



# Towards negotiable SLA-based QoS Support for Data Services

ESSI Seminar

UPC, 11<sup>th</sup> May 2010

**Jesús Bisbal**

*Universitat Pompeu Fabra, Barcelona, Spain*

Published in Proceedings of the 10th IEEE International Conference on Grid  
Computing (Grid 2010), pages 259-265.



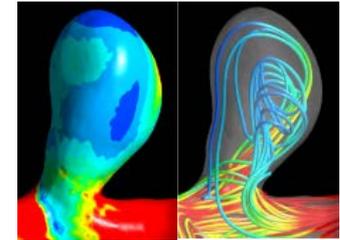
# Outline

- Motivation for domain-specific data QoS
- Quality of Service (QoS) – Service Level Objectives (SLOs)
- QoS Model
- QoS Negotiation and QoS SLAs
- QoS Management in Data Mediation
- Experimental Evaluation
- Conclusions and future directions

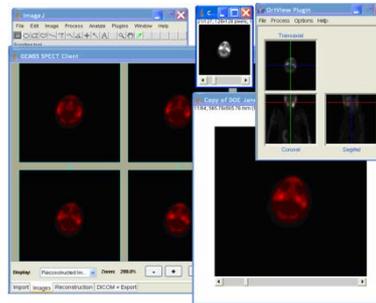
# QoS Scenario – Traditional Objectives

I want to pay less than 10 €,  
I can start simulation today  
at noon and I need the  
results by 3 pm

Blood flow Simulation



Medical  
practitioner



CFD Client

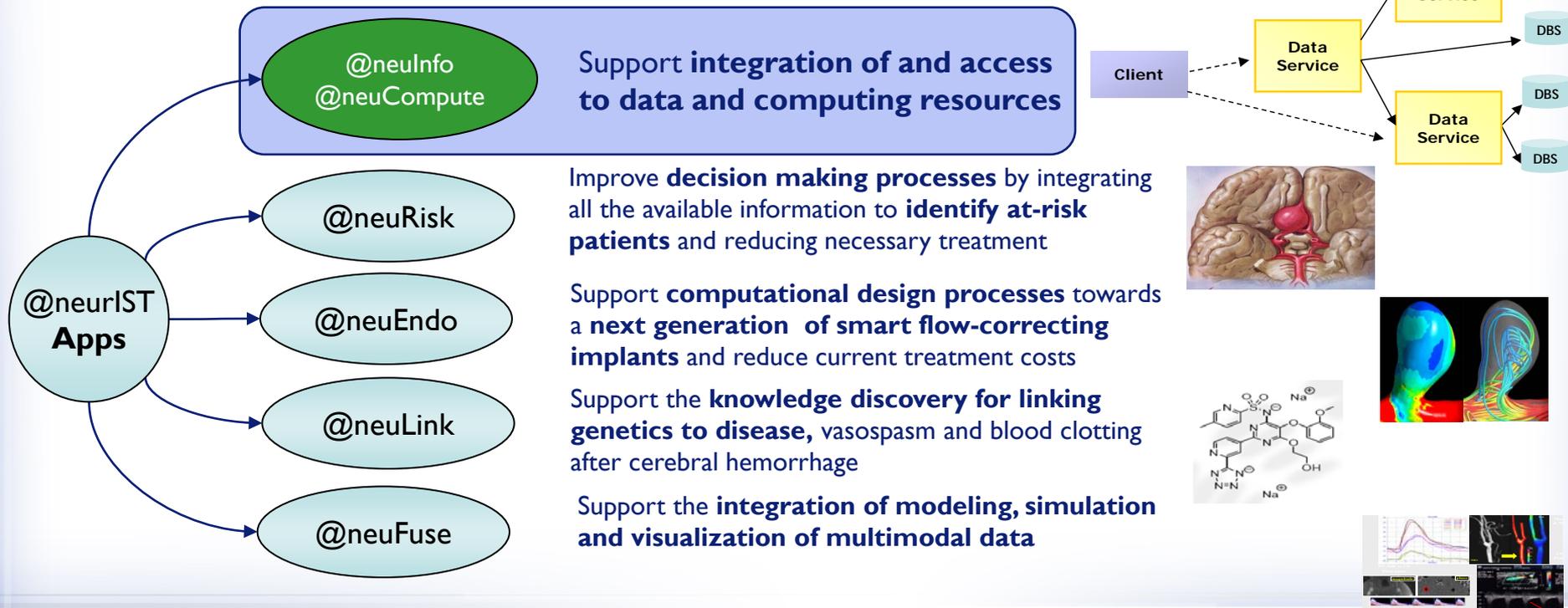


QoS-aware  
Grid  
Service

- Remote HPC facilities to be used by many different customers/clients
- Guaranteed **response times** and **price**
  - Resource reservation
  - Capacity/resource estimation
- Need to go beyond **time** and **price** guarantees: QoS in data services

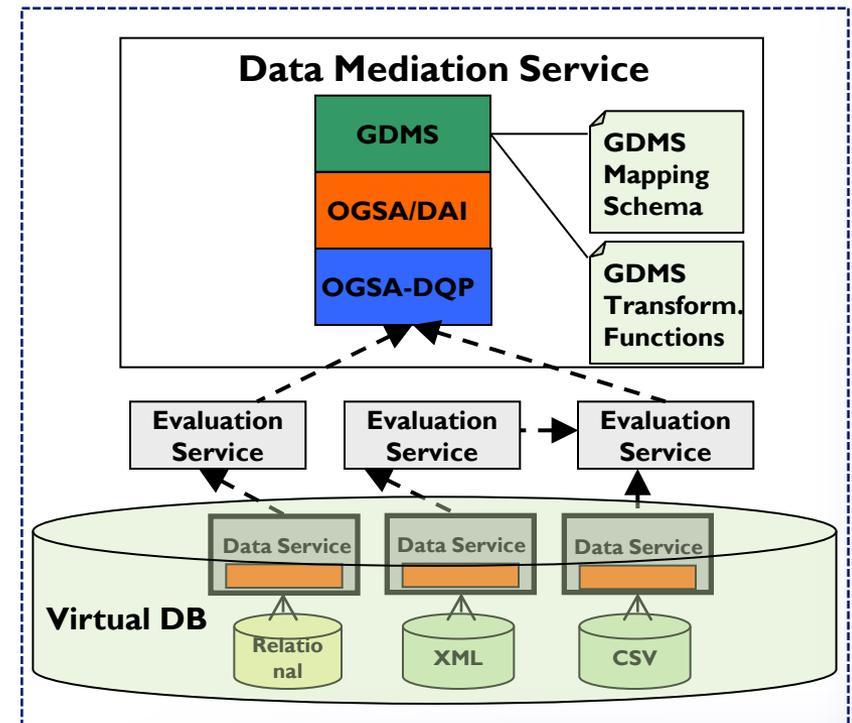
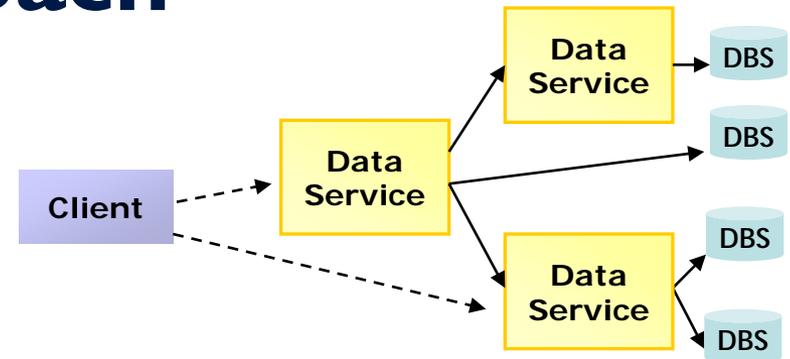
# Motivation – QoS on Biomedical Data

