

Multi-centric studies in computational neurosciences

Johan Montagnat

CNRS, I3S lab, Modalis team
on behalf of the NeuroLOG and
the CrEDIBLE consortiums

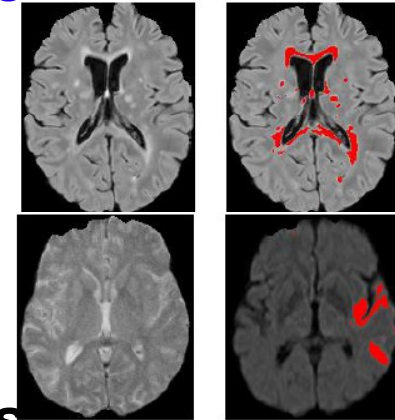


EGI Technical Forum

Madrid, Sept. 18, 2013

- **Neurodata**

- High heterogeneity: images, clinical data, biomarkers, biology....
- Increasing amount / number of (open) sources – **Big Data**
 - Large-scale medical studies (Alzheimer's...)
- Need for cross-factors analysis – **Linked Data**
 - Large-scale medical studies (epidemiology...)
 - Data (re)analysis opportunities
 - Translational research



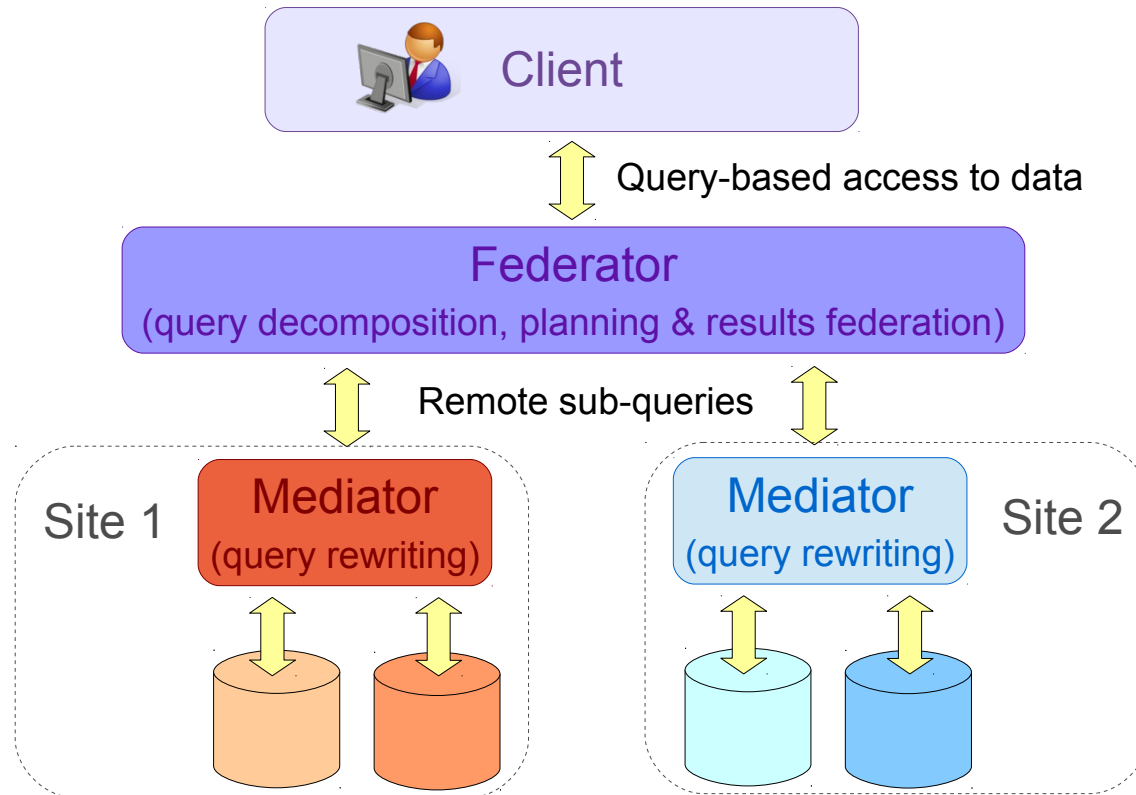
- **Centralized approaches encounter limitations**

- Large data volumes to transfer / archive / search
- Sensitive patient data / complex access control policies
- Need to adopt uniform data model & format

- **Data is *de facto* distributed over acquisition centers**

- **Approach: federate existing resources in a distributed, collaborative platform**

- Heterogeneous databases schema mediation
- Data federation through distributed querying and query rewriting



- Scientific experiment support platform:
 - raw data + models + processing results + models + provenance...

- **Application ontology**

- 3-levels structure
 - one Foundational ontology: i.e. DOLCE
 - Several Core ontologies
 - Several Domain ontologies

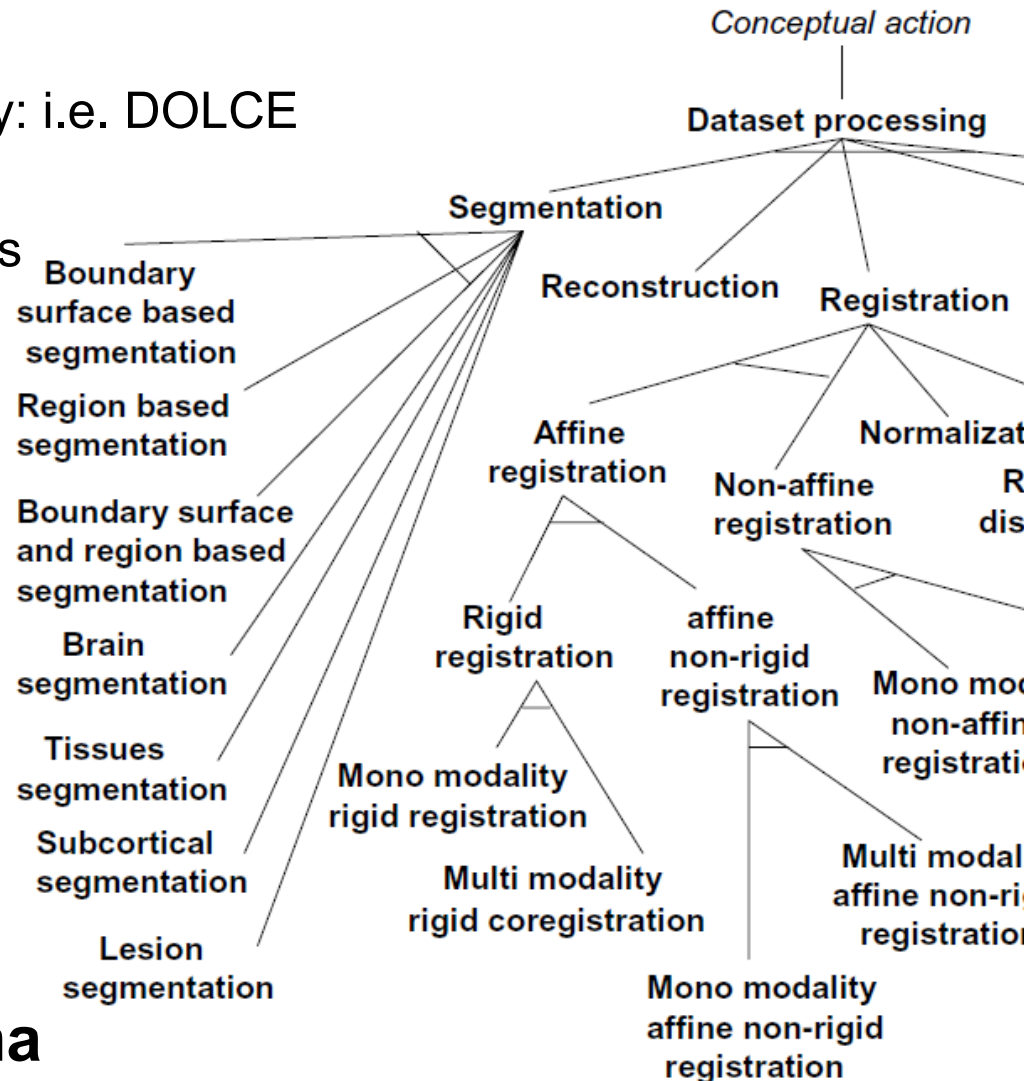
- **Covering**

- Data sets
- Data processing tools
- Scientific measure
- Medical context
- Data provenance
- ...

