NeuroLOG Software technologies for integration of process and data in medical imaging

A transitional middleware to support neurosciences on the EGEE grid

Alban Gaignard, Javier Rojas Balderrama,
Franck Michel, David Godard,
Diane Lingrand, Johan Montagnat

CNRS – I3S Laboratory
Sophia-Antipolis, France

4th EGEE User Forum
Catania – Sicily - Italy

Financé par ANR

Motivations

- **Target neuroscientist community**
  - Neuroscience has a major societal impact
    - Understanding and improving healthcare of brain diseases
    - Discovering brain functions
  - “Good candidate” for grid technologies
    - Huge consumer of computational resources (neuroimaging)
    - Collaborative work, multidisciplinarity → transdisciplinarity
- **Ease multi-centric studies**
  - Large population datasets, growing size and heterogeneity
  - Complex analysis pipeline, inter-operability
  - Distribution in a wide scale environment
- **Consider existing environments !**
  - Processing tools, databases, practices

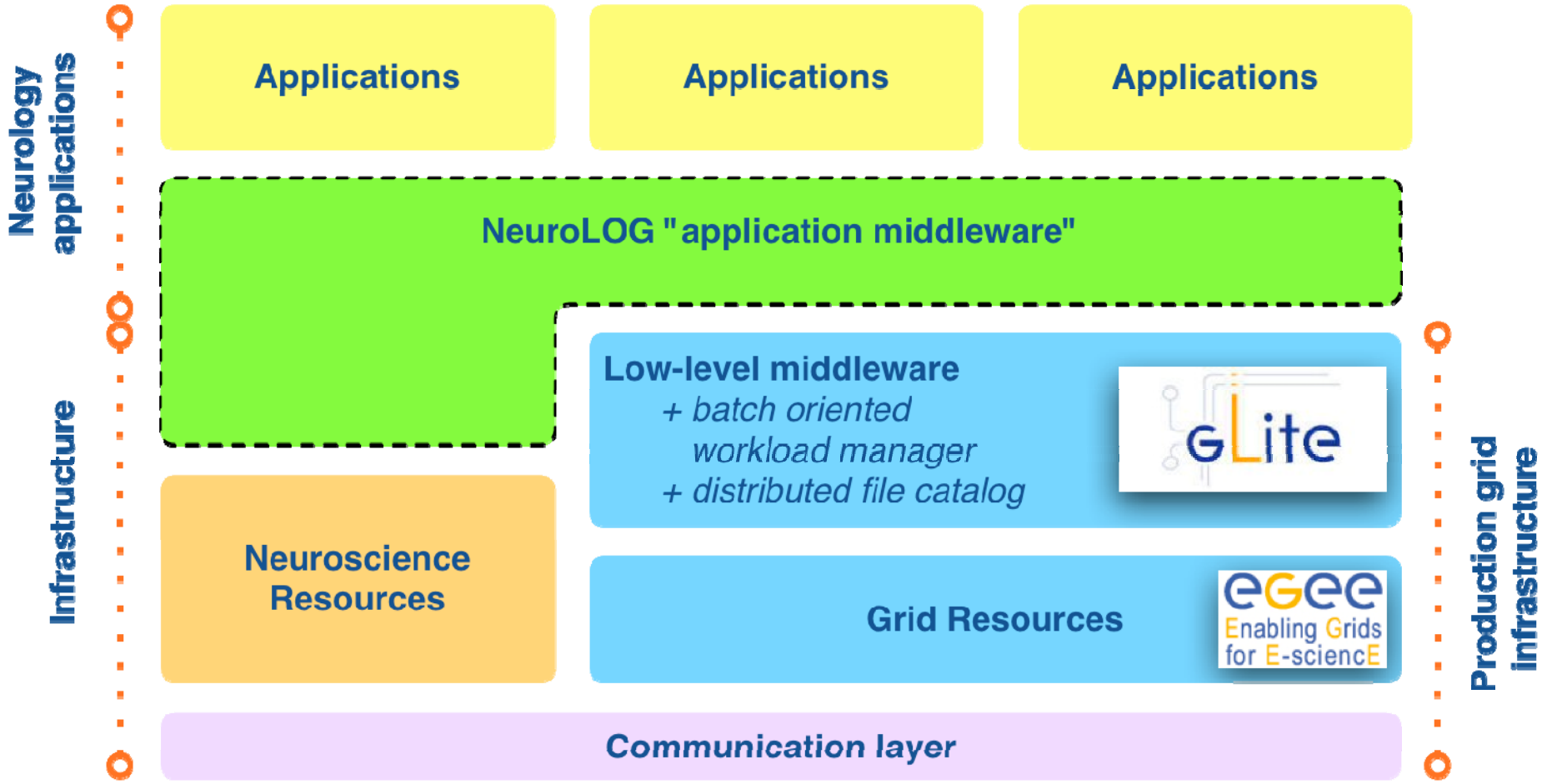
Approach 1/2

- **Leverage grid technologies:**
  - Bridge between EGEE and neuroscience research centers
  - Federating existing data
  - Repository of processing tools
  - Virtual Organizations concepts
- **Knowledge representation**
  - Design of a domain ontology (rich queries and representation)
- **Software integration**
  - Data integration (raw files, metadata, semantic data)
  - Application workflows
    - Perspectives of Interaction between WFs and data
  - Usability of tools by non computer experts … and even less grid technologies experts

