

Entrepôts de données distribués. Partage de chaînes de traitement de données.



Johan Montagnat
CNRS, I3S lab, Modalis team
on behalf of the NeuroLOG consortium

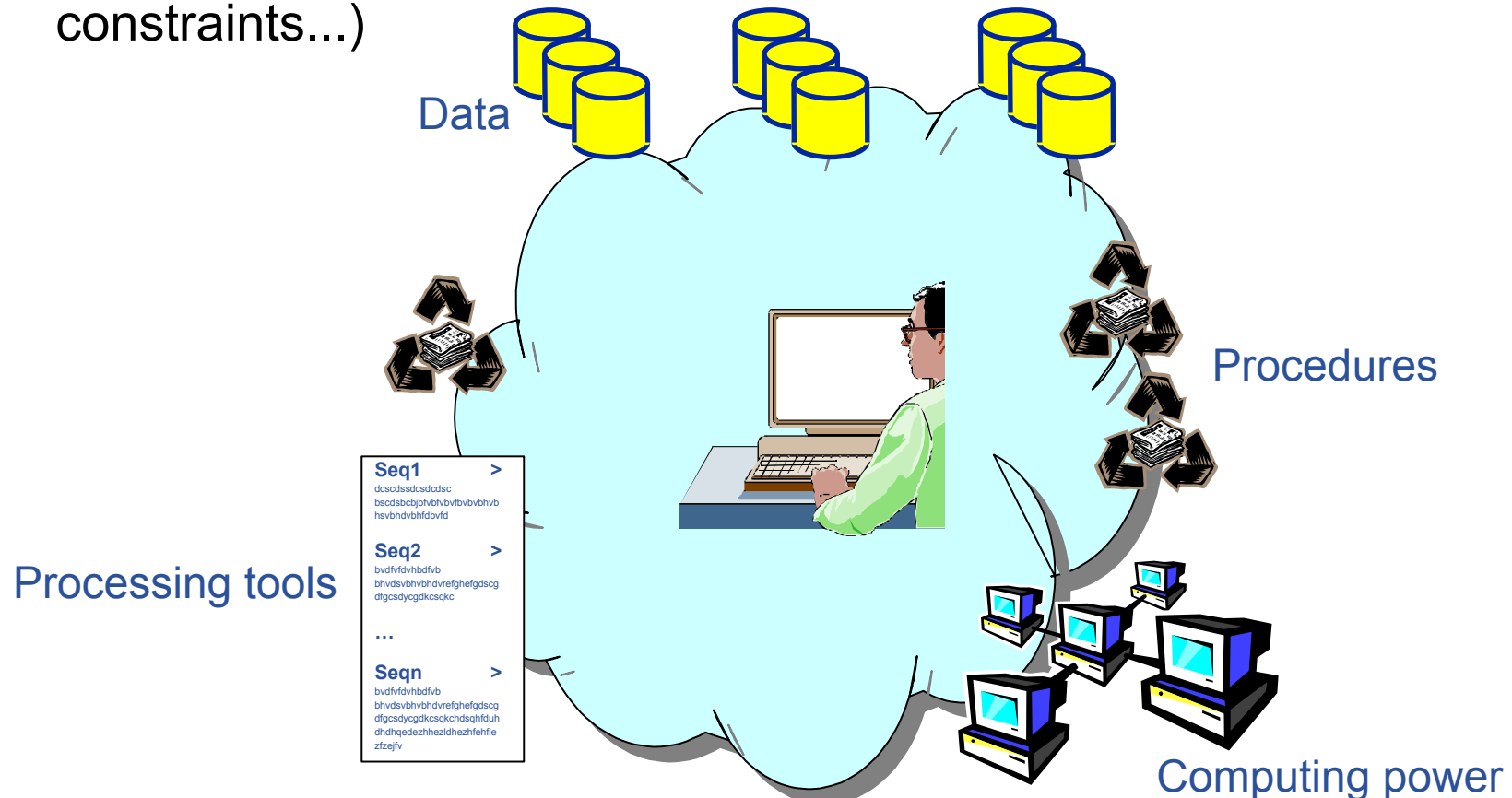


Journées IS & IDM
12-13 décembre 2011
Sophia Antipolis

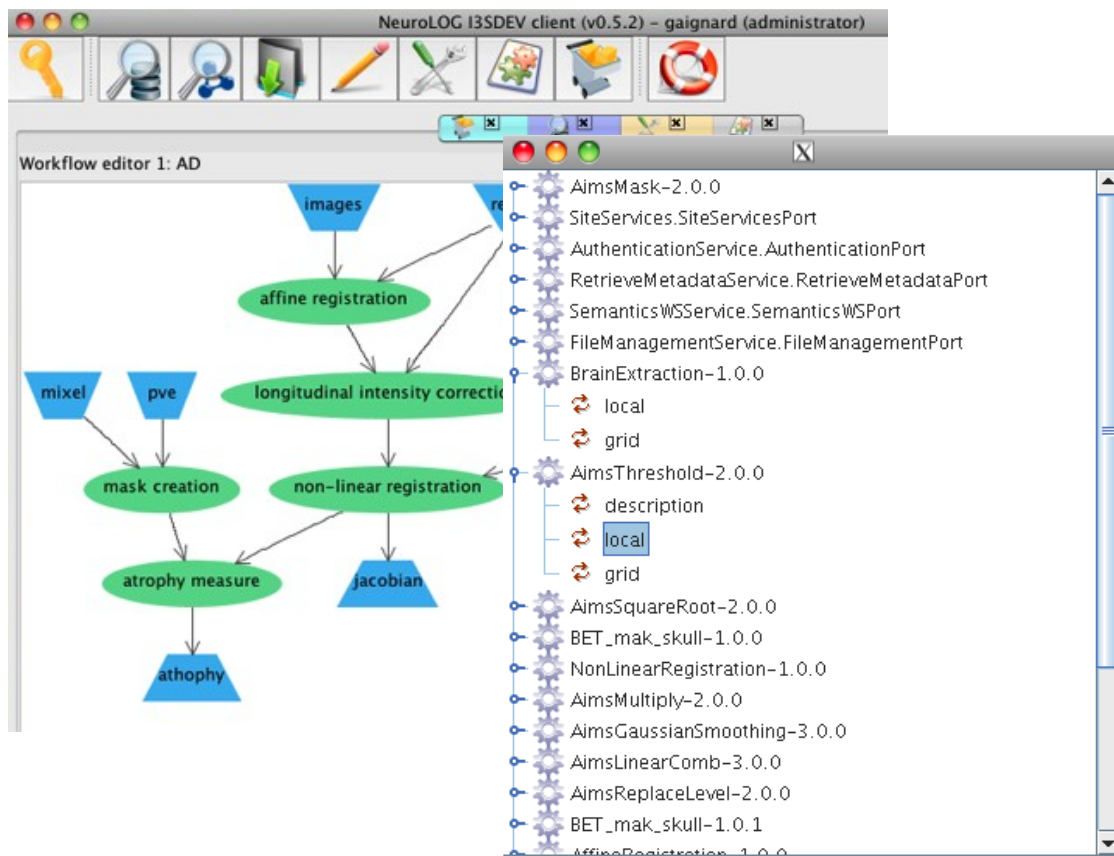
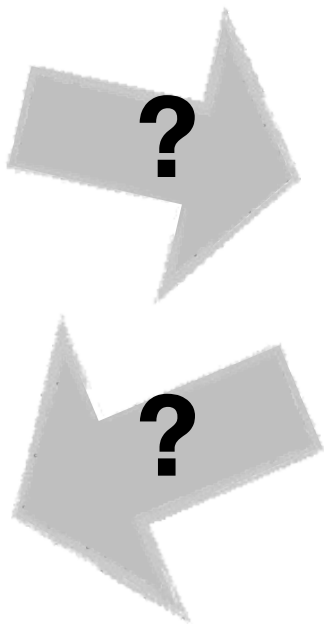
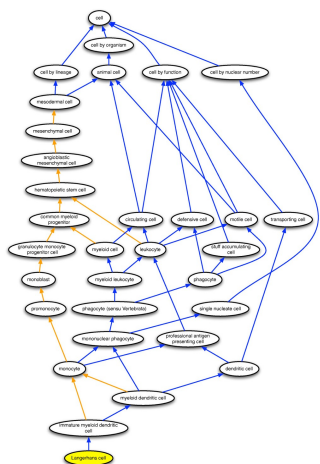


- **Sharing computing resources and algorithms**

- Research (populations studies, models design, validation, statistics)
- Complex analysis (compute intensive image processing, time constraints...)

