:name }}
...

$$\forall U \forall I \left(W_i(U, \text{"Prélat catholique"}, I) \wedge W_d(I, \text{"nom"}, X) \rightarrow foaf{:}name(U, X) \right)$$

{{**PropertyMapping** | templateProperty = devise | ontologyProperty = motto }}

$$\forall U \forall I \left(W_i(U, \text{"Prélat catholique"}, I) \wedge W_d(I, \text{"devise"}, X) \rightarrow dbp{:}motto(U, X) \right)$$

}}

TGDs with existentially quant. variables

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- $\forall \vec{x}\left(\psi(\vec{x}) \rightarrow \exists \vec{y}\, CQ(\vec{x}, \vec{y})\right)$
- $\forall x\left(Cleric(x) \rightarrow \exists z\, PraysFor(x, z)\right)$
- Chase with non-full TGDs:
  - For each assignment $\vec{a}$ for $\vec{x}$ satisfying ψ, construct an assignment $\vec{b}$ of **fresh distinct values** for $\vec{y}$ (often special **labeled null** values);
  - Yield the atoms in $CQ(\vec{a}, \vec{b})$

IntermediateNodeMapping via (non full) TGDs

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

$$\forall U \forall I \big(W_i(U, \text{"Prélat catholique"}, I) \wedge W_d(I, \text{"ministère 1"}, X) \rightarrow \exists Z \ \ dbo\colon title(Z, X) \big)$$

$$\forall U \forall I \big(W_i(U, \text{"Prélat catholique"}, I) \wedge W_d(I, \text{"prédécesseur1"}, X) \rightarrow \exists Z \ \ dbo\colon predecessor(Z, X) \big)$$

Chase does not enforce that the same value instantiates Z in both cases.

Nested TGDs

- $\forall \overrightarrow{x_1} \left(\psi_1(\overrightarrow{x_1}) \rightarrow \exists \overrightarrow{y_1} \left(CQ_1(\overrightarrow{x_1}, \overrightarrow{y_1}) \wedge \forall \overrightarrow{x_2} \left(\begin{array}{l} \psi_2(\overrightarrow{x_1}, \overrightarrow{x_2}) \rightarrow \\ \exists \overrightarrow{y_2} \, (CQ_2(\overrightarrow{x_1}, \overrightarrow{x_2}, \overrightarrow{y_1}, \overrightarrow{y_2}) \wedge \cdots) \end{array} \right) \right) \right)$

- $\forall x_1 \left(Pastor(x_1) \rightarrow \exists y \left(PraysFor(x_1, y) \wedge \forall x_2 \left(ParishMember(x_2, x_1) \rightarrow PraysFor(x_2, y) \right) \right) \right)$

- Chase with nested TGDs vs. standard TGDs:
  - Preconditions of nested implications can access the values bound in the preconditions of ancestor implications: e.g., $ParishMember(x_2, x_1)$
  - Fresh values generated in the ancestor implications can be reused in the conclusions of nested implications, e.g.: $PraysFor(x_2, y)$

IntermediateNodeMapping via Nested TGDs

{{PropertyMapping
| templateProperty = devise | ontologyProperty = motto}}

{{**IntermediateNodeMapping**
| nodeClass = PersonFunction
| correspondingProperty = occupation
| mappings =

$$\forall U \forall I \left(W_i(U, \text{"Prélat catholique"}, I) \rightarrow \exists Z \left(dbr{:}\, PersonFunction(Z) \wedge dbo{:}\, occupation(U, Z) \wedge \left(W_d(I, \text{"ministère 1"}, X) \rightarrow dbo{:}\, title(Z, X) \right) \right. \right.$$

$$\ldots\ldots\ldots$$

$$\left. \left. \wedge \left(W_d(I, \text{"prédécesseur 1"}, X) \rightarrow dbo{:}\, predecessor(Z, X) \right) \wedge \left(W_d(I, \text{"successeur 1"}, X) \rightarrow dbo{:}\, successor(Z, X) \right) \right) \right)$$

