SPARQL and relational databases

Richard Cyganiak

Agenda

- ~~D2RQ~~
- Relational algebra for SPARQL
- Translation to SQL
- sparql2sql
- Querying *virtual graphs*

SPARQL example

```
SELECT ?name, ?pic
WHERE {
    ?person rdf:type foaf:Person .
    ?person foaf:name ?name .
    OPTIONAL {
       ?person foaf:depiction ?pic
    }
    FILTER REGEX(?name, “^C”)
}
```

Relational algebra

“RDF Relations”

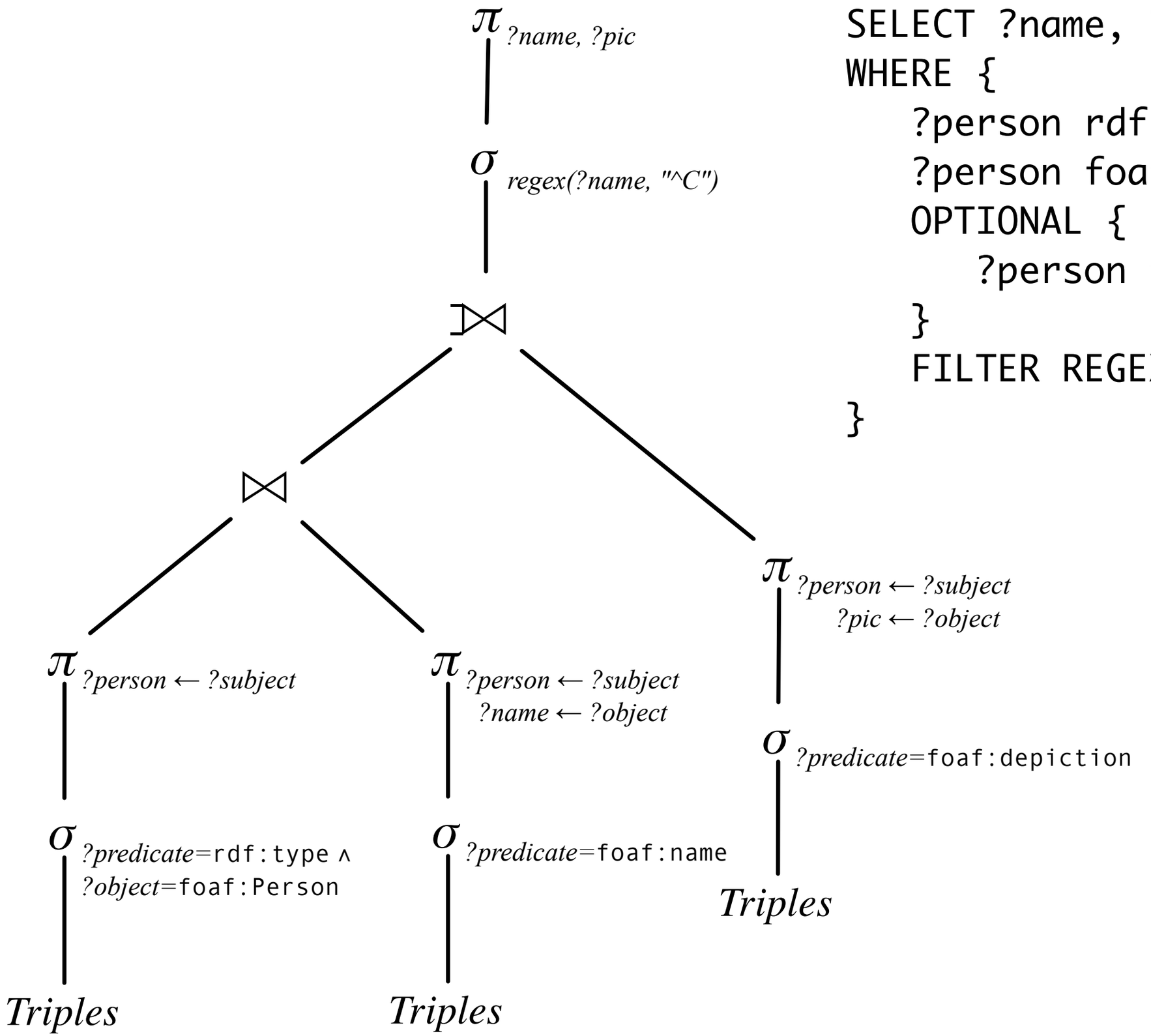
?subject	?predicate	?object
:s1	:p1	“1”^^xsd:int
_:s2	:p2	:o2