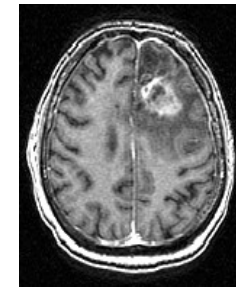
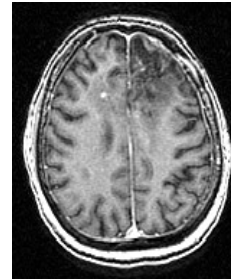
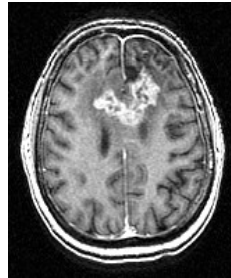


Domain ontologies (syntactic) Scientific Workflows



- **Pathologies**

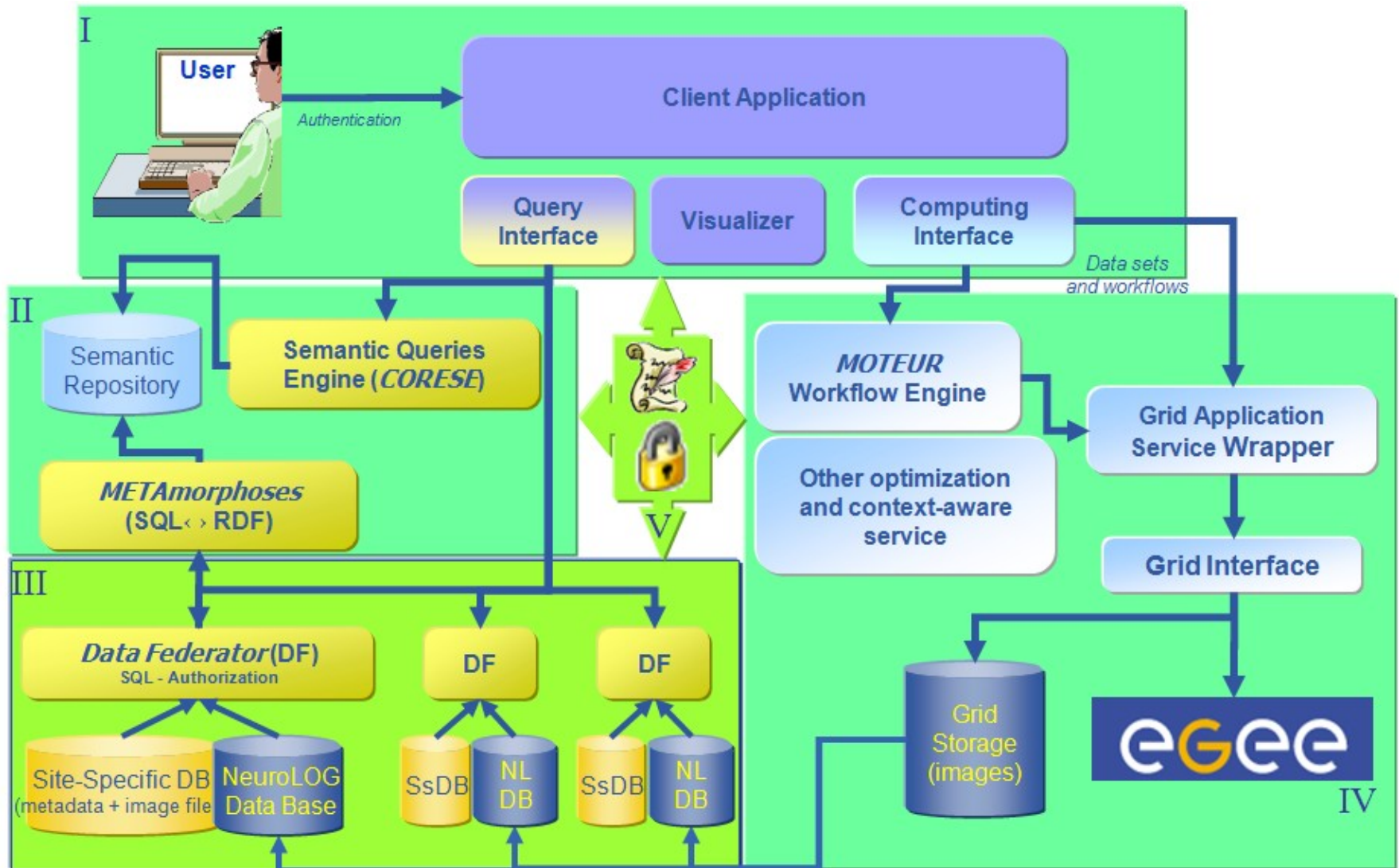
- Multiple Sclerosis
- Brain strokes
- Brain tumors
- Alzheimer's



