



Grid-wide neuroimaging data federation in the context of the NeuroLOG project

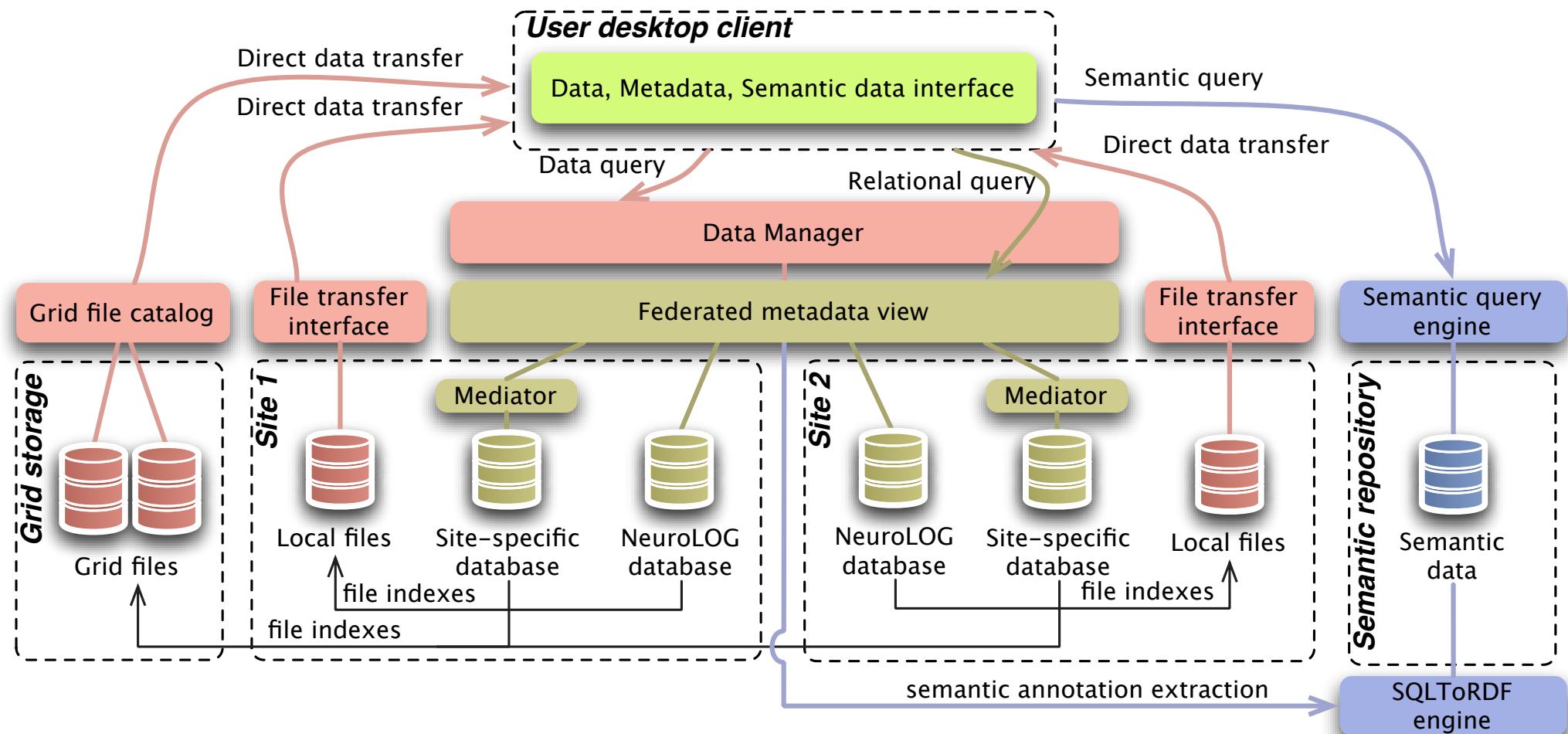
Franck MICHEL, Alban GAIGNARD, Farooq AHMAD, Christian BARILLOT, Bénédicte BATRANCOURT, Michel DOJAT, Bernard GIBAUD, Pascal GIRARD, David GODARD, Gilles KASSEL, Diane LINGRAND, Grégoire MALANDAIN, Johan MONTAGNAT, Mélanie PÉLÉGRINI-ISSAC, Xavier PENNEC, Javier ROJAS BALDERRAMA, Bacem WALI.