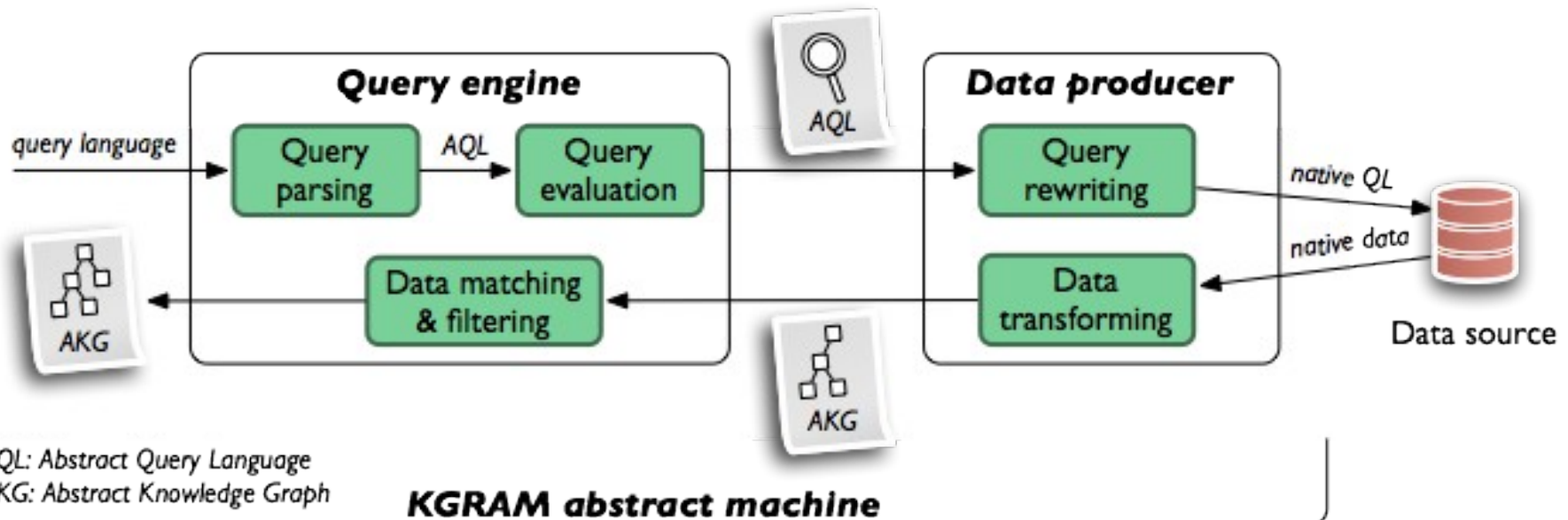
- **Domain-specific rules**

- Inference abilities

- **Derived relational schema**



- **Databases federation**
 - Requires relational model mediation
 - Based on a semantic reference to derive the federated schema
- **Based on KGRAM (Knowledge Graph Abstract Machine)**
 - Semantic query engine enabling:
 - Heterogeneous data models
 - Fully supports SPARQL v1.1



- **KGRAM query processing**

```
Q SELECT ?name ?date
   WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
           FILTER (CONTAINS (?name, 'Bob')) }
```

- **KGRAM query processing**

```
Q SELECT ?name ?date
WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
      Q1 FILTER (CONTAINS (?name, 'Bob')) } Q2
```

- KGRAM query processing**

```

Q SELECT ?name ?date
  WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
         Q1 FILTER (CONTAINS (?name, 'Bob')) } Q2
    
```

- Asynchronous execution**

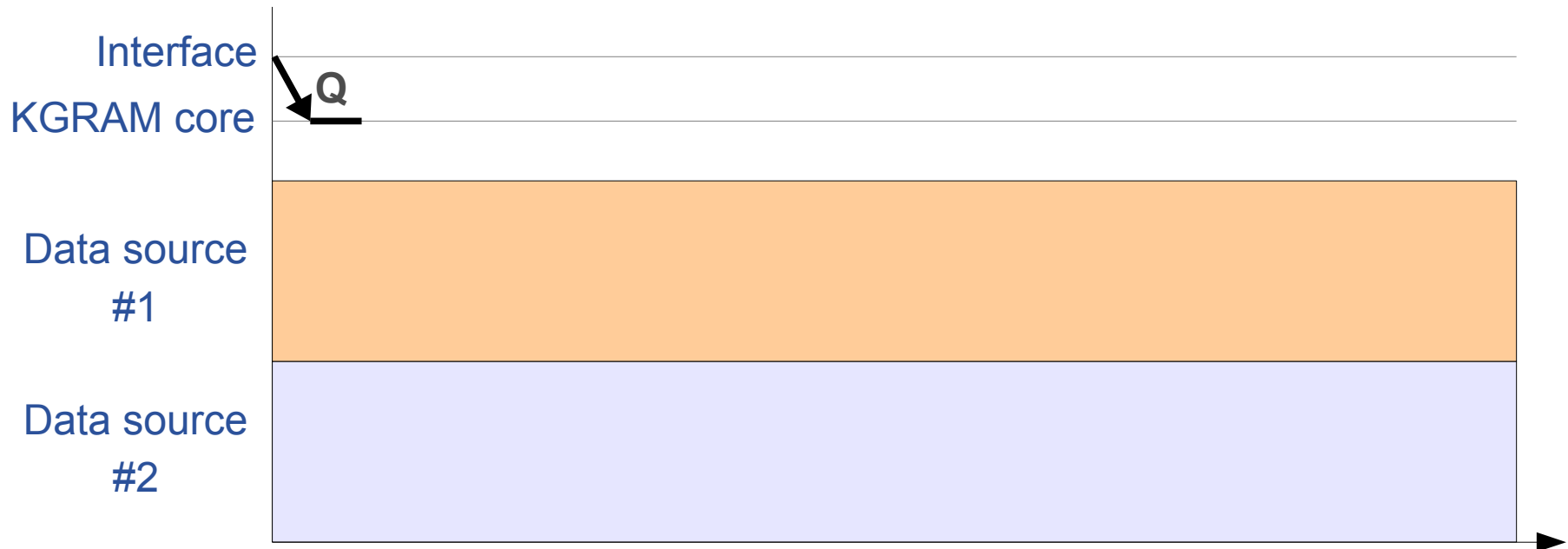


- KGRAM query processing**

```

Q SELECT ?name ?date
  WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
         Q1 FILTER (CONTAINS (?name, 'Bob')) } Q2
    
```