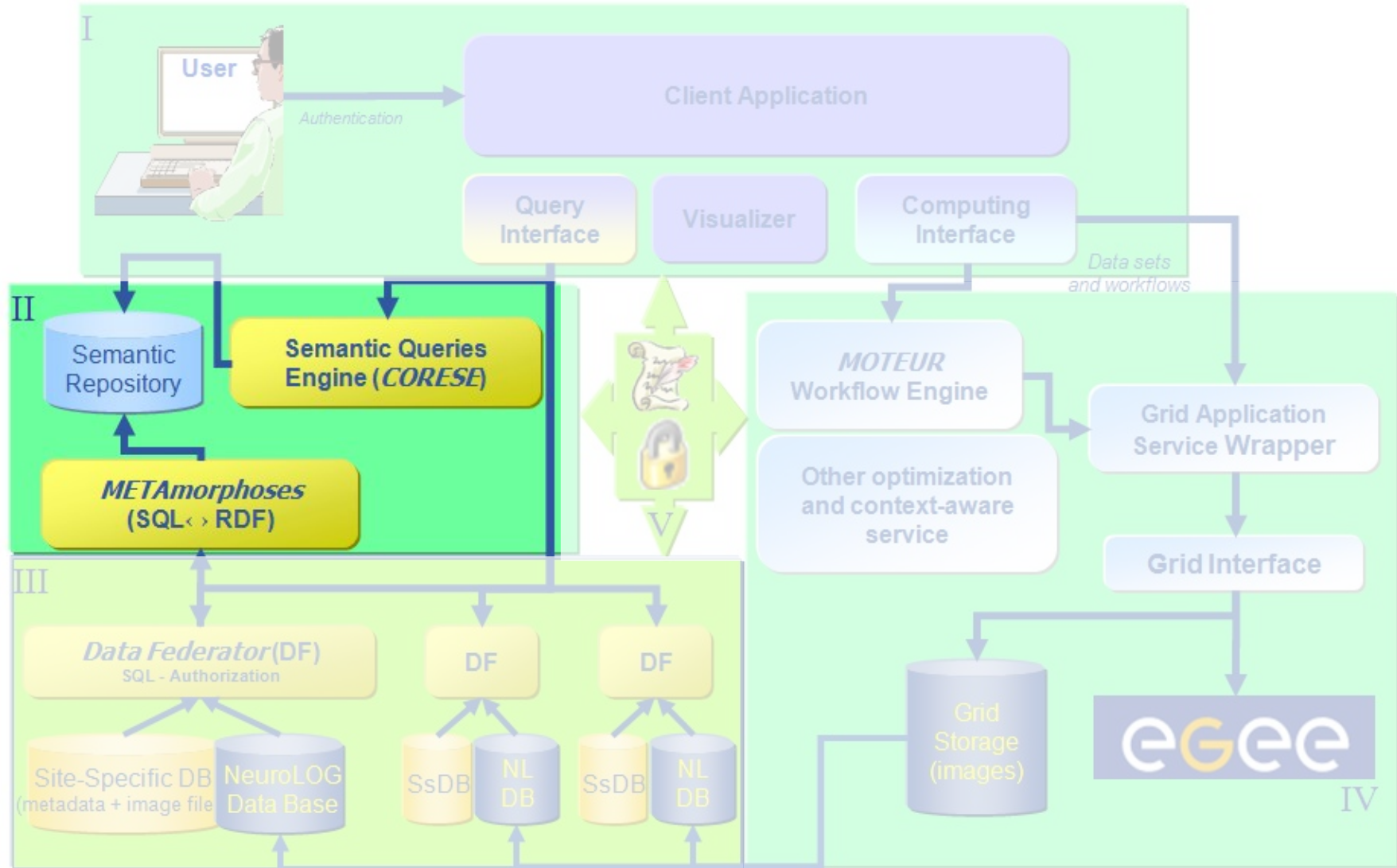
- **Data stores**

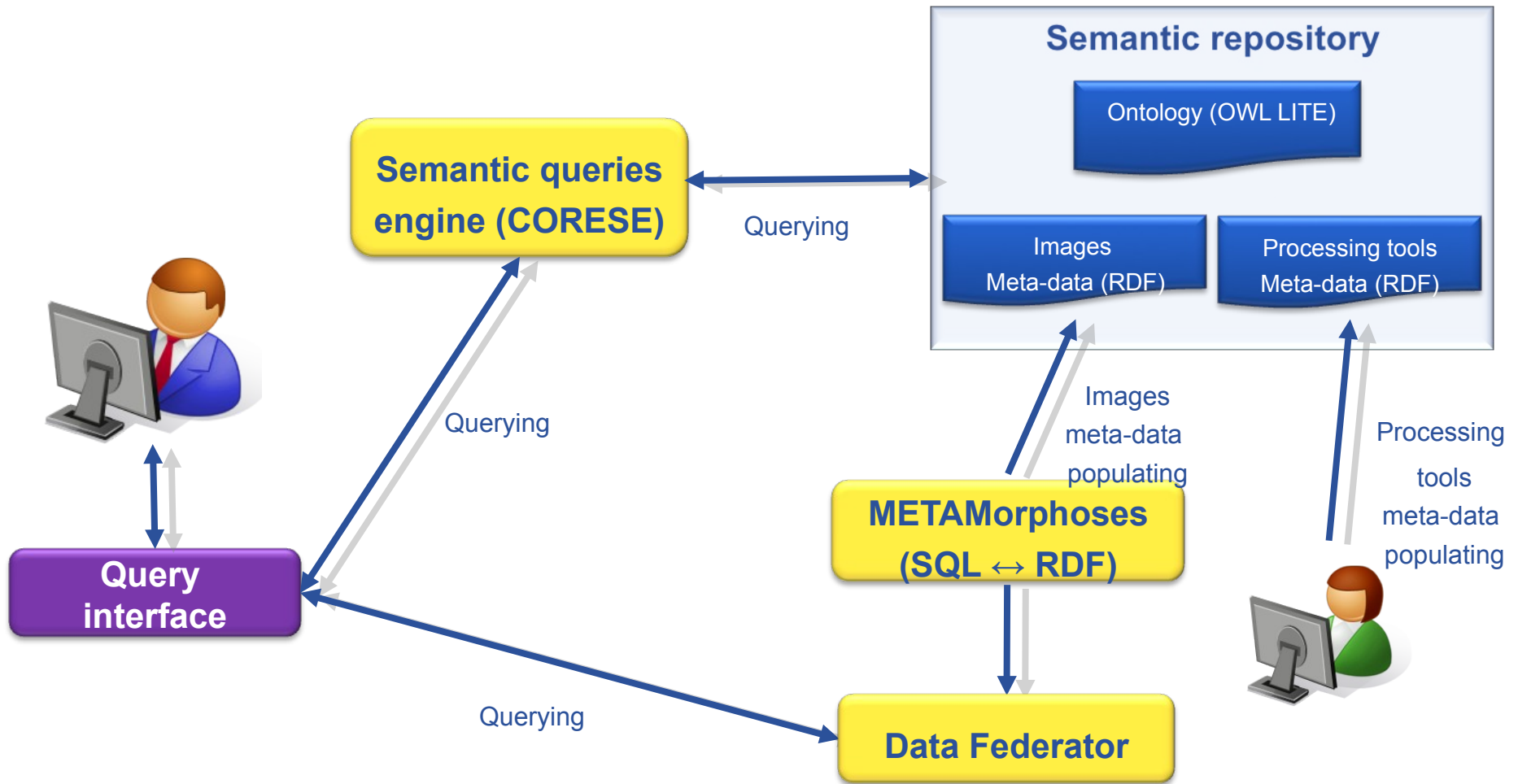
- Mainly MR data
 - Various modalities (T1, T1 Gado, T2, Flair, Diffusion, PD)
 - Various processed images (Registered, Segmented, ...)
- Neurophysiological and Neuroclinical data
- Clinical context, Patient assessment, Acquisition...

Software technologies for integration of process, data and knowledge in medical imaging

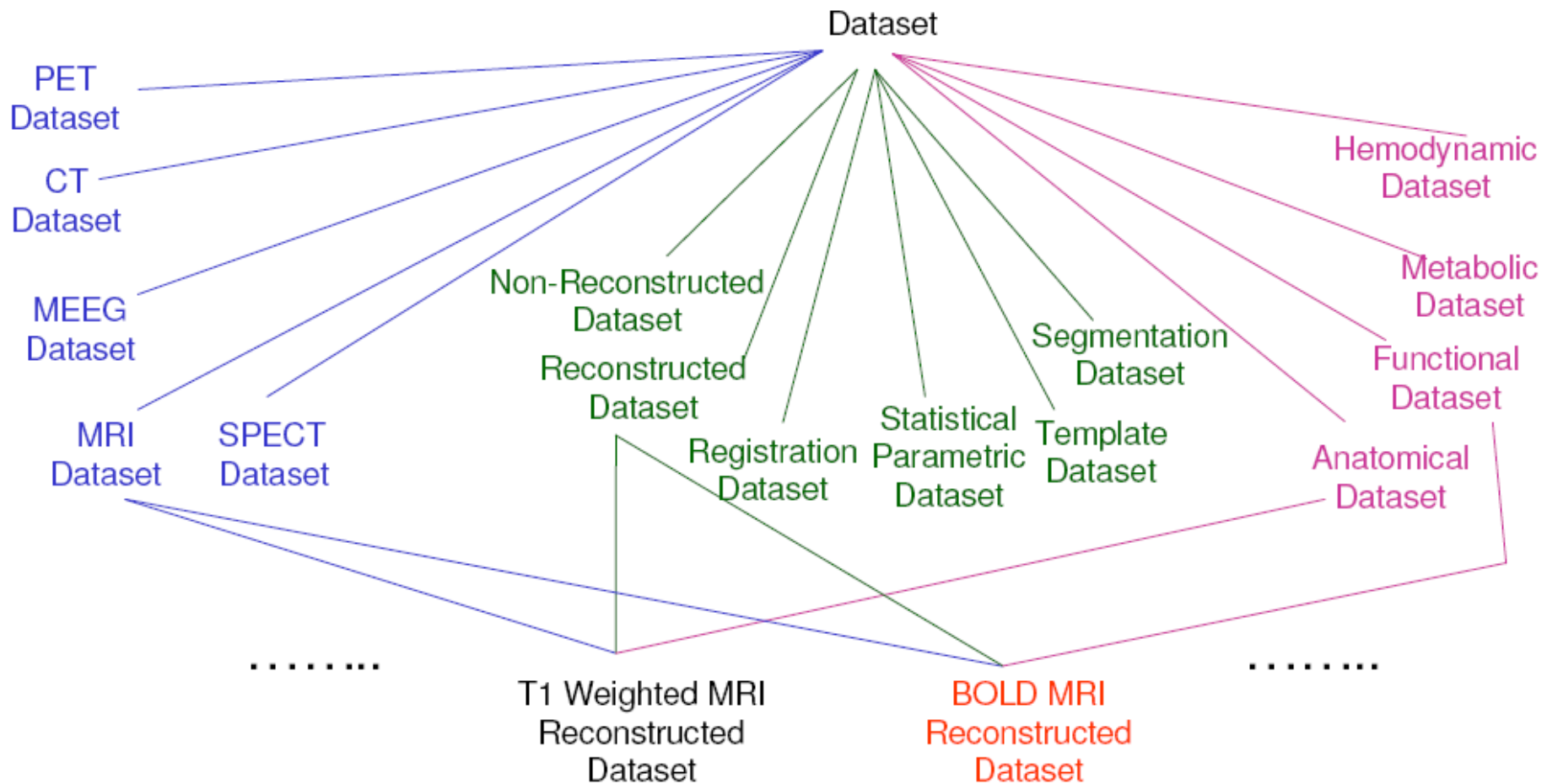


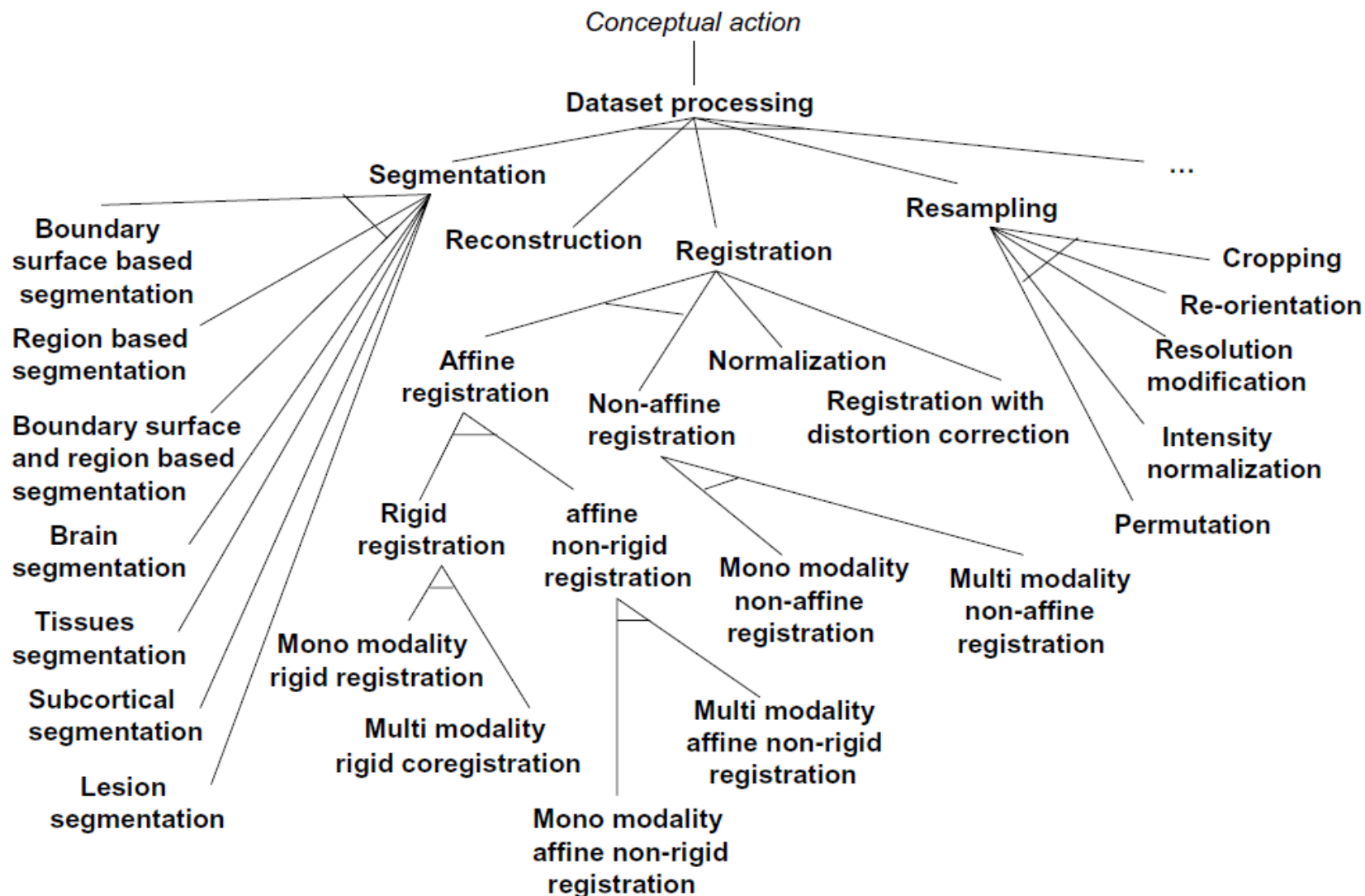
Software technologies for integration of process, data and knowledge in medical imaging