Approach 2/2

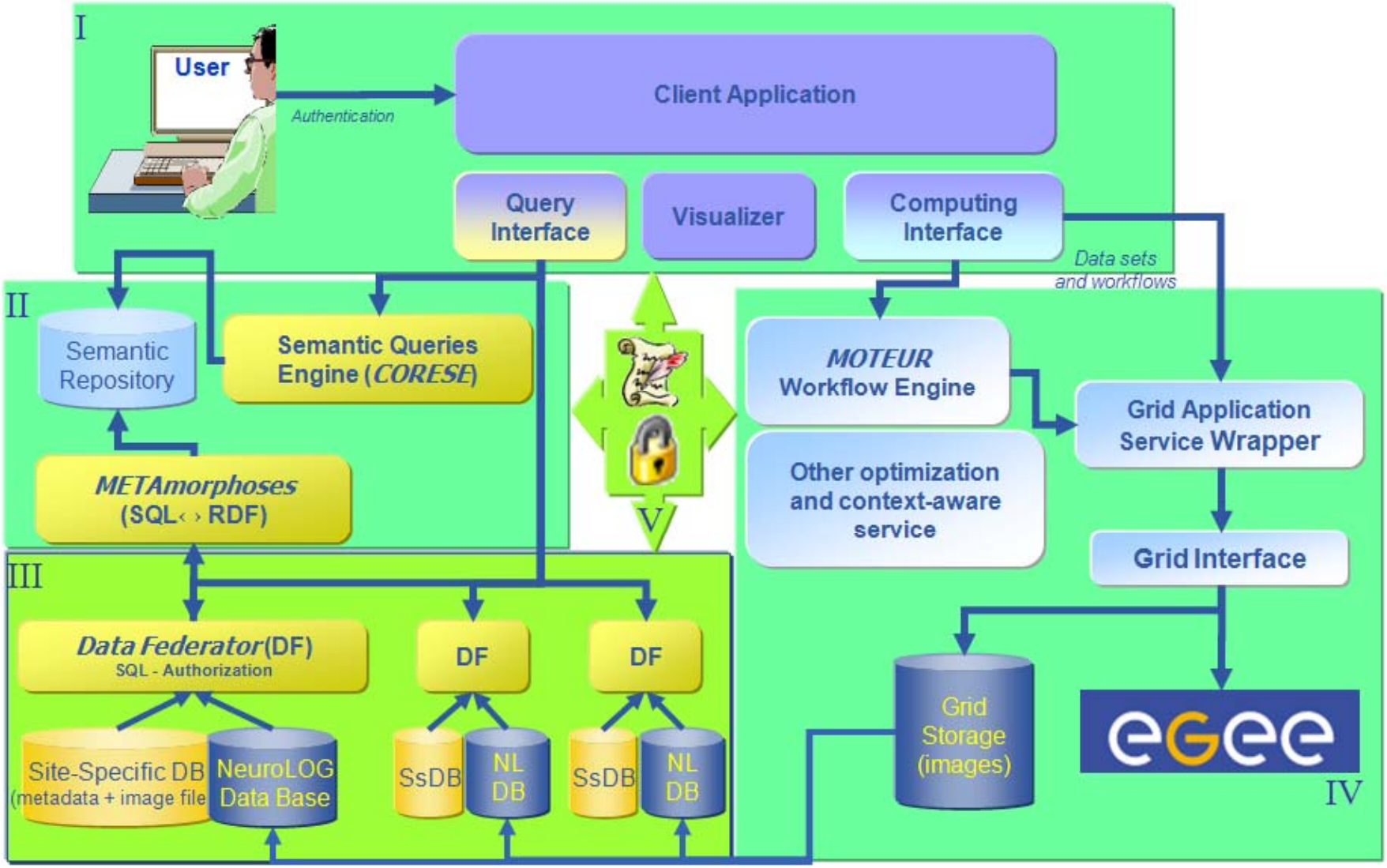
- **Collaborative Neuroscience**
  - Focus on 3 clinical applications
    - Multiple sclerosis
    - Brain stroke
    - Brain tumours
  - Commonalities
    - Neuroimages (MR)
    - Descriptive information
    - Processing tools: Registration, Skull stripping, Normalization, Tissues classification
  - Infrastructure used to
    - share databases
    - share and evaluate processing tools

