Toward ontology-based federated systems for sharing medical images: lessons from the NeuroLOG experience

Bernard Gibaud

MediCIS, LTSI, U1099 Inserm
Faculté de médecine, Rennes
bernard.gibaud@univ-rennes1.fr

Overview

- Introduction - scope
- Part 1. NeuroLOG
- Part 2. Lessons from NeuroLOG
- Part 3. Some key issues to be addressed
- Conclusion

Introduction

Scope of this presentation

- Sharing (i.e. query and retrieval) of bulky medical data
  - e.g. images, video, signals, etc.
  - Using associated metadata
- For research applications
  - Clinical research
  - Translational research
- Out of scope
  - Data integration for clinical care
  - Non-technical aspects of data sharing

Scope of this presentation

- Data: bulky medical data
  - Images, video, signals, etc.
  - Acquired or processed (segm., registration, etc.)
- Metadata, describing
  - Studies
  - Data acquisition context and provenance
  - Subjects from which data was taken
    - Scores obtained in various assessments
    - Biological data, etc.
  - Measurements derived from image data

Part I - NeuroLOG

Goals of the NeuroLOG project

- To set up a federated system, allowing the sharing and re-use of:
  - Neuroimaging data (images and related technical, demographical and medical metadata)
  - Processing tools published by cooperating partners
  - Computer processing resources (local, GRIDs)
- Three-year project (mid-2007 ➔ end-2010)

- This presentation focuses on the data sharing part of the project

Major design choices

- Federated system
  - Federating independent legacy systems
  - A solution that provides flexibility for data organization
  - … but brings heterogeneity
- Mediation
  - Use of a common ontology
  - Consistent with the « local as views » integration approach

→ Come up with a **global federated view** that hides data distribution and heterogeneity from the end-user

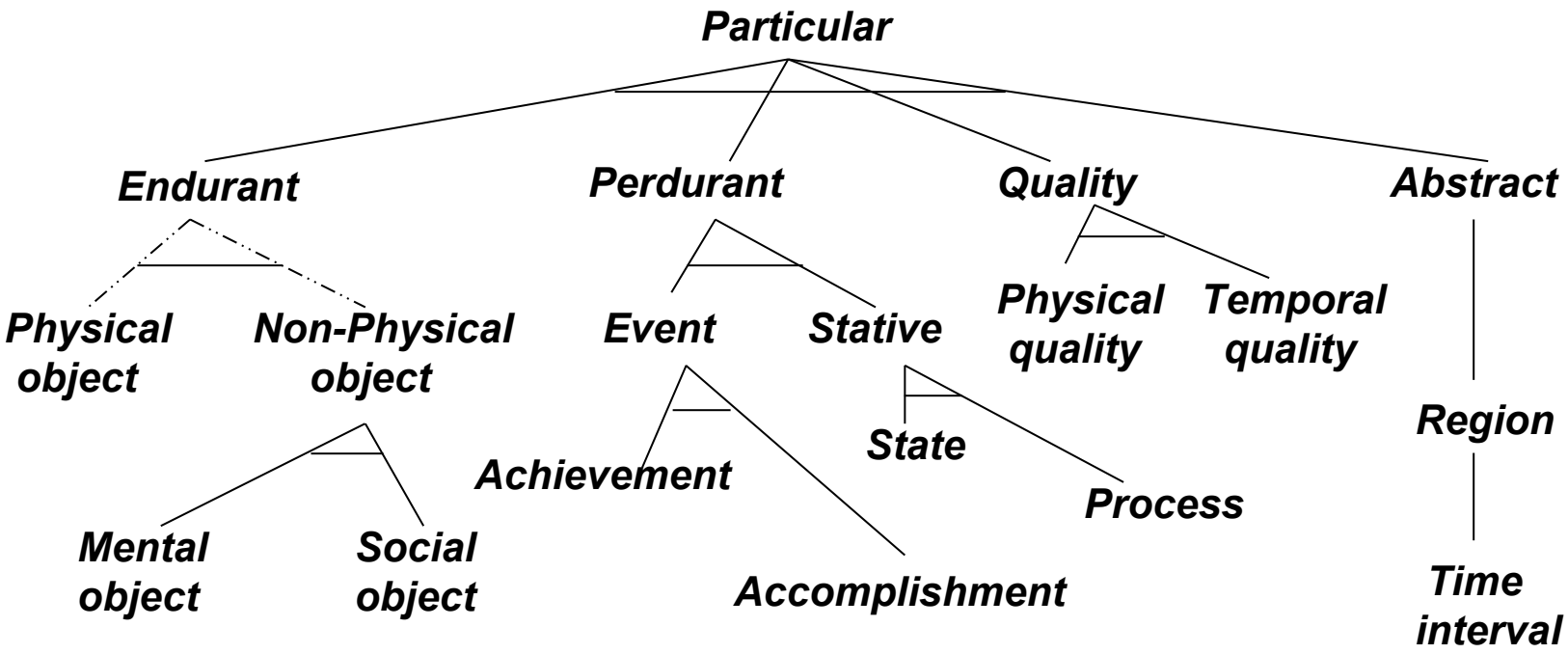
Ontology design

Ontology: general approach

- Application ontology (called OntoNeuroLOG)
  - Based on a common modelling framework
  - 3-level structure
    - one **Foundational ontology**: i.e. DOLCE
    - Several **Core ontologies**
    - Several **Domain ontologies**
  - Major concerns
    - Re-use of existing ontologies (when applicable)
    - Documentation

DOLCE: an ontology of particulars

(Masolo et al., 2003)



Ontology: scope

- To assemble a common application ontology to provide a uniform and consistent modelling of shared information, e.g. :
  - Images (*Datasets*)
  - Image acquisition and image processing (*Dataset processing*)
  - Context of acquisition and exploitation of the images (*Studies, Subjects*, *Examinations, Centers*, etc.)
  - Results of other kinds of explorations (*Subject data acquisition instruments*, *Instrument variables*, *Assessments*, *Scores*, etc)
- Use of this ontology to integrate heterogeneous data
  - Common relational schema

Ontology: 3-level structure

- Application ontology (called OntoNeuroLOG)
  - one **Foundational ontology** (DOLCE)
  - Several **Formal and core ontologies**
  - Several **Domain ontologies**

Major Formal and Core Ontologies	Major Domain ontologies
Particular (i.e. DOLCE)	Study
Action	Examination and Subject
Artefact	Neuroimaging Dataset
Participant role	Medical image expresssion
Capacity	Medical image file
Discourse, Message, and Discourse act	Medical image format
Number, Scalar quale, and Unit of measure	Dataset processing
Inscription, Expression, Conceptualization	Dataset acquisition
Language and Computer language	MR protocol
Computer language expression	MR sequence
Assessment-Instrument	Specific Assessment-Instruments (MMS, EDSS, etc.)

Ontology: 3 representations

1. OntoSpec representation (Kassel, 2005)
   - Semi-formal notation (rich semantics)
   - Numerous axioms
2. OWL-Lite
   - Edited with PROTÉGÉ
   - Tailored to perform inferences with CORESE (search engine)
3. Federated relational schema
   - Entities and relations are closely linked to concepts and relations of the ontology

(http://neurolog.i3s.unice.fr/public_namespace/ontology)

Example of OntoSpec representation

Instrument-based assessment, Instrument administration, Testing

Meta-properties

INSTRUMENT-BASED ASSESSMENT is RIGID (+**R**). INSTRUMENT-BASED ASSESSMENT is EXTERNALLY-DEPENDENT (+**D**). TEST-BASED ASSESSMENT and QUESTIONNAIRE-BASED ASSESSMENT *is a non-trivial partition of* INSTRUMENT-BASED ASSESSMENT.