HealthGrid'2010 Conference
29 of june 2010, Paris, France

Objectives and methods

- NeuroLOG: middleware for neuro-sciences supporting multi-centric studies
 - Data federation is a key component
- Objectives of this work: design a **distributed neuroradiological data manager**
 - Neuro-radiology data stores are pre-existing and should not be altered (**sites autonomy**);
 - Radiology image files are completed with extensive descriptive metadata (**heterogenous data schemas** requiring a **common view**);
 - Data schemas are site-specific and do not share a common semantic a priori.
- Methods:
 - A common semantics is implemented through a domain Ontology, which is derived into a relational **Federated Schema** ;
 - A **dynamic mediation interface** maps local schemas to the federated schema ;
 - A **global federated view** hides data distribution to the end-user.

Data management layer architecture 1/2



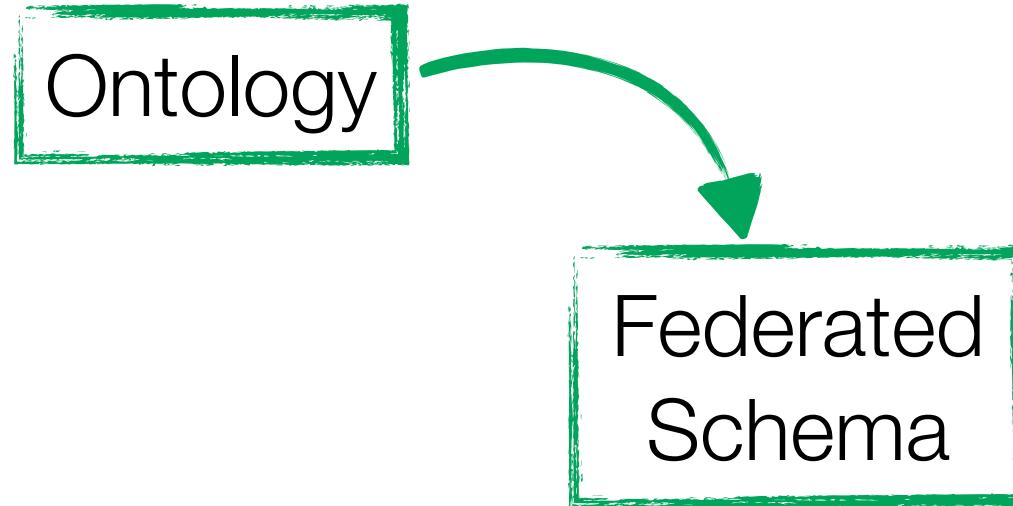
Federated Schema 1/2

- Federating heterogeneous databases => common semantics ?
- The *OntoNeuroLOG* ontology as a cornerstone
 - Designed through a sound methodological approach (using DOLCE foundational ontology)
- Implemented through OWL-Lite and a relational schema.
 - Limitation of relational representation.