Design: bridging the gap



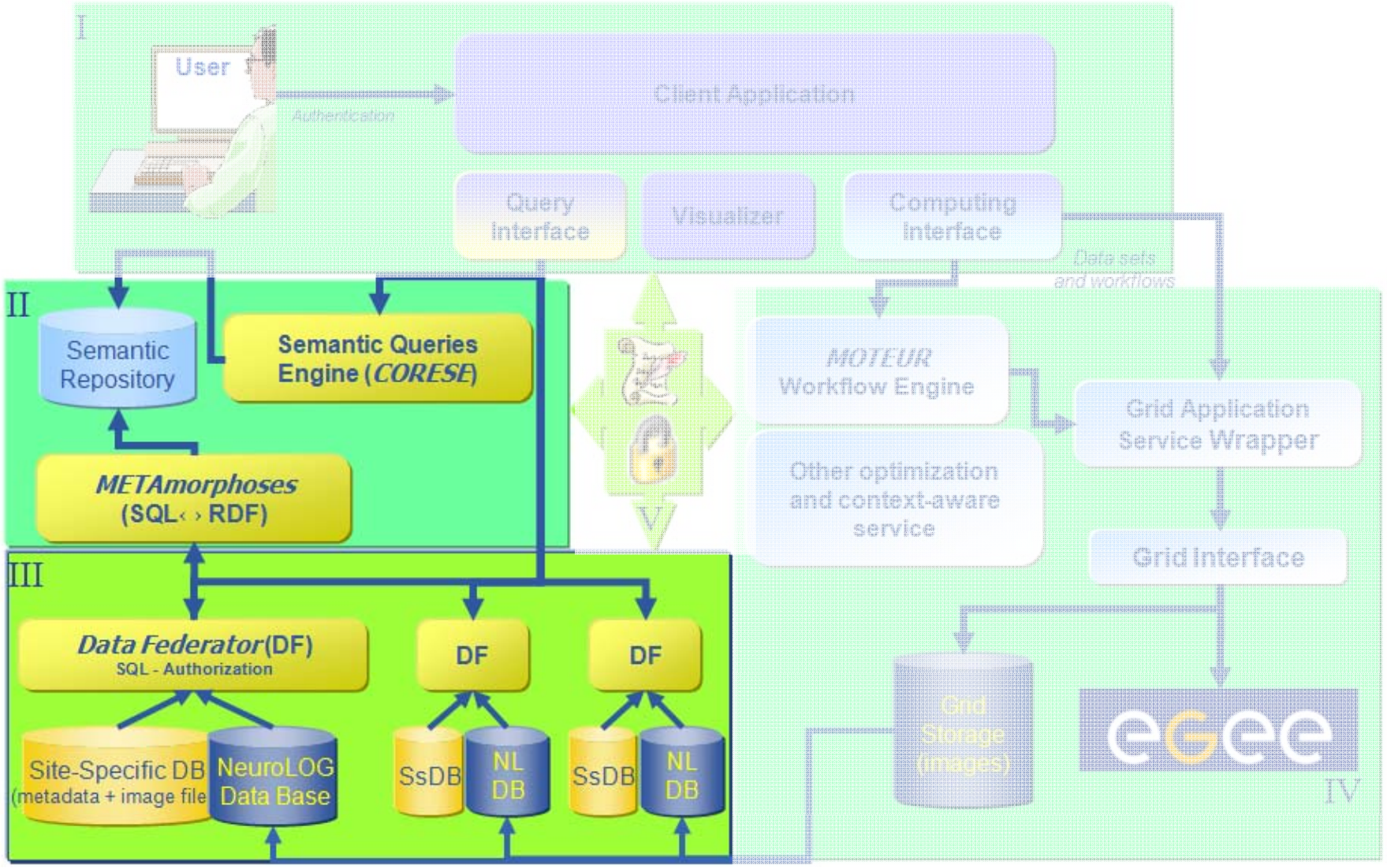
NeuroLOG Design: General Architecture

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



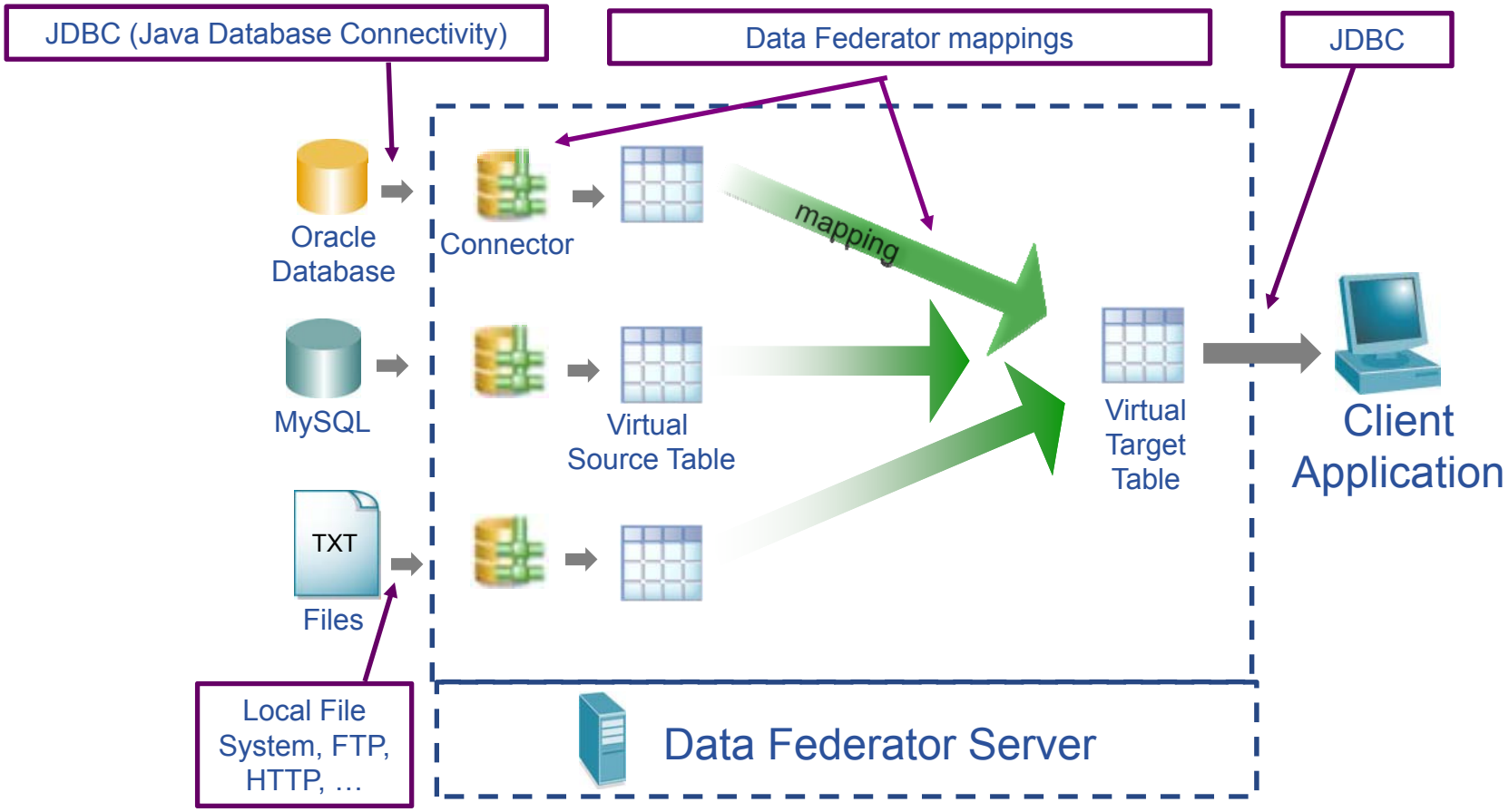
Data Management 1/5

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