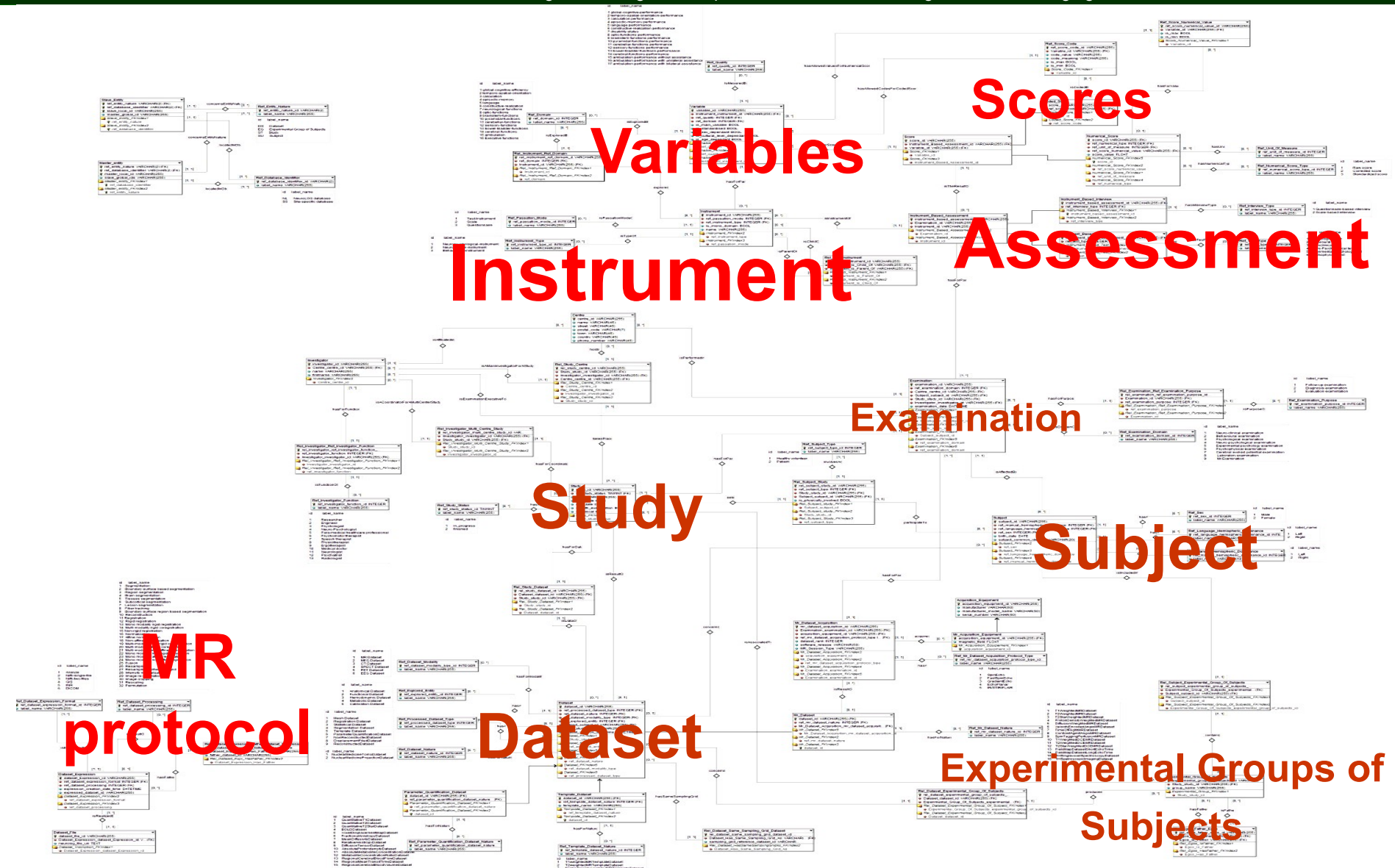


- On acquisition (**isAcquiredFrom**), on processings (**isResultOf**) and on structure/function (**explores**)