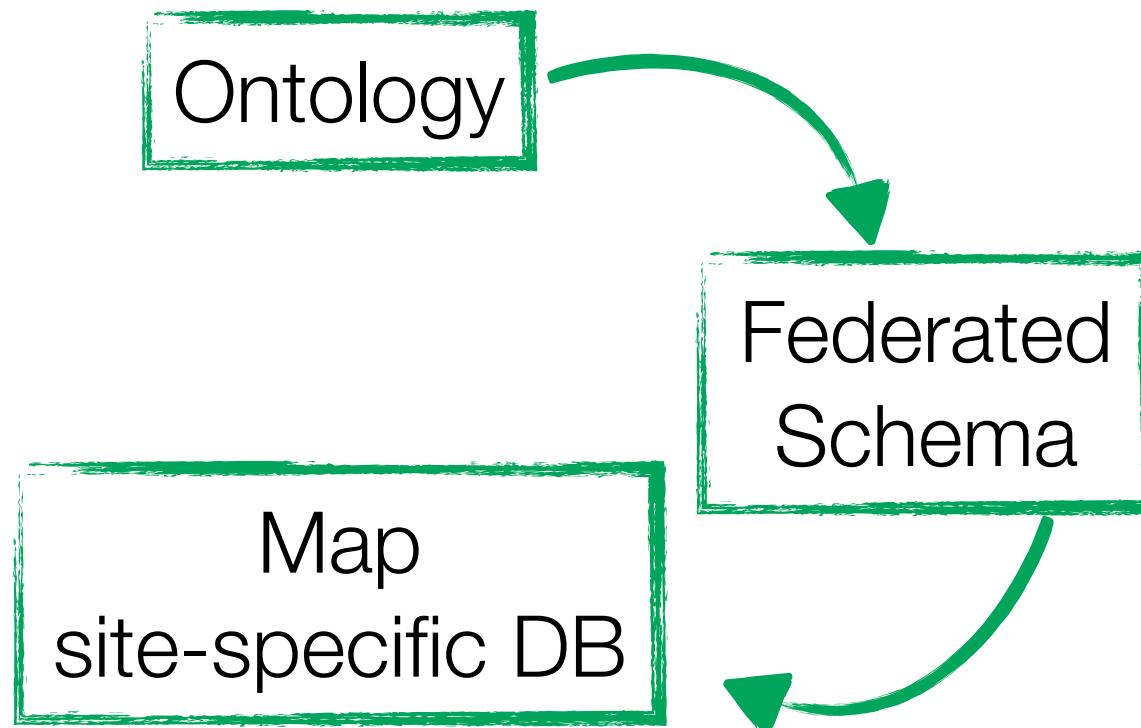
Federated Schema 2/2

Ontology

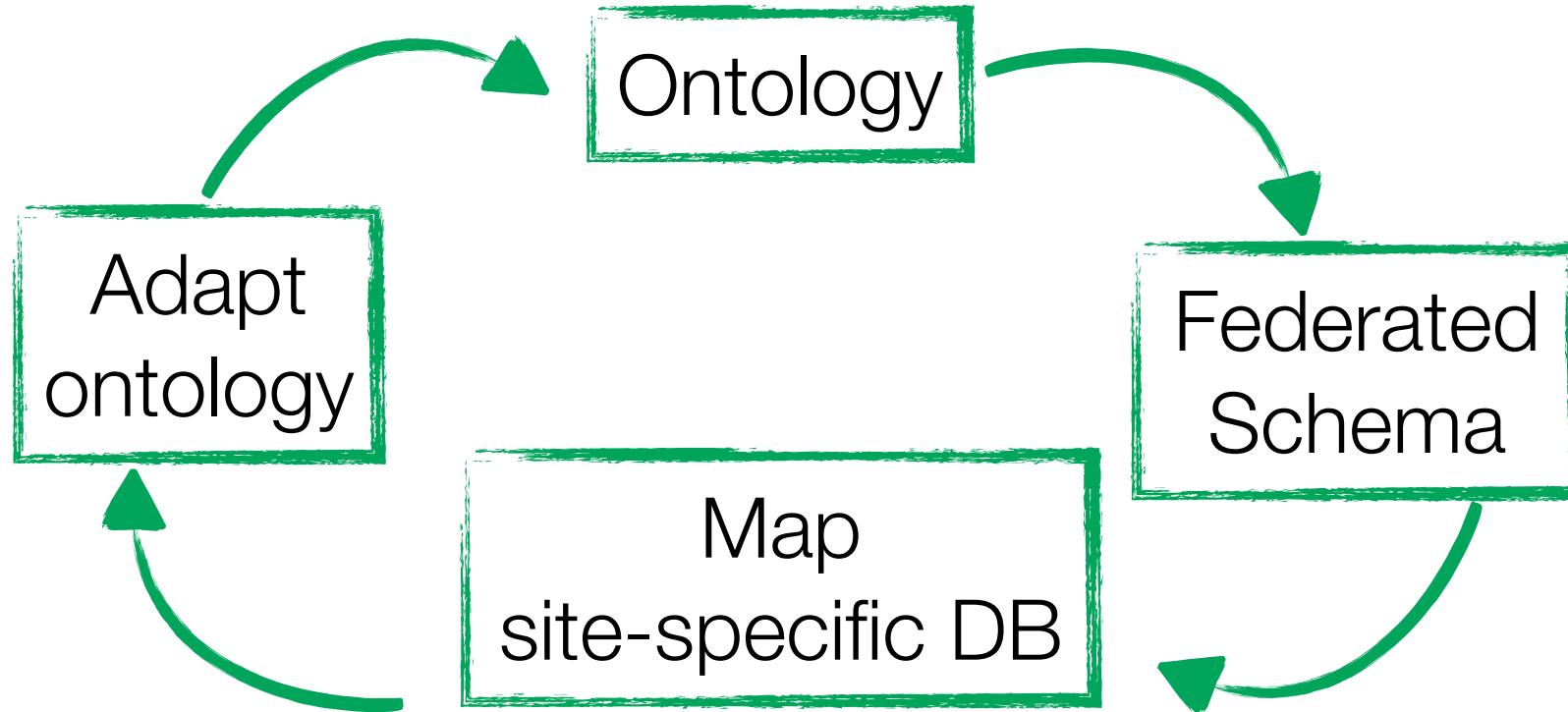
Federated Schema 2/2



Federated Schema 2/2

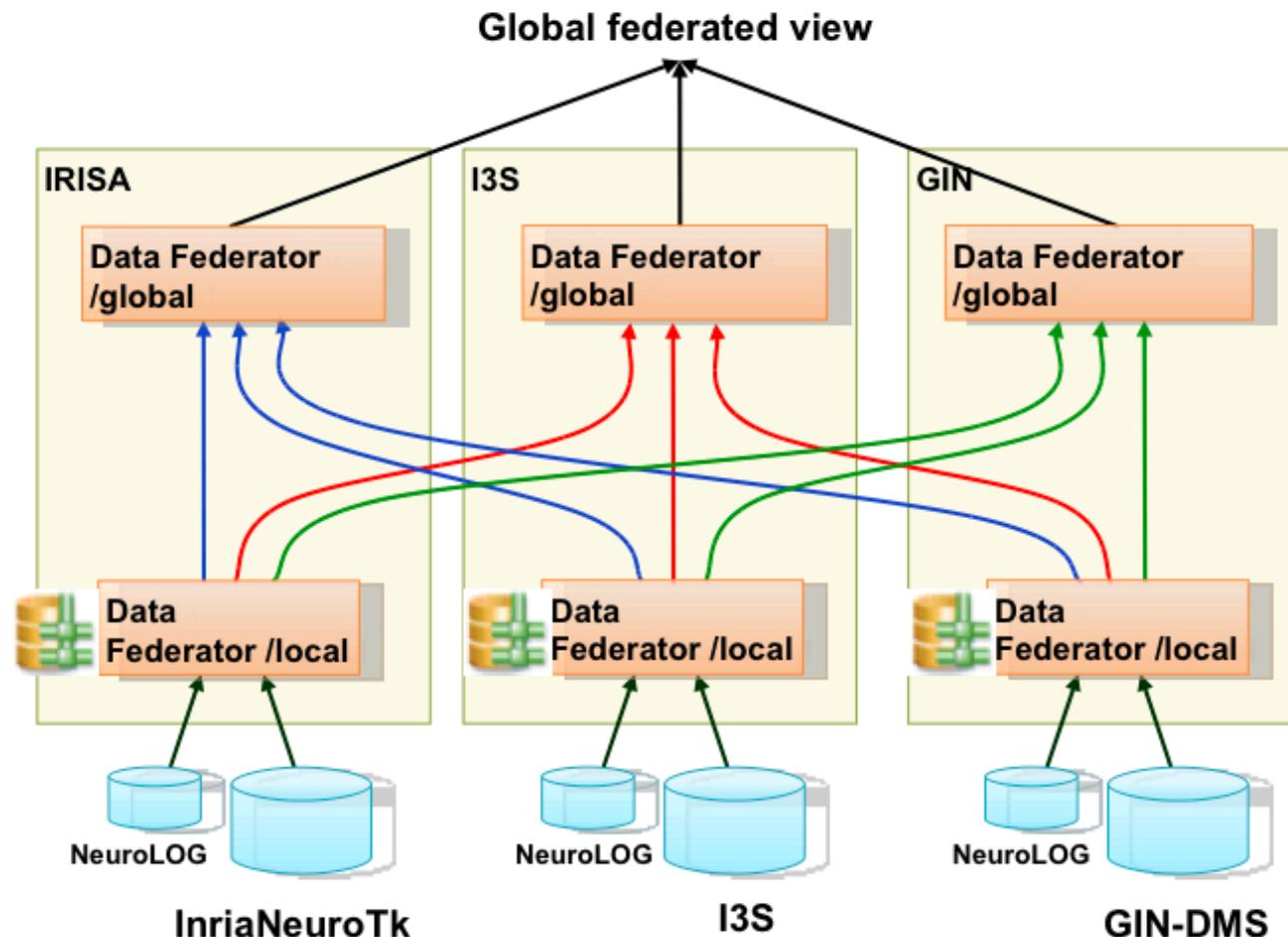