}}

DBpedia Mappings: syntactic restrictions

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Only unary interpreted predicates in conditions
  (e.g., “if name = full_name” not possible)
- NO nested conditions
  - At most two atoms on the left-hand side of implications
    (test condition + atomic query of a property value)
- IntermediateNodeMappings are not nested currently.
- Fixed structure of generated triples:
  - Subject: page URI or an intermediate node
  - Predicate: defined by the property mapping
  - Object: infobox property value

Dbpedia mappings: Expressive features

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Limited forms of negation:
  - only the first satisfied condition triggers,
  - the final „otherwise“ branch
- Generation of new IRIs (existentially quantified variables in the encoding via tgds)
- Special functions:
  (“contains”, “exists”, “not exists” in conditions), string & geocoordinate manipulation, arithmetic operations for property generation

DBpedia update translation

Semantics of updates

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Input: Dbpedia update $u = (u^-, u^+)$
- u^- atoms to be deleted
- u^+ atoms to be inserted
- Semantics: a set of possible Wiki updates (e_i^-, e_i^+) that

Consistency-oblivious semantics

- Comply with the Infobox schema
- Minimal (e.g., w.r.t. the subset relation)
- Together with the mappings and the TBox, derive a DBpedia instance which contain u^+ and does not contain u^-

Consistency-aware semantics

- Updated instance is consistent with the TBox

Existence of solutions (consistency-oblivious)

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

For a fixed mapping:
Given an insert u^+, is its consistency-oblivious semantics empty?

Standard NP completeness proof (in DB context Abiteboul&Duschka 98)

Underlying data source (Database): **colored** graph.

View (ontology): non-colored graph, colors, and **invalid coloring marker** set to false: require a valid coloring

$$(\exists C_1 \exists C_2 \exists A\ \exists B$$
$$Edge(A,B) \wedge Color(A,C_1) \wedge Color(B,C_2) \wedge \boldsymbol{C_1 = C_2}) \rightarrow Invalid$$

No binary predicates (e.g. $\boldsymbol{C_1 = C_2}$) in the conditions of DBpedia mappings!

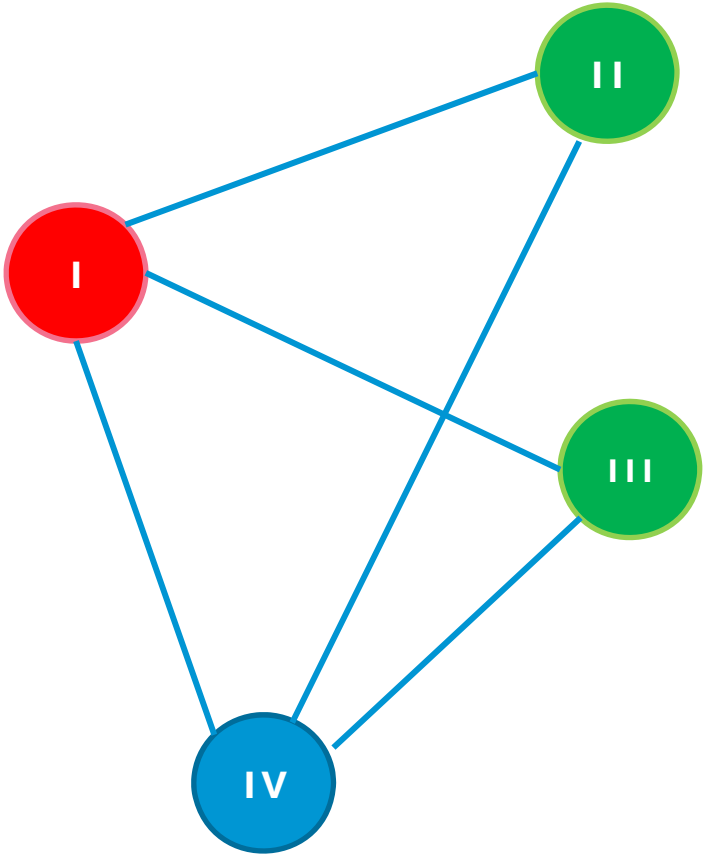
Source consistency of the ABox: Reduction from 3COL

