

The Role of MetaData in Querying Grid-Resident Medical Images

Tamás Hauer

University of West England – CERN

IST Workshop on Metadata Management in Grid and P2P Systems

16th December 2003

Outline



- The MammoGrid project
- Meta-Data and domain ontology (flexibility)
- Meta-Data, services and grid (openness)
- Query negotiator
- Outlook

Outline



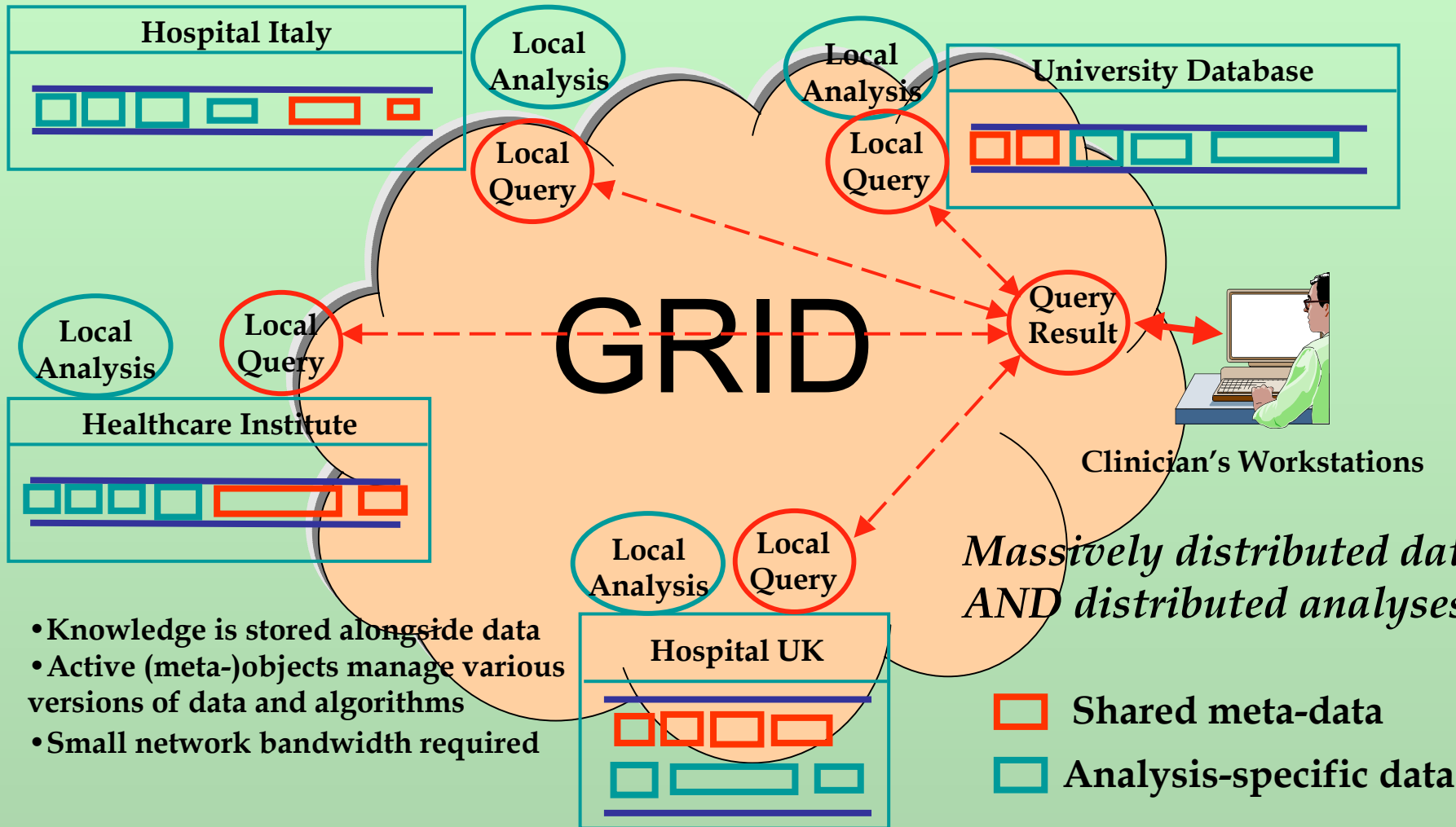
- **The MammoGrid project**
- Meta-Data and domain ontology
- Meta-Data, services and grid
- Query negotiator
- Outlook