Federated Schema 2/2



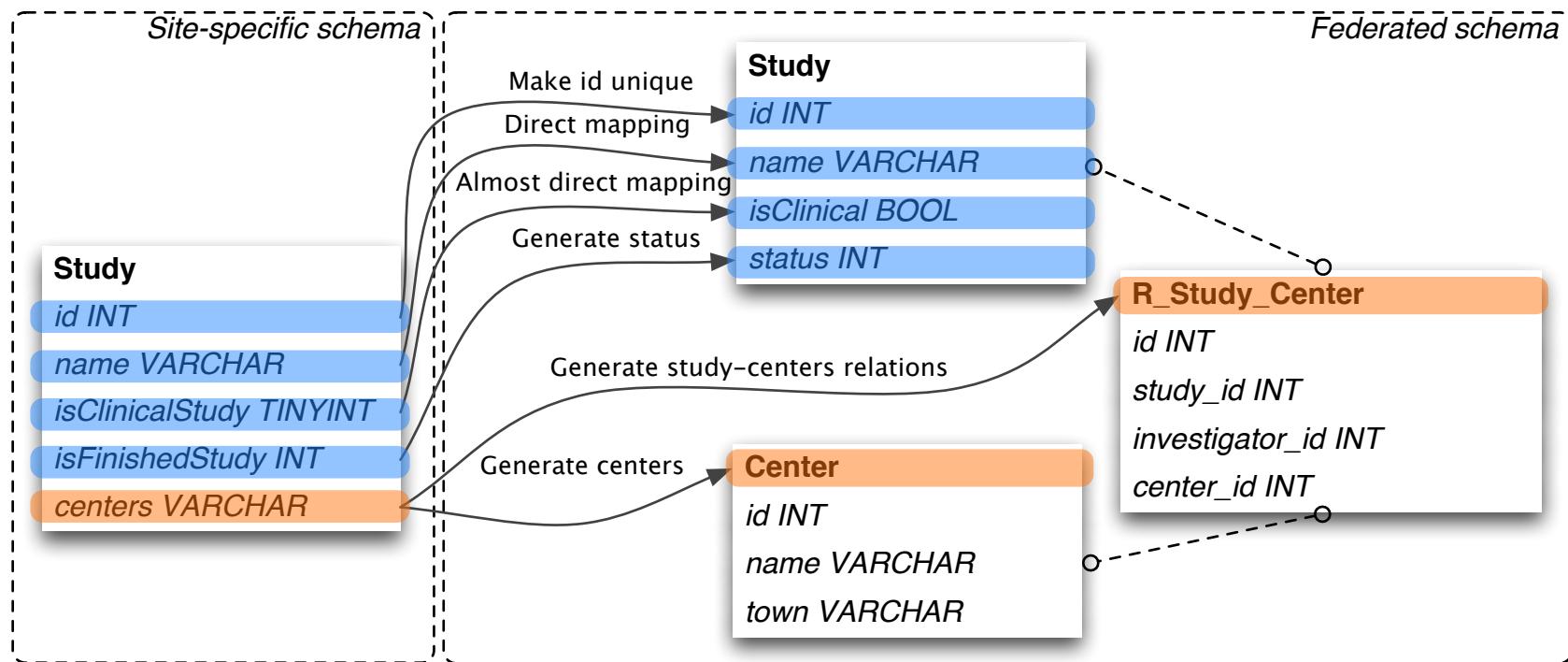
Mediation Layer 1/2

- DataFederator tool (*BusinessObject - SAP*) to distribute and rewrite (+optimization), relational queries:



Mediation Layer 2/2