- ❑ **@neurIST project** – EU Integrated Project for the ‘Integrated Biomedical Informatics for the Management of Cerebral Aneurysms’
  - ❑ **Service-oriented ICT infrastructure** providing
  - ❑ **On-demand simulation, analysis and data-integration services**
  - ❑ Handling **multi-scale, multi-modal** information at distributed resources



# Data Mediation approach

- ❑ Data access and integration
- ❑ Virtualization of heterogeneous data sources as services
  - ❑ Hierarchical composition of data services
  - ❑ Integration of multiple data sources
  - ❑ Based on **OGSA-DAI**, de-facto standard for data access **on the Grid**
  - ❑ Distributed Query Processing (DQP)
- ❑ Data mediation services set up manually - Mapping Schemas
  - ❑ Large efforts required
  - ❑ Future semantic mediation ...



# Mapping Schema overview

```
<VDSConfig>
<logicalSchema>
  <table name="treatment">
    ...
  </table>
</logicalSchema>
<MappingSchema>
  <VDSTable name="treatment">
    <Join>
      <Select resource="http://..">
        <column name="id"
          source="did"/>
        ...
      </Select>
      <Select/>
      <JoinInfo type="parallel"/>
    </Join>
  </VDSTable>
</MappingSchema>
</VDSConfig>
```

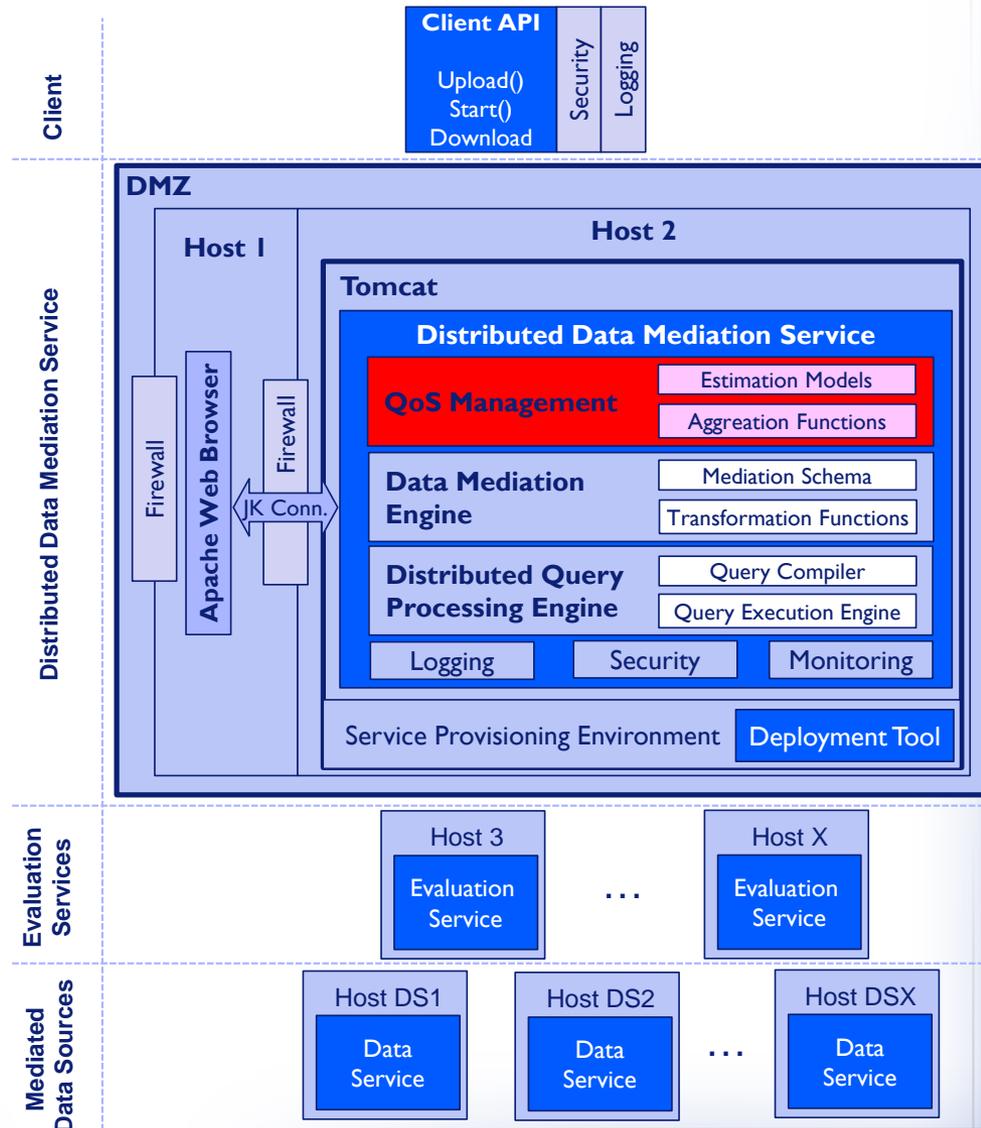
□ Global-as-View (GAV)  
mediation approach

1. Definition of Global  
Schema

2. Mapping rules between the  
global schema and the  
integrated schemas

# Data Mediation Architecture

- ❑ Architecture of the **Vienna Grid Environment (VGE)**
- ❑ QoS Management for data - *new*
- ❑ Data Mediation Engine and Distributed Query Processing (DQP) run on a service hosting environment (Tomcat + Axis)
- ❑ Query Evaluation Services set up on several hosts (DQP)
- ❑ Data Sources to be integrated run on separated hosts