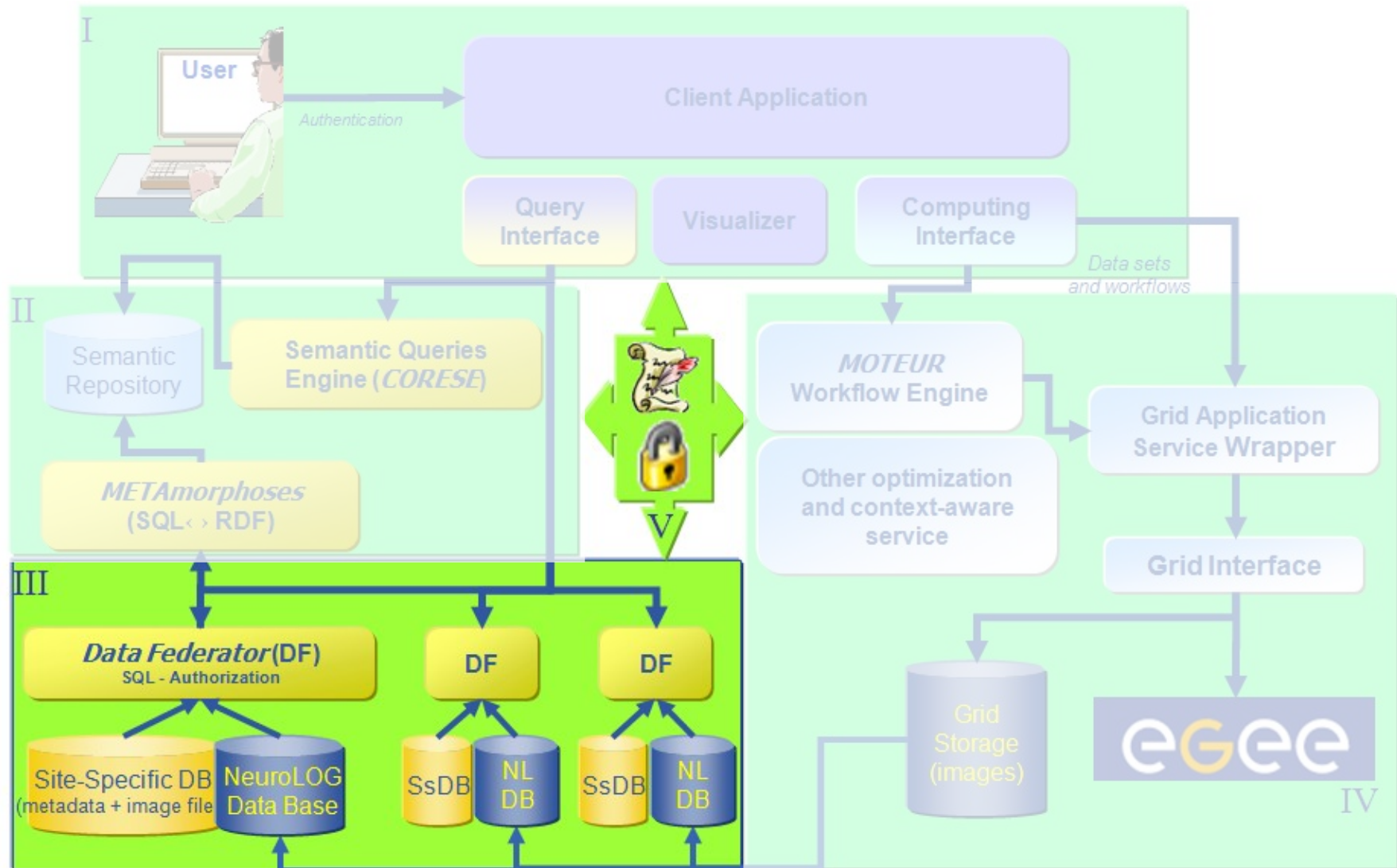




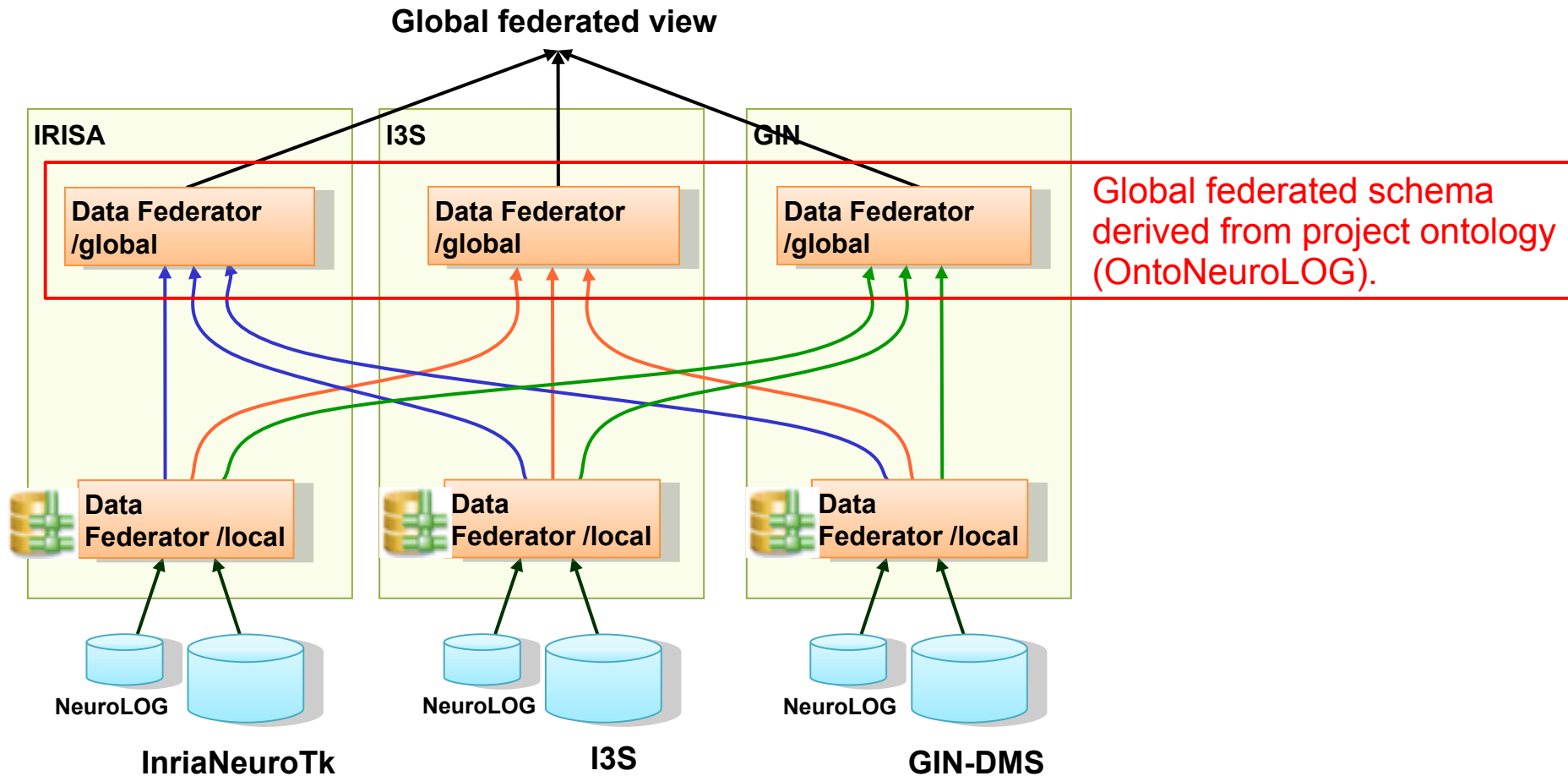
Software technologies for integration of process, data and knowledge in medical imaging



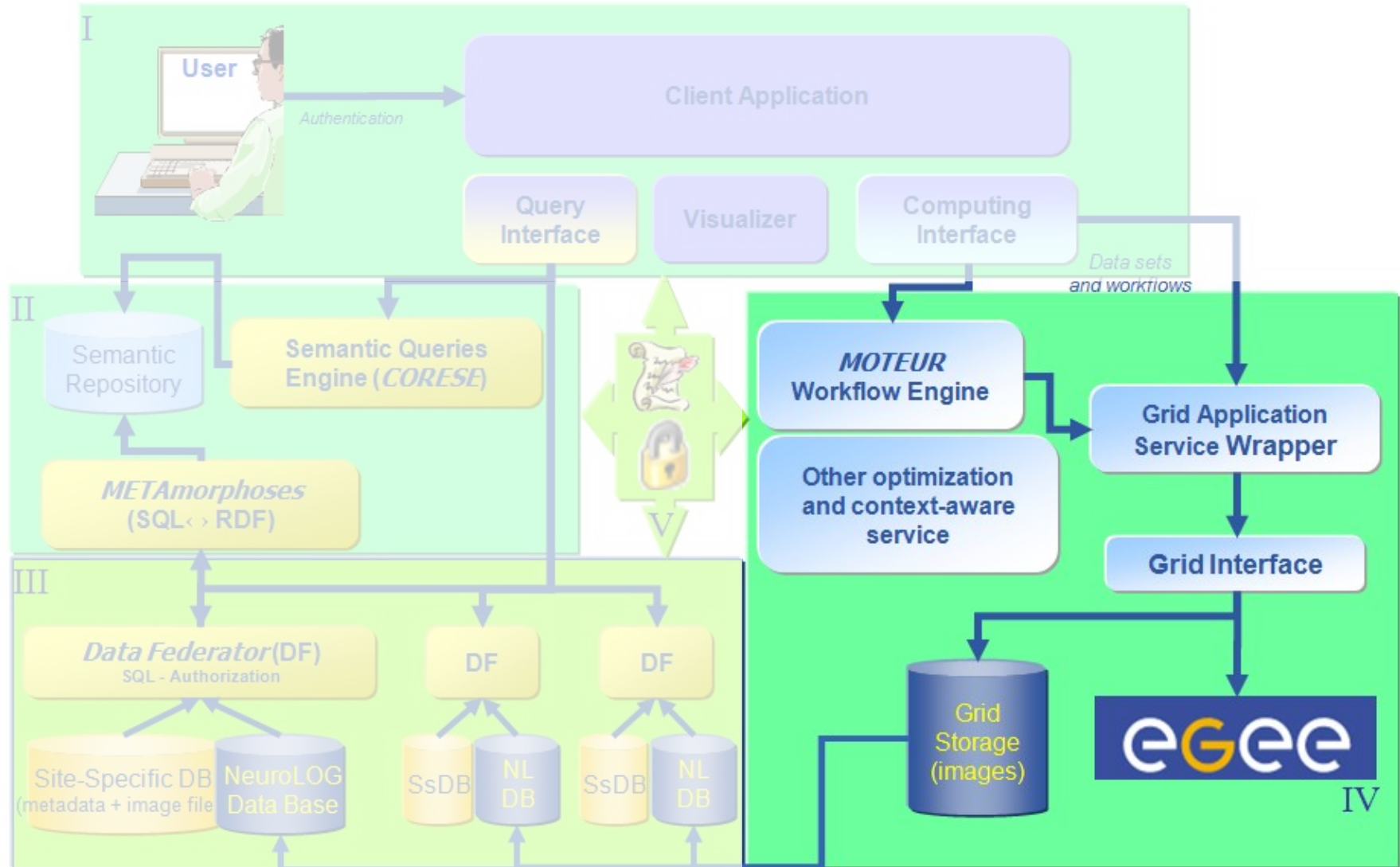
Software technologies for integration of process, data and knowledge in medical imaging



- **Data Federator: relational data mapping and federation tool (Business Object / SAP)**



Software technologies for integration of process, data and knowledge in medical imaging



NeuroLOG Embedding image analysis tools

Software technologies for integration of process, data and knowledge in medical imaging

- Rich description of tools and execution procedure
- Generates personalized service description for each application
- Dependencies manipulation
- Hot deployment
- Instrumentation of service execution according to the interface (grid or local)
- Strong mapping between types of services and arguments of application