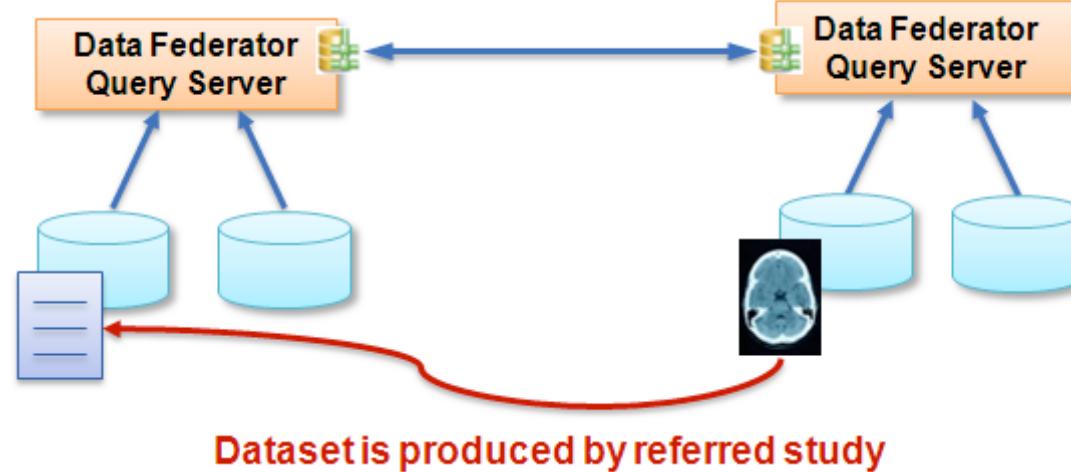
- Alignment of site-specific relational schema to the federated schema.
- Manual process, by local database expert, can be straightforward or rather complex::