# Data Mediation Practice

- ❑ Follows a **Best Effort** strategy for data services
  - ❑ Queries all services available
  - ❑ Applies mapping rules
  - ❑ Compiles result
- ❑ Recall that “The Grid ...
  - ✓ *uses standard, open, general purpose protocols and interfaces*
  - ✓ *coordinates resources that are NOT subject to centralized control*
  - *delivers non-trivial qualities of service”*

*Foster, Kesselman (2002)*

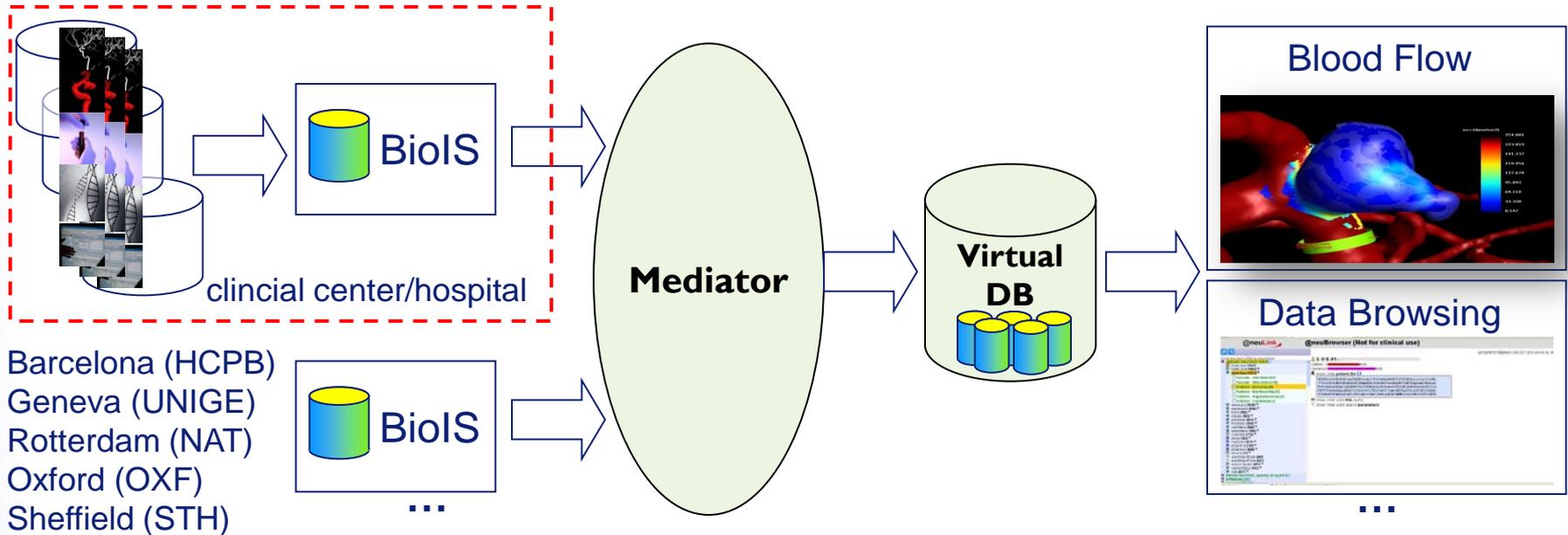
- ❑ Explore the specificities of **Qualities of Service within Data Mediation Services**
  - ❑ Common requirement for advanced scientific applications
  - ❑ Defines path to Business Model for typical (scientific) usages
  - ❑ Experimentation using the VGE-based data mediation middleware
  - ❑ QoS Management prior to initiating data mediation and QDP

# Usage of Data Grid Services

Clinical Sites

Data Mediation

Applications



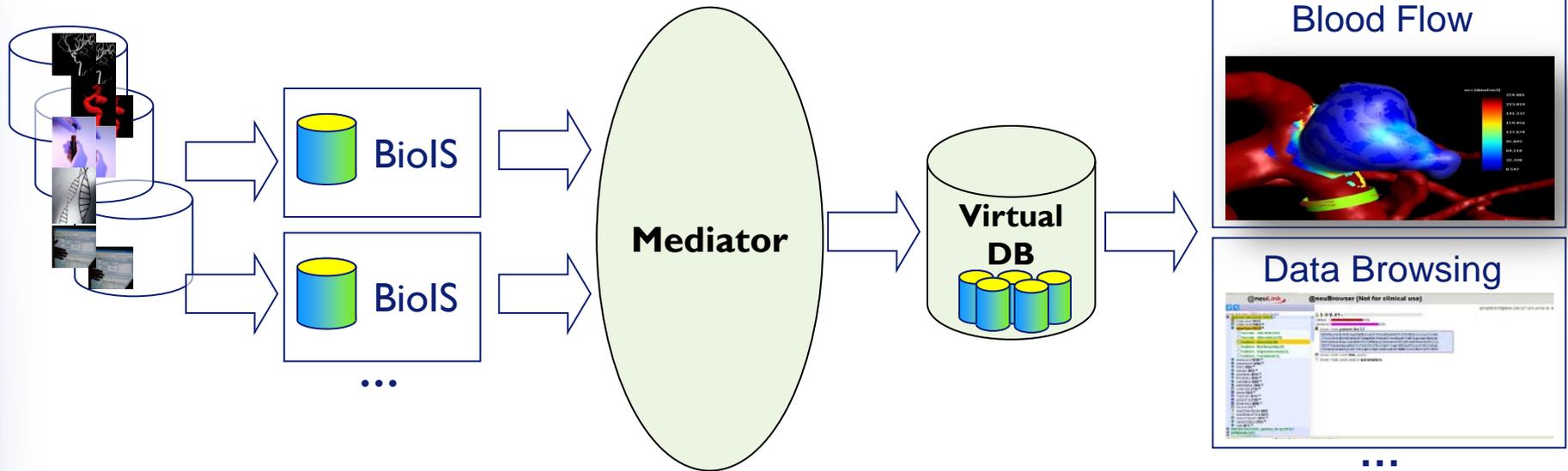
Virtualization of **distributed** and **heterogeneous data sources** as a large single **virtual database** (federation of data access)

# Why QoS for Data in this Context?

ClinicalSites

Data Mediation

Applications



- Data is fragmented
- Amount of relevant data
- Cost of data access
- Security/Privacy

## Biomedical research use-cases

- Data mining (epidemiology)
- Content-Based Information Retrieval (decision support)
- Atlas generation (population variability)