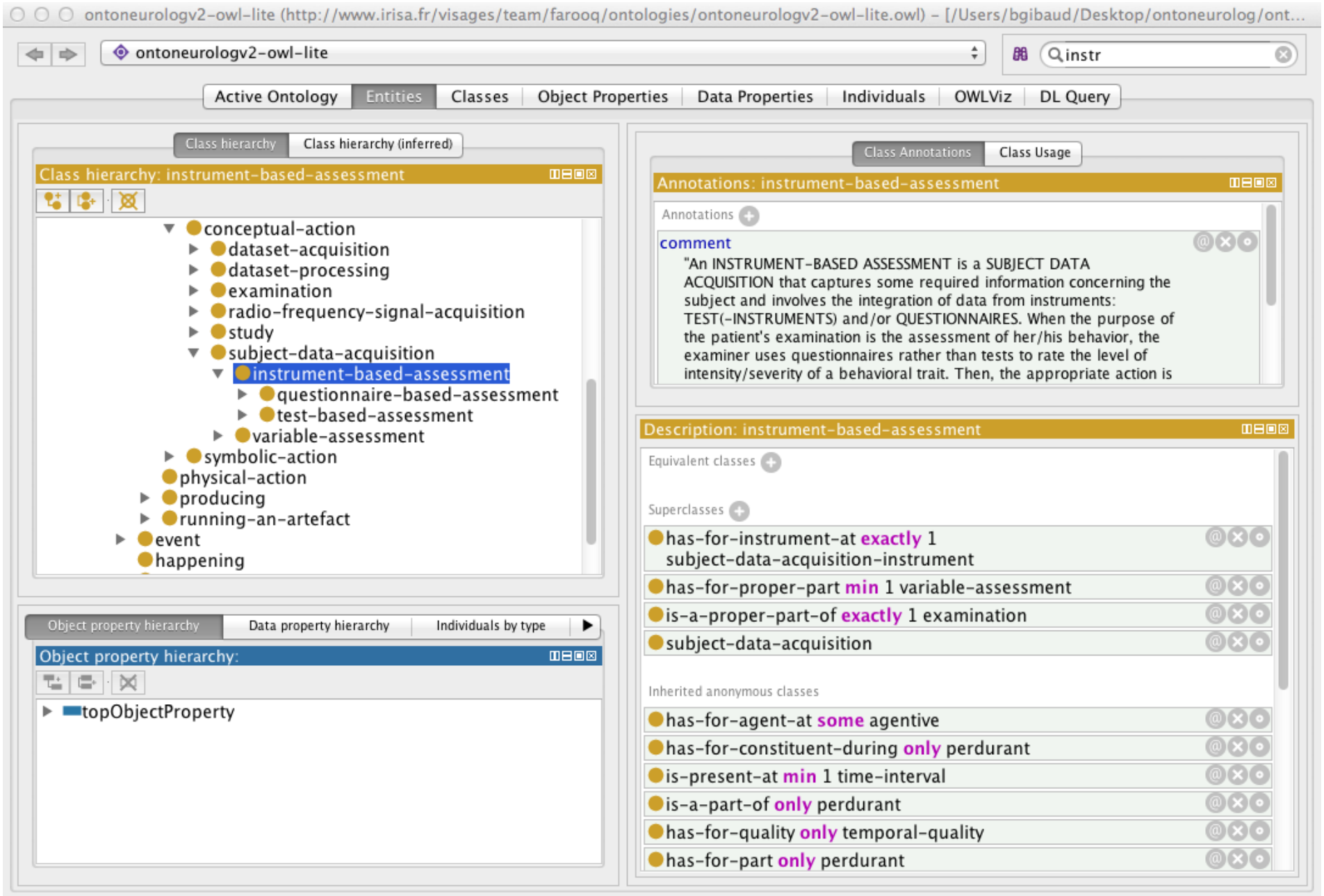
Properties

[EP/SL] An INSTRUMENT-BASED ASSESSMENT, or INSTRUMENT ADMINISTRATION, or TESTING, is a SUBJECT DATA ACQUISITION. [EP/ER] Every INSTRUMENT-BASED ASSESSMENT *has for instrument* exactly one SUBJECT DATA ACQUISITION INSTRUMENT *at* a TIME INTERVAL. [EP/ER] Every INSTRUMENT-BASED ASSESSMENT *is a proper part of* exactly one EXAMINATION. [EP/ER] Every INSTRUMENT-BASED ASSESSMENT *has for proper part* at least one VARIABLE ASSESSMENT.

Comment

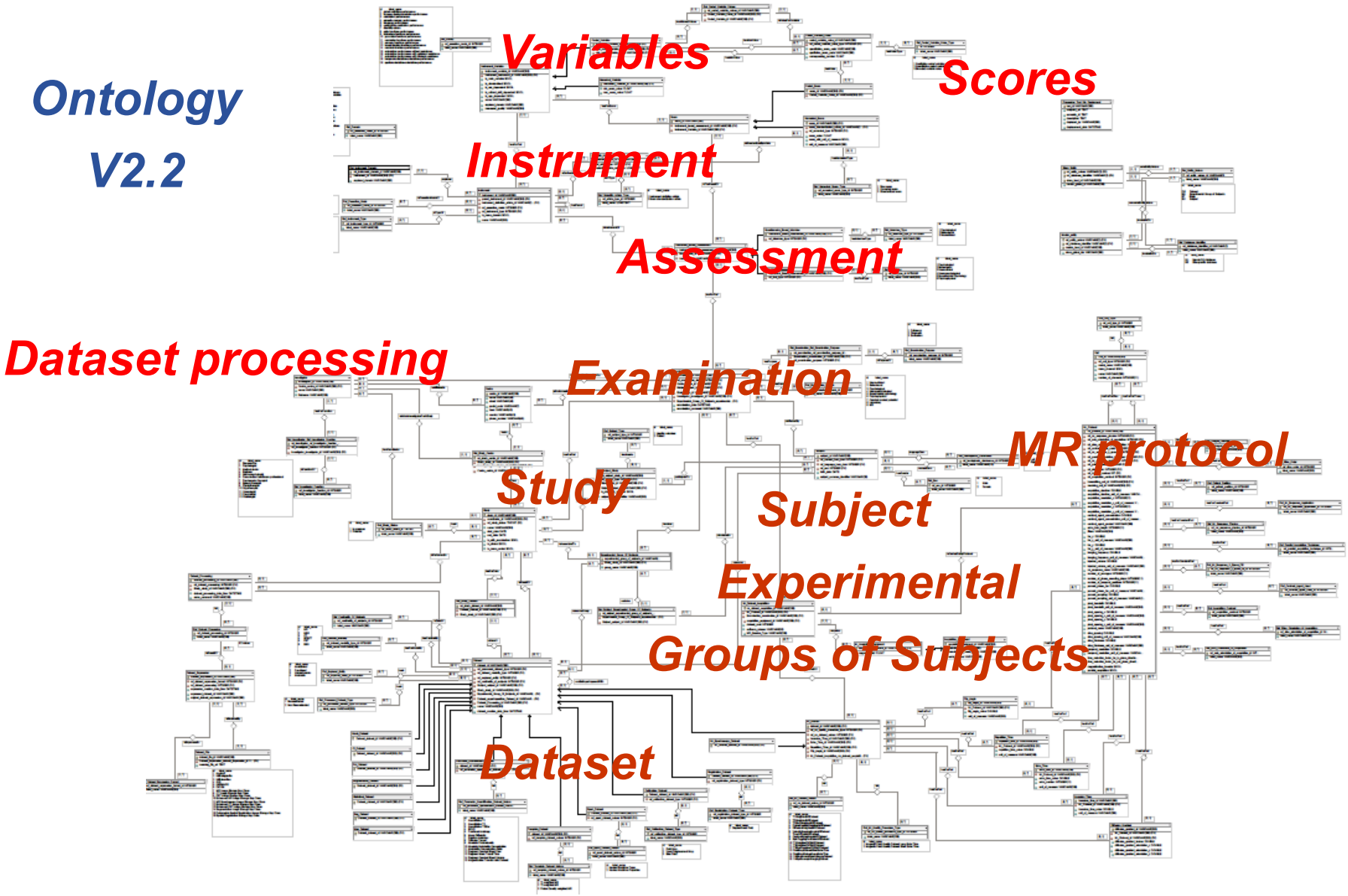
[DEF] An INSTRUMENT-BASED ASSESSMENT is a SUBJECT DATA ACQUISITION that captures some required information concerning the subject and involves the integration of data from instruments: TEST(-INSTRUMENTS) and/or QUESTIONNAIRES. When the purpose of the patient's examination is the assessment of her/his behavior, the examiner uses questionnaires rather than tests to rate the level of intensity/severity of a behavioral trait. Then, the appropriate action is a BEHAVIOURAL INTERVIEW rather than a BEHAVIOURAL TEST which is less adapted.

[SA] INSTRUMENT-BASED ASSESSMENTS are divided among TEST-BASED ASSESSMENTS and QUESTIONNAIRE-BASED ASSESSMENTS according to the kind of instrument which is administrated and therefore to the specific roles played by the subject and the healthcare professional in the assessment.


ontoneurologv2-owl-lite (http://www.irisa.fr/visages/team/farooq/ontologies/ontoneurologv2-owl-lite.owl) - [/Users/bgibaud/Desktop/ontoneurolog/ont...
ontoneurologv2-owl-lite
instr
Active Ontology
Entities
Classes
Object Properties
Data Properties
Individuals
OWLViz
DL Query
Class hierarchy
Class hierarchy (inferred)
Class hierarchy: instrument-based-assessment
conceptual-action
dataset-acquisition
dataset-processing
examination
radio-frequency-signal-acquisition
study
subject-data-acquisition
instrument-based-assessment
questionnaire-based-assessment
test-based-assessment
variable-assessment
symbolic-action
physical-action
producing
running-an-artefact
event
happening
Class Annotations
Class Usage
Annotations: instrument-based-assessment
Annotations
comment
"An INSTRUMENT-BASED ASSESSMENT is a SUBJECT DATA ACQUISITION that captures some required information concerning the subject and involves the integration of data from instruments: TEST(-INSTRUMENTS) and/or QUESTIONNAIRES. When the purpose of the patient's examination is the assessment of her/his behavior, the examiner uses questionnaires rather than tests to rate the level of intensity/severity of a behavioral trait. Then, the appropriate action is
Description: instrument-based-assessment
Equivalent classes
Superclasses
has-for-instrument-at exactly 1 subject-data-acquisition-instrument
has-for-proper-part min 1 variable-assessment
is-a-proper-part-of exactly 1 examination
subject-data-acquisition
Inherited anonymous classes
has-for-agent-at some agentive
has-for-constituent-during only perdurant
is-present-at min 1 time-interval
is-a-part-of only perdurant
has-for-quality only temporal-quality
has-for-part only perdurant
Object property hierarchy
Data property hierarchy
Individuals by type
Object property hierarchy:
topObjectProperty