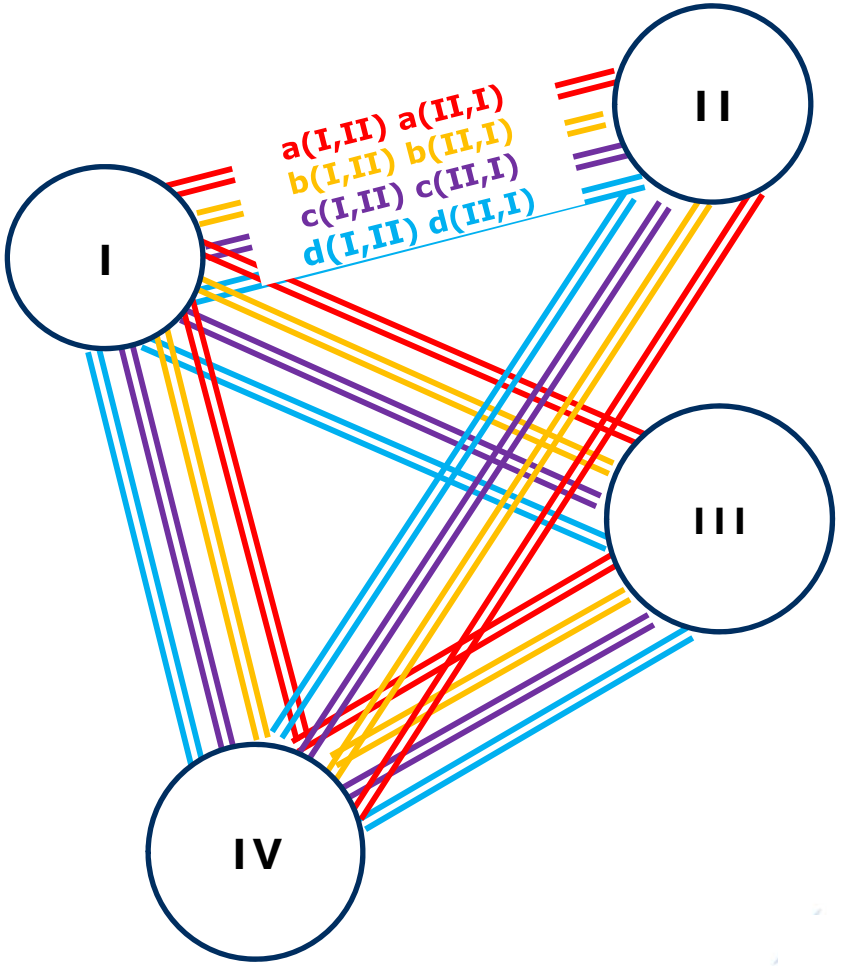
- Graph of degree ≤ 4 (intractable 3COL)
- Proper ABox shape: IRIs connected by symmetric labeled properties **a, b, c, d**

  $\Rightarrow$ 8 property assertions to represent a single graph edge in the materialized ABox.