# QoS Objectives → SLOs for Data

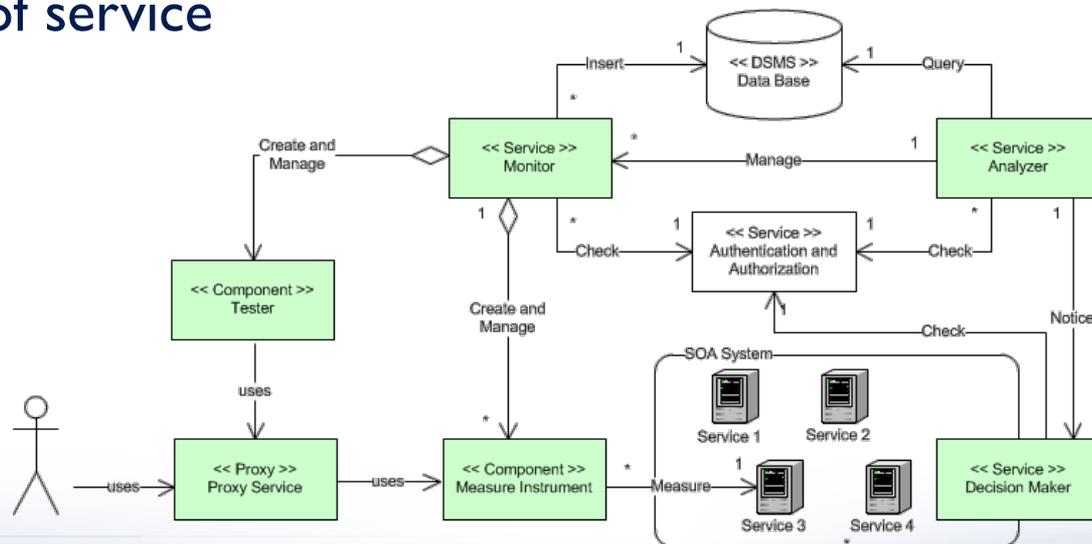
Adapt QoS management from computing services to data services

Service Level Objective (SLO)	Description
Cost	Price of query execution, based on pricing model (e.g. constant, function of result size)
Response Time	Guarantee response time to retrieve all results, depends on size of query result
Data Cardinality	Cardinality of total subjects (e.g. tuples) returned
	Cardinality of reliable / quality (complete) subjects, or level of <i>constraints satisfaction</i> acceptable
	Cardinality of queried subjects
Data Diversity	Maintain a certain diversity of data sources being queried (providers) – epidemiology
Data Locality	Specify the Locality of data access (legal constraints)

# SLOs for Data: Monitoring

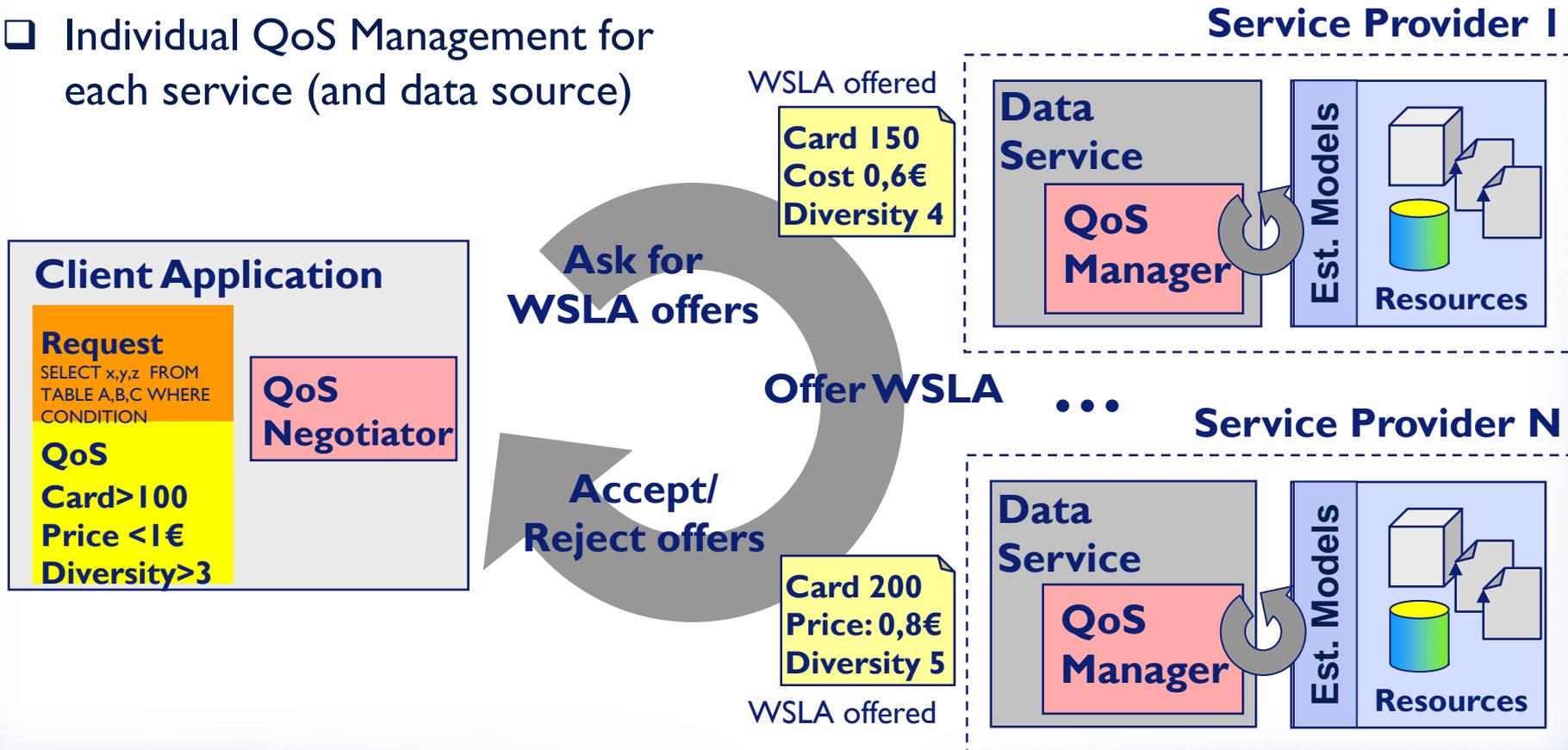
New SLOs require novel Monitoring – **SALMon** to identify degradation

- ❑ Identify Response Time degradation after SLA have been accepted
- ❑ Data-intensive scientific domains with QoS beyond response time
- ❑ Need to monitor the satisfaction of agreed SLAs for these other qualities of service