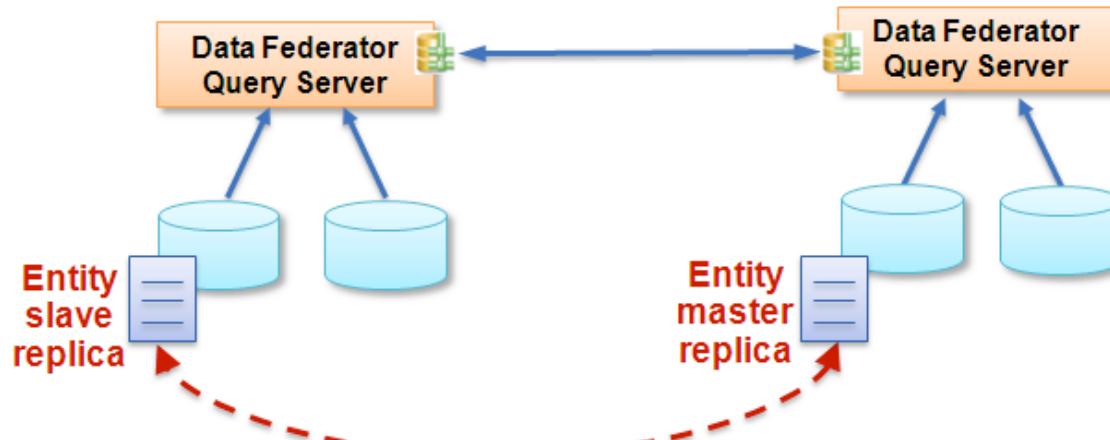
- Questions raised about loose of information in mappings, or lack of semantics in the site-specific schema. Potentially new iteration in the Federated Schema design.