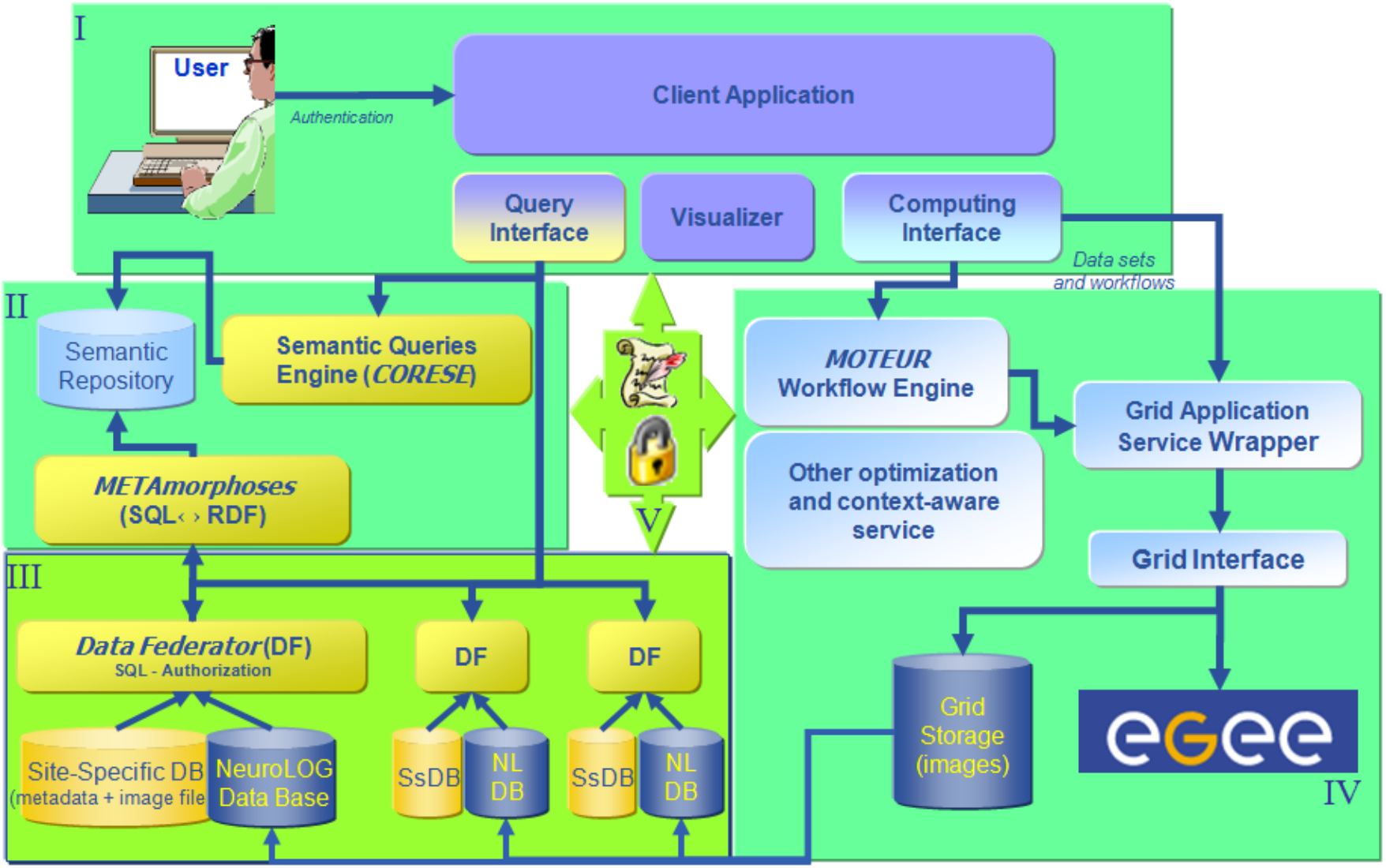
Instruments' descriptions

F	G	H	I	J	K
Type Instrument type d'instrument	**Hierarchy /Hiérarchie**	**Ref to parent-instrument in hierarchy**	**mono / multi**	**Instrument model acronym**	**Instrument model name**
neuropsychological/ neuroclinical/ behavioural	**ROOT/ sub-instrument**			**Instrument**	
ıeuropsychological	ROOT		mono	**MMS**	**Mini-Mental-State**
ıeuropsychological	sub-instrument	**MMS**	mono	**MMS-1**	MMS-orientation
ıeuropsychological	sub-instrument	**MMS-1**	mono	**MMS-1-1**	MMS-orientation-to-time
ıeuropsychological	sub-instrument	**MMS-1**	mono	**MMS-1-2**	MMS-orientation-to-place
ıeuropsychological	sub-instrument	**MMS**	mono	**MMS-2**	MMS-registration
ıeuropsychological	sub-instrument	**MMS**	mono	**MMS-3**	MMS-attention-and-calculation
ıeuropsychological	sub-instrument	**MMS**	mono	**MMS-4**	MMS-recall
ıeuropsychological	sub-instrument	**MMS**	mono	**MMS-5**	MMS-language-tests
ıeuropsychological	sub-instrument	**MMS-5**	mono	**MMS-5-1**	MMS-language-naming
ıeuropsychological	sub-instrument	**MMS-5**	mono	**MMS-5-2**	MMS-language-repetition
ıeuropsychological	sub-instrument	**MMS-5**	mono	**MMS-5-3**	MMS-language-3-stage-command
ıeuropsychological	sub-instrument	**MMS-5**	mono	**MMS-5-4**	MMS-language-reading
ıeuropsychological	sub-instrument	**MMS-5**	mono	**MMS-5-5**	MMS-language-writing
ıeuropsychological	sub-instrument	**MMS**	mono	**MMS-6**	MMS-copy-design
ıeuroclinical	ROOT		mono	**EDSS**	**Expanded-Disability-Status-Scale**
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-1**	visual-optic-functions-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-2**	cranial-nerve-examination-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-3**	pyramidal-functions-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-4**	cerebellar-examination-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-5**	sensory-examination-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-6**	bowel-bladder-functions-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-7**	mental-status-examination-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-8**	ambulation-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-8**	ambulation-EDSS
ıeuroclinical	sub-instrument	**EDSS**	mono	**EDSS-8**	ambulation-EDSS
ıeuropsychological	ROOT		mono	**CFT-C**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Copy-administration**
ıeuropsychological	ROOT		mono	**CFT-C**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Copy-administration**
ıeuropsychological	ROOT		mono	**CFT-C**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Copy-administration**
ıeuropsychological	sub-instrument	**CFT-C**	mono	**CFT-C-1**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Copy-with-points-of-reference**
ıeuropsychological	sub-instrument	**CFT-C**	mono	**CFT-C-2**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Copy-with-program-of-realization**
ıeuropsychological	ROOT		mono	**CFT-M**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Recall-administration**
ıeuropsychological	ROOT		mono	**CFT-M**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Recall-administration**
ıeuropsychological	ROOT		mono	**CFT-M**	**Rey-Osterrieth-Complex-Figure-Test-(CFT)-Recall-administration**

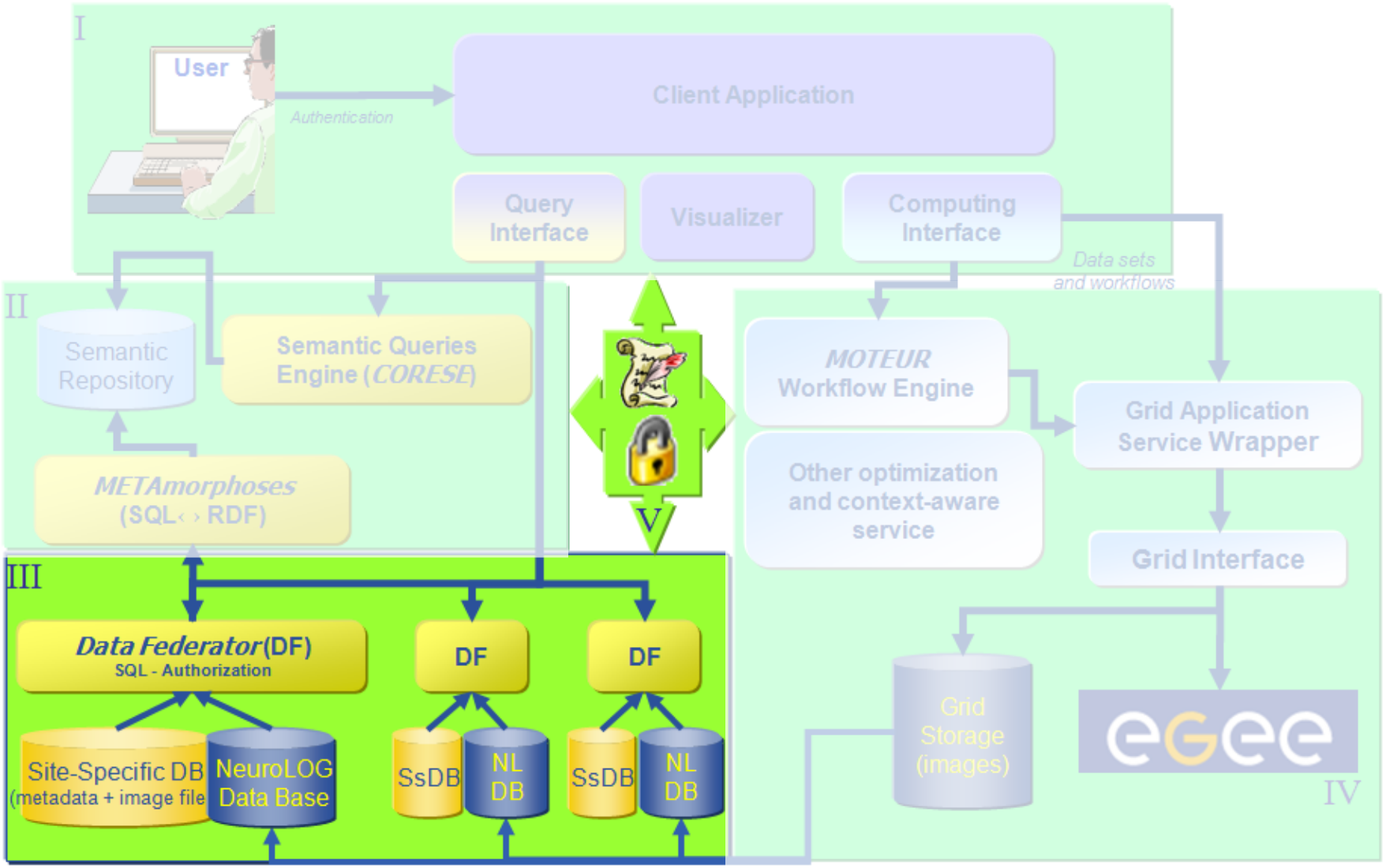

Ontology
V2.2
Variables
Scores
Instrument
Assessment
Dataset processing
Examination
MR protocol
Study
Subject
Experimental
Groups of Subjects
Dataset