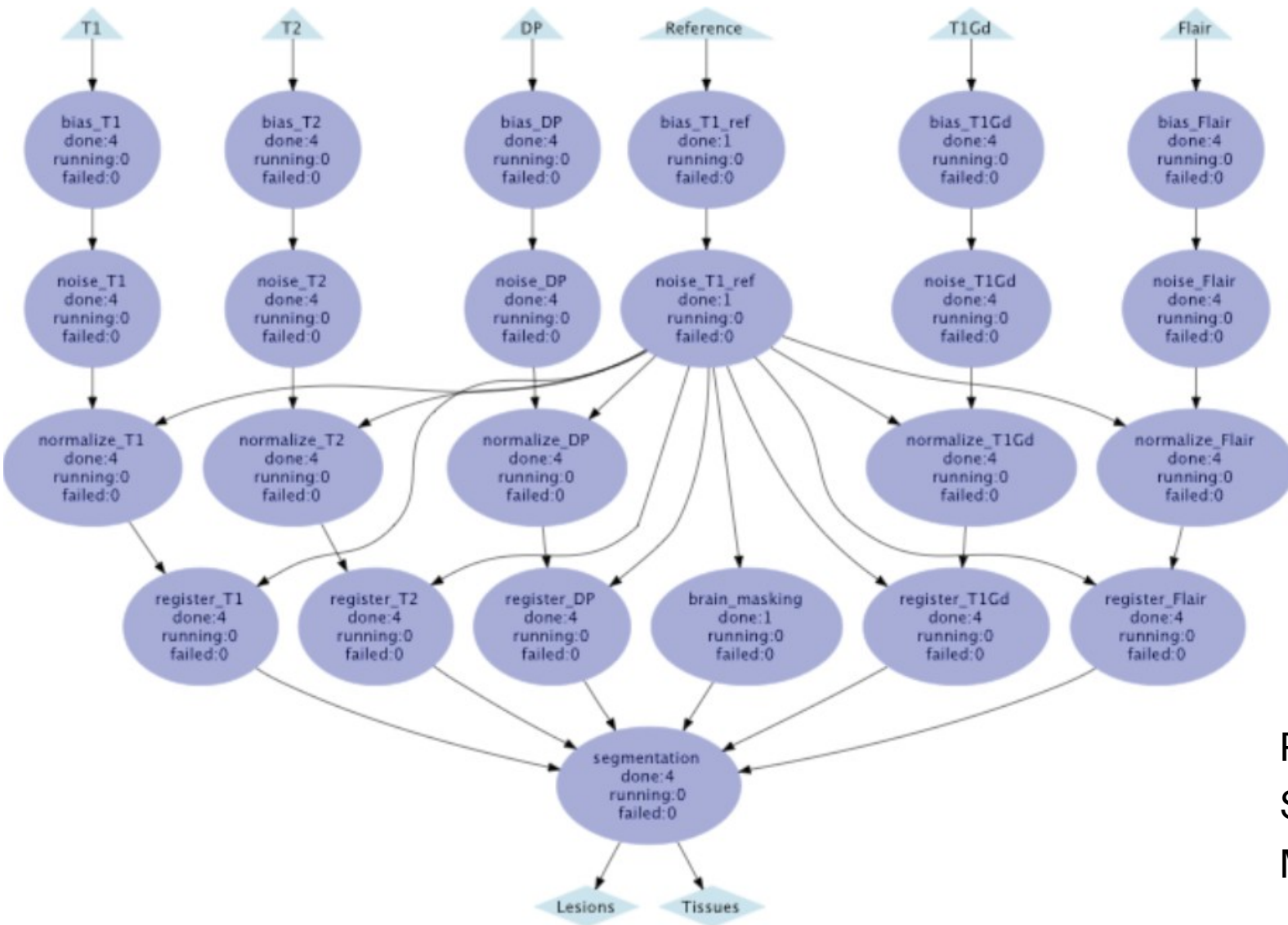
The screenshot displays the 'NeuroLOG I3S client (v0.3.0) - Javier Rojas (administrador)' window. The 'Add new tool' dialog is open, showing the following details:

- Executable: /opt/I3s/MOTEUR/data/baladin
- Service name: baladin
- Version: 1.0.0
- Vendor: (empty)
- Copyright: CeCILL-B
- Key: (empty)
- Address: (empty)
- Category: Application

The 'Argument' tab is selected, showing a list of arguments: reference [IN], image [IN], and output [OUT]. The 'Dependency' tab is also visible. At the bottom right, there are buttons for '+', '-', 'Grid Configuration', and 'Create Service'.

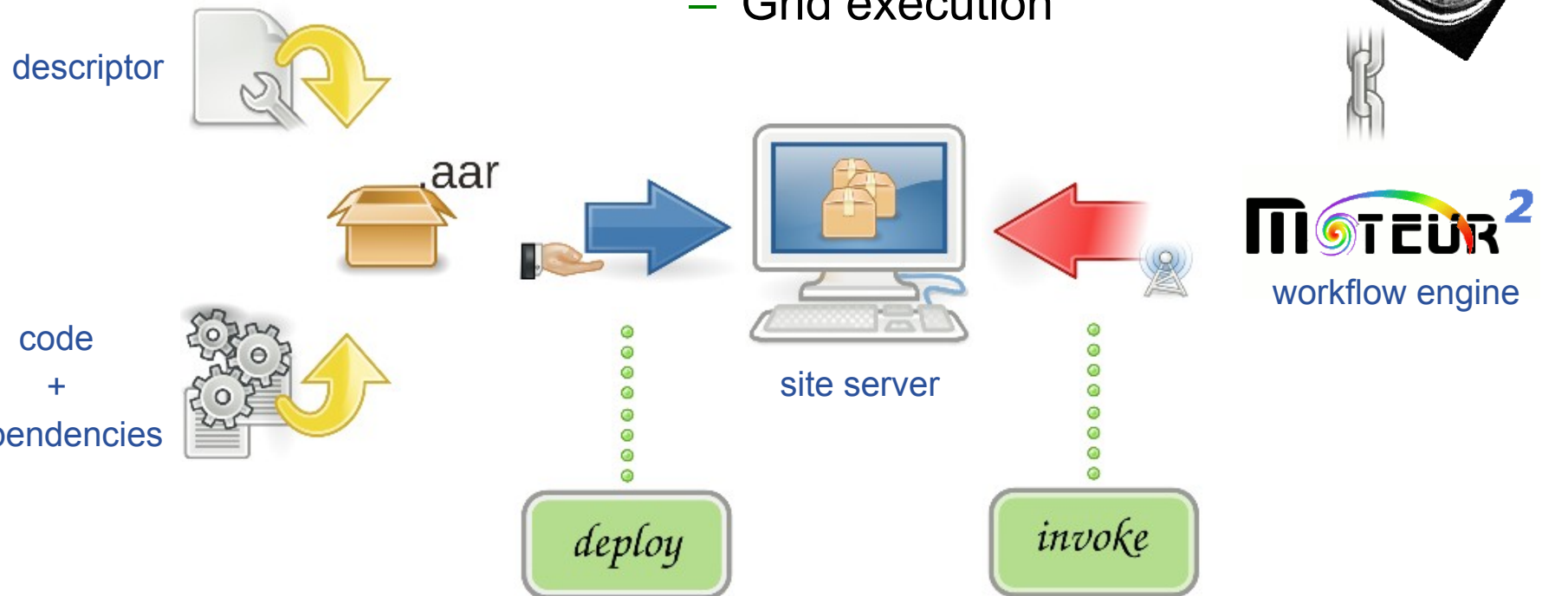
NeuroLOG Processing pipelines description

Software technologies for integration of process, data and knowledge in medical imaging