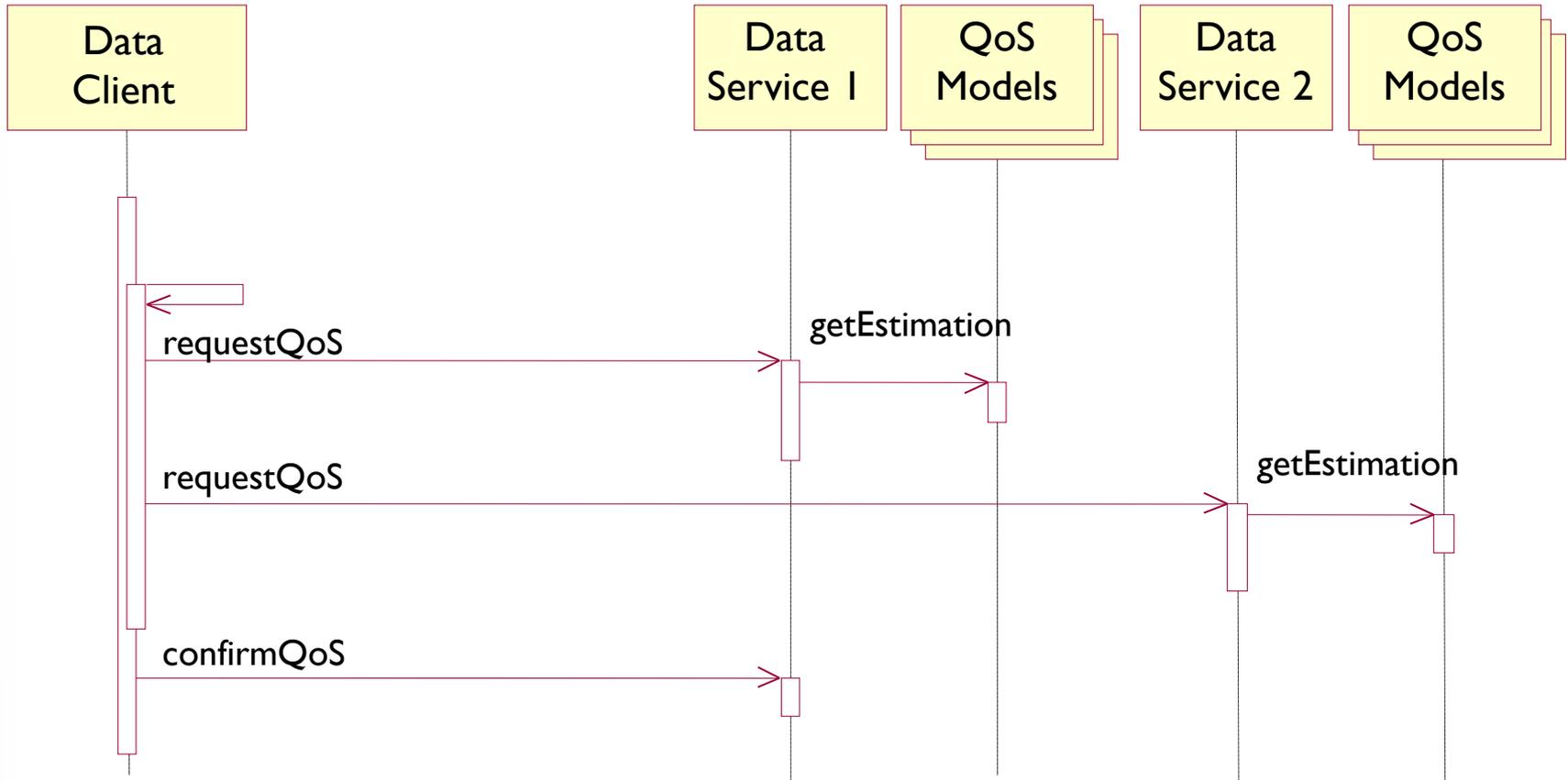


# QoS Model for Data Services

- ❑ Client driven **QoS negotiation** with potential service providers  
Client supplies: QoS requirements (e.g. **data quality**) and data request
- ❑ Request/Offer are **Web Service Level Agreements (WSLA)**
- ❑ Individual QoS Management for each service (and data source)



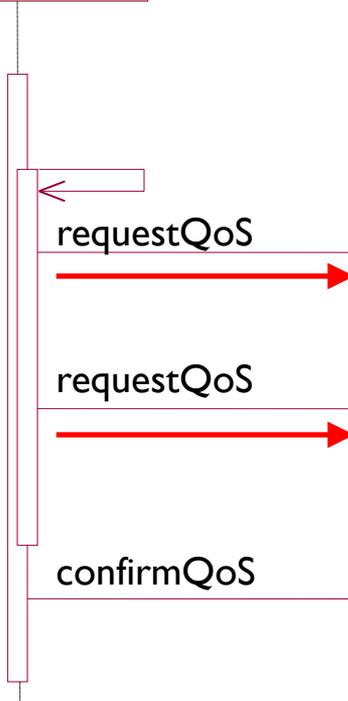
# QoS Negotiation and WSLAs



- ❑ Negotiation follows (multiple rounds of) Request-Offer and finally a confirmation
- ❑ Based on Web Service Level Agreement (WSLA)

# QoS Negotiation and WSLAs

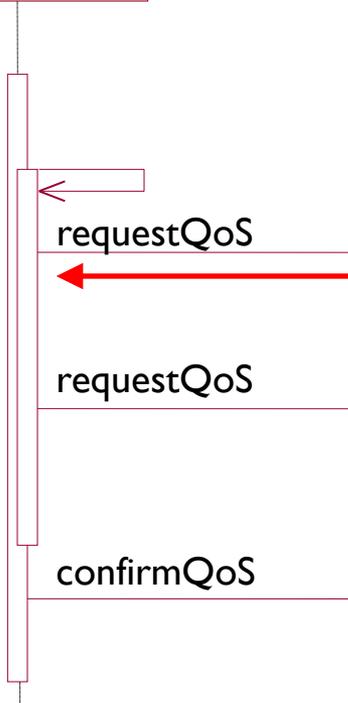
Data Client



```
<SLA xmlns="http://www.ibm.com/wsla" ... >
  <Parties> <ServiceConsumer> <!-- from certificate -->
    </ServiceConsumer>
  </Parties>
  <ServiceDefinition ... name="BioIS">
    <SLAParameter name="cost" ...>
    <SLAParameter name="cardinality" ...>
    <SLAParameter name="diversity" ...>
    ...
    <!-- Metrics for each SLA parameter --> ...
    ...
  </ServiceDefinition>
  <Obligations>
    <ServiceLevelObjective name="cost"> ...
      <Expression><Predicate xsi:type="LessEqual">
        <SLAParameter>price</SLAParameter>
        <Value>1</Value> <!-- 1 Euro -->
      </Expression>
      ...
    <ServiceLevelObjective name="cardinality"> ...
      <Expression><Predicate xsi:type="GreaterEqual">
        <SLAParameter>cardinality</SLAParameter>
        <Value>100</Value> <!-- 100 result sets -->
      </Expression>
      <!-- other objectives --> ...
    </ServiceLevelObjective>
  </Obligations>
</SLA>
```