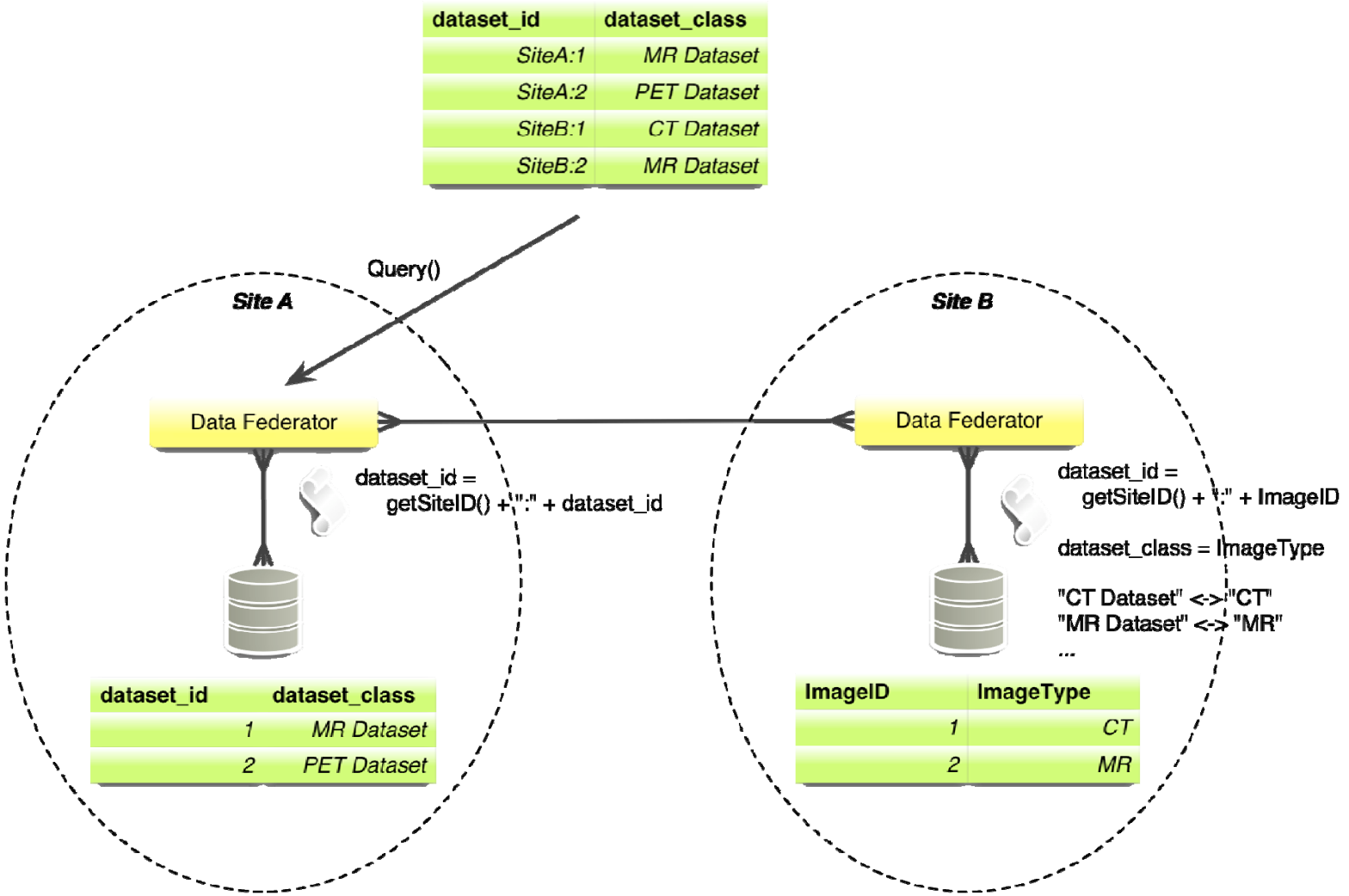
Data Management 2/5

- **Data Federator tool (Business Object - SAP)**



Data Management 3/5

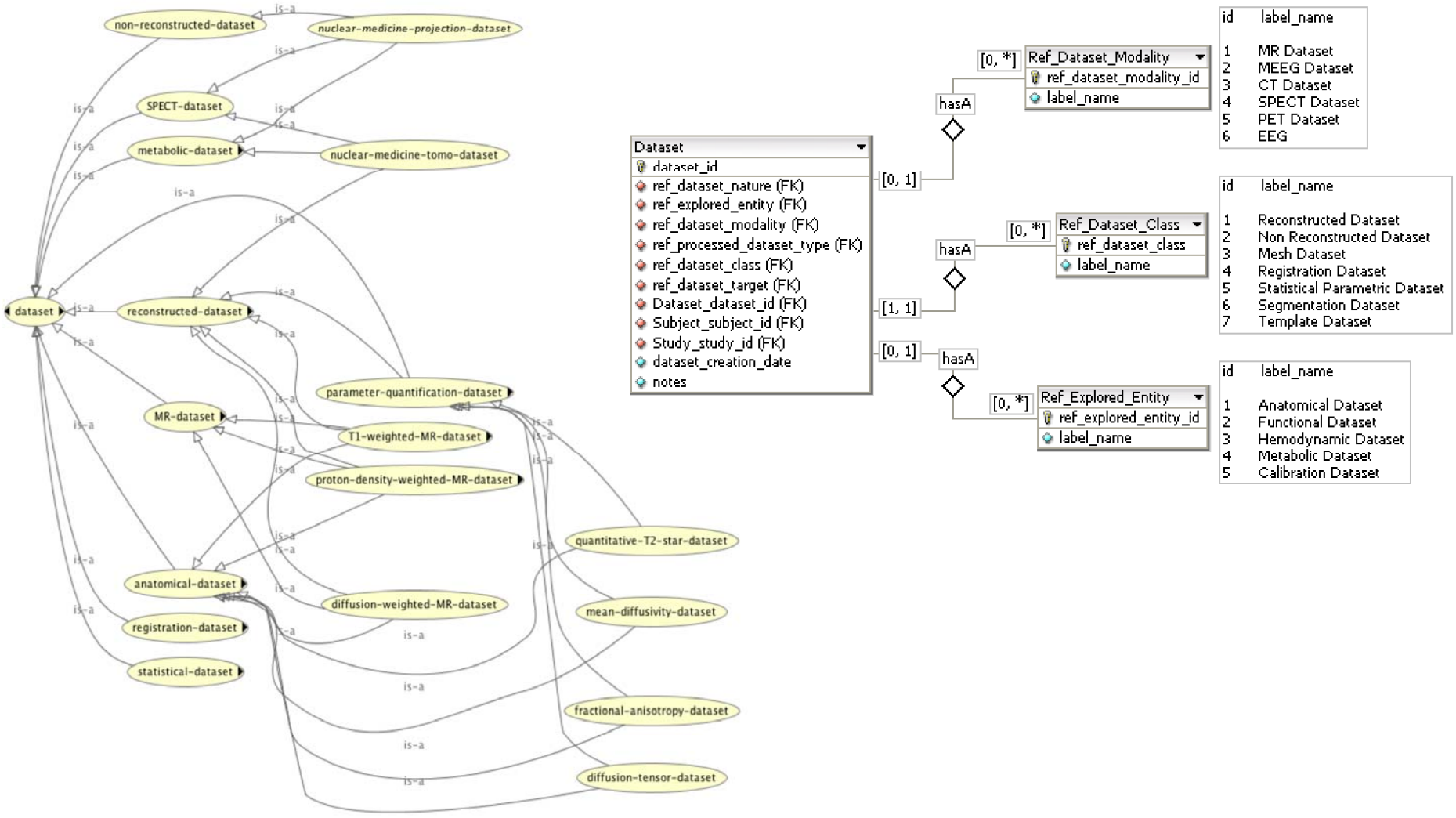
- **Interoperability: variability in metadata schemas**