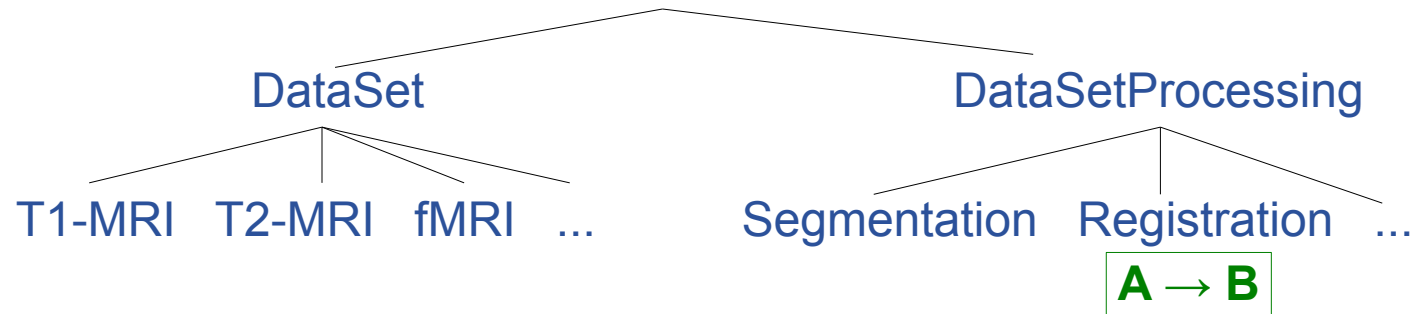


Pipeline
Segmentation
MS Virage

- **Interfaces for**
 - Local execution
 - Grid execution



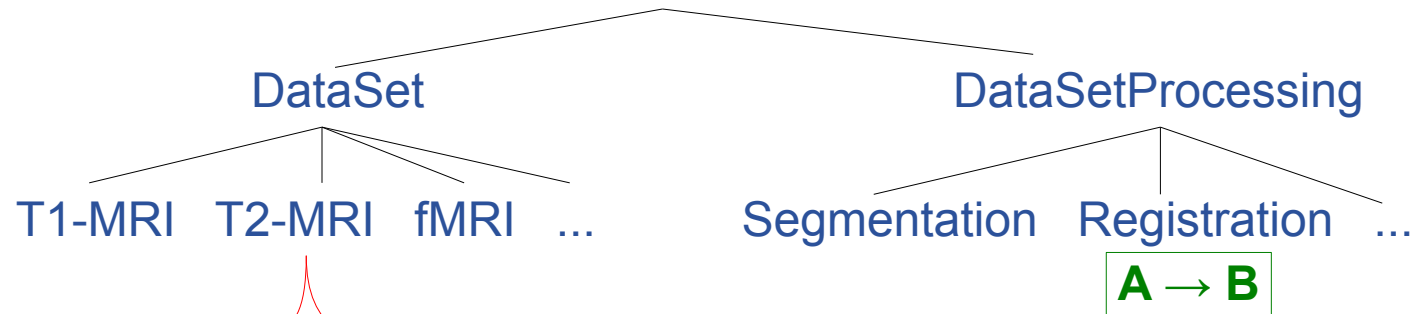
- **Ontology**
 - Concepts & **Rules**



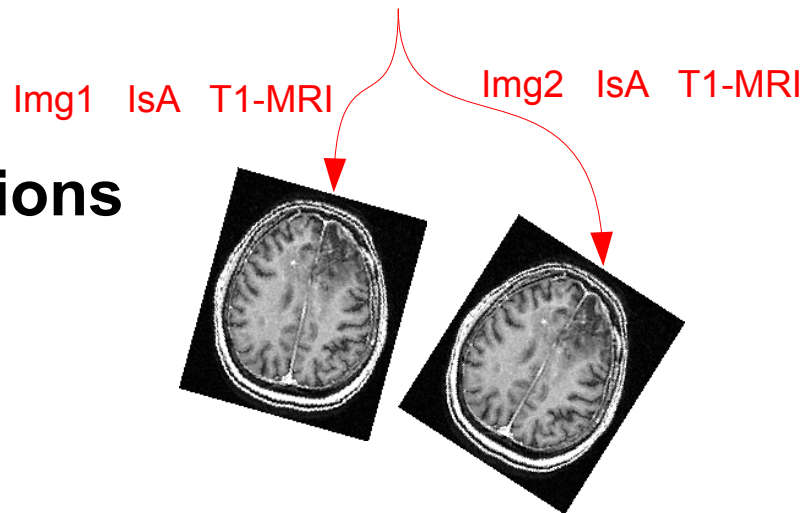
- **Annotations**

- **Processing**

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



- **Annotations**



- **Processing**

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



- **Annotations**

Img1 IsA T1-MRI

Img2 IsA T1-MRI

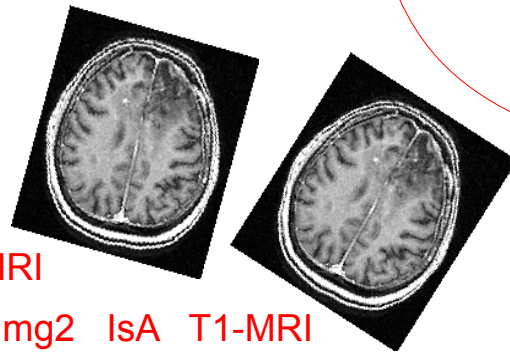
- **Processing**

Tool1 HasInput T1-MRI

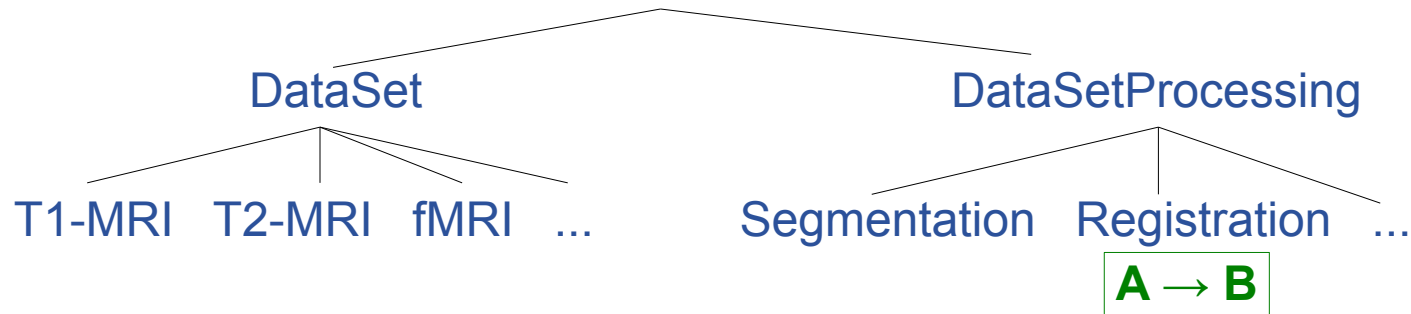
Tool1 IsA Registration

Tool1 HasOutput Transfo

A → B

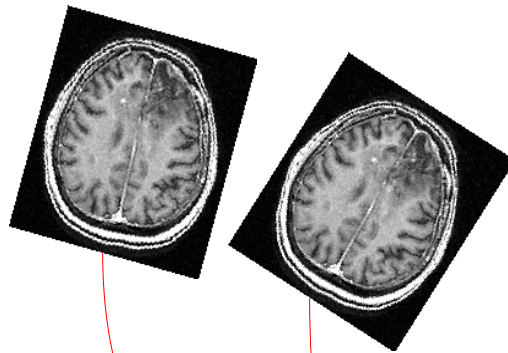


- **Ontology**
 - Concepts & **Rules**



- **Annotations**

Img1 IsA T1-MRI
 Img2 IsA T1-MRI

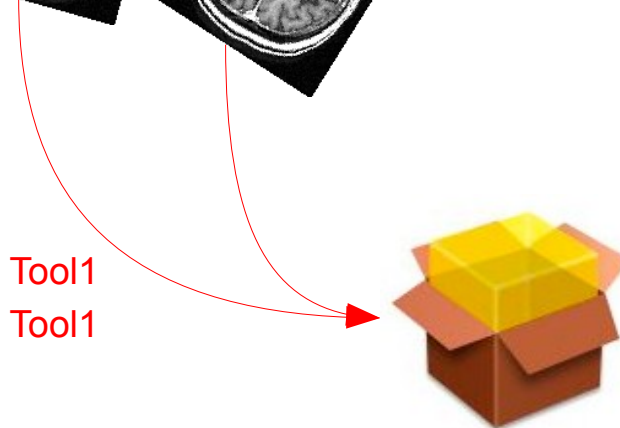


Tool1 HasInput T1-MRI
 Tool1 HasOutput Transfo
 Tool1 IsA Registration



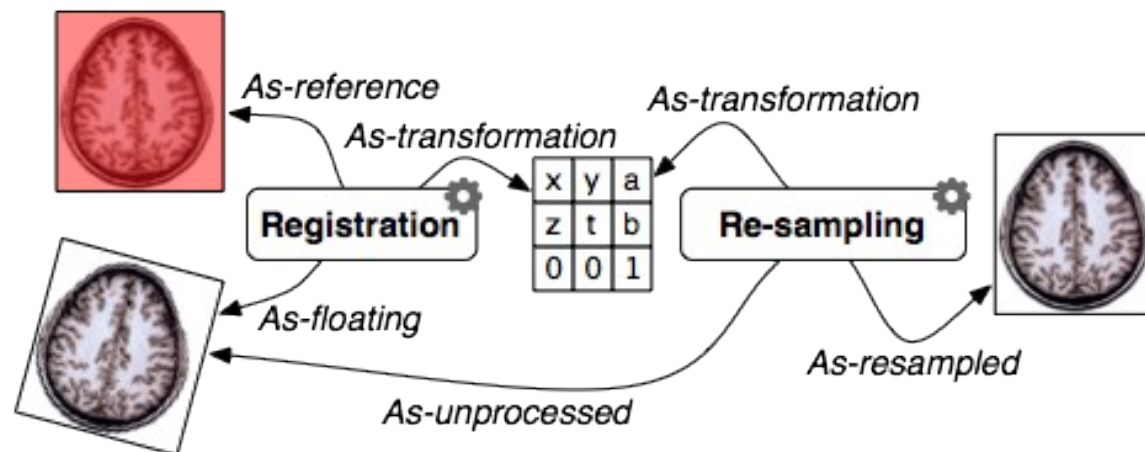
- **Processing**

Img1 IsProcessedBy Tool1
 Img2 IsProcessedBy Tool1

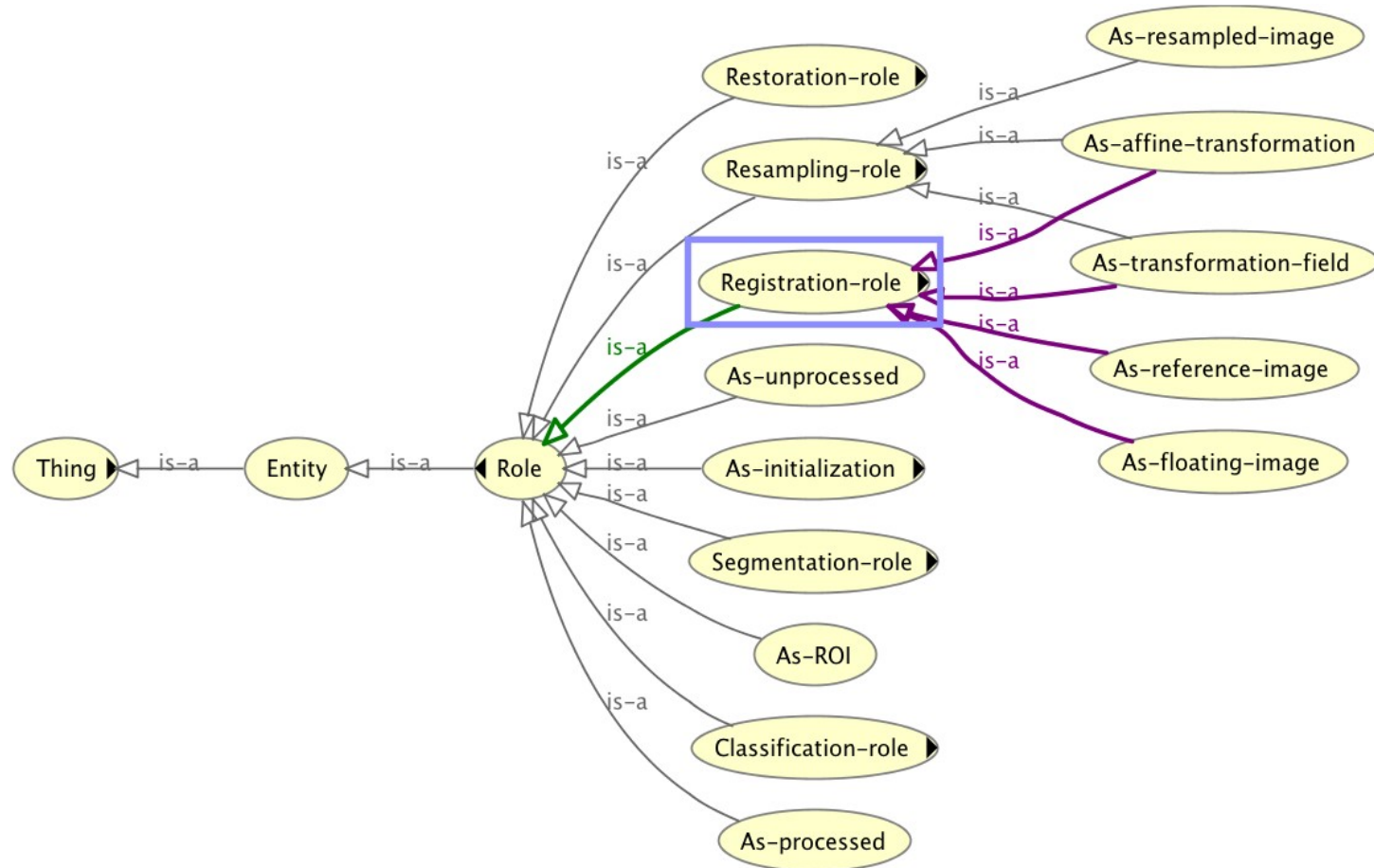


Tool1 Produced Transfo1
 Transfo1 IsA GlobalTransfo

- **Several levels of semantic information:**
 1. **Technical** information (OWL-S, OPM) ;
 2. Domain-specific **Nature** (data and services) ;
 3. Domain-specific **Role** of data from the service point of view.
- Roles to disambiguate the annotation of service parameters.
 - Example: the registration inputs may share the same modality, the same format but they differ by their relationship to the registration process.

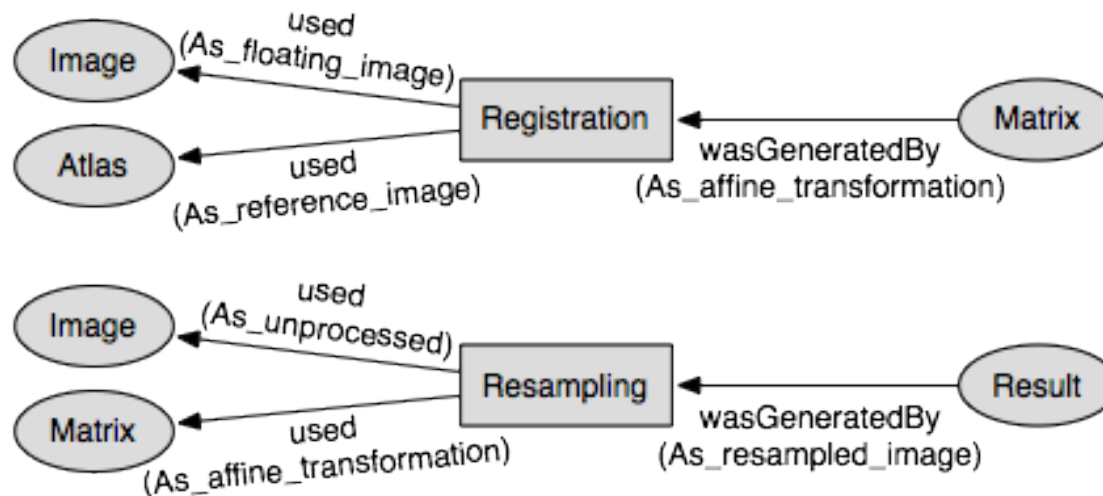


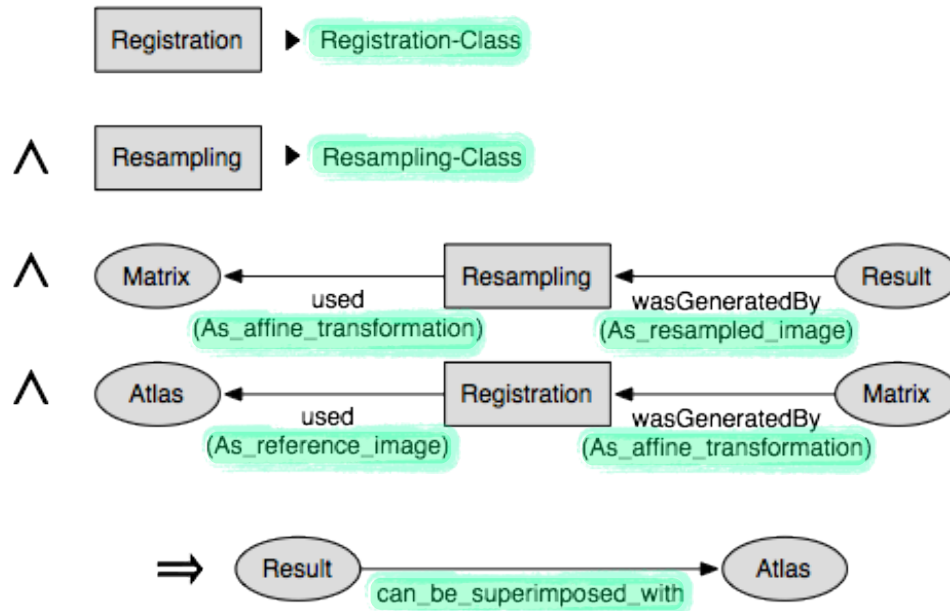
- Extension of the Open Provenance Model (OPM) Role class
- Taxonomy of domain-specific roles



- OPM instrumented workflow engine to **populate on-the-fly** the knowledge base with OPM **provenance** annotations.

- Sample invocation of the registration workflow:





```

<cosrule cosdebug="true">
<cos:if>
  PREFIX opav: <it;opav:igdt;
  PREFIX opao: <it;opao:igdt;
  PREFIX vs: <it;opav:igdt;
  PREFIX ds: <it;ds:igdt;
  PREFIX dp: <it;ds:igdt;
  PREFIX lcc: <it;ds:igdt;
  PREFIX neuro: <it;neuro:igdt;

  (
    ?usedReg rdf:type opav:Used
    ?usedReg opav:cause ?a1
    ?genReg rdf:type opav:WasGeneratedBy
    ?genReg opav:effect ?a2

    ?a1 rdf:type opav:Artifact
    ?a1 opav:ovalue ?v1
    ?v1 opav:content ?a1C