The MammoGrid Project



- A grids solution for mammography
- EU-funded, 2002-2005
 - CERN, Mirada Solutions
 - Universities: UWE, Oxford, Pisa, Sassari
 - Hospitals: Addenbrookes (Cambridge), Policlinico Universitario, Udine
- Proof-of-concept R&D:
 - Grid + medicine
 - Pan-European distributed X-ray image database
 - Possible target areas:
 - Education
 - Quality control
 - Epidemiology research
 - Standardized screening
 - Communication aid

Federated System Solution



- Knowledge is stored alongside data
- Active (meta-)objects manage various versions of data and algorithms
- Small network bandwidth required

Outline

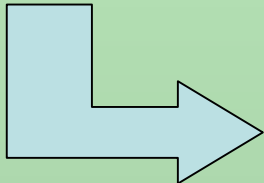


- The MammoGrid project
- **Meta-Data and domain ontology**
- Meta-Data, services and grid
- Query negotiator
- Outlook

The Medical Domain

➤ Requirements:

- User community:
 - process-oriented rather than information-oriented
 - Heterogeneous (technician, radiologist, researcher,...)
- Research area → hard-to-predict requirements
- Regional differences
- Constantly changing knowledge-base



Flexibility, Extensibility is top priority
Management of domain information

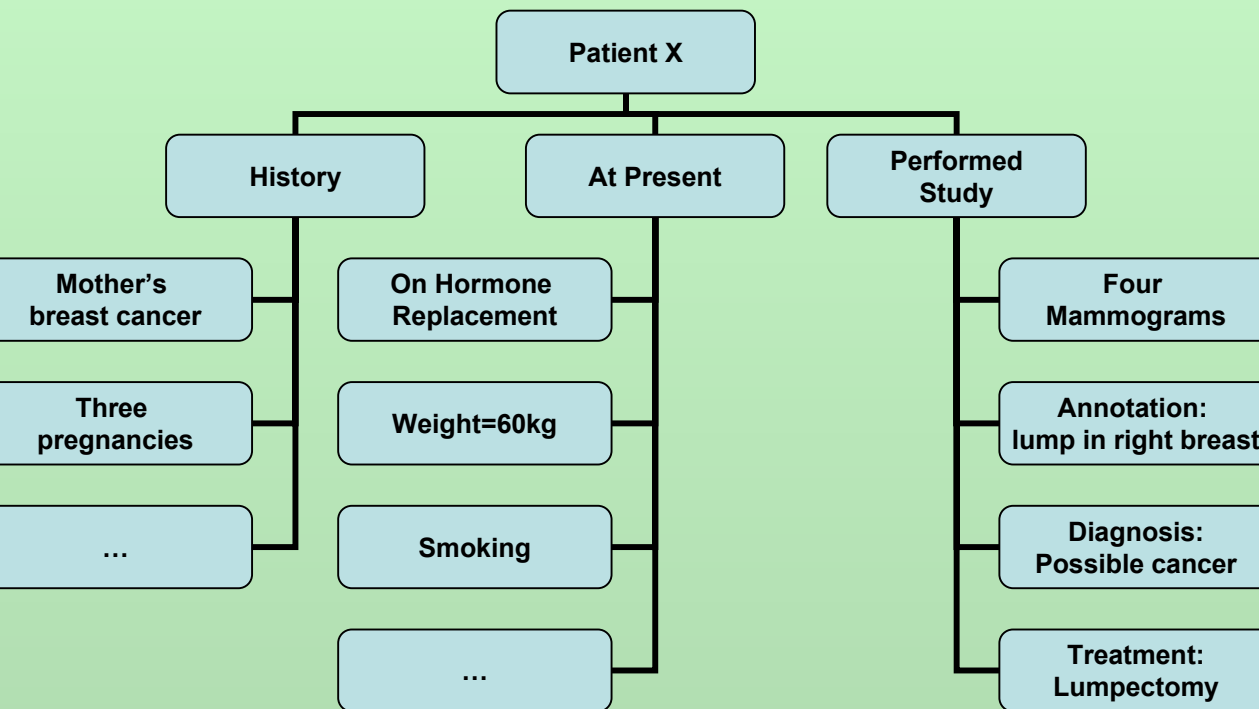
The Medical Domain



➤ Entities

- Patients, Physicians, Hospitals, Equipment, ...
- Medical History (Drugs, Surgery, Pregnancy,...)
- Studies, Images, Pathology results, ...
- Annotation, Diagnosis, Treatment, ...
- Treatment protocol, screening, ... (workflow)
- Epidemiological study, research, ...

The Medical Domain



Define semantics so that elements are queryable

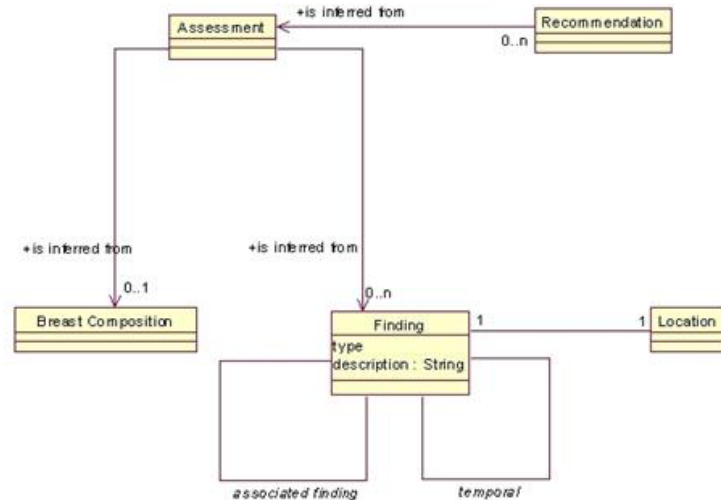
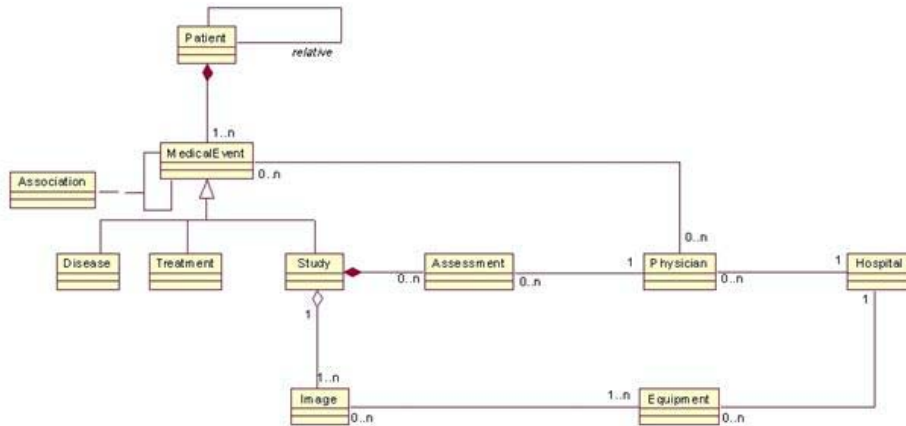
Bridge discrepancies between different standards

Associations between pixel data of digitized images and structured information

Extension to other modalities

CAD, "Find-one-like-it"

The Medical Domain



Typical workflow:

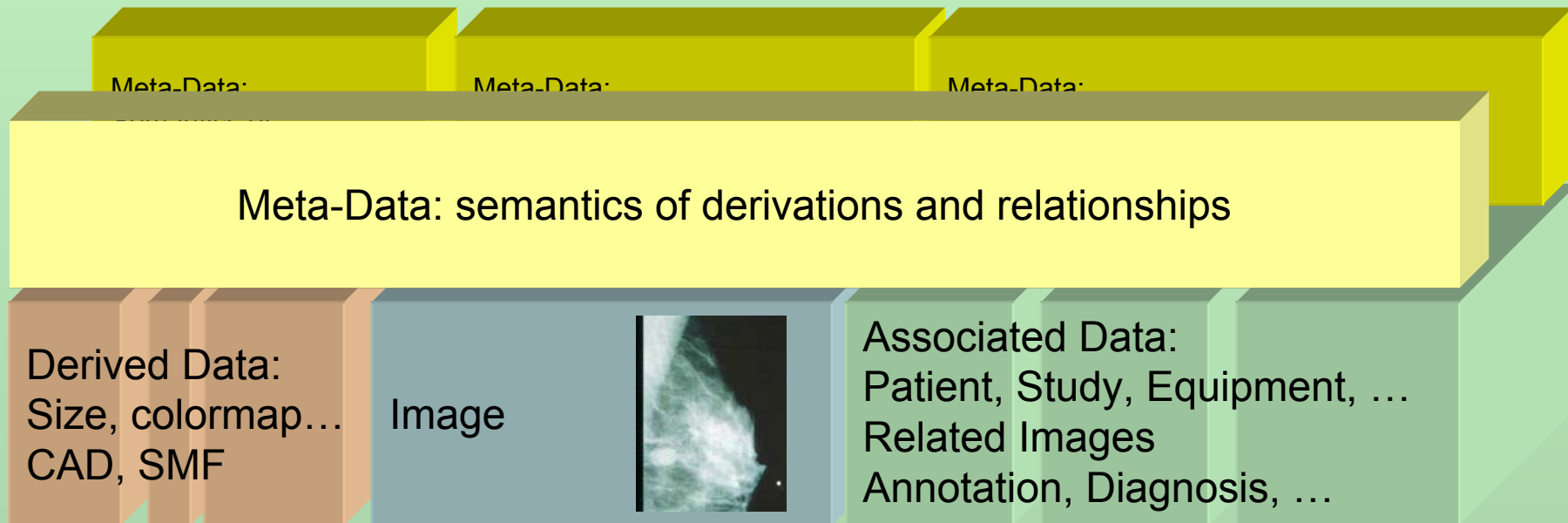
1. Register patient, scan image, add pathology data
2. Run SMF to bring image(s) to standardized format (possibly automatic)
3. Annotate image (define location, nature, size, etc... of finding(s))
4. Run CAD
5. Define relationship between findings (temporal, spatial pathological)
6. Write assessment of image, series, study. Associate to findings



The Medical Domain

Image Centric Approach:

- Digitized images (e.g. raw file)
- Semistructured information about the image
- Associate non-derived information (make it useful and interesting)
- Describe the image format
- Define “classes” of derived and associated information
- Describe the relationships between the entities (of a record, of records)



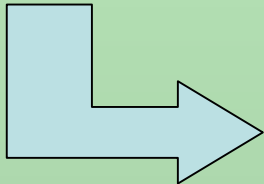
Outline



- The MammoGrid project
- Meta-Data and domain ontology
- **Meta-Data, services and grid**
- Query negotiator
- Outlook