Data Management 4/5

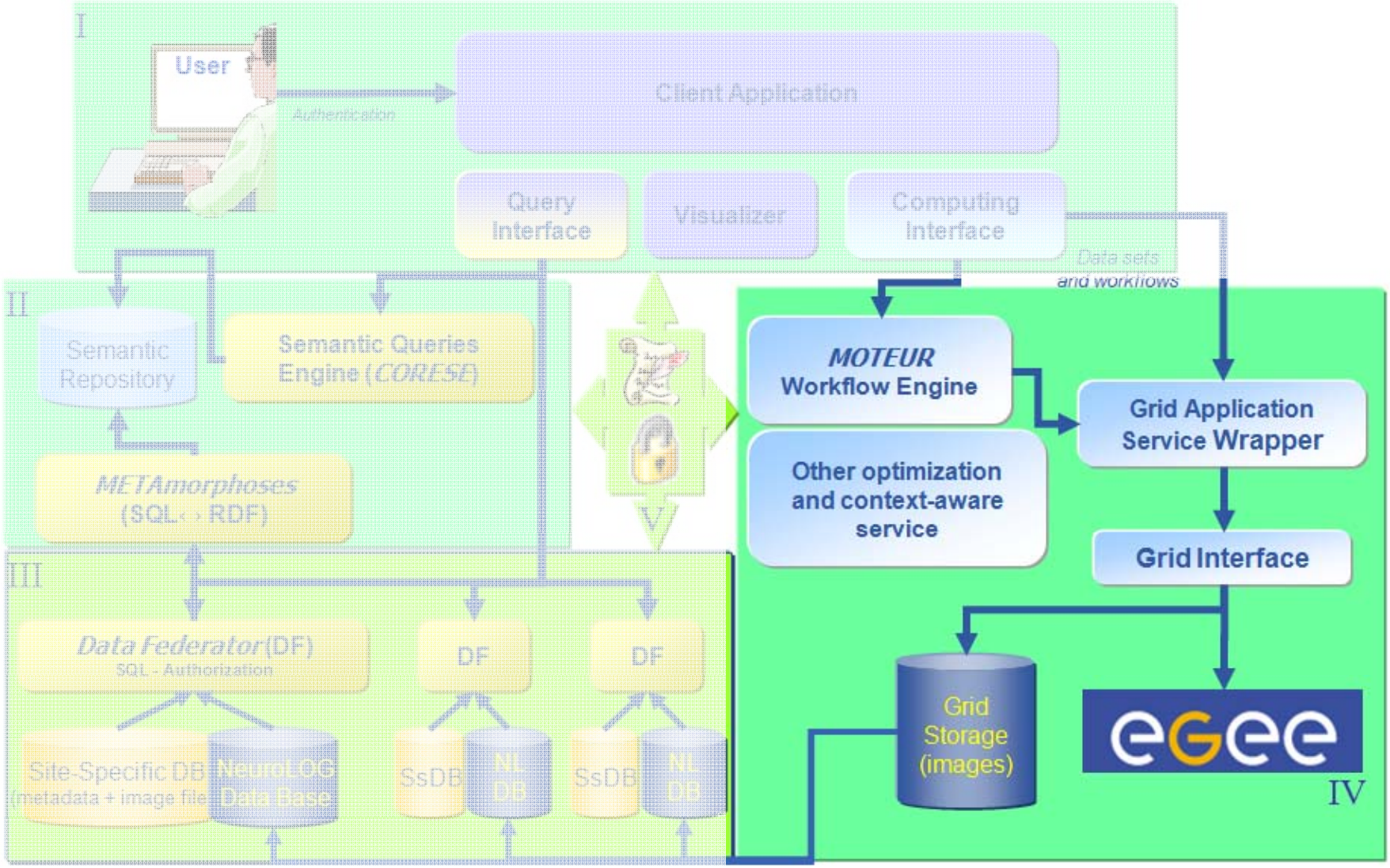
- **Domain ontologies**
  - Medical data (e.g. image type, associated medical record, neuro-psychological tests)
  - Processing tools (e.g. algorithm kinds, inputs and outputs)
  - Related relational database schemas
- **Semi-automatic semantic data collections**
  - Analyze DICOM headers
  - METAmorphose: relational DB to RDF conversion
  - OWL Lite representation
- **Semantic queries**
  - CORESE: conceptual graph-based query engine
  - SPARQL query language