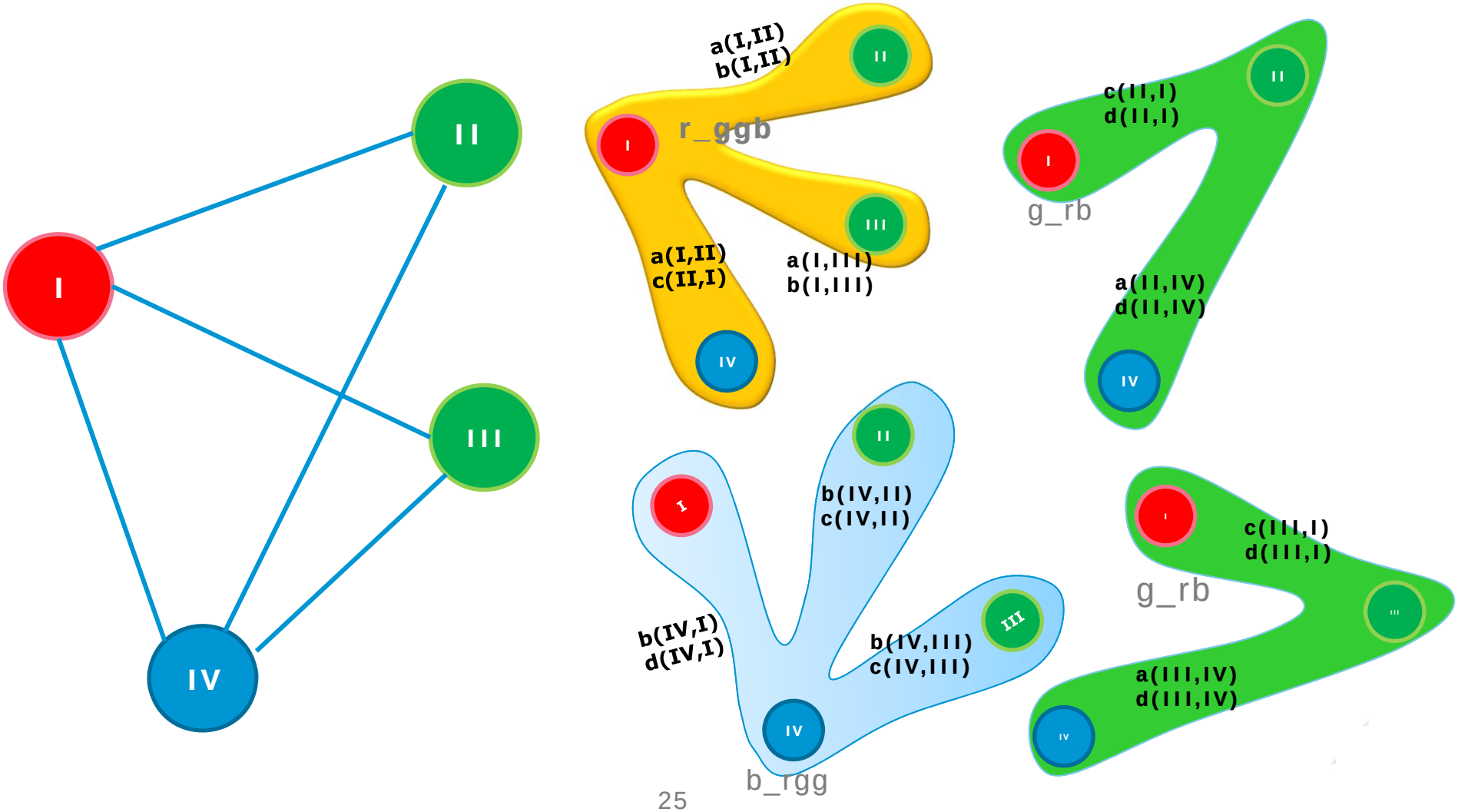
- TBox: $\mathbf{a \sqsubseteq a^-}$, $\mathbf{b \sqsubseteq b^-}$, $\mathbf{c \sqsubseteq c^-}$, $\mathbf{d \sqsubseteq d^-}$