Data integration

Overall architecture

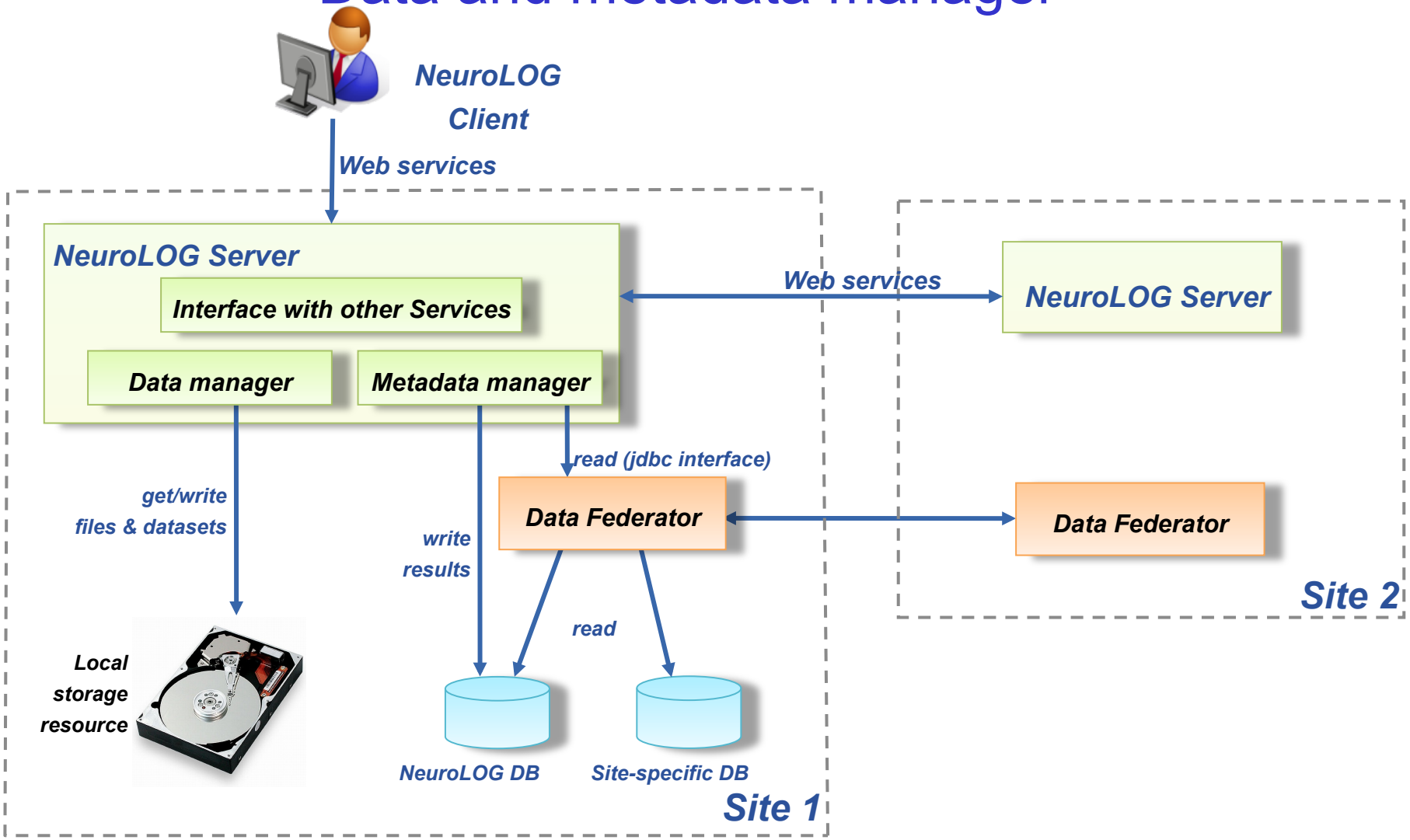


Overall architecture: data integration

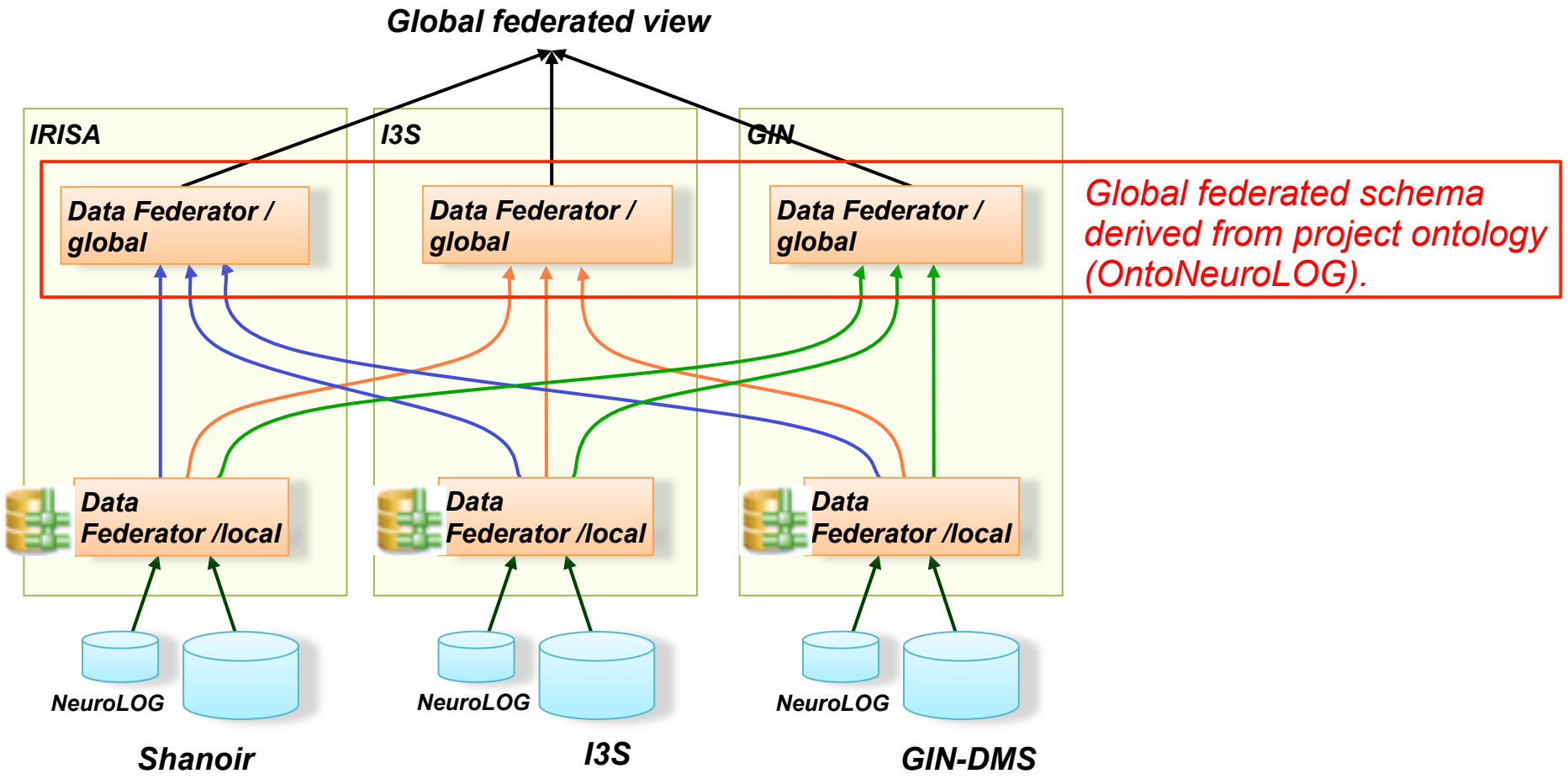


NeuroLOG server

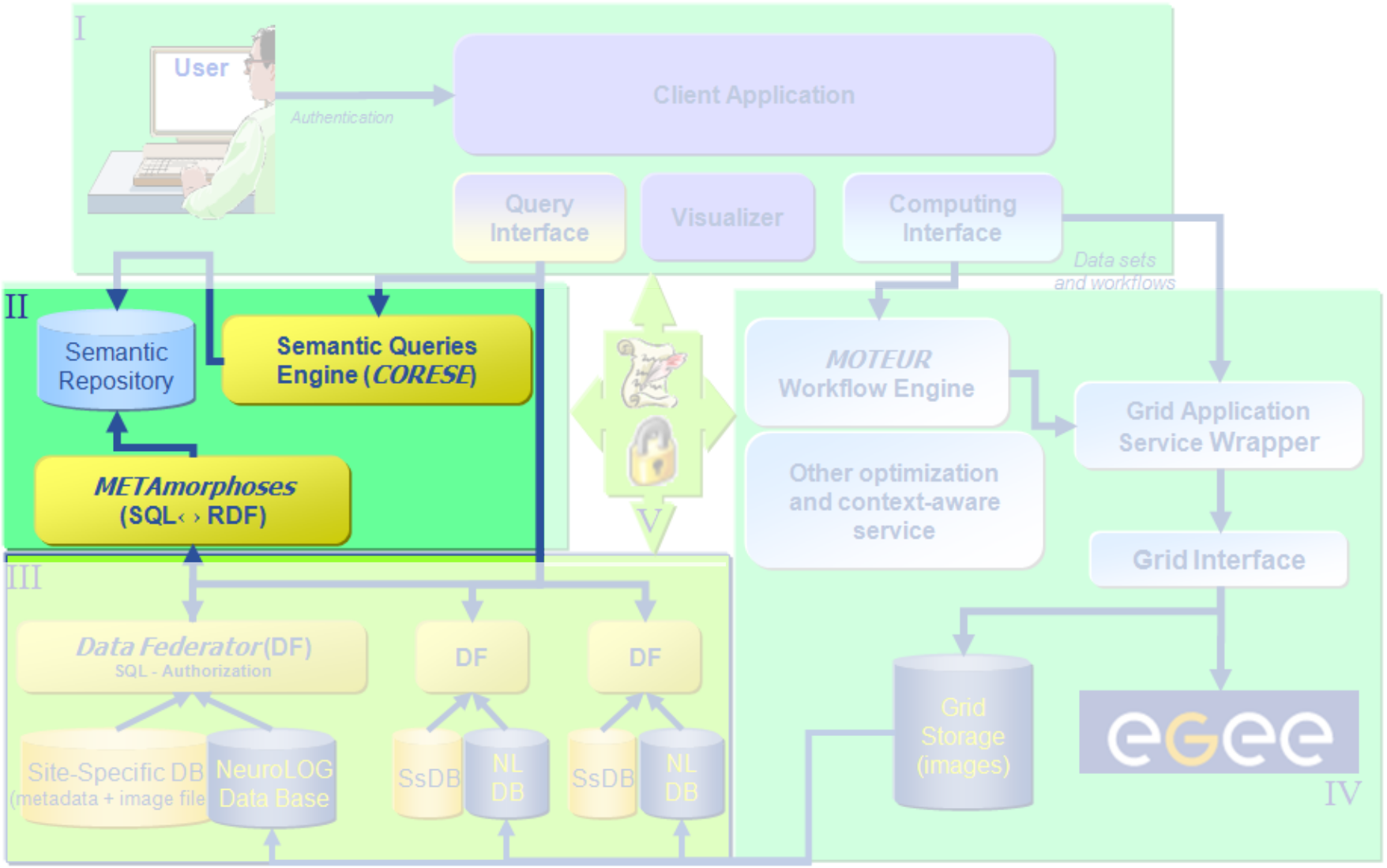
Data and metadata manager



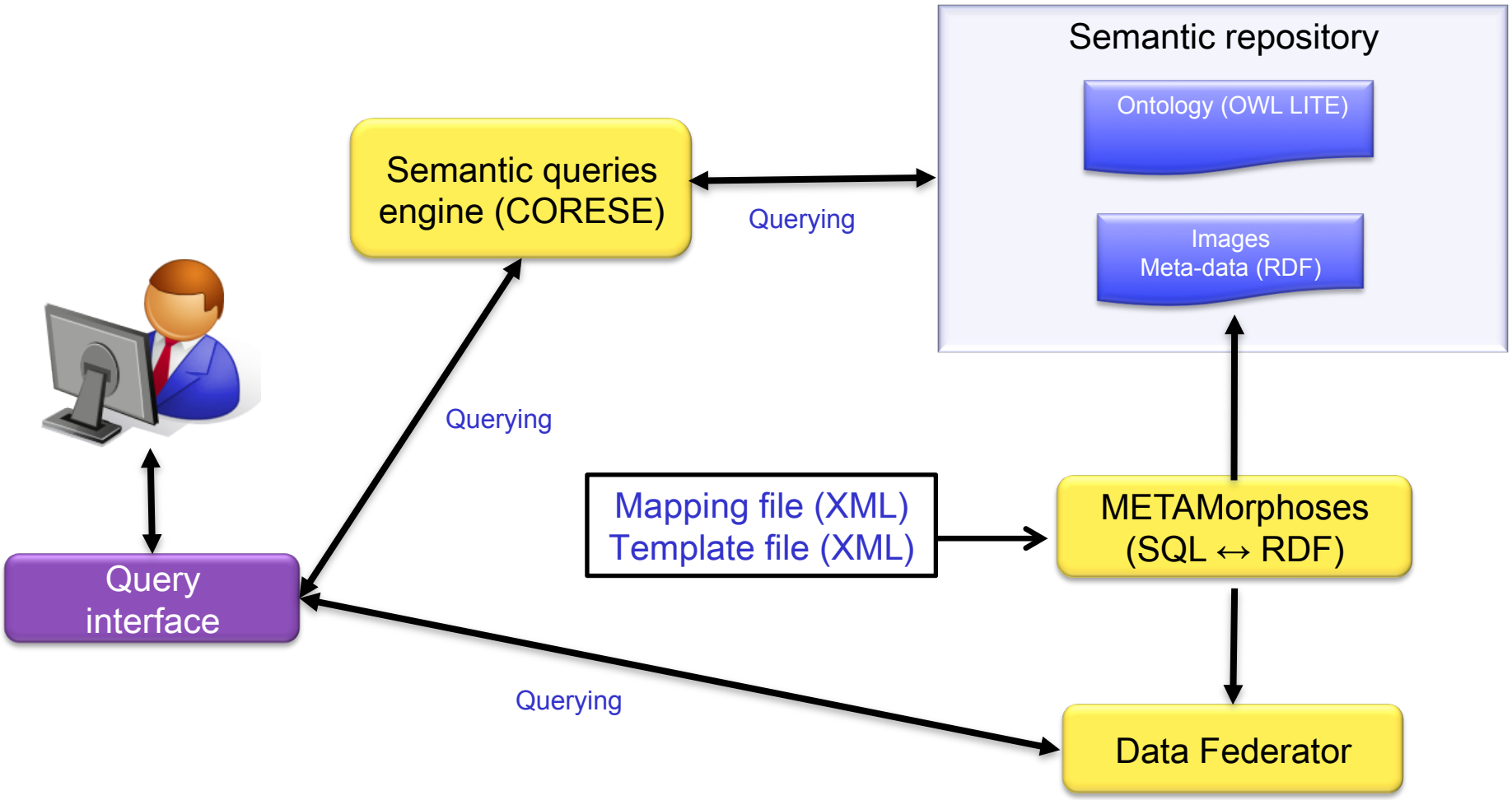
Metadata mapping
using Data Federator (SAP)