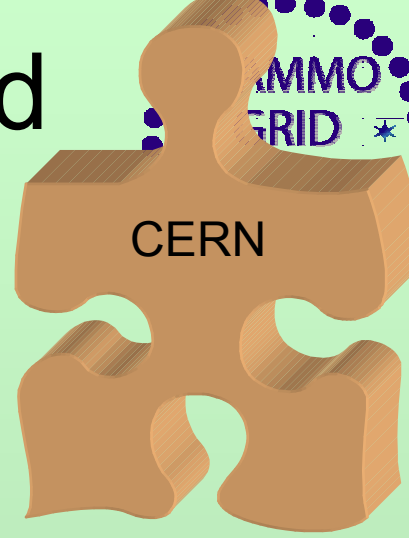
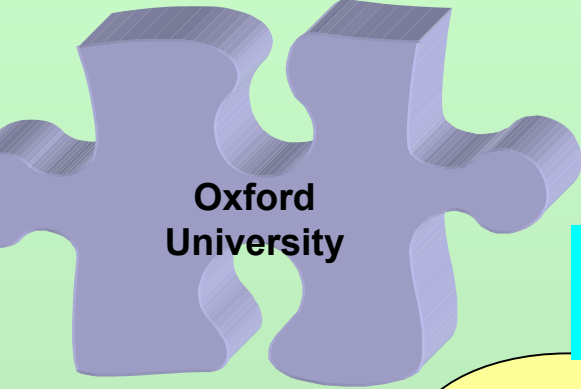
➤ Requirements:

- Data heavy (20MB/image) – storage, bandwidth
- CPU heavy (image processing: CAD, SMF,...)
- Geographically distributed, inhomogeneous,...
- “Plug-and-play” → discovery
- Scalability, adaptability
- Confidentiality



- Service-oriented architecture
- Use grid-middleware

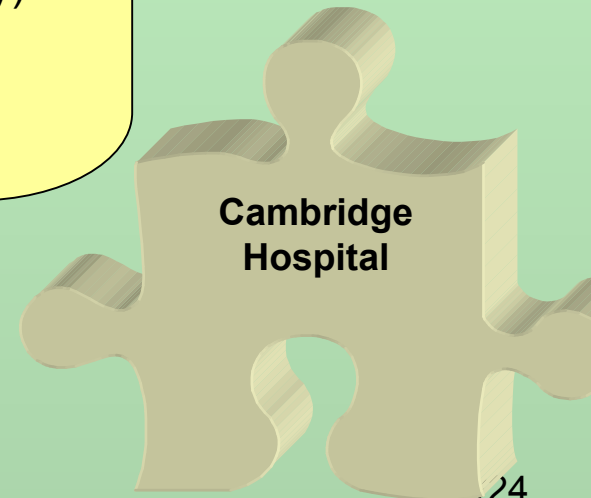
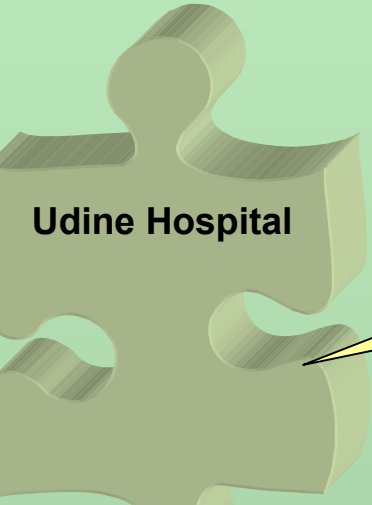
Meta-Data services and grid



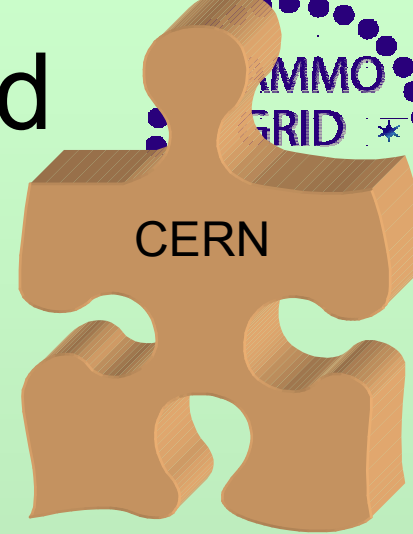