    ?a2 rdf:type opav:Artifact
    ?a2 opav:ovalue ?v2
    ?v2 opav:content ?a2C

    ?usedReg opav:role ?usedRole
    ?usedRole rdfs:label ?usedRoleLabel

    ?usedReg opav:effect ?regP
    ?genReg opav:cause ?regP
    ?a1bReg rdf:type opav:WasControlledBy
    ?a1bReg opav:effect ?regP
    ?a1bReg opav:cause ?regService
    ?regService lcc:refers-to ?IDP
    ?IDP rdf:type dp:Registration

    ?regService vs:has-output ?outReg
    ?outReg lcc:refers-to ?outRoleReg
    ?outRoleReg rdf:type neuro:light:AS-offline-transformation

    ?regService vs:has-input ?inReg
    ?inReg lcc:refers-to ?inRoleReg
    ?inRoleReg rdf:type neuro:light:AS-reference-image

    ?inReg rdfs:comment ?usedRoleLabel

    ?usedRes rdf:type opav:Used
    ?usedRes opav:cause ?a3
    ?genRes rdf:type opav:WasGeneratedBy
    ?genRes opav:effect ?a4

    ?a3 rdf:type opav:Artifact
    ?a3 opav:ovalue ?v3
    ?v3 opav:content ?a3C
    ?a4 rdf:type opav:Artifact
    ?a4 opav:ovalue ?v4
    ?v4 opav:content ?a4C

    ?usedRes opav:role ?resInPort
    ?resInPort rdfs:label ?resInPortLabel
    ?genRes opav:role ?genResRole
    ?genResRole rdfs:label ?genResRoleLabel

    ?usedRes opav:effect ?resP
    ?genRes opav:cause ?resP
    ?a1bRes rdf:type opav:WasControlledBy
    ?a1bRes opav:effect ?resP
    ?a1bRes opav:cause ?resService
    ?resService lcc:refers-to ?IDP2
    ?IDP2 rdf:type dp:Resampling

    ?resService vs:has-input ?inRes
    ?inRes lcc:refers-to ?inRoleRes
    ?inRoleRes rdf:type neuro:light:AS-offline-transformation

    ?resService vs:has-output ?outRes
    ?outRes lcc:refers-to ?outRoleRes
    ?outRoleRes rdf:type neuro:light:AS-resampled-image

    ?outRes rdfs:comment ?genResRoleLabel

    FILTER (?a2C = ?a3C)
  )
</cos:if>
<cos:then>
  (
    ?a4 ds:can-be-superimposed-with ?a1
  )
</cos:then>
</cos:rules>
  
```

*Verbose (95 lines) and error-prone !
Reusability !*

- **Semantic-only Information System**
 - Feasibility study
 - Map relational data to ontology concept
 - Based on Semantic Web standards (OWL / RDF / SPARQL)
 - Using the CORESE inference engine
 - KGRAM semantic data producers
 - Can mix SQL and SPARQL queries
- **Semantic data store distribution**
 - Implementation of a Web Service interface to KGRAM producers to enable remote data stores
 - Optimization of the query master to take into account distribution
 - Feasibility demonstrated, performance study on-going