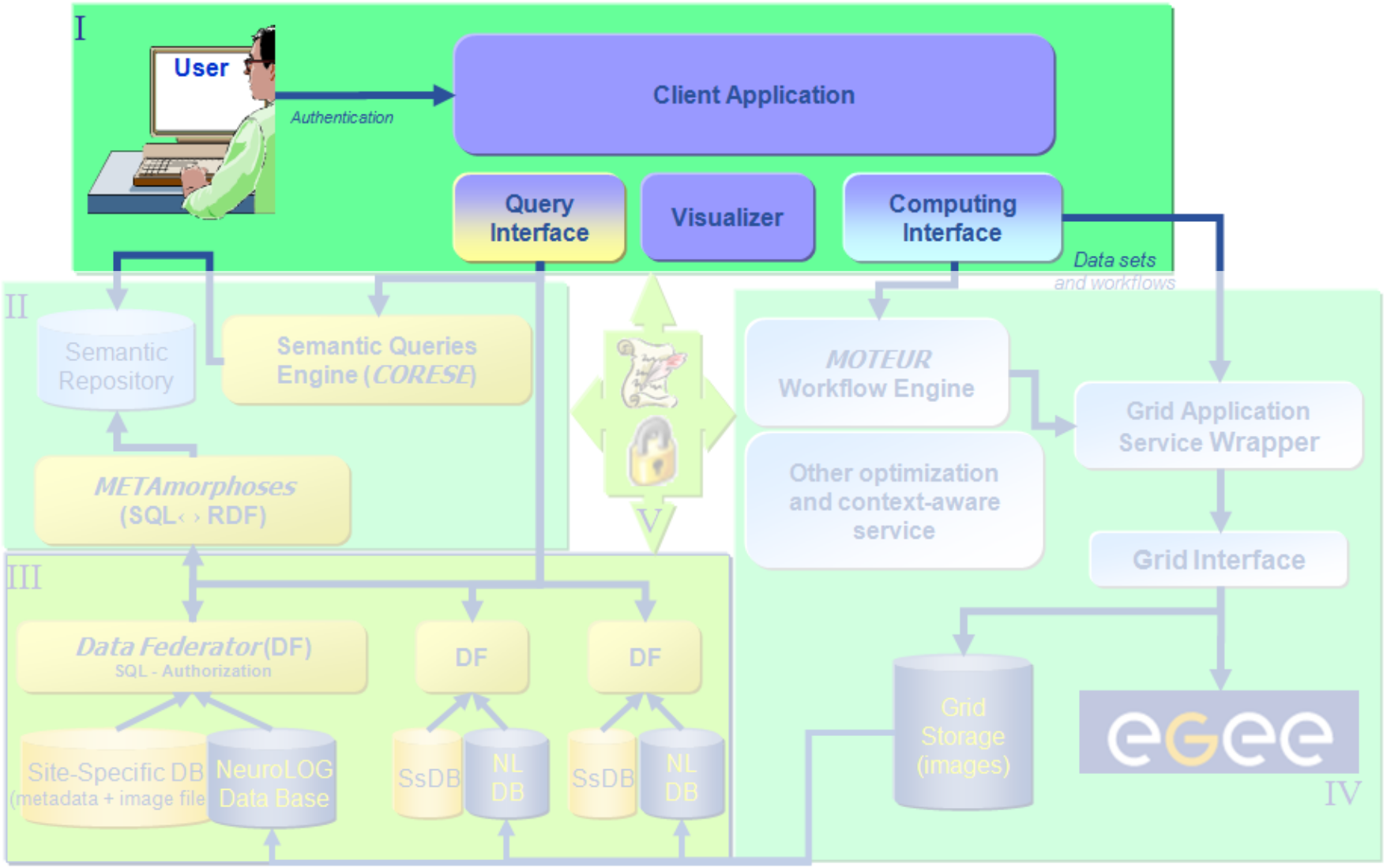
Overall architecture: Semantic data



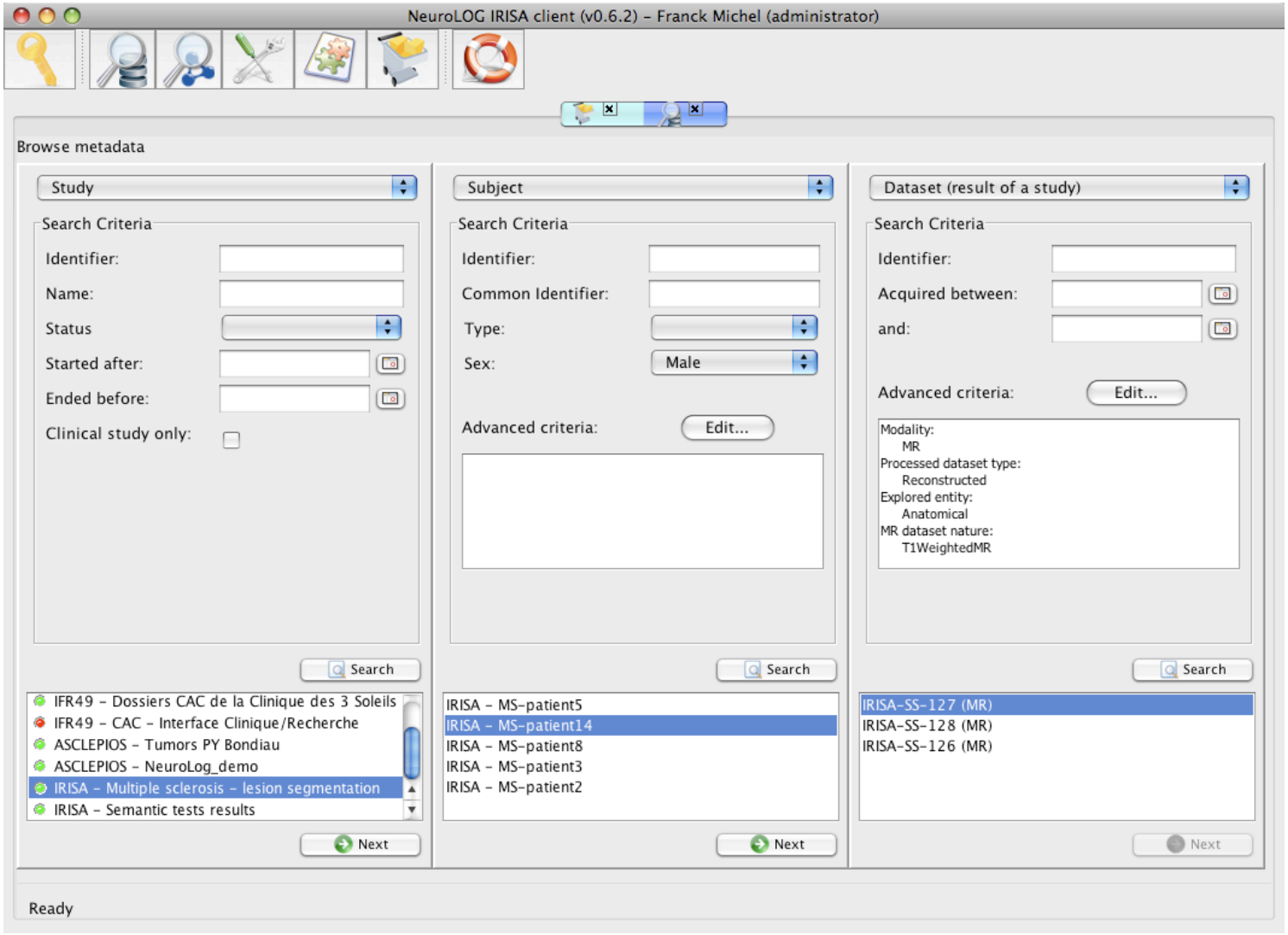
Architecture: semantic module



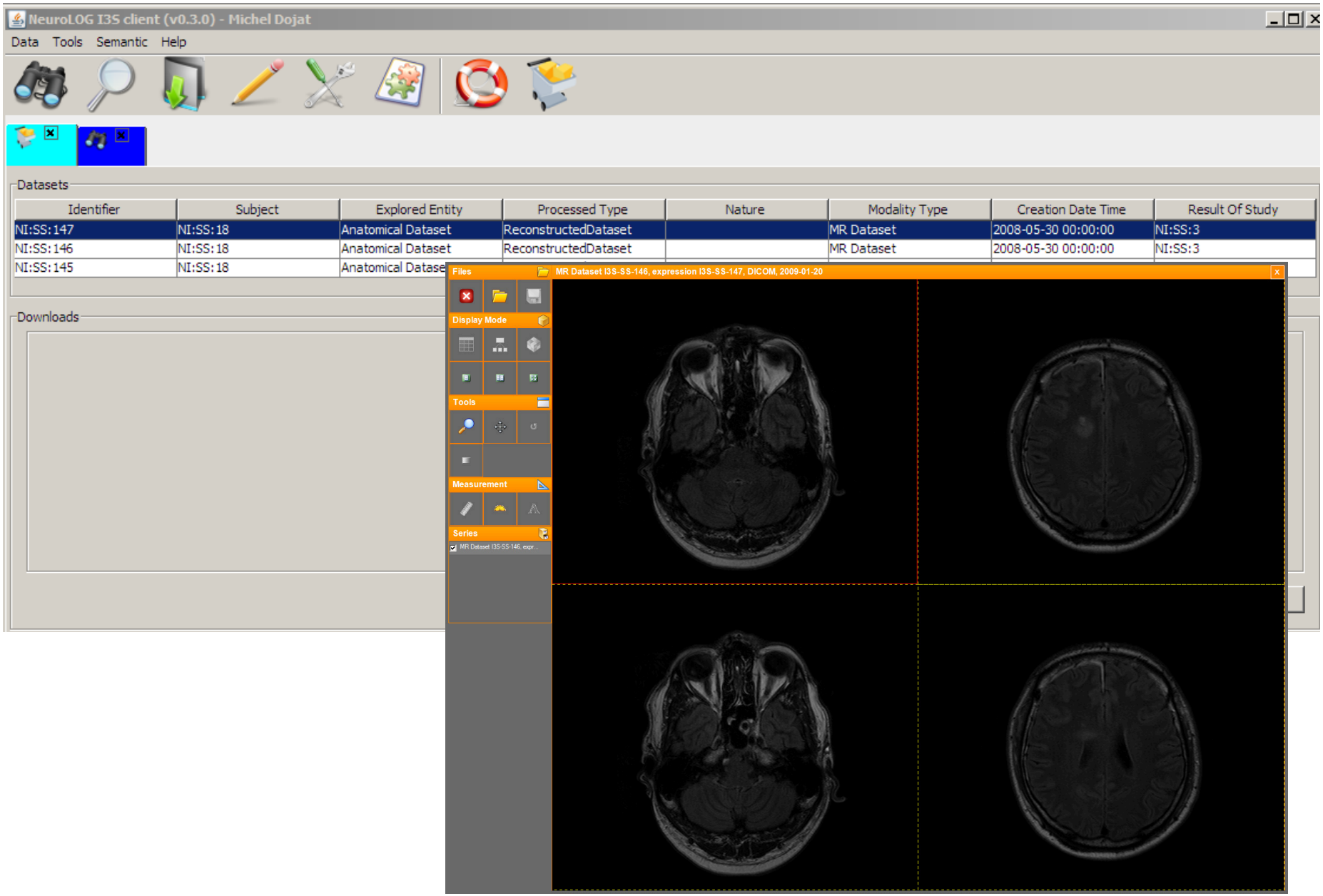
Overall architecture: client software



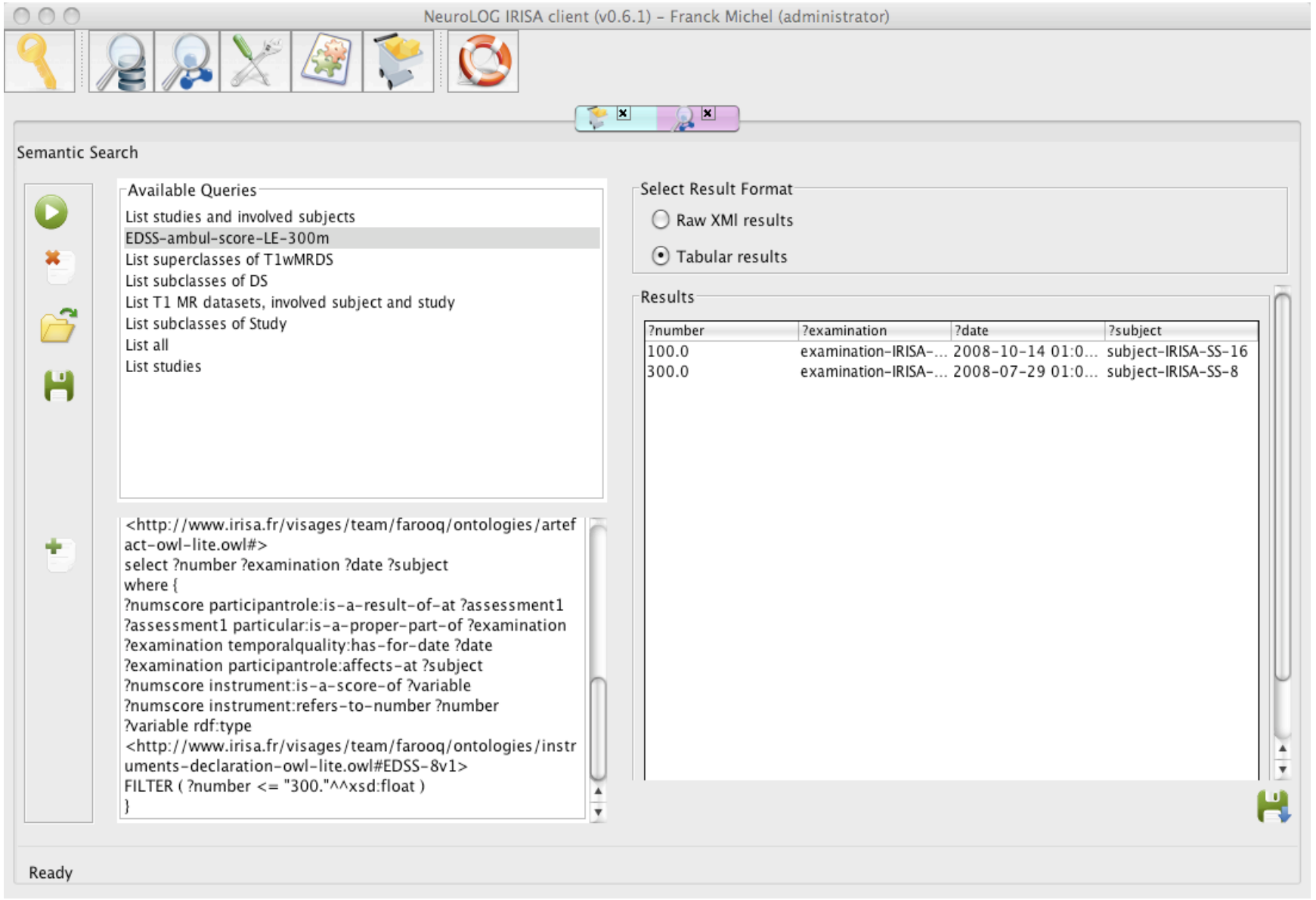
Client: querying metadata and accessing images



Client: image viewer



Semantic query example (SPARQL)
« EDSS scores with ambulation scores <=300 »



System deployment

NeuroLOG platform (2009-2011)

- 5 sites federated
  - 4 legacy databases
  - 12 studies
  - > 70 subjects
    - MS
    - Brain tumors
    - AD
  - > 500 datasets

Part 2. Lessons from NeuroLOG

Middleware for data integration

-  NeuroLOG: Data Federator (SAP)
  - « Comfortable » : well-documented, efficient, reliable