Coherency enforcement of distributed metadata

- Cross-site references between entities: federation-wide identifiers

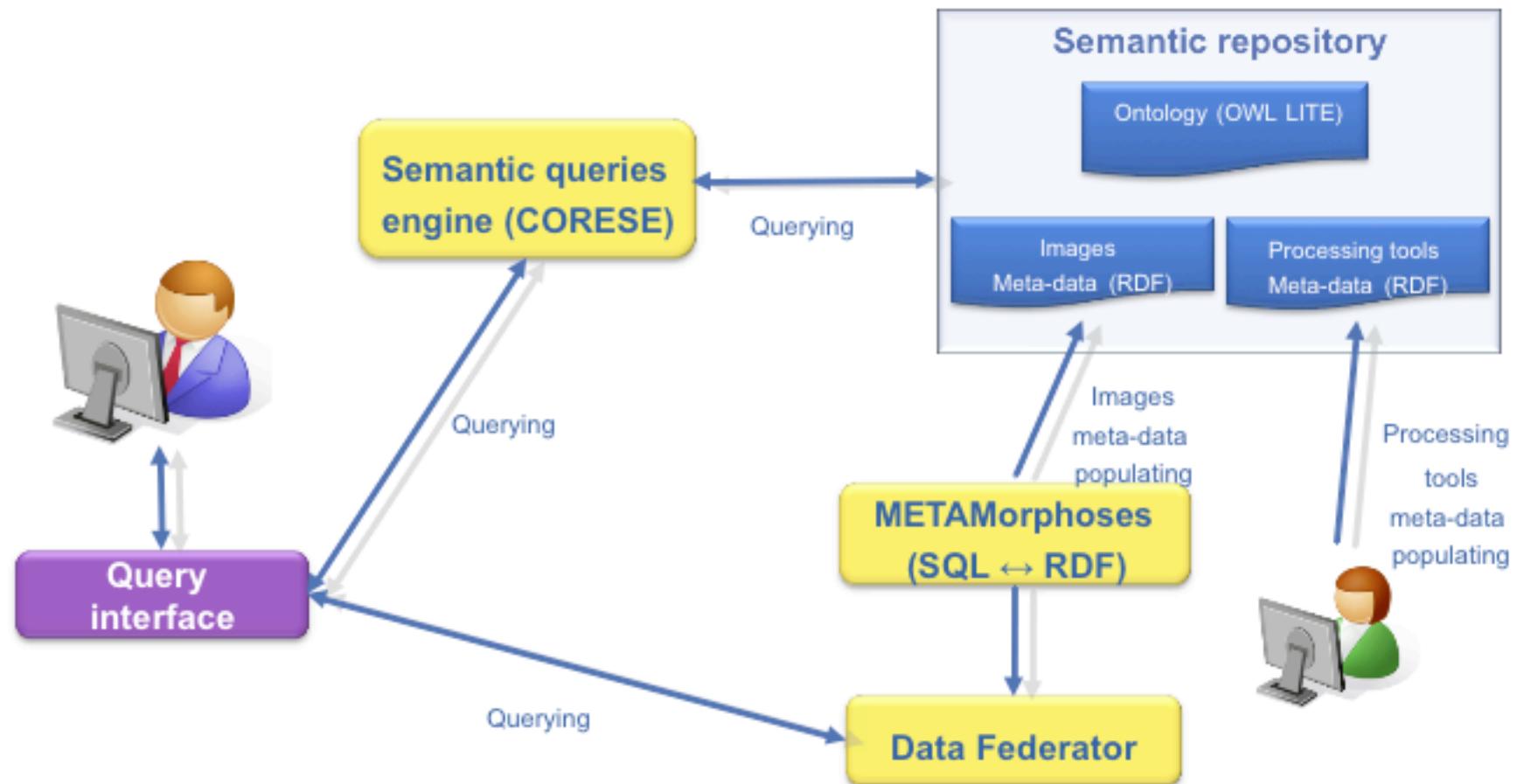


- Multiple instance of entities: master/slave entities



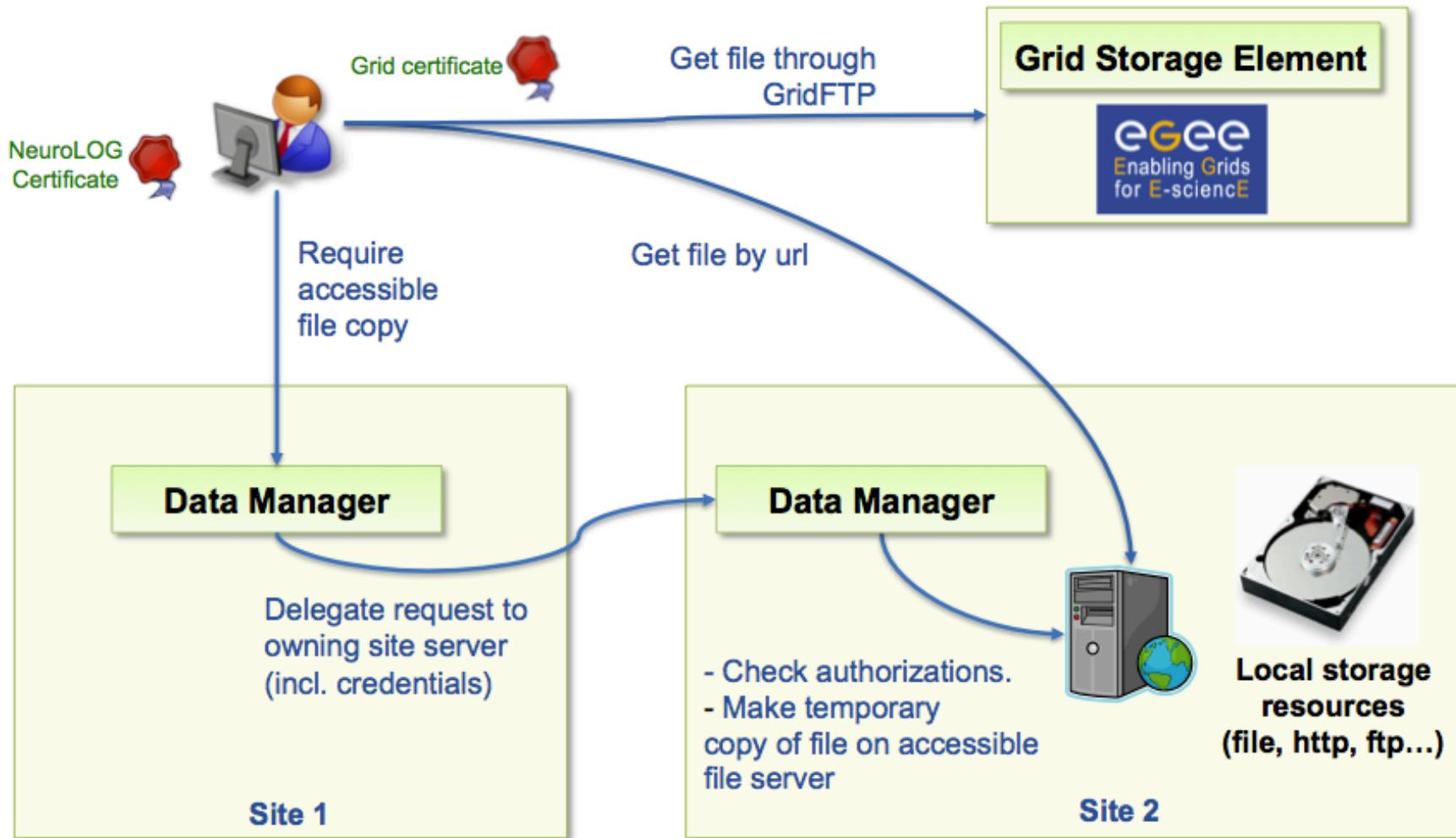
- Curation procedure: periodically detect and resolve inconsistencies

Semantic repository



Distribution of files

- Files are exposed for a direct access (NeuroLOG / GLite GUID identifiers)