?name	?pic
“Alice”	
“Bob”	<http://example.com/bob.png>

- Tables of RDF nodes
- Example: *graph relation* (?s, ?p, ?o)
- Example: SPARQL or RDQL query result
- Operators on RDF Relations: σ, π, ⋈, ⟕, ∪

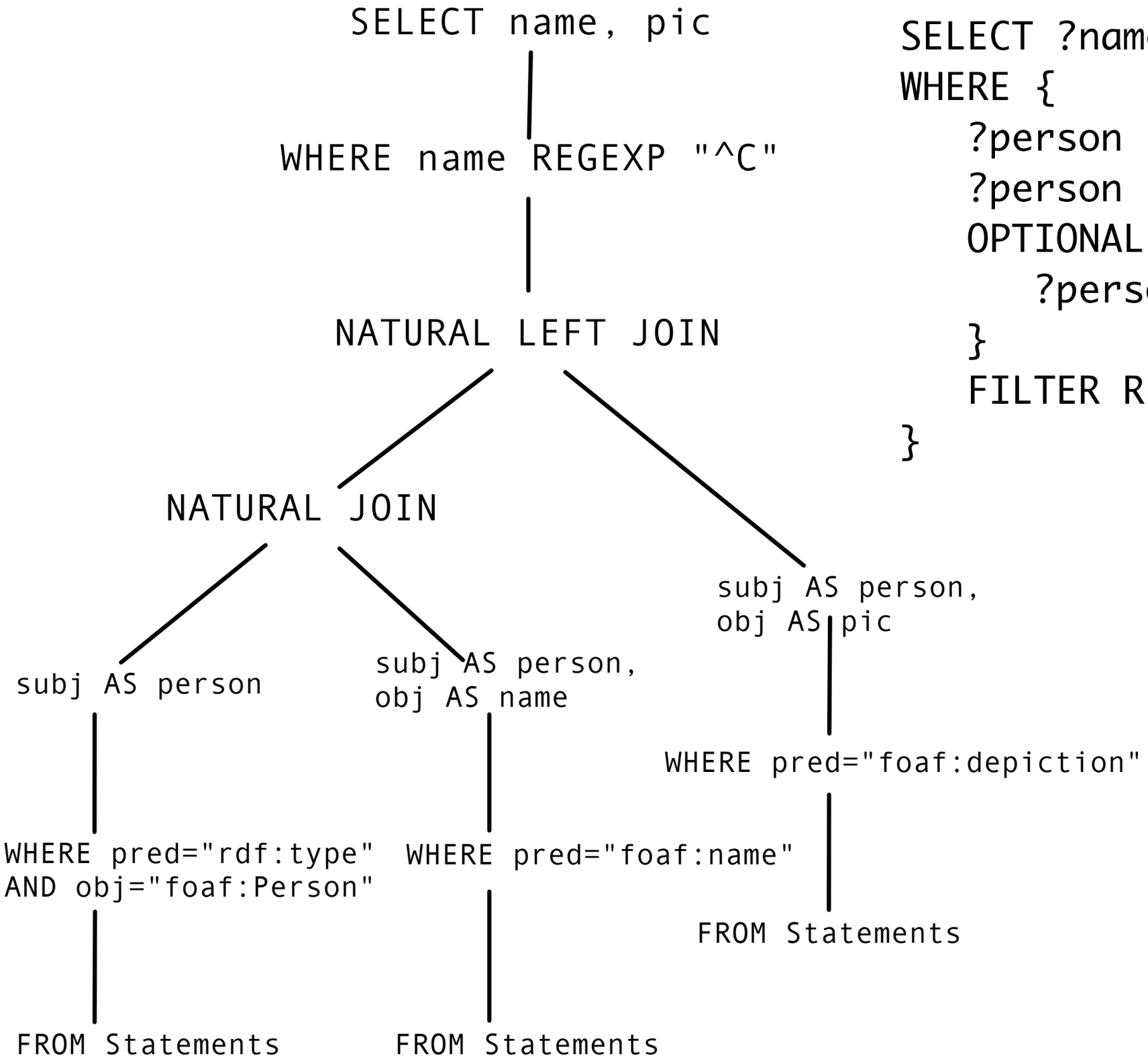
Relational operators

- $\pi_{?a,?b}(R)$: Restrict R to columns ?a and ?b
- $R_1 \bowtie R_2$: Join rows from R_1 and R_2 where shared variables match
- $R_1 \cup R_2$: All rows from R_1 and all from R_2
- $\sigma_p(R)$: find rows from R matching expression p




```
SELECT ?name, ?pic
WHERE {
   ?person rdf:type foaf:Person .
   ?person foaf:name ?name .
   OPTIONAL {
      ?person foaf:depiction ?pic
   }
   FILTER REGEX(?name, “^C”)
}
```

Benefits: Expression transformation, query planning




```
SELECT ?name, ?pic
WHERE {
   ?person rdf:type foaf:Person .
   ?person foaf:name ?name .
   OPTIONAL {
      ?person foaf:depiction ?pic
   }
   FILTER REGEX(?name, “^C”)
}
```

Details in upcoming tech report

Nesting SQL queries