-  License keys expensive
  - Primarily marketed for business and finance
  - but, not affordable for regular academic research

→ Need for open source alternatives and to refine strategy w.r.t. metadata integration & alignment

Definition of an application ontology

-  Significant results
  - Modeling framework (DOLCE) satisfactory though incomplete
  - Significant achievements: e.g., instruments and scores
-  All objectives not met
  - e.g., annotate images with quantitative measurements derived from image data
  - Relation to biological structures and processes

→ Still needs substantial work (discussed hereafter)

Deployment and adoption

-  Successful deployment of the platform
-  Limited impact
  - Exploitation of the platform stopped in 2011
  - So, limited use
  - Application for a new ANR Grant (Alzheimer's disease and Epilepsy)
    → (NOT successsful)

→ Need to find new financing, possibly international

Part 3. Some key issues to be addressed

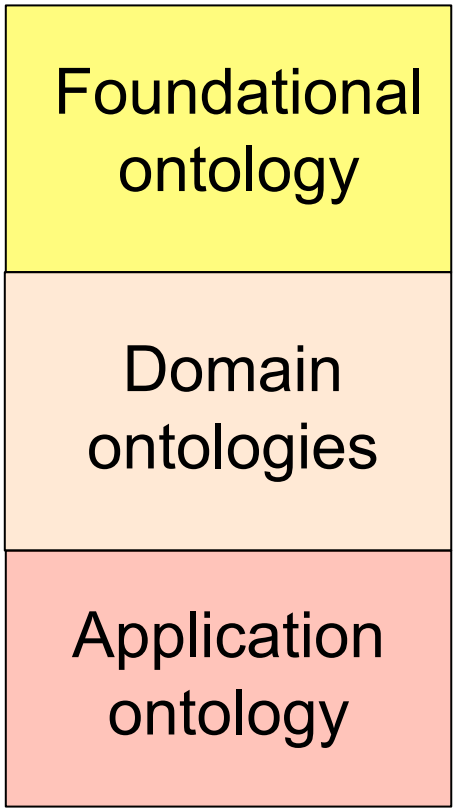
1st. key issue: what middleware for metadata integration & alignment ?

- Are we ready to move to semantic mediation solutions ?
  - Can applications produce semantic annotations, natively ? (since human annotation irrealistic)
  - Should semantic annotations be derived from legacy relational DBs ? (maintenance of mappings)

How to ensure that the instances are properly identified across the federated system (URI)?

2nd key issue: which foundational ontology ?

- DOLCE or BFO ?
  - New BFO 2.0 (2012)
  - a new version of DOLCE coming soon, too (called DOLCE-core)
- Will they fill the gaps ?
  - e.g. Theory of observation and measurement
- How ? Ontology Design Patterns
  - Adoption ? Maintenance ?



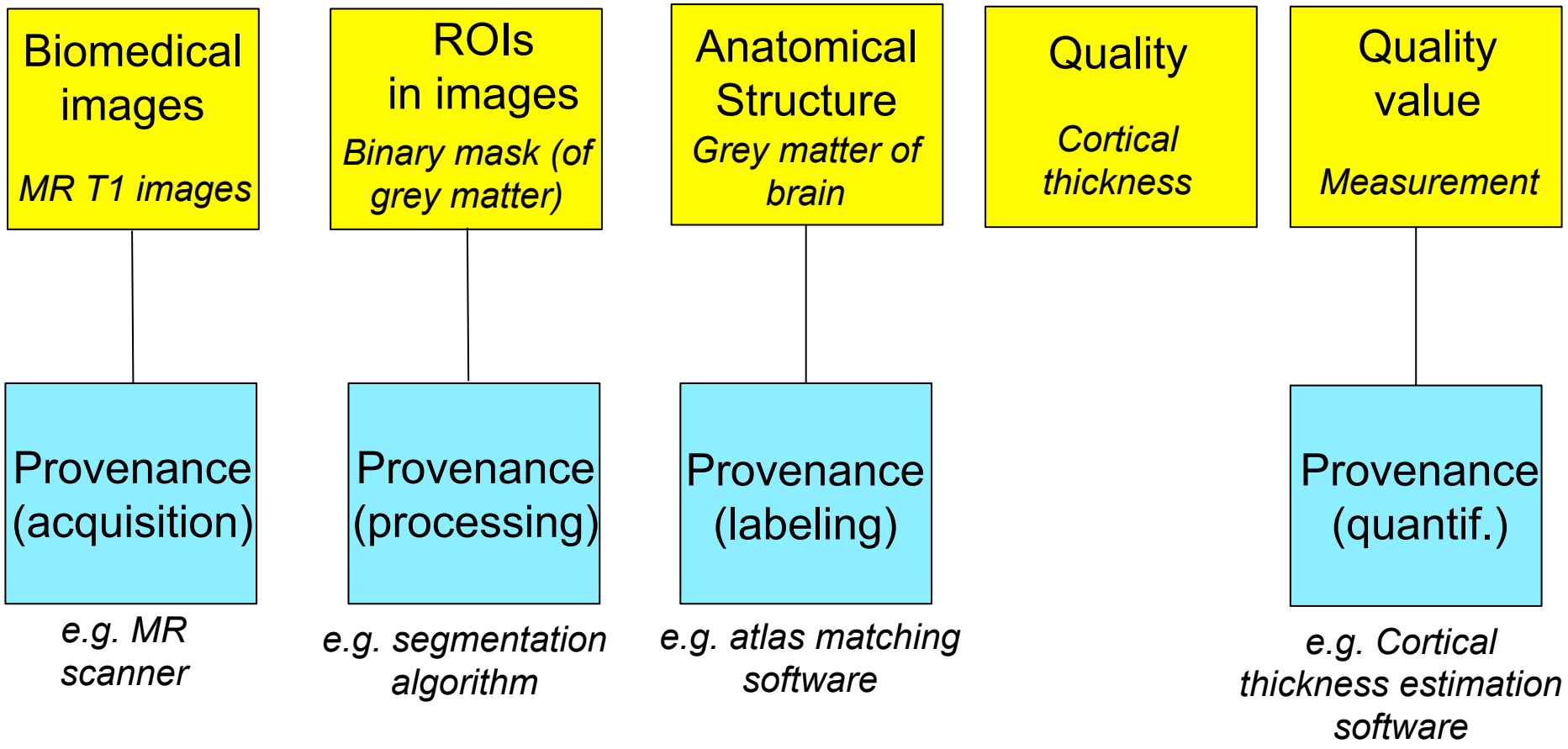
3rd key issue: acutely needed ontologies

- Medical images
- Imaging features
- Derived measurements
- Provenance in image processing
- Relation to biological entities

→ in order to link the measurements made using image processing to the image data

→ to facilitate the automatic production of provenance metadata from the knowledge about processing tools

Linking the measurements made using image processing to the images
Example: cortical thickness



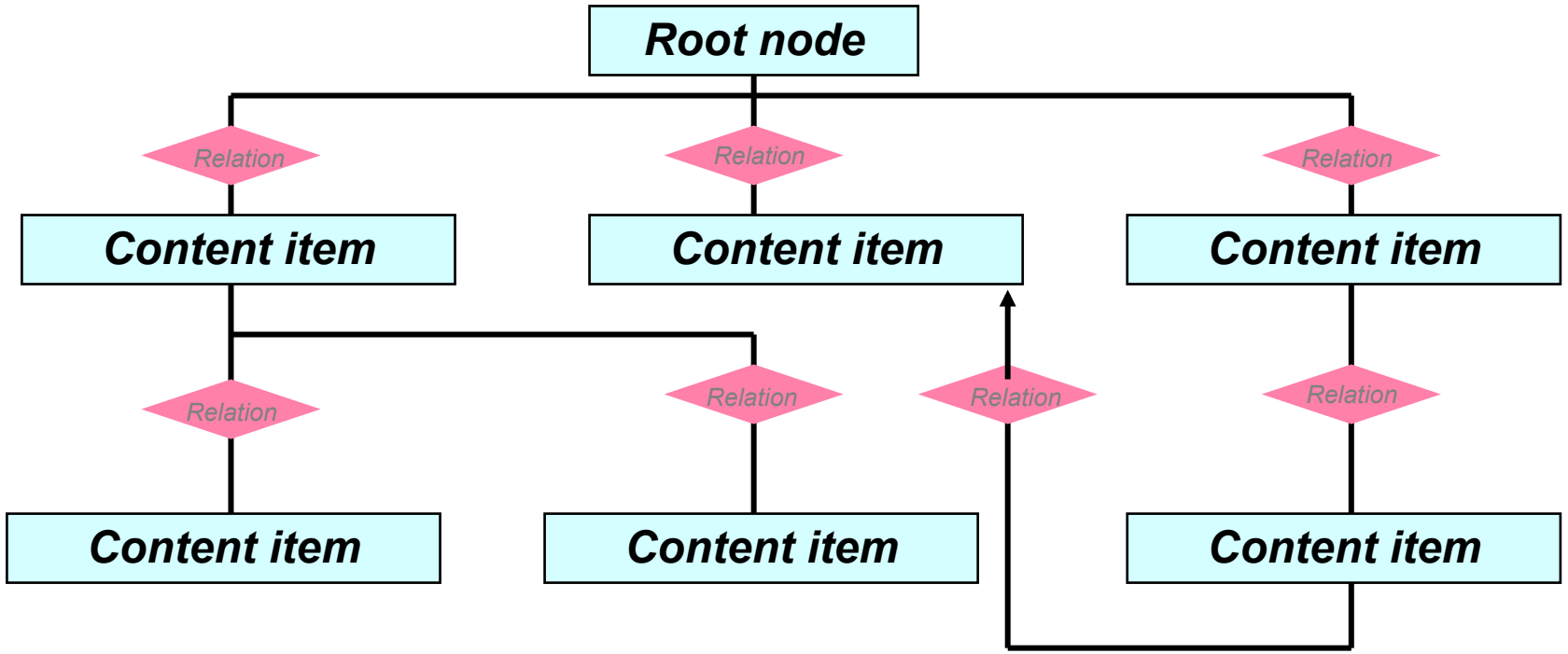
Linking the measurements made using image processing to the images