- Asynchronous execution**

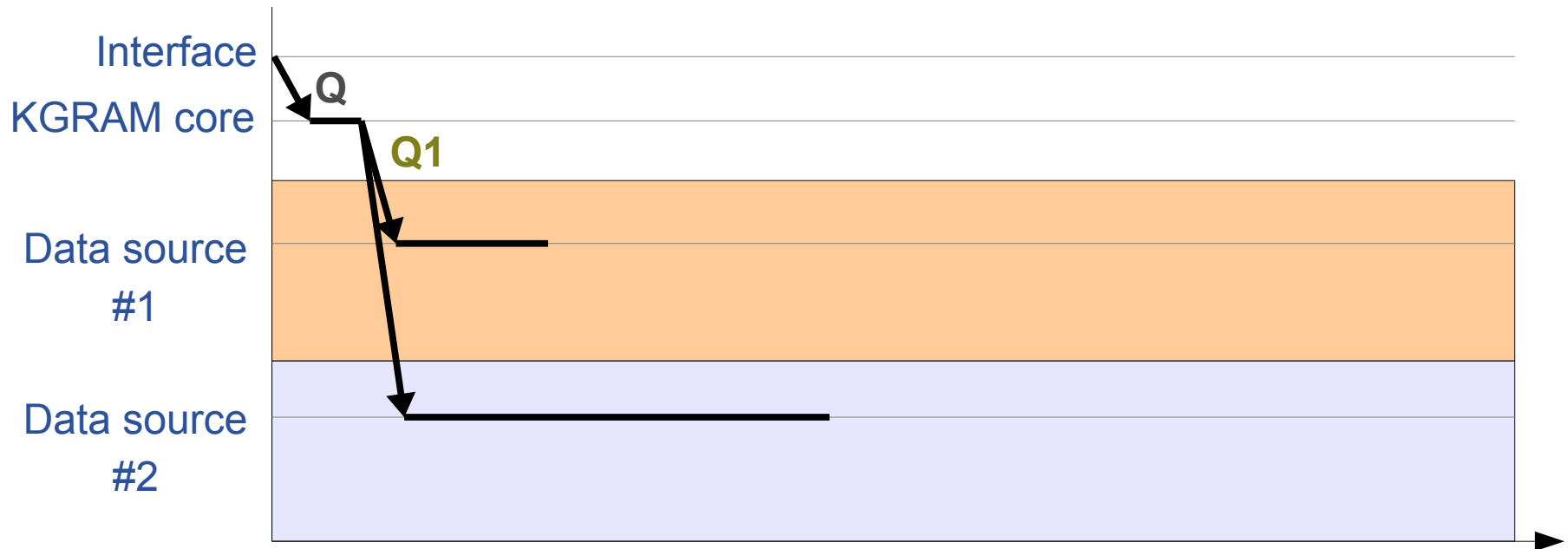


- KGRAM query processing**

```

Q SELECT ?name ?date
  WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
         Q1 FILTER (CONTAINS (?name, 'Bob')) }           Q2
    
```

- Asynchronous execution**

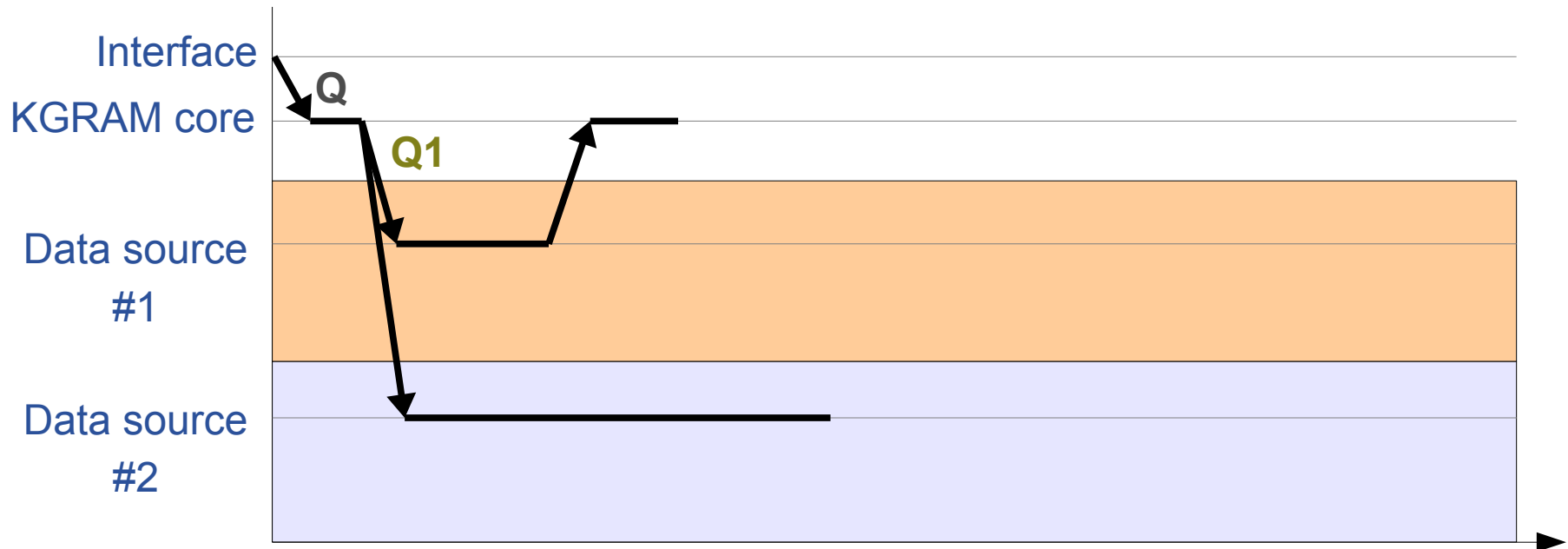


- KGRAM query processing**

```

Q  SELECT ?name ?date
   WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
           Q1 FILTER (CONTAINS (?name, 'Bob')) }           Q2
    
```

- Asynchronous execution**

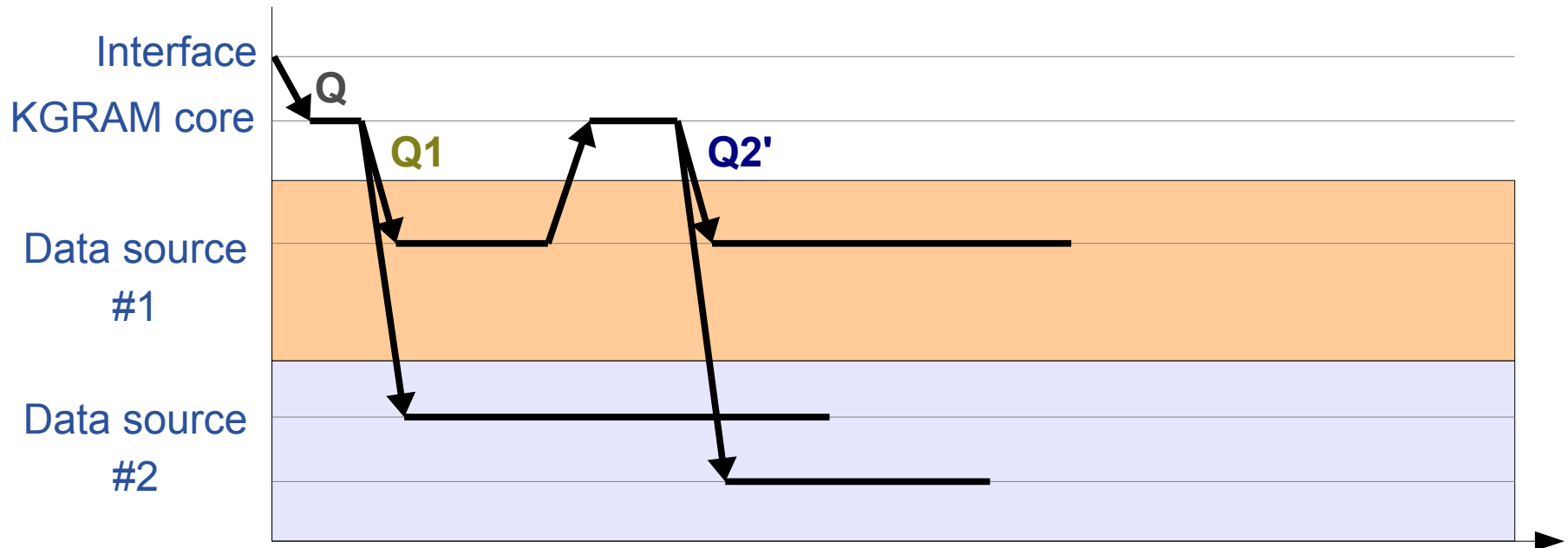


- KGRAM query processing**

```

Q SELECT ?name ?date
  WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
         Q1 FILTER (CONTAINS (?name, 'Bob')) } Q2
    
```