Options:

- ... FROM (SELECT ...) JOIN (SELECT ...)
- ... FROM A JOIN (B LEFT JOIN C)
- Temporary tables

DB layout

Triples		
subj_h	pred_h	obj_h

Quads			
graph_h	subj_h	pred_h	obj_h

Graphs
graph_h

Nodes						
hash	lexical	datatype	lang	int	float	datetime

- Steve Harris: *SPARQL query processing with conventional relational database systems*
- Richard Cyganiak: *Note on database layouts for SPARQL datastores* (coming soon)

sparql2sql

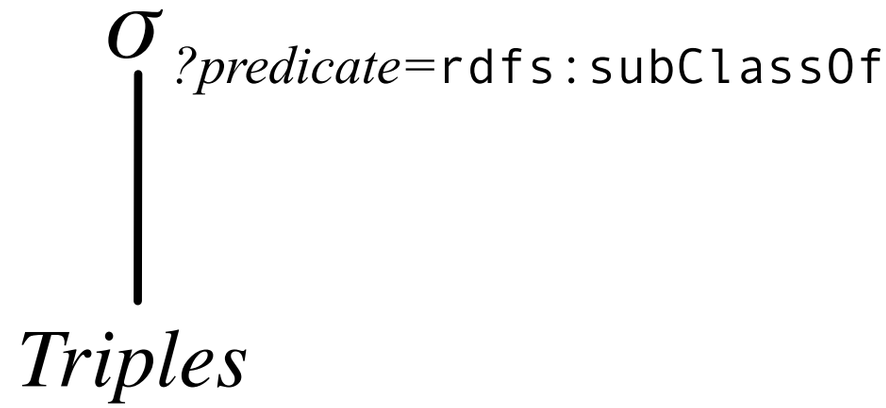
- Uses ARQ parser and API
- DB schema: ModelRDB (bad for SPARQL)
- Does most of optional, union, graph
- Doesn't do value testing in DB
- Used in gnowsis (plus MySQL fulltext search)
- http://jena.sourceforge.net/sparql2sql

Lessons learnt

- MySQL prior to V5 makes job harder – better start with good DB and scale down
- Experimentation and performance testing with RDF DBs is difficult
- # of joins can become an issue (32 join limit in MySQL 3; slow query planner in Postgres)