Relational implementation of the ontology

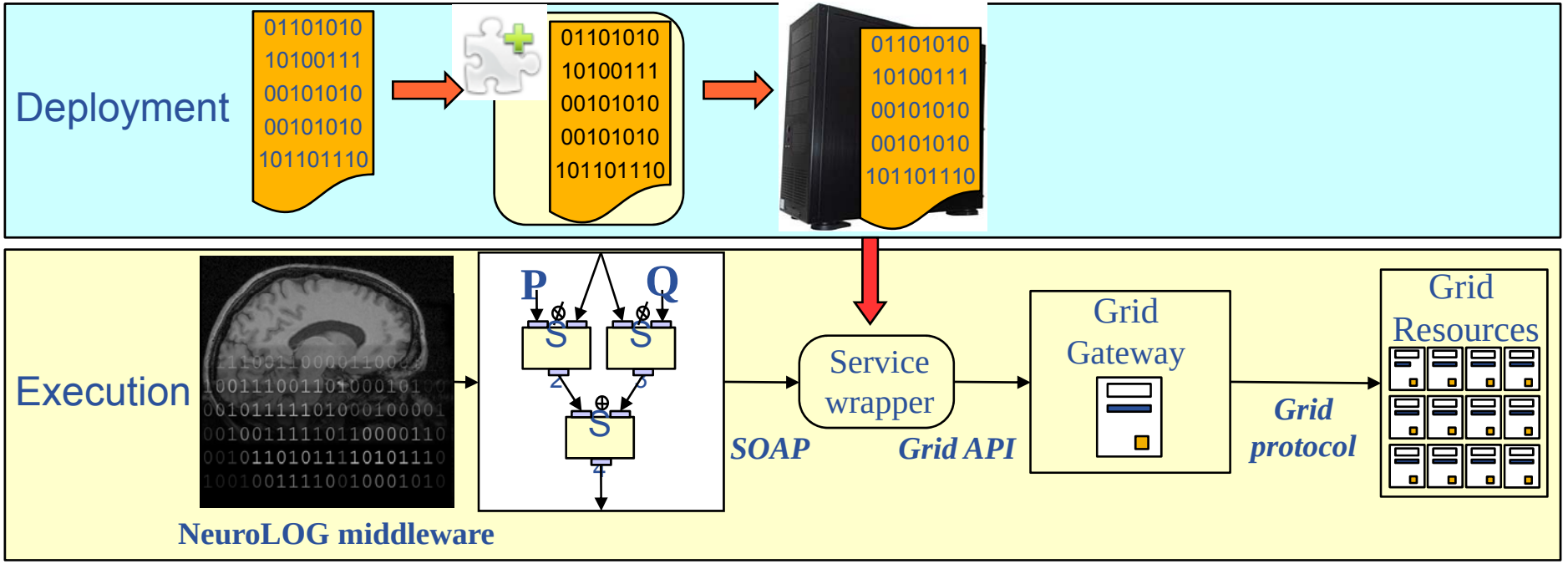


Computing Distribution 1/3

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



Computing Distribution 2/3

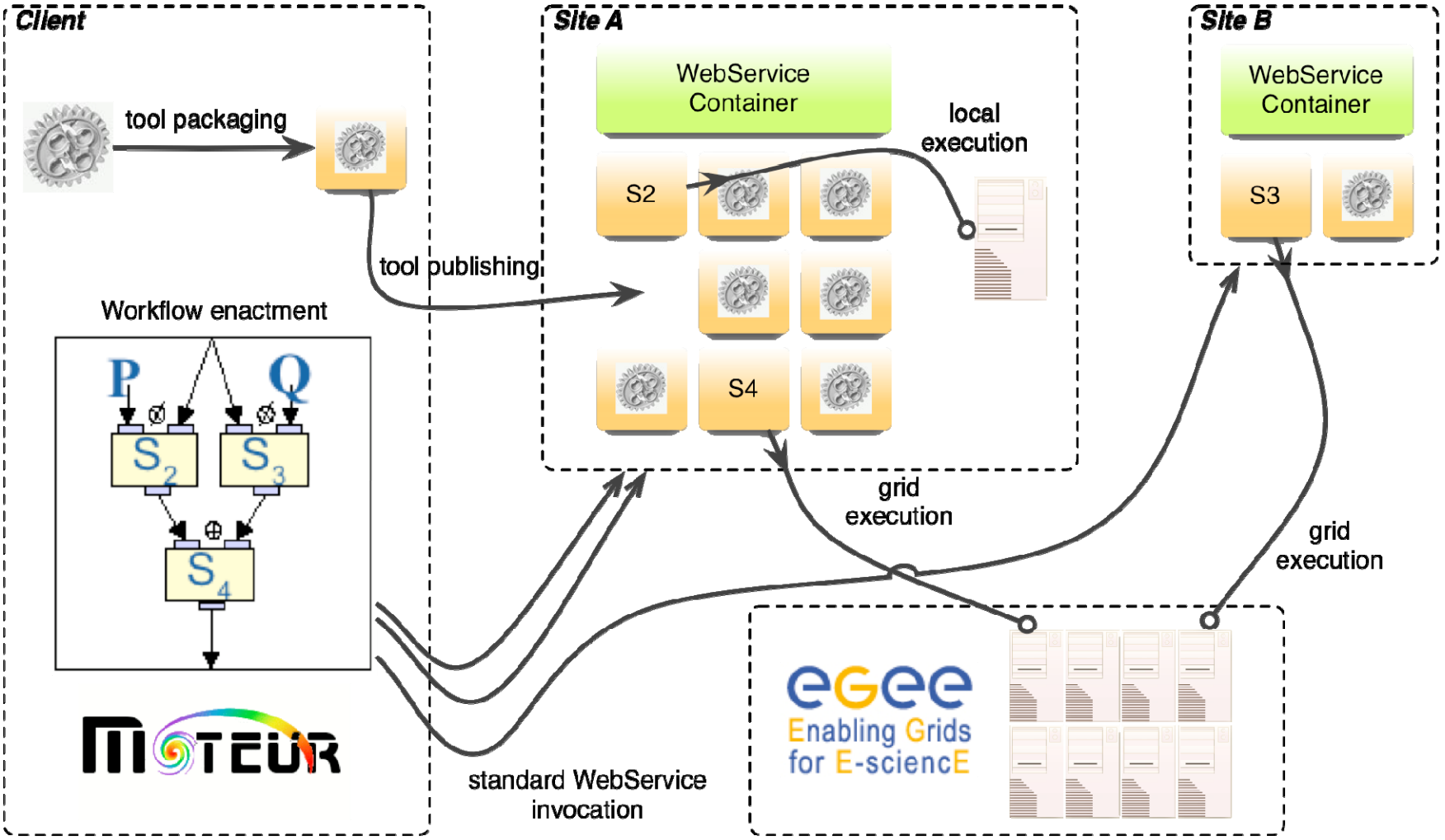


MOTEUR gLite Enabling Grids for E-sciencE

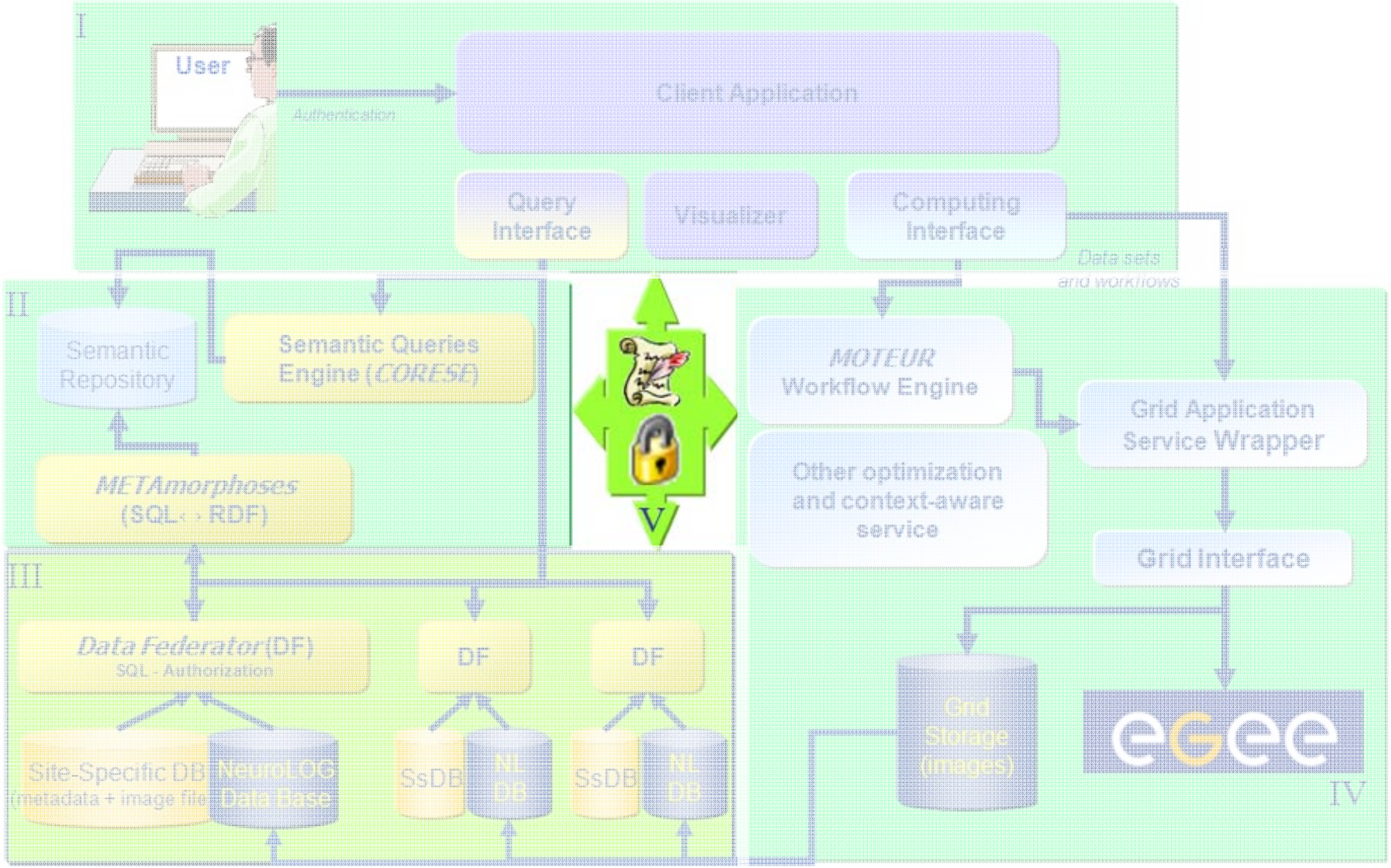
- **Workload management: gLite**
- **Application description & enactment: MOTEUR**
- **Application codes deployment: WS wrapper + WS containers**
- **Application codes packaging & publication: OSGi-like**