- Asynchronous execution**

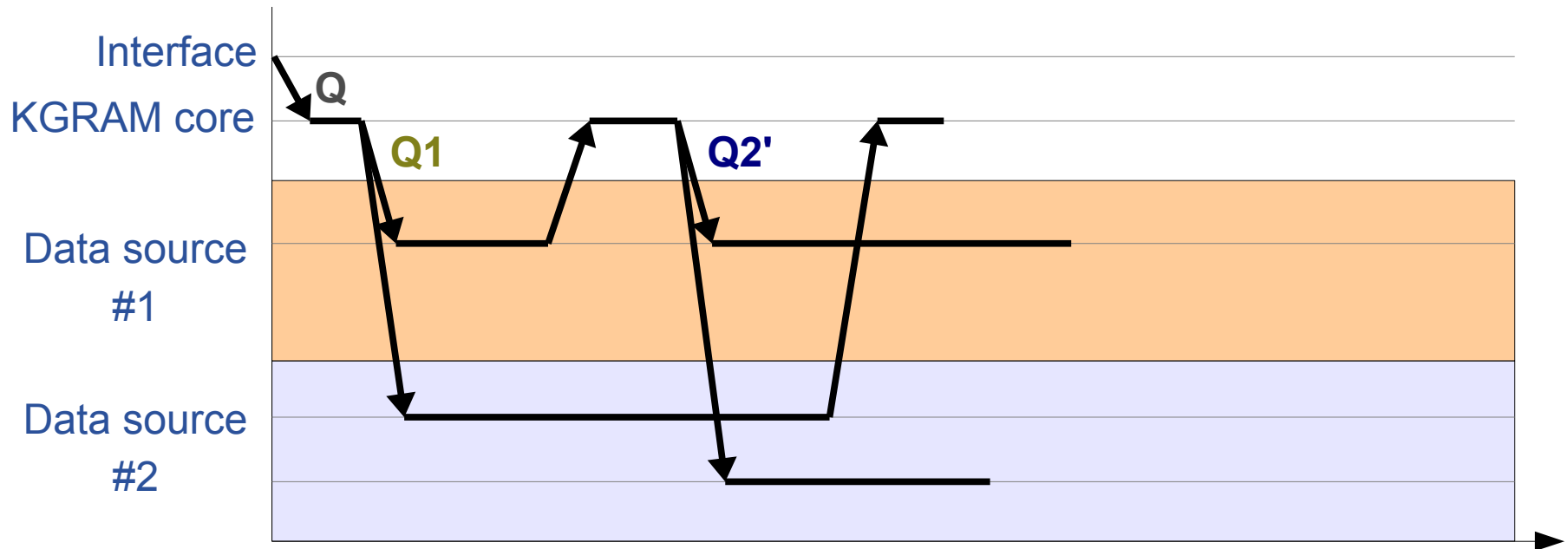


- KGRAM query processing**

```

Q SELECT ?name ?date
  WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
         Q1 FILTER (CONTAINS (?name, 'Bob')) } Q2
    
```

- Asynchronous execution**

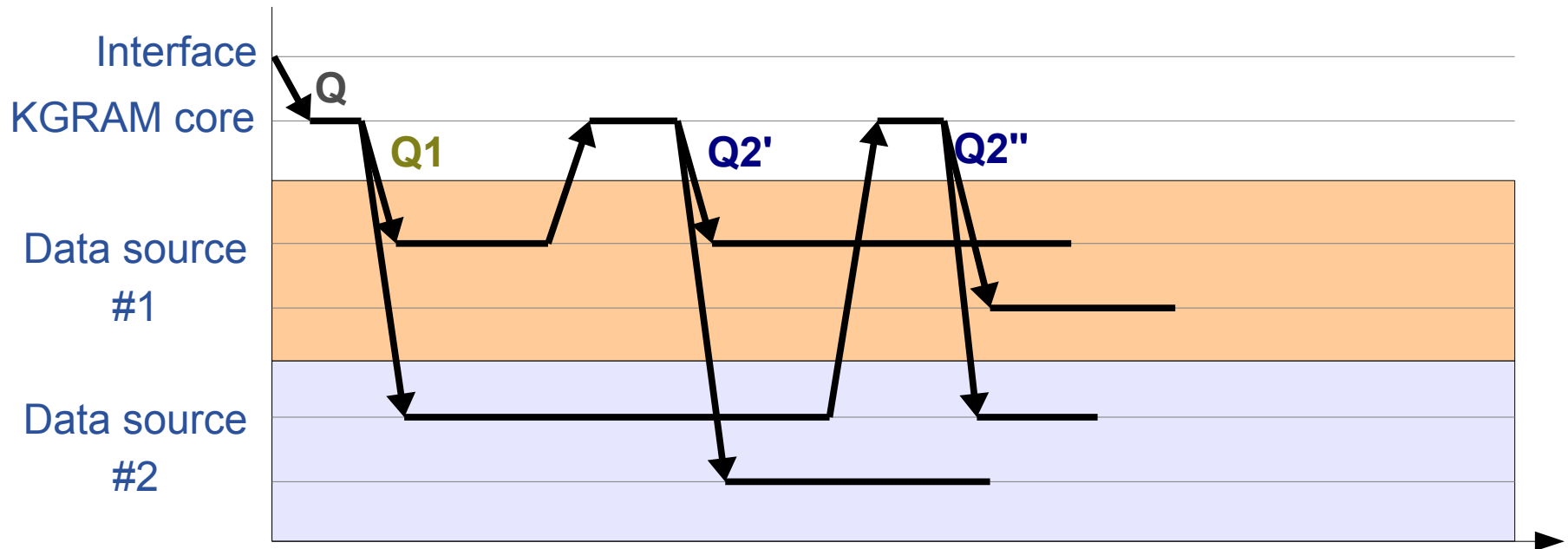


- KGRAM query processing**

```

Q  SELECT ?name ?date
   WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
          Q1 FILTER (CONTAINS (?name, 'Bob')) }           Q2
    
```