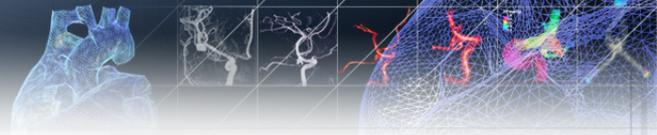
# QoS Negotiation and WSLAs

Data Client



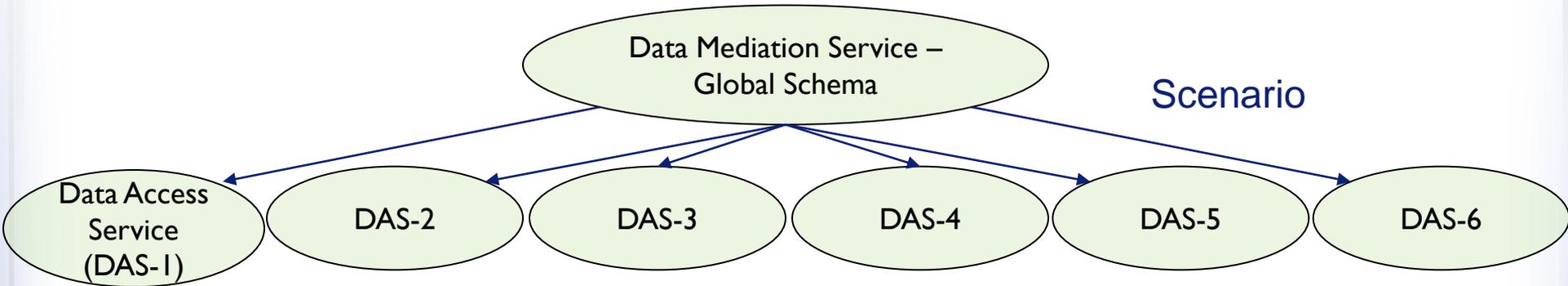
Client

```
<SLA xmlns="http://www.ibm.com/wsla" ... >
  <Parties> <ServiceConsumer>...
            <ServiceProvider>...
  </Parties>
  <ServiceDefinition ... name="BioIS_UPF">
    <SLAParameter name="cost" ...>
    <SLAParameter name="cardinality" ...>
    <SLAParameter name="diversity" ...>
    <!-- Metrics for each SLA parameter --> ...
    <WSDLFile>https://datanode.upf.edu/.../ds?wsdl ...
    <!-- Definition of service operations --> ...
  </ServiceDefinition>
  <Obligations>
  <ServiceLevelObjective name="cost"> ...
    <Expression><Predicate xsi:type="Equal">
      <SLAParameter>cost</SLAParameter>
      <Value>0,6</Value> <!-- 0,6 Euro -->
    ...
  <ServiceLevelObjective name="cardinality"> ...
    <Expression><Predicate xsi:type="Equal">
      <SLAParameter>cardinality</SLAParameter>
      <Value>150</Value> <!-- 150 result sets -->
    <!-- other objectives --> ...
  </Obligations>
</SLA>
```



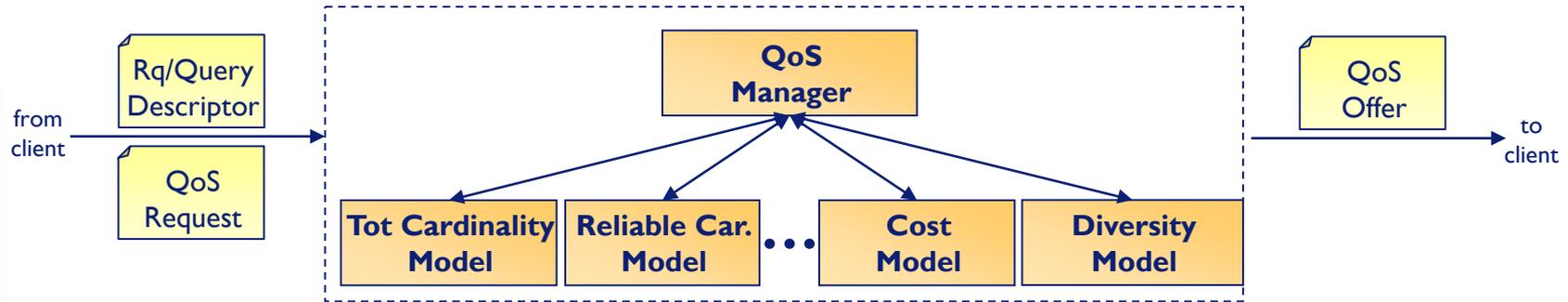
# QoS Aggregation of Federated Data Services

- ❑ Client aggregates QoS from **several** Data providers to meet SLO
- ❑ Data mediation/federation services aggregate QoS offers

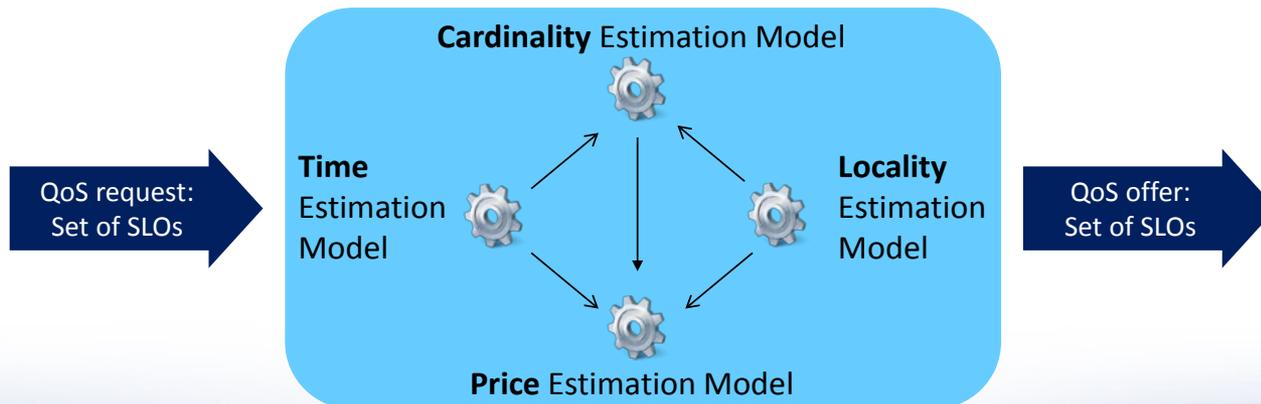


SLO	Satisfaction condition	Aggregation Function
Cost	$\leq$	$\Sigma \text{cost}(\text{DAS}_i)$
Response time	$\leq$	$\mathbf{max} \text{resp}(\text{DAS}_i)$
Cardinality	$\geq$	$\Sigma \text{card}(\text{DAS}_i)$
Diversity	$\geq / =$	$\Sigma \text{dive}(\text{DAS}_i)$
Locality	$=$	$\Lambda \text{loca}(\text{DAS}_i)$