Colored graph and its representation as a Dbpeida ontology





Infoboxes: 1-Neighborhoods of nodes

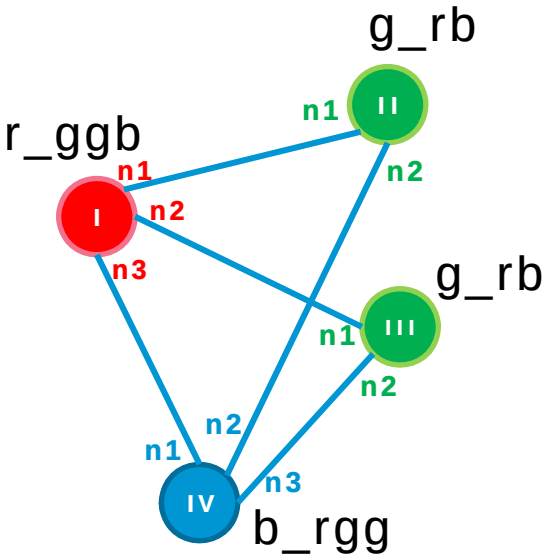


Mapping conditions: ensure consistent color assignment across Infoboxes

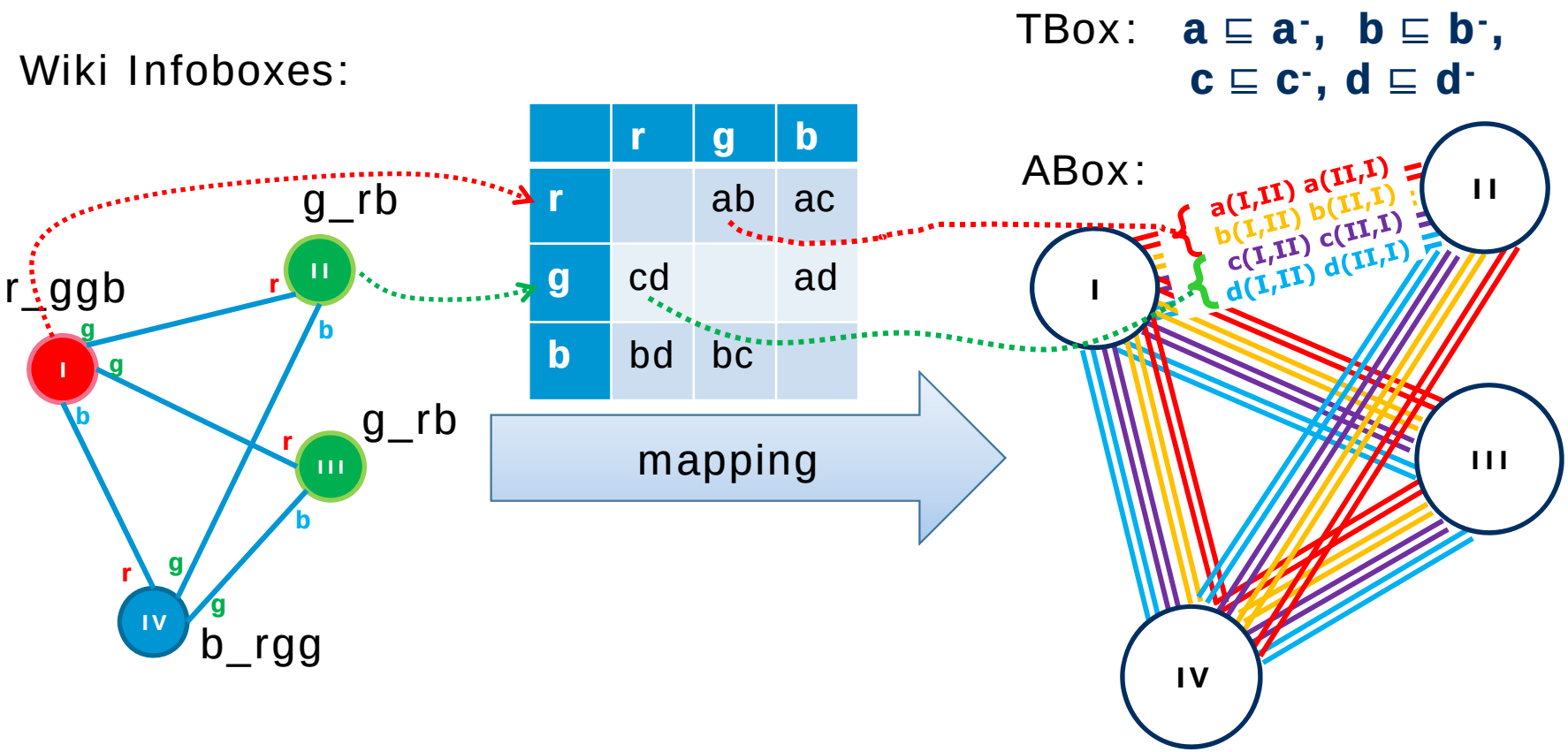
if neighbourhood type = „r_ggb“
<this> dbo:a <neighbor n1>.
<this> dbo:b <neighbor n1>.
<this> dbo:a <neighbor n2>.
<this> dbo:b <neighbor n2>.
<this> dbo:a <neighbor n3>.
<this> dbo:c <neighbor n3>.