- Asynchronous execution**

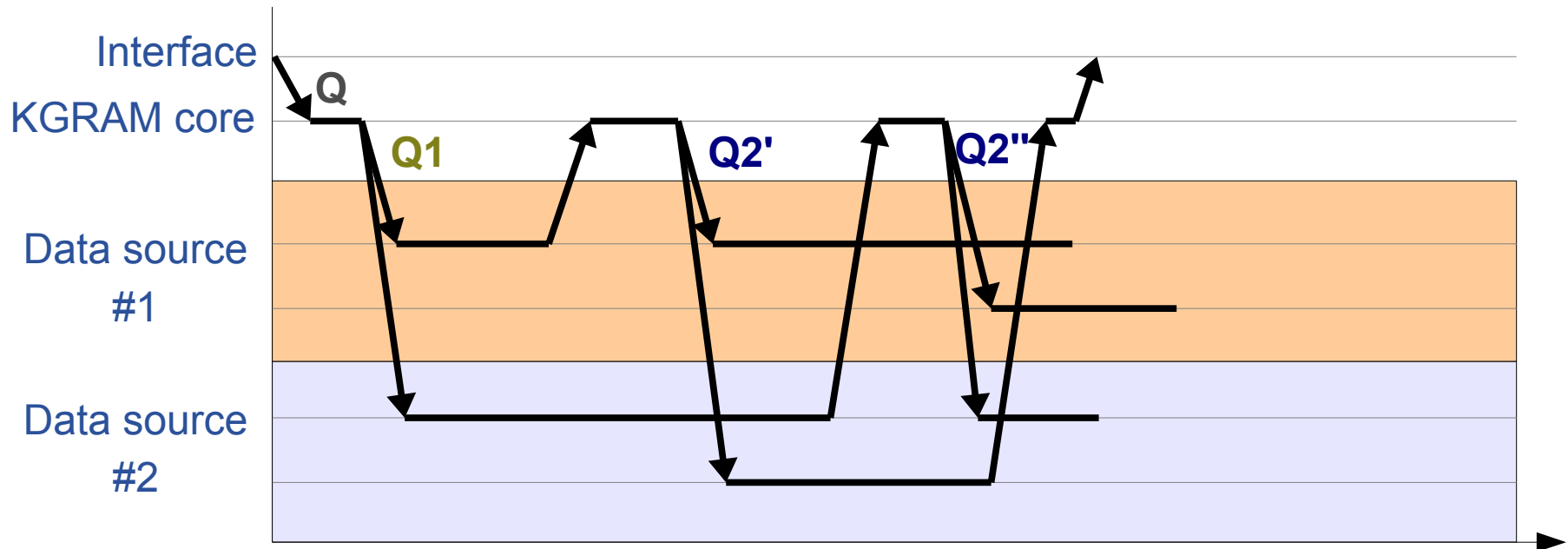


- KGRAM query processing**

```

Q  SELECT ?name ?date
    WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
           Q1 FILTER (CONTAINS (?name, 'Bob')) }           Q2
    
```

- Asynchronous execution**

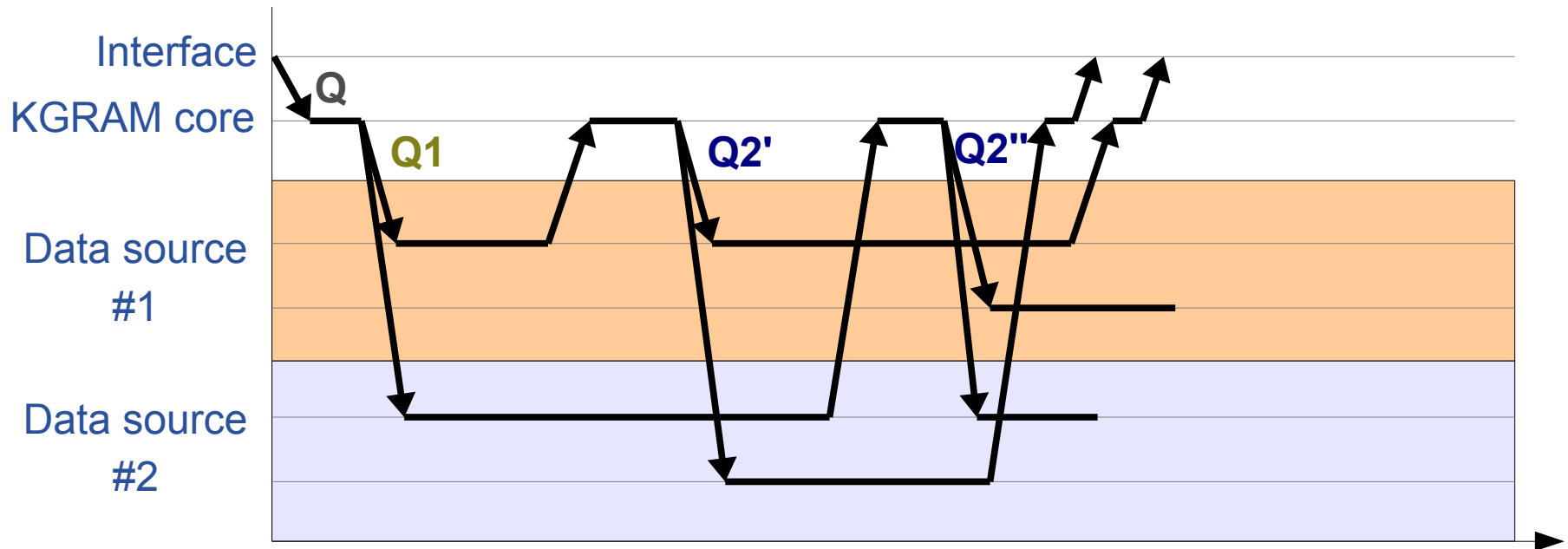


- KGRAM query processing**

```

Q  SELECT ?name ?date
    WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
           Q1 FILTER (CONTAINS (?name, 'Bob')) }           Q2
    
```