Computing Distribution 3/3

Interoperability: variability in tools



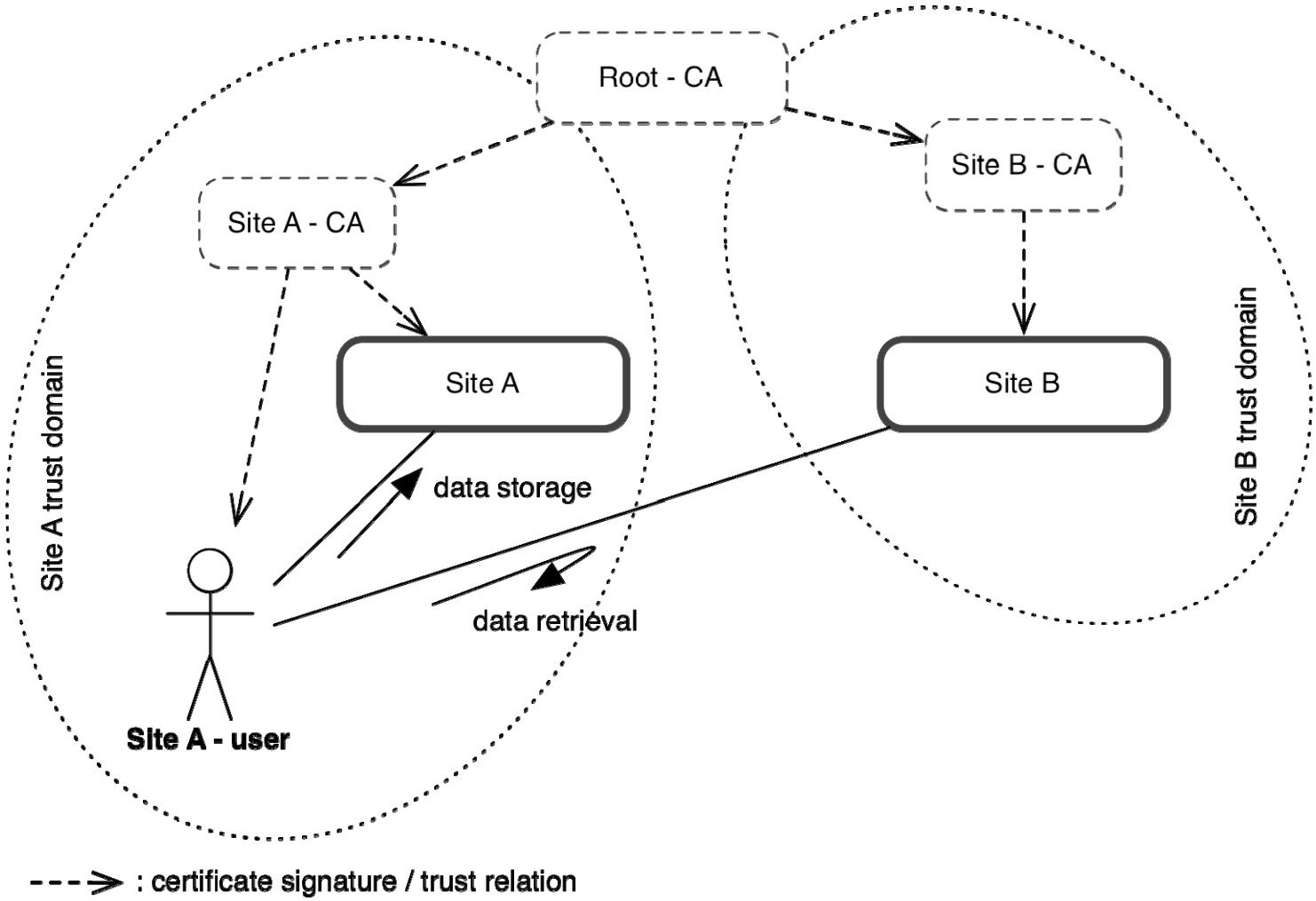
Security and Administration



Secured Collaborations 1/4

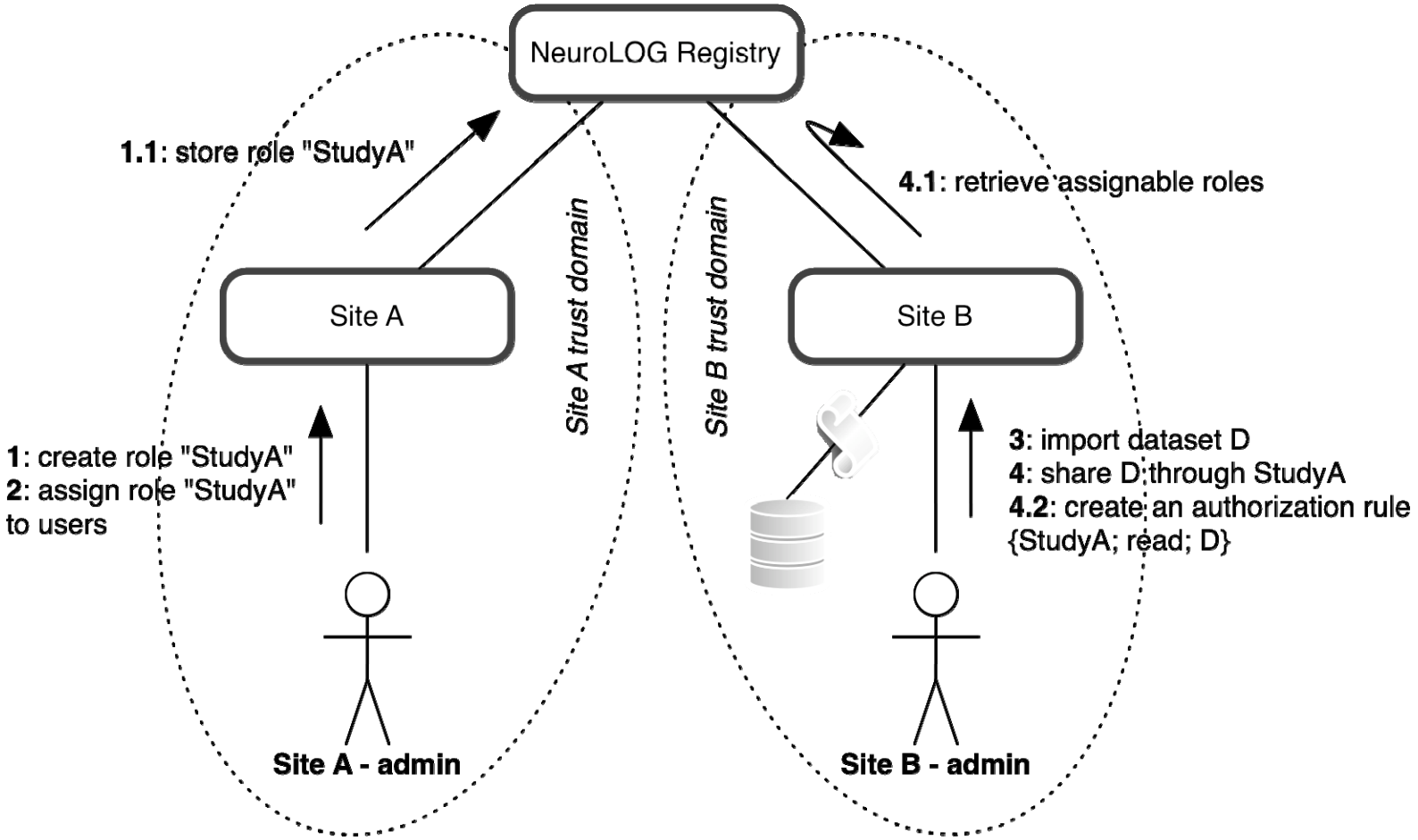
- **Collaboration in a Competitive/Sensitive environment**
- **Security Requirements**
  - Medical data protection
  - Autonomous site administration (no super-admin)
  - Multi-centric studies → Distributed access control
- **Authentication:**
  - Multiple X509 credentials per user (EGEE / NeuroLOG)
  - User registration through each Site Certification Authority
- **Data access control**
  - Partly conflicting interests : Autonomy vs Collaboration
  - Sharing through federation-wide roles
  - Owners of data keep the control over its data