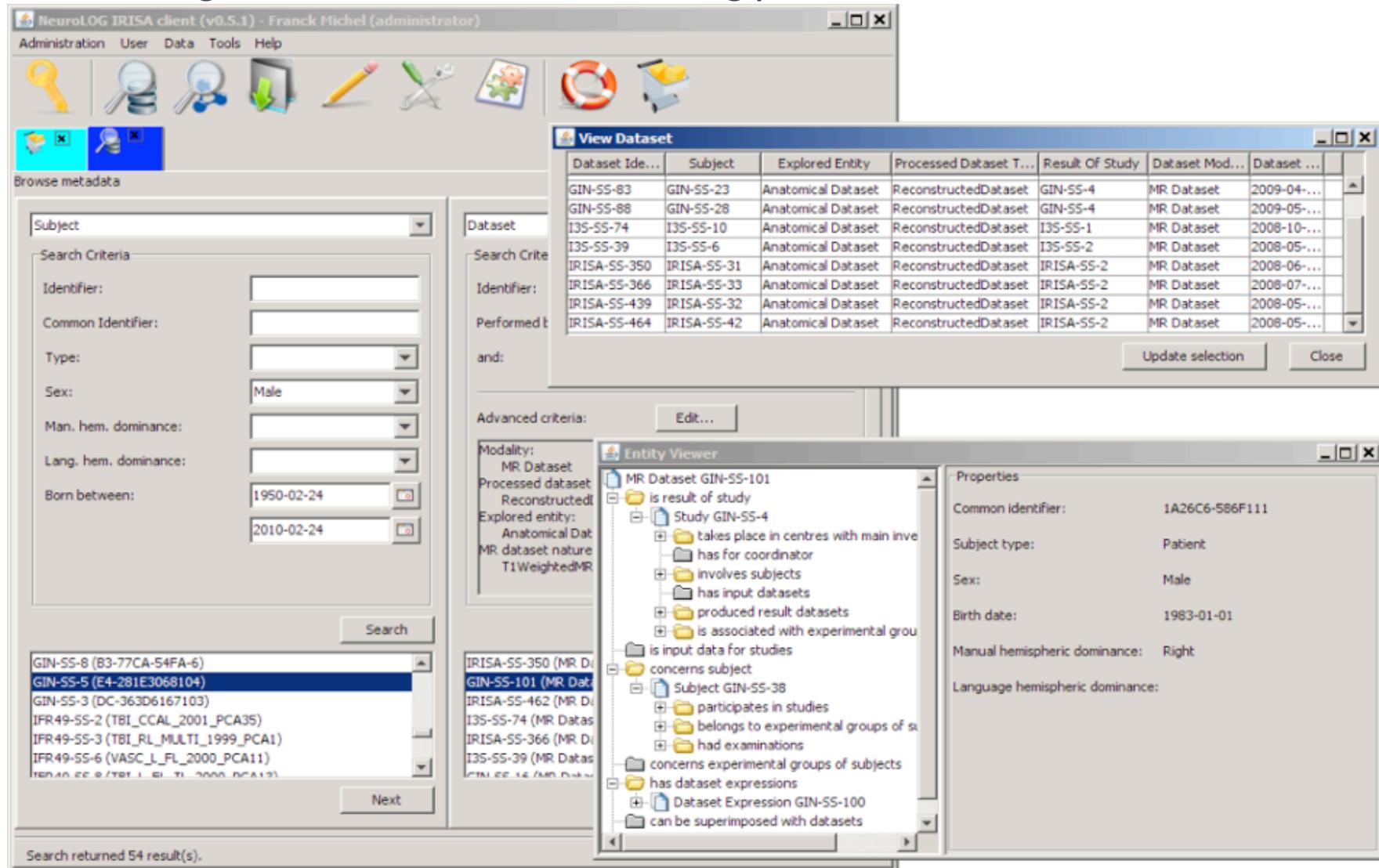
Platform deployment 1/2



- 5 sites interfaced:
 - I3S (Sophia Antipolis) core technical site ;
 - IRISA (INRIA Rennes), collaborating with the University Hospital of Rennes ;
 - IFR49 (INSERM affiliated neuroscience group in Paris La Pitié Salpêtrière Hospital) ;
 - GIN (INSERM affiliated neurosciences institute of Grenoble, Michalon Hospital) ;
 - INRIA Sophia Antipolis (Centre Antoine Lacassagne).
- Clients connect from anywhere.

Platform deployment 2/2

- Platform usage scenario / metadata browsing path :



Conclusion

- Ease the setup of multi-centric studies:
 - Common **semantics** (domain ontology design) ;
 - Hide the **heterogeneity** (metadata mediation and federation) ;
 - Hide the **distribution** (metadata browsing) ;
- Autonomy property is necessary to foster HealthGrid adoption by clinical user communities.
 - Non intrusive, preserve **legacy environments**
 - Yet enabling semantically aligned data sharing