- Asynchronous execution**

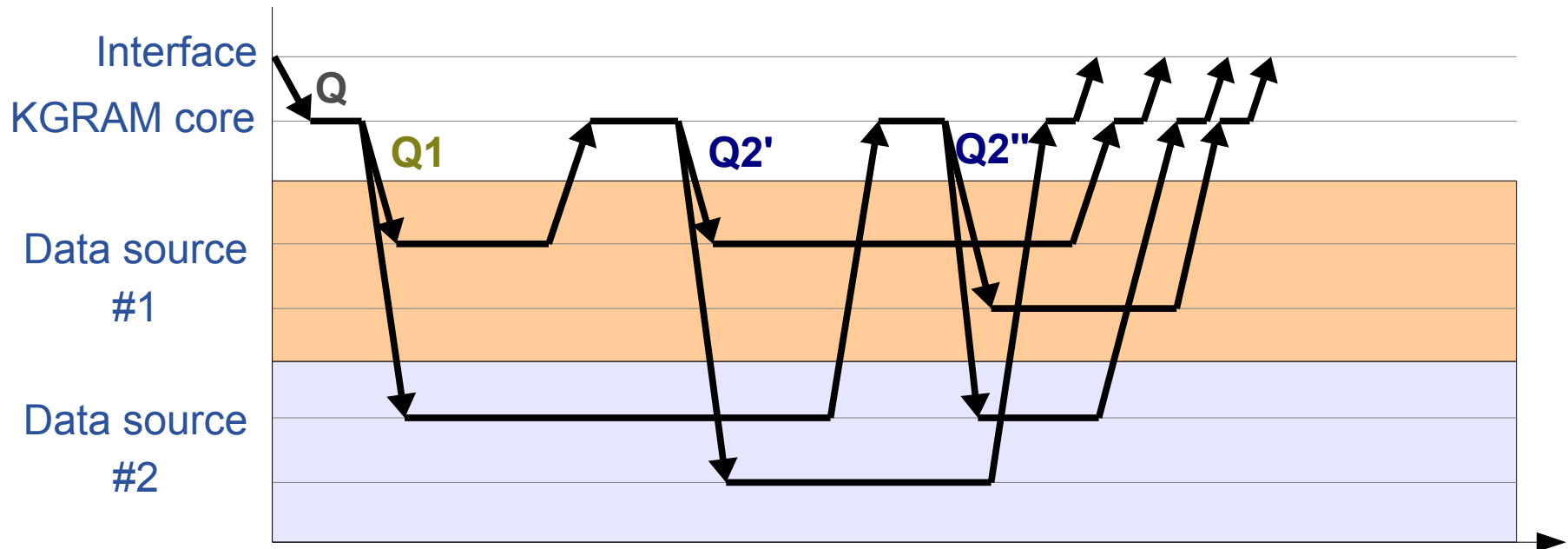


- KGRAM query processing**

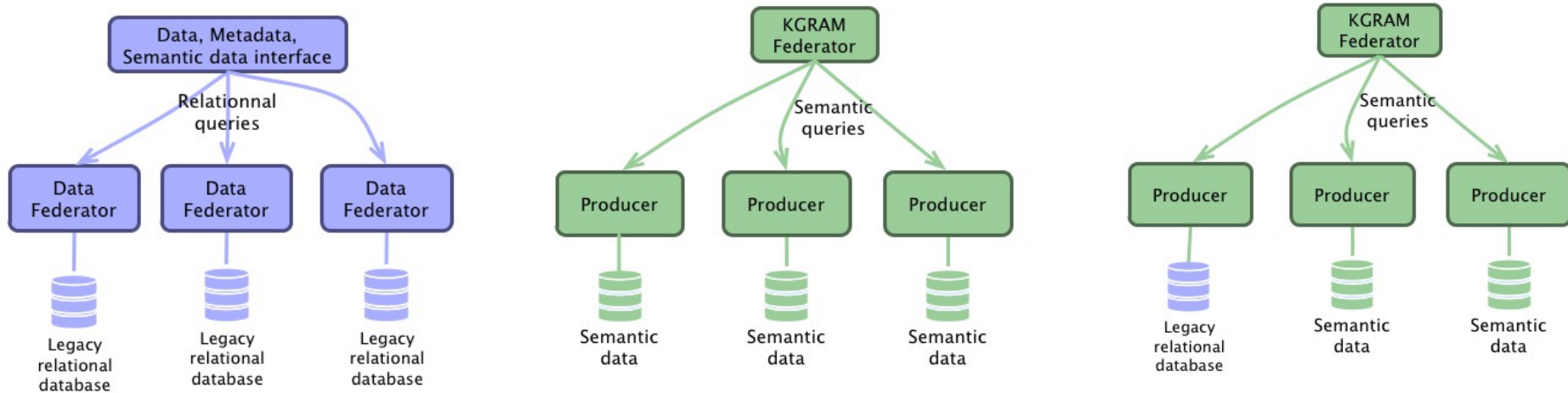
```

Q  SELECT ?name ?date
    WHERE { ?x foaf:name ?name . ?x dbpedia:birthDate ?date .
           Q1 FILTER (CONTAINS (?name, 'Bob')) }           Q2
    
```

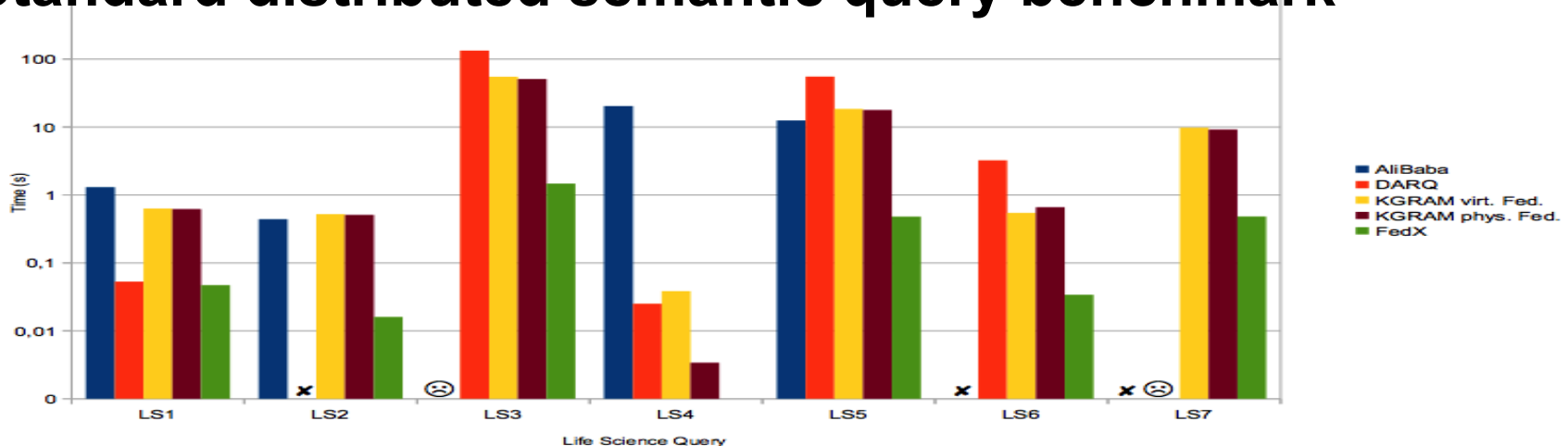
- Asynchronous execution**



- **Heterogeneous (relational / semantic) stores querying**
 - Close results with DF relational federation and KGRAM-DQP

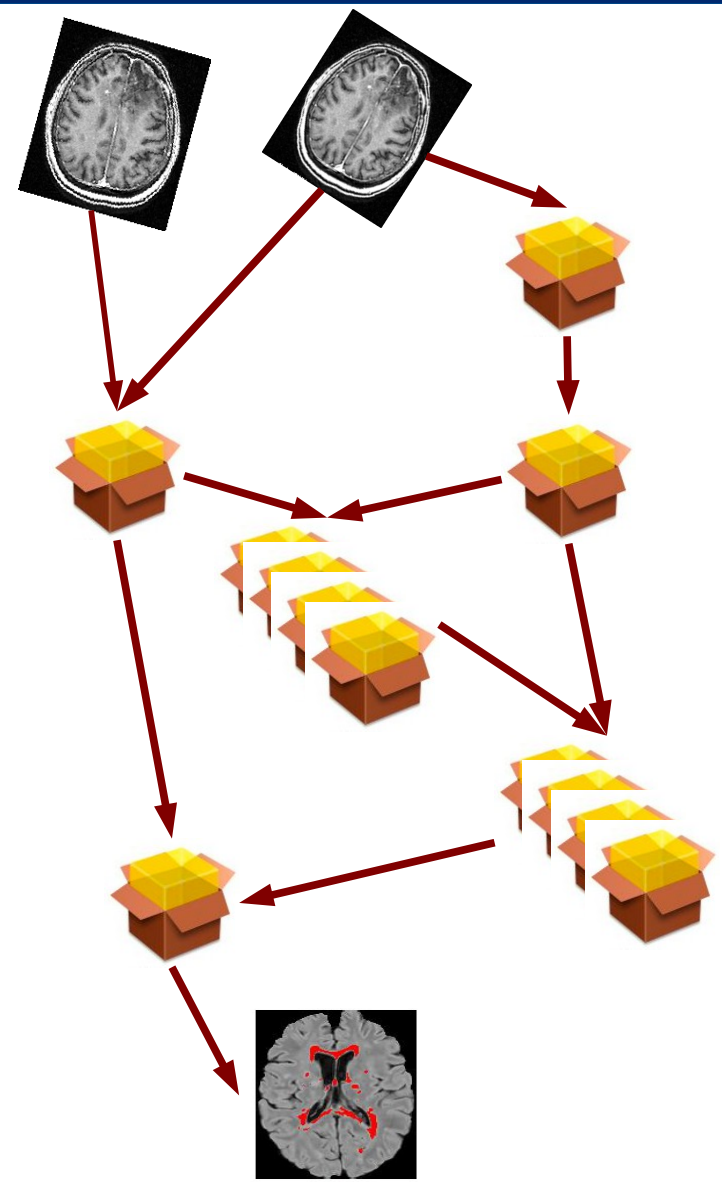


- **Standard distributed semantic query benchmark**

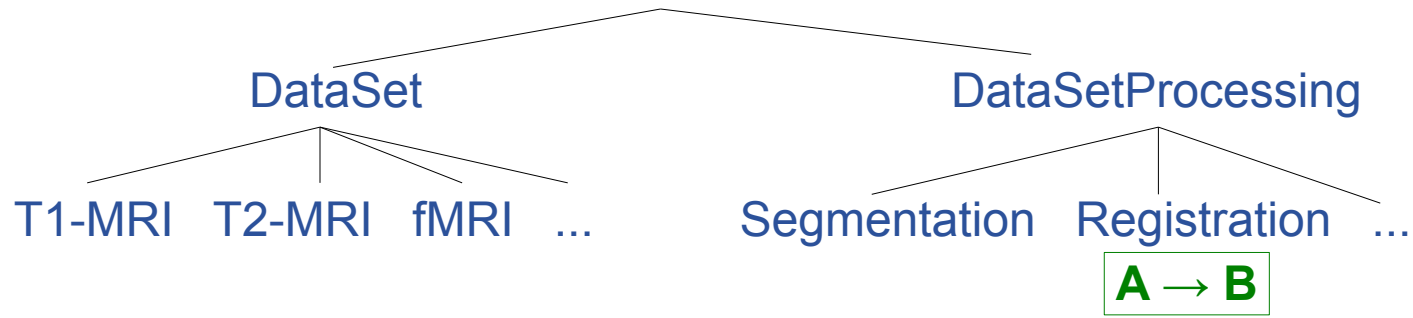


- **Data-intensive distributed computing workflow engine**

- Abstract description of data analysis pipelines
- Interface to computing infrastructure
 - Data processing code wrapper component (jigsaw)
 - Interface to grid(s) workload management API
- <http://modalis.polytech.unice.fr/moteur2>