if neighbourhood type = „b_rgg “
<this> dbo:b <neighbor n1>.
<this> dbo:d <neighbor n1>.
<this> dbo:b <neighbor n2>.
<this> dbo:c <neighbor n2>.
<this> dbo:b <neighbor n3>.
<this> dbo:c <neighbor n3>.

	r	g	b
r		ab	ac
g	cd		ad
b	bd	bc	



3COL of max. degree 4 graphs via Source consistency of the ABox



3COL iff consistent coverage by 1-neighborhoods (Infoboxes) exist
iff ABox has regular shape.

Worst-case vs. actual mappings

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- DBpedia and many Linked Open Data sources [Glimm,Hogan,Krötzsch,Polleres LDOW'12] is economic in terms of the TBox features used:
  - Especially with respect to inconsistency derivation:
  - Few disjoint classes / properties.
  - Functionality only for data properties.

- Computational intractability is less of a practical issue than infeasibility for verification by curators (Wikipedia maintainers)!

Our Approach

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Consider each subject (Wiki page) separately
- Find the most likely translation of properties based on evidence, applying update resolution policies
- Construct sound re-usable **patterns of update translation**
- Use patterns to implement batch updates

Update resolution policies

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

```
{{PropertyMapping | templateProperty = state1 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state2 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state3 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state4 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state5 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state6 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state7 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state8 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state9 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district1 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district2 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district3 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district4 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district5 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district6 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district7 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district8 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district9 | ontologyProperty = region }}
```

Update resolution policies

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Infobox frequency first

- Prefer most frequently used properties
- Based on precomputed statistics across all infoboxes

```
{{PropertyMapping | templateProperty = state7 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state8 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = state9 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district1 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district2 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district3 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district4 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district5 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district6 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district7 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district8 | ontologyProperty = region }}
{{PropertyMapping | templateProperty = district9 | ontologyProperty = region }}
```

Update resolution policies

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Infobox frequency first

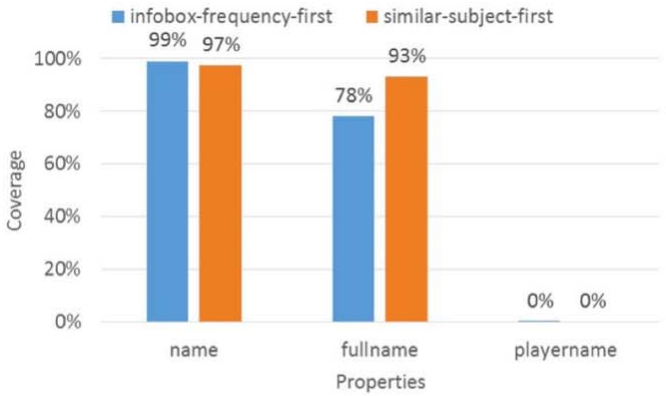
- Prefer most frequently used properties
- Based on precomputed statistics across all infoboxes