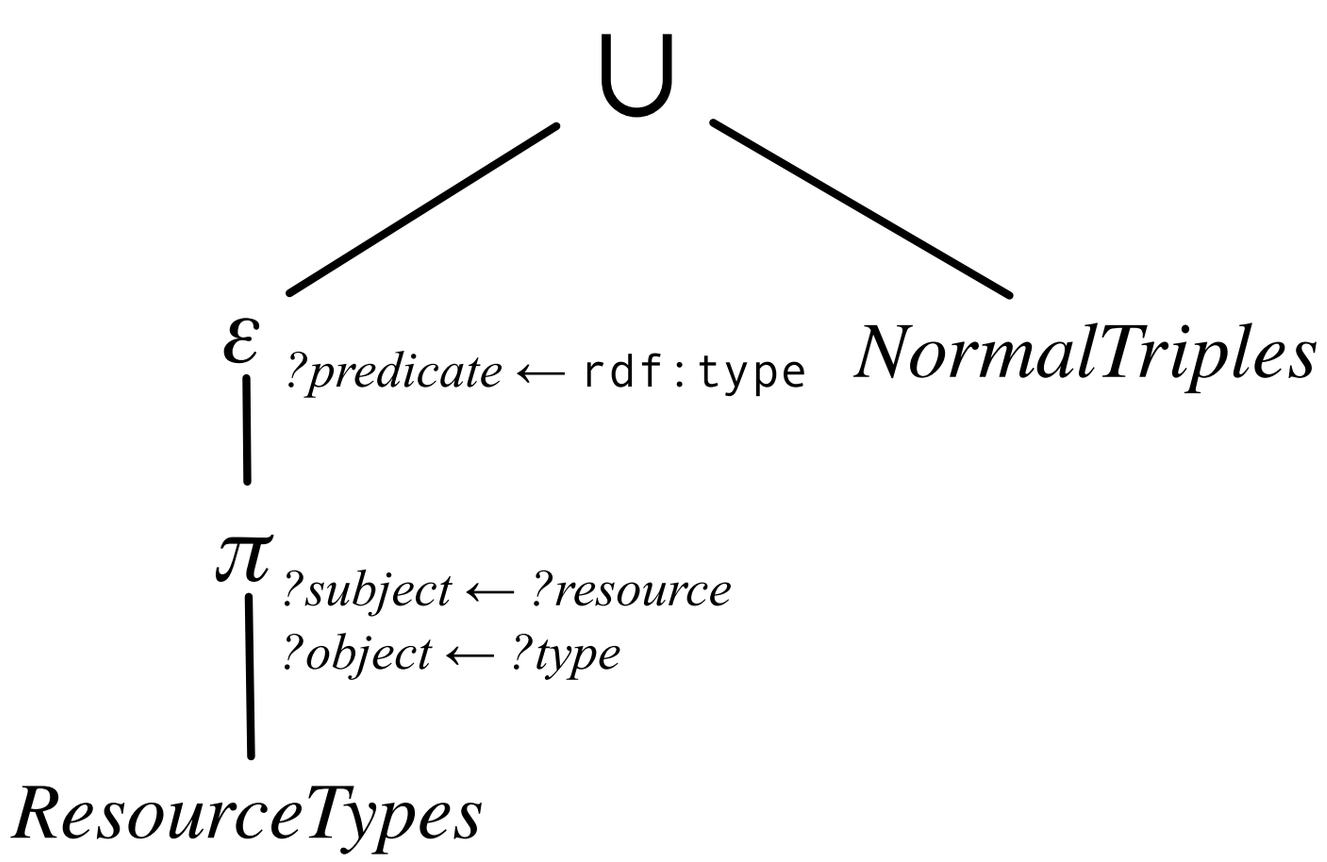
Querying *virtual graphs*

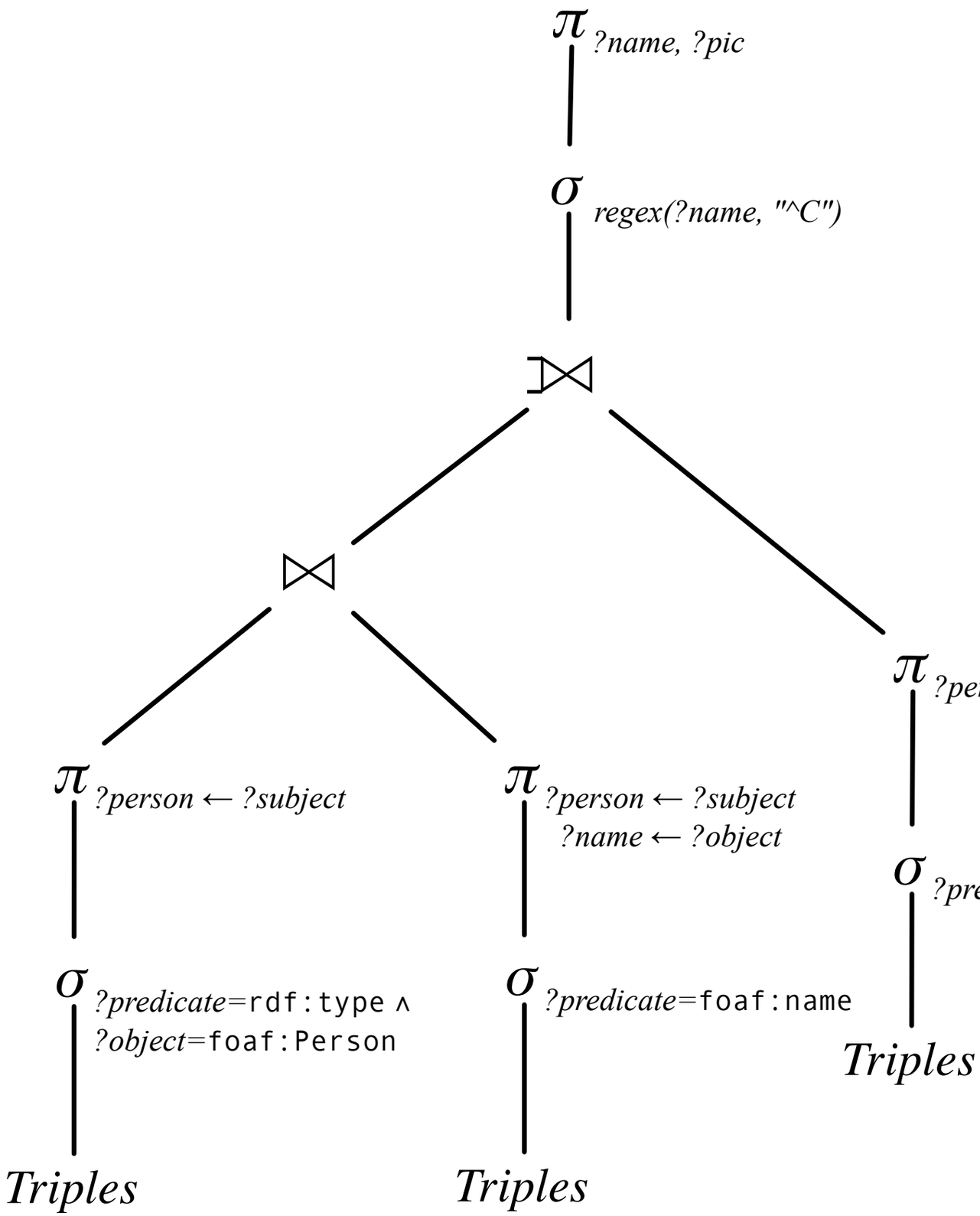
Virtual graph: Triples don't exist physically, but are generated from a data source through relational operations



Component graph: Special kind of virtual graph; union of several "kinds of triples"

Example: SPARQL CONSTRUCT result





Querying a virtual graph:

- Substitute graph definition for *Triples*
- Simplify tree by expression transformation
- Execute simplified tree

What I will do next

- Back to university in Berlin
- Thesis in the SPARQL (+DB?) area
- Work part time in Chris Bizer's group (Trust, Named Graphs, RDF with PHP)

Thank you!