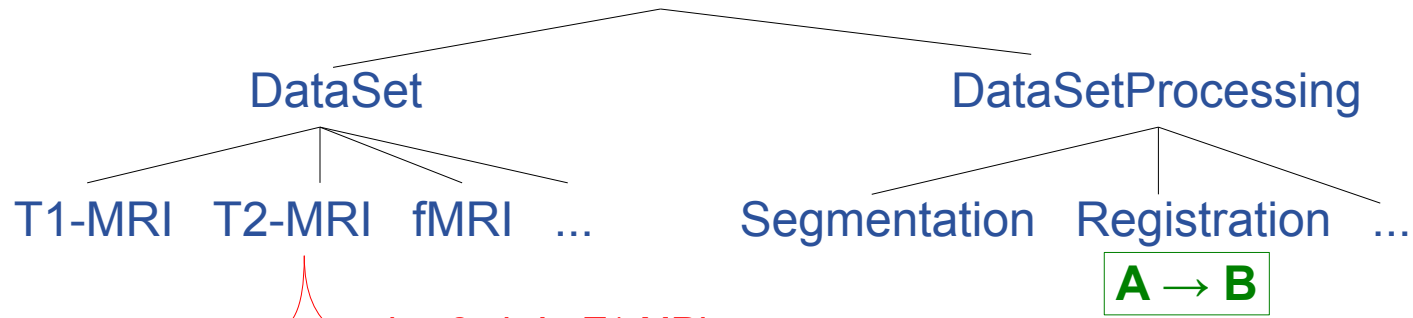
- **Ontology**
 - Concepts & **Rules**



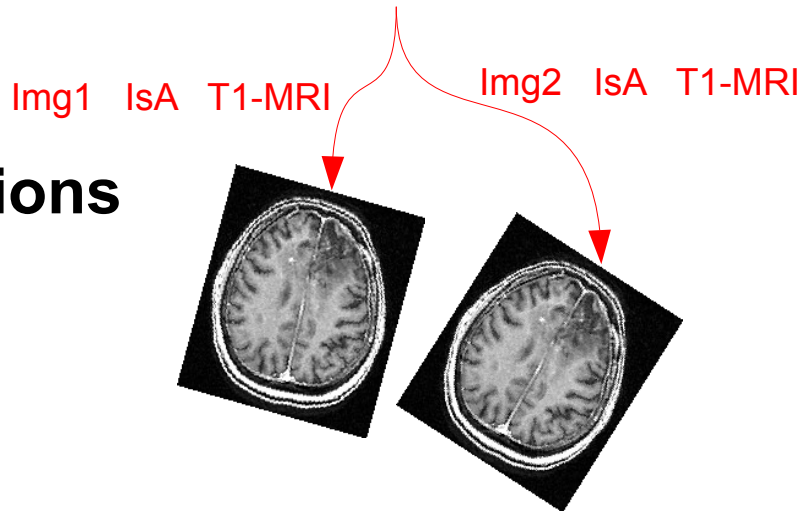
- **Annotations**

- **Processing**

- **Ontology**
 - Concepts & **Rules**

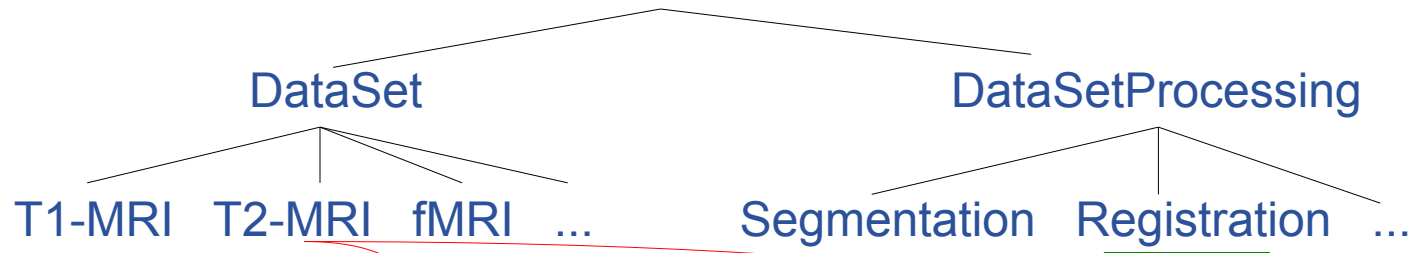


- **Annotations**

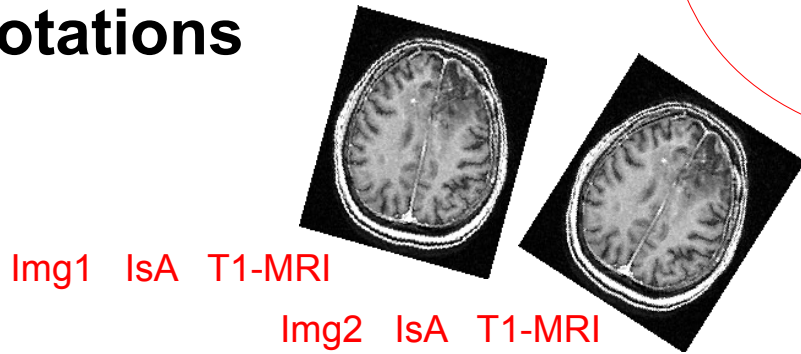


- **Processing**

- **Ontology**
 - Concepts & **Rules**



- **Annotations**



- **Processing**

Tool1 HasInput T1-MRI

Tool1 IsA Registration

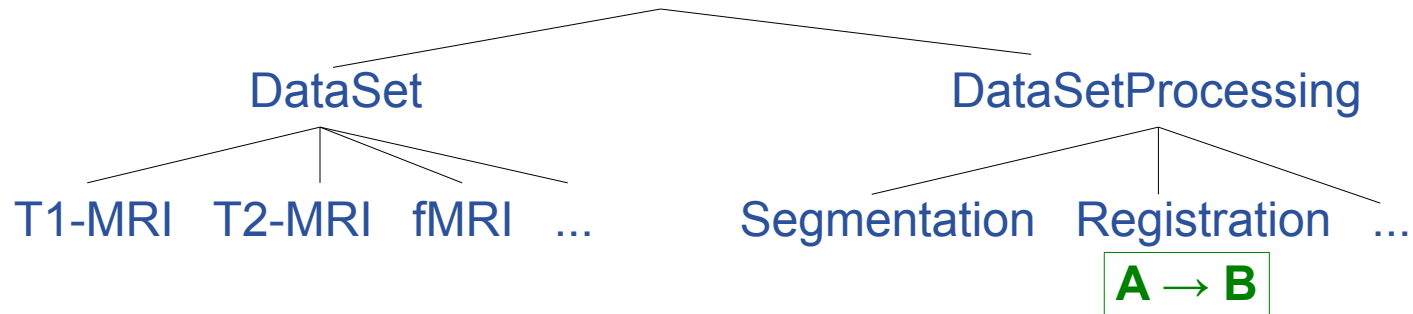
A → B



Tool1 HasOutput Transfo

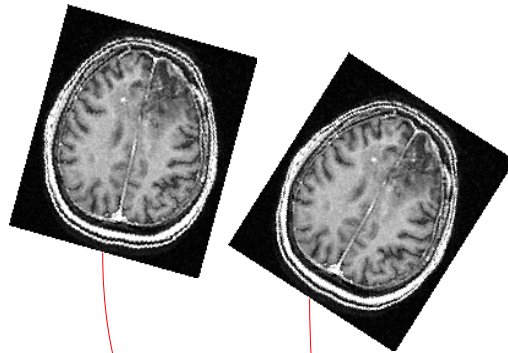
- Ontology**

- Concepts & **Rules**



- Annotations**

Img1 IsA T1-MRI
 Img2 IsA T1-MRI

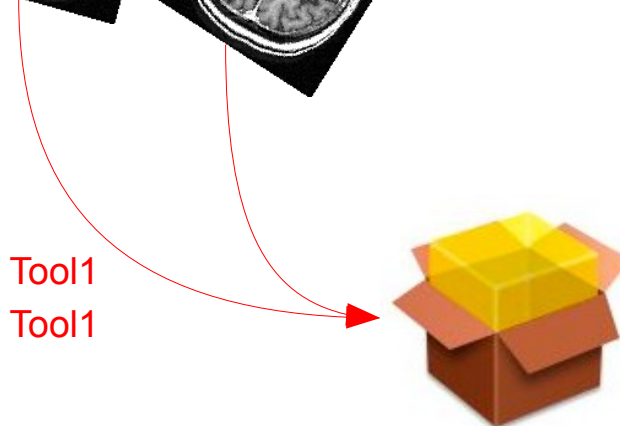


Tool1 HasInput T1-MRI
 Tool1 HasOutput Transfo
 Tool1 IsA Registration



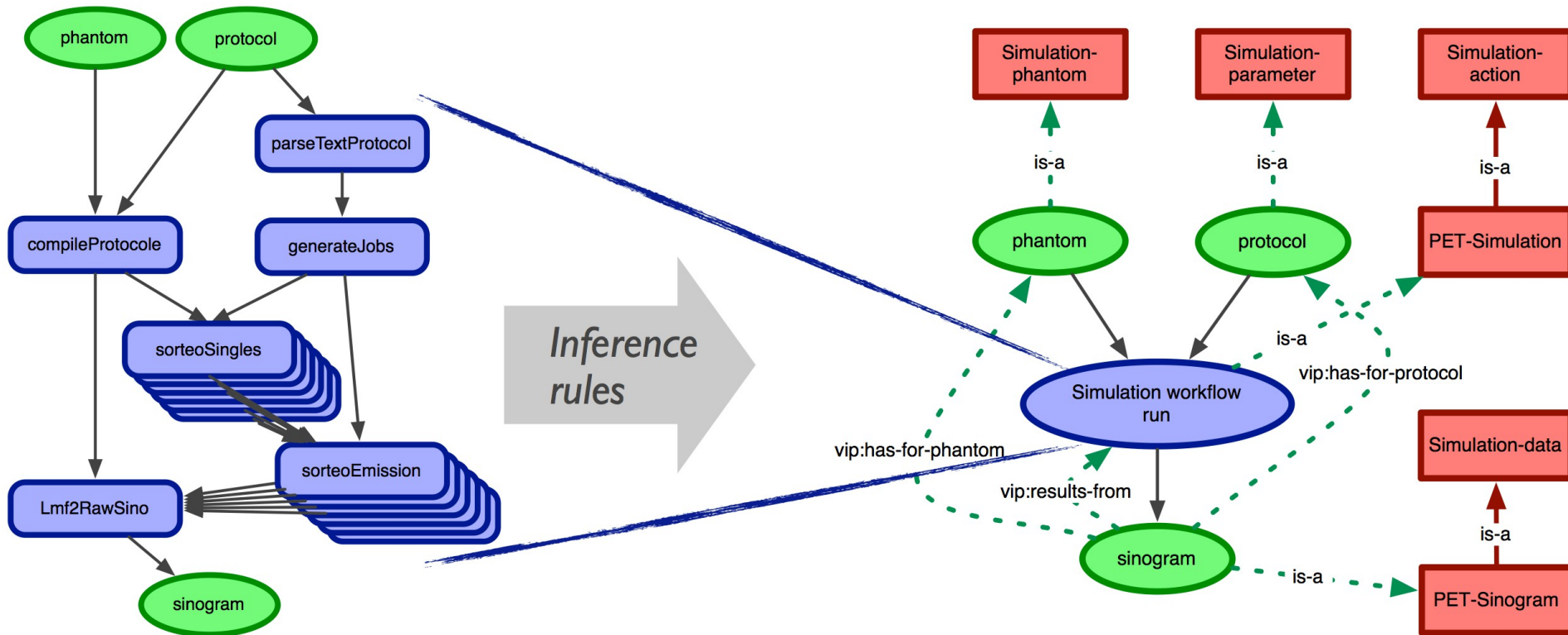
- Processing**

Img1 IsProcessedBy Tool1
 Img2 IsProcessedBy Tool1



Tool1 Produced Transfo1
 Transfo1 IsA GlobalTransfo

- Fine-grained annotation traces generated at run-time
- Summary generated by inference rules application
 - Produce relevant and human-tractable experiment summaries



- Example in the context of the VIP (Virtual Imaging Platform) project

- **Neurodata federation**
 - Relational data federation requires semantic reference
 - Well-founded semantic data model, high quality of data
 - Leverages the *Web of Linked Data*
 - Domain-specific ontology, generic methods & tools
- **Semantic technologies**
 - Powerful semantic query and inference engine (SPARQL 1.1)
 - Trade-off between query language expressivity and performance