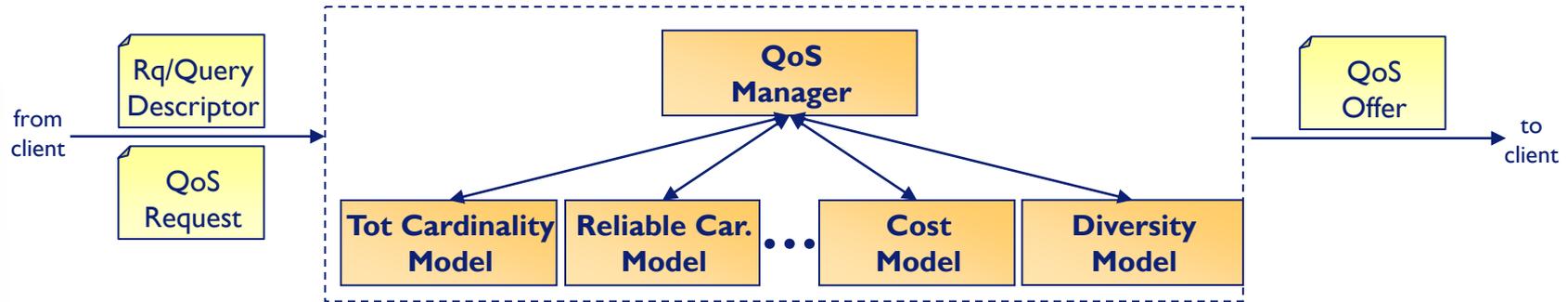
# QoS Management



- ❑ **Estimation model** predicts one or more **SLOs**  
Data source specific (relational DBs vs. PACs/DICOM images)
- ❑ Estimation Models may dependent on prediction of another model



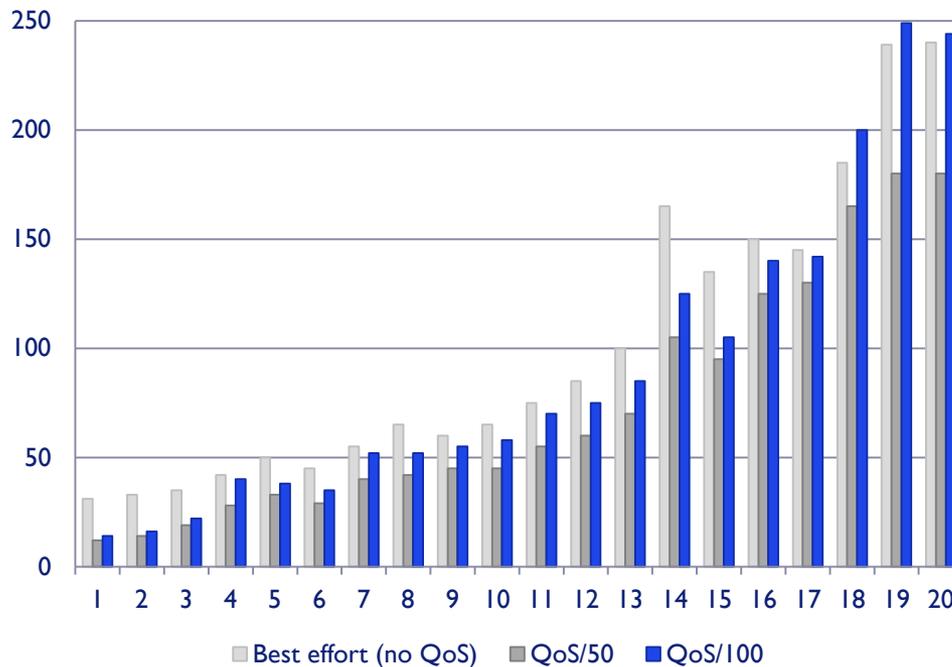
# QoS Management



- ❑ Estimation Models may depend on prediction of another model  
→ Challenge of **orchestrating** the models (direct acyclic graph of models)
  - ❑ **Brute force**: executing all permutations of models (<5 SLOs)
  - ❑ **Topology sort** to identify model invocation sequence (>5 SLOs)
- ❑ Conflicting objectives, cyclic dependencies - potential solutions:
  - ❑ Genetic algorithms
  - ❑ Mixed integer programming and linear programming (MIP/LP)
  - ❑ Answer set programming (ASP)

# Experimental Evaluation

- ❑ Sample queries against @neurIST existing (best effort) data services
- ❑ Execute with QoS constraints (cardinality 50 or 100) and without constraints
- ❑ Measure query execution time