Network interoperability: cross-domain accessibility

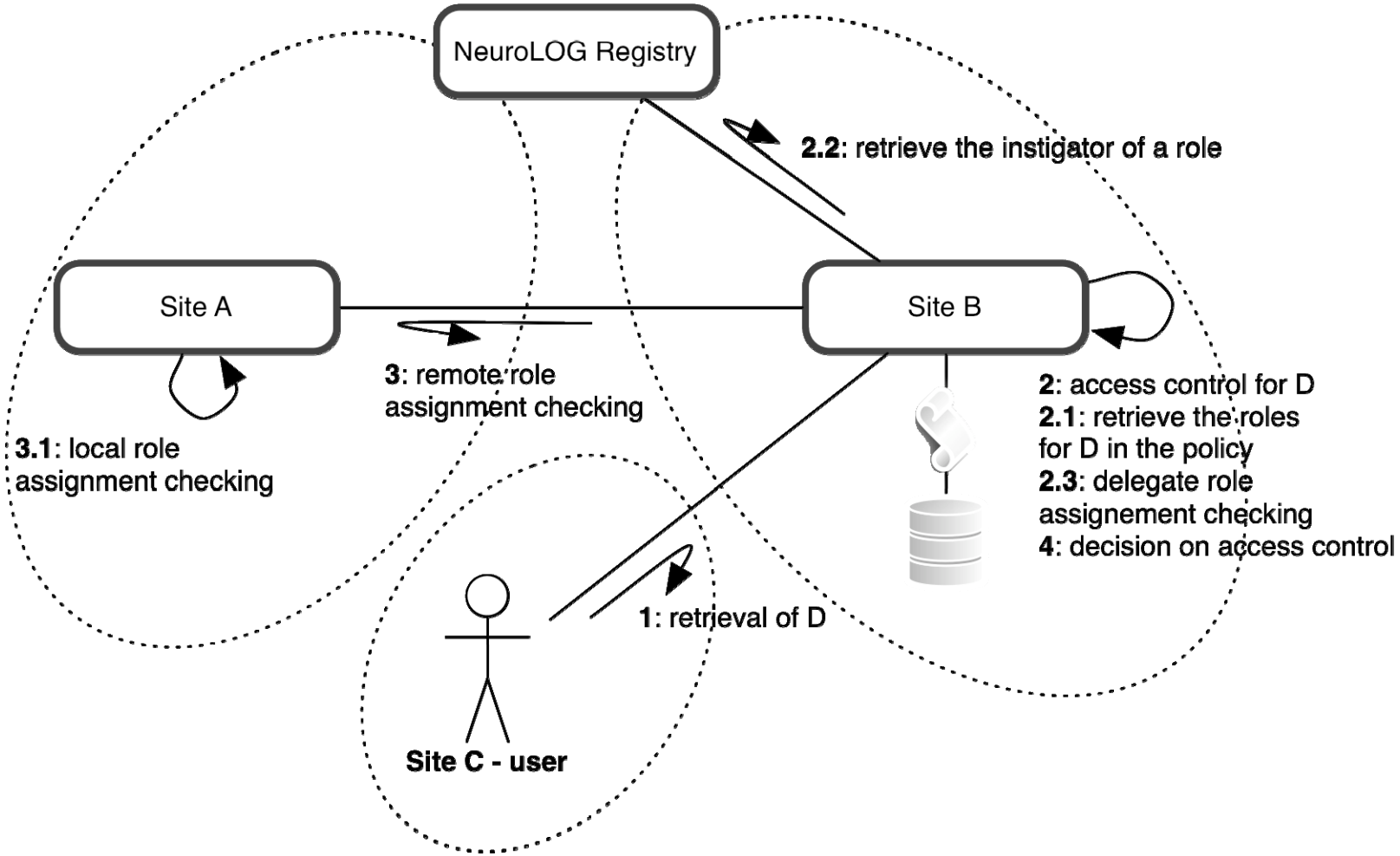


Secured Collaborations 3/4

Application interoperability: secured sharing use case



Application interoperability: distributed RBAC



Conclusion

- **Meeting neuroscientists expectations**
  - Cope with legacy environment (interoperability issues)
  - Preserve sites autonomy, while enabling distributed studies
  - Strong data access control
- **Transitional model towards HealthGrids**
  - Foster the adoption of grid technologies
  - Bring grid infrastructures closer to the clinical centers
  - Bridging local and global resources
- **Major challenge**
  - Keep it simple... for installation, configuration and use

http://neurolog.polytech.unice.fr