- This is a domain that was addressed by the DICOM Standard in DICOM SR (Structured Reports)
- Especially for representing CAD results

DICOM
Digital Imaging and Communications in Medicine

Structure of a SR document

- Tree structure (different types of nodes)



Nodes of a SR document

- CONTAINER
- TEXT
- PNAME (person name)
- DATETIME
- DATE
- TIME
- NUM (numerical value)
- IMAGE
- WAVEFORM
- COMPOSITE
- UIDREF
- SCOORD (spatial)
- SCOORD3D (3D)
- TCOORD (temporal)
- CODE

Nodes of a SR document

- CONTAINER
- TEXT
- PNAME (person name)
- DATETIME
- DATE
- TIME
- NUM (numerical value)
- IMAGE
- WAVEFORM
- COMPOSITE
- UIDREF
- SCOORD (spatial)
- SCOORD3D (3D)
- TCOORD (temporal)
- CODE

Red: context of observation ; *Blue: image evidence ;* *Black: other*

Relations within a SR document

- Contains
- Has Observation Context
- Has Acquisition Context
- Has Properties
- Inferred From
- Selected From
- Has Concept Modifier

Relations within a SR document

- Contains
- Has Observation Context
- Has Acquisition Context
- Has Properties
- Inferred From
- Selected From
- Has Concept Modifier

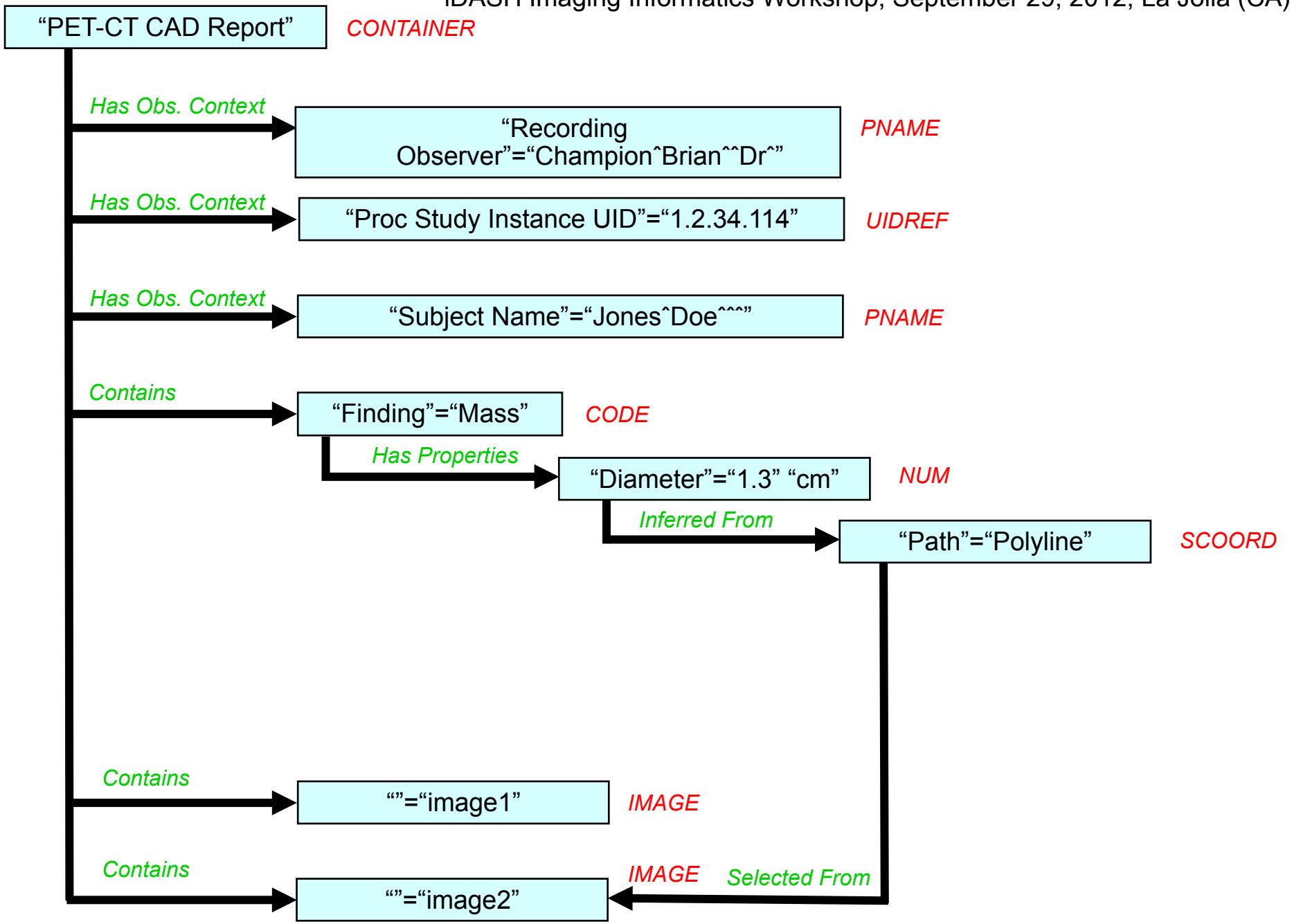
Red: context of observation ; *Blue: image evidence ;* *Black: other*

Contrainsts on relationships

ex: Chest CAD SR

Table A.35.6-2
RELATIONSHIP CONTENT CONSTRAINTS FOR CHEST CAD SR IOD

Source Value Type	Relationship Type (Enumerated Values)	Target Value Type
CONTAINER	CONTAINS	CODE, NUM, IMAGE[1], CONTAINER.
TEXT, CODE, NUM, CONTAINER	HAS OBS CONTEXT	TEXT, CODE, NUM, DATE, TIME, PNAME, UIDREF, COMPOSITE[1].
IMAGE, WAVEFORM	HAS ACQ CONTEXT	TEXT, CODE, DATE, TIME, NUM.
CONTAINER, CODE, COMPOSITE, NUM	HAS CONCEPT MOD	TEXT, CODE[2].
TEXT, CODE, NUM	HAS PROPERTIES	CONTAINER, TEXT, CODE, NUM, DATE, IMAGE[1], WAVEFORM[1], SCOORD, TCOORD, UIDREF.
CODE, NUM	INFERRED FROM	CODE, NUM, IMAGE[1], WAVEFORM[1], SCOORD, TCOORD, CONTAINER, TEXT.
SCOORD	SELECTED FROM	IMAGE[1].
TCOORD	SELECTED FROM	SCOORD, IMAGE[1], WAVEFORM[1].



(Adapted from D. Clunie SPIE 2001)

Possibilities / limitations of DICOM SR

- Rich possibilities
  - Control of content
  - Rich content (images, ROI, measurements, codes)
- … but, limitations
  - in DICOM syntax (binary, with DICOM tags)
  - in general, no formal semantics (codes, constraints on relationships)
  - Specific software for querying / reasoning on SR data

Linking the measurements made using image processing to the images

- So, what needs to be done is to revisit DICOM SR (as well as other relevant sources) to produce the needed ontologies
- Some works made in this direction by Daniel Rubin, in 2010

Missing ontologies: Medical images

- What needs to be modeled
  - The nature of signal (i.e. what the signal actually represents) and how it is represented: scalar, vector, tensor, etc.
  - What the variables represent : space, time, etc.
- Useful sources
  - General ontologies:
    - Notion of *Field*: Werner Kuhn (Univ Münster)
  - Specific:
    - *Abstract multidim. image model,* DICOM WG23 (Sup 118)

Missing ontologies: Imaging features

- What need to be modeled
  - ROIs (any dimension), meshes, paths (e.g. DTI)
- Useful sources
  - General ontologies
    - Spatial information: Werner Kuhn (Univ Münster)
  - Other:
    - AIM model, Daniel Rubin & coll., CaBIG
    - DICOM SR SCOORD, TCOORD

Missing ontologies: Derived measurements

- What need to be modeled
  - Well-defined quantities, derived from imaging features, especially imaging biomarkers
- Useful sources
  - General ontologies:
    - *Engineering mathematics*, Thomas Gruber 94
    - *Observation* and *Measurement*: Florian Probst, Werner Kuhn (Univ Münster)
  - Specific:
    - Catalog of imaging biomarkers (MGH)
    - DICOM SR specifications & codes (Snomed, RadLex)

Missing ontologies: Provenance in image processing

- What needs to be modeled
  - Image processing actions
  - Roles of data and parameters in image processing
  - Tools that are used
- Useful sources
  - General ontologies:
    - *Open Provenance Model*
  - specific:
    - *Provenir*, Satya Sahoo
    - DICOM SR

Relation to biological entities

- Ontologies are available
  - Anatomy, e.g. FMA
  - Physiological processes, e.g. SNOMED
  - Qualities and traits e.g. PATO
  - Pathology, e.g. NCIT, SNOMED, MPATH

- But their integration remains non-trivial
  - Integration efforts: Neuroscience Information Framework (Maryann Martone)

Relation to biological entities
key problems

- Overlapping ontologies
- Insufficient modularity
- Based on different incompatible modeling frameworks and foundational ontologies
- E.g., Problems with « canonical » anatomy
  - Pathology: → inconsistencies (Robert Hoehndorf 2007)
- Unequal quality

Conclusion

- The NeuroLOG project allowed us to get a practical experience of ontology-based sharing of heterogeous distributed images

- I summarized here some of the lessons we learned, and reviewed some of the issues for developing such systems

Conclusion

- As a continuation, two approaches are possible
  - A top-down one, in which you analyze the requirements of translational research and develop the necessary components (ontologies, automated annotation tools, etc.)
  - A bottom-up one, in which you get experience from focused applications, based on existing components
- Both are needed
  - The latter is the only one to convince researchers of the feasibility and the added value of such data sharing
  - The former is indispensable with respect to large-scale multi-domain data integration and data mining

Acknowledgements

- ANR, who supported NeuroLOG under contract ANR-06-TLOG-024
- Former partners of the NeuroLOG project, especially:
  - Johan Montagnat
  - my colleagues of the ontology task force: Gilles Kassel, Michel Dojat and Benedicte Batrancourt
- Current partners of the CrEDIBLE project
- Colleagues of the DICOM WG6 and WG23, especially Larry Tarbox