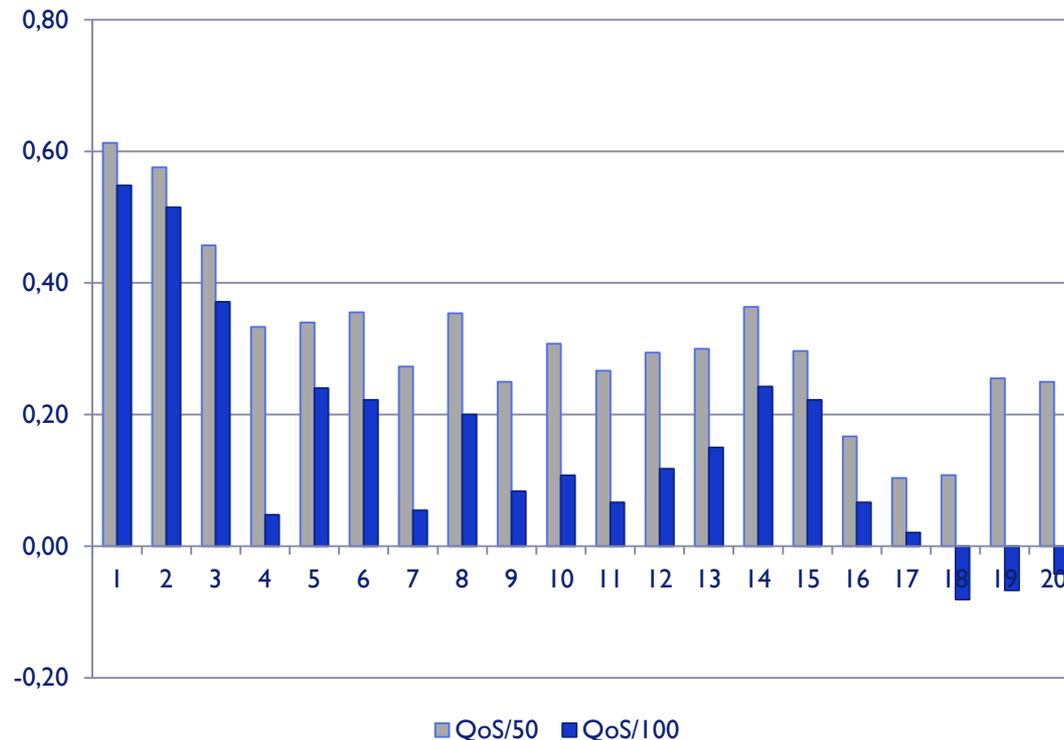


- ❑ Samples queries sorted by size of their results
- ❑ Ranging from:
  - Q1 few KBytes to
  - Q20 few MBytes

- ❑ QoS Support saves up to 60% query execution time

# Experimental Evaluation (II)

- ❑ Compare gain with respect to 'best effort' query execution policy



- ❑ QoS guarantees the specified constraints (i.e. cardinality of results)
- ❑ But... QoS/100 can be worse... Thus *efficient QoS Management and Negotiation* remains challenging



# Conclusions

- ❑ Domain driven QoS approach, tested in @neurIST sources
- ❑ QoS Negotiation
  - ❑ Request-Offer-Confirmation workflow
  - ❑ Aggregation of Service Level Objectives (SLOs)
- ❑ QoS Management
  - ❑ QoS Estimation Models
  - ❑ Different orchestration approaches



# Future Work

- ❑ Identify synergies with Earth Observation applications (ESA, [www.esa.int/esaEO](http://www.esa.int/esaEO)) for SLOs for data services
- ❑ Evaluate guarantee of other data-SLOs (data diversity, quality, locality)
- ❑ QoS Support for more heterogenous data resources (different image modelities, simulation results/models, genetics, etc.)
- ❑ Investigation of more sophisticated QoS Mgmt models
  - ❑ Evaluate resolution of conflicting objectives
- ❑ *Cloud* infrastructure provision

**Questions?**

**Thank You**



# The @neurIST Project



## Integrated Biomedical Informatics for the Management of Cerebral Aneurysms

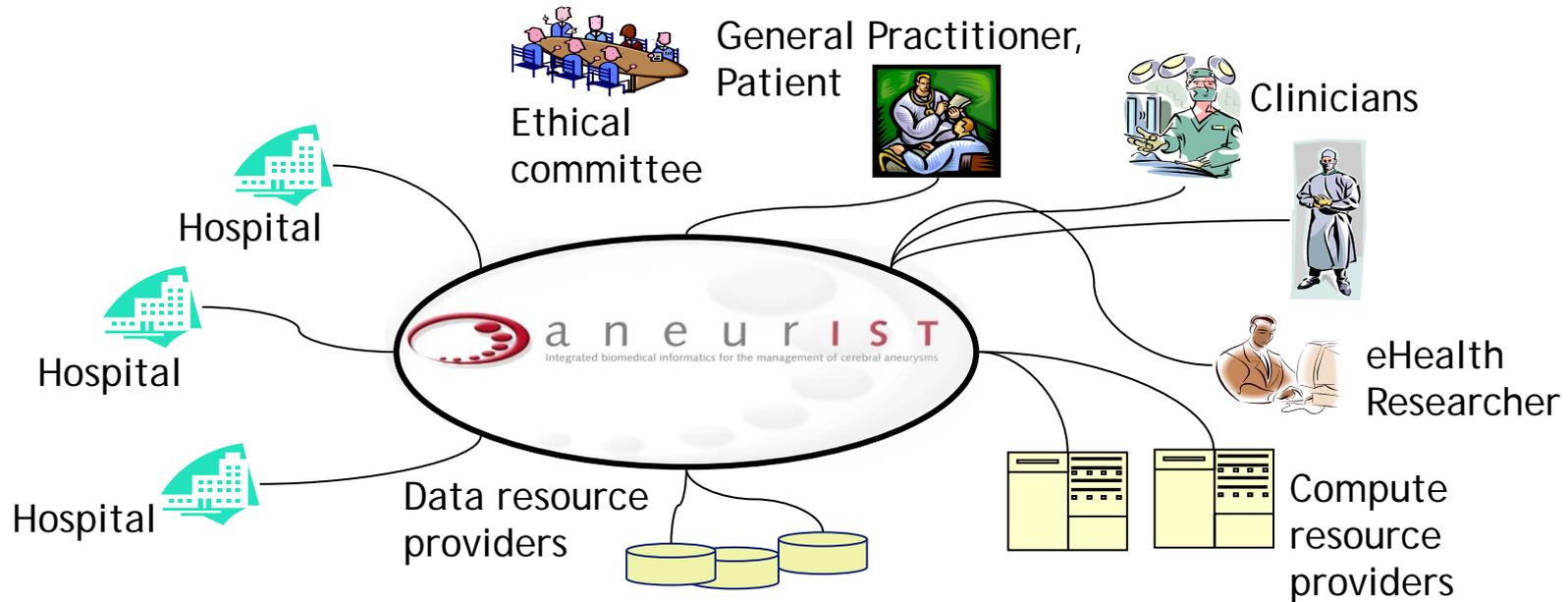
- ❑ Project duration: 2006-2010 (FP 6)
- ❑ 30 Partners
- ❑ Budget: ~17,5 MEuro



### Objectives:

- ❑ Development of a **generic IT infrastructure** for the **management & processing of heterogeneous data** associated with the diagnosis & treatment of cerebral aneurysms.
- ❑ Transform the management of cerebral aneurysm by providing new insight, **personalised risk assessment** and methods for the **design of improved medical devices and treatment protocols**.

# Motivation – QoS on Biomedical Data



## Generic Processes:

- Obtaining relevant clinical information of patients (EHR - Electronic Health Record)
  - Providing clinical decision support
  - Offering simulation services
  - Creating normalized population-based datasets
  - Providing knowledge discovery services
- Compute power for simulations
  - Patient data confidentiality
  - Data access and integration