{{PropertyMapping | templateProperty = state7 | ontologyProperty = region }}

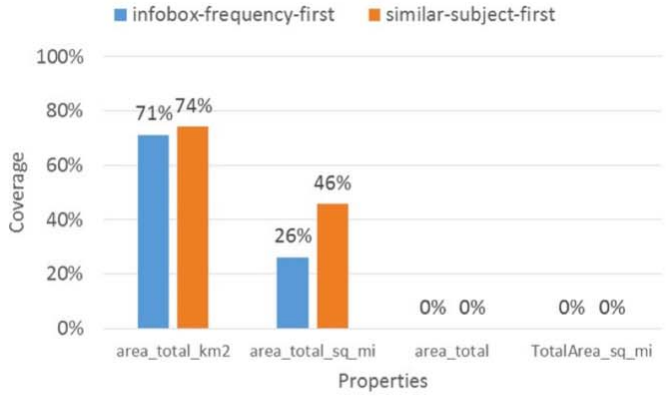
Similar subject first

- Taking only similar subjects into account: same infobox type and having same properties as being insterted.
- “How are other football players are represented in the infobox?”

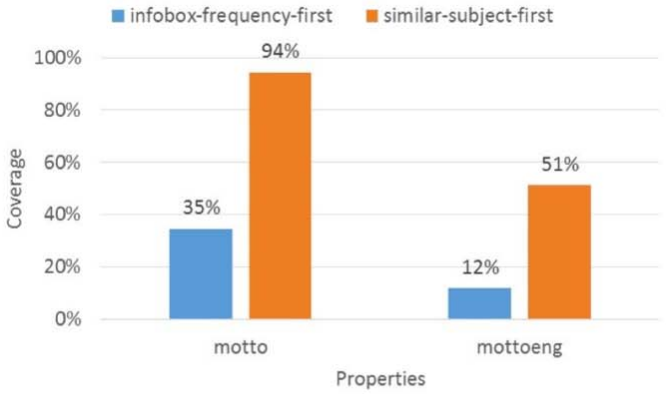
Examples of resolution policies



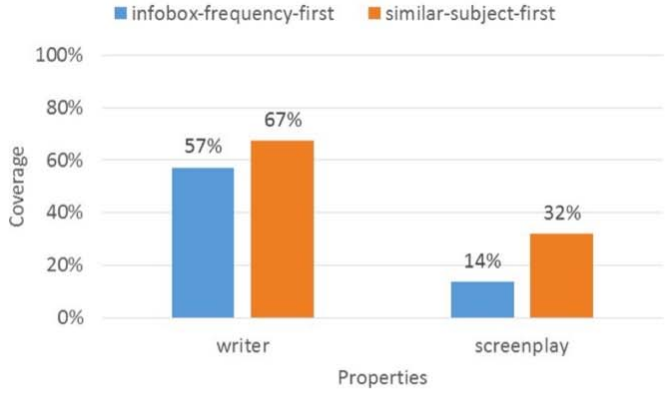
(a) Infobox Football biography



(b) Infobox Settlement



(c) Infobox Universities



(d) Infobox Film

Update resolution patterns

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- GOAL: input update (u^-, u^+) for a single subject
- CONTEXT: relevant pre-existent facts that need to be removed to avoid clashes
- ACTION: single wiki update (e^-, e^+)

IRIs and data values replaced by variables, to enabe reuse across Wiki pages in bulk updates.

Summary and conclusions

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- DBPedia mappings: fixed-depth nested tgds with negation and unary interpreted predicates in preconditions.
- Existence of solutions (DBpedia update feasibility) is intractable even not taking consistency w.r.t. TBox into account.
- Evidence-based update resolution policies.
- Introduce update resolution patterns as a means to formulate **reusable examples of update translation** in OBDM.