 - Coupling data and processing semantics to produce post-processing experiment summary
- **Application to other domains**
 - VIP Virtual Imaging Platform <http://www.creatis.insa-lyon.fr/vip>
 - GINSENG epidemiology network <http://www.e-ginseng.org>

- **Reports & publications available on-line**
 - <http://credible.i3s.unice.fr> & <http://neurolog.i3s.unice.fr>
- **Publications**
 - O. Corby, A. Gaignard, C. Faron-Zucker, J. Montagnat.
KGRAM Versatile Inference and Query Engine for the Web of Linked Data
IEEE/WIC/ACM International Conference on Web Intelligence, Macao, China, Dec. 2012.
 - A. Gaignard, J. Montagnat, C. Faron-Zucker, O. Corby.
Semantic Federation of Distributed Neurodata
MICCAI Workshop on Data- and Compute-Intensive Clinical and Translational Imaging Applications, pages 41-50, Nice, France, October 2012.
 - B. Gibaud, G. Kassel, M. Dojat, B. Batrancourt, F. Michel, A. Gaignard, J. Montagnat
NeuroLOG: sharing neuroimaging data using an ontology-based federated approach
AMIA, vol. 2011, pages 472–480, Washington DC, USA, October 2011.
 - F. Michel, A. Gaignard, F. Ahmad, C. Barillot, B. Batrancourt, M. Dojat, B. Gibaud, *et al.*
Grid-wide neuroimaging data federation in the context of the NeuroLOG project
HealthGrid'10, pages 112-123, IOS Press, Paris, France, 28-30 June 2010.
- **Research reports**
 - CrEDIBLE-12-1-v1: multi-disciplinary workshop report
 - CrEDIBLE-12-2-v1: distributed semantic query engines
 - CrEDIBLE-12-3-v1: sémantique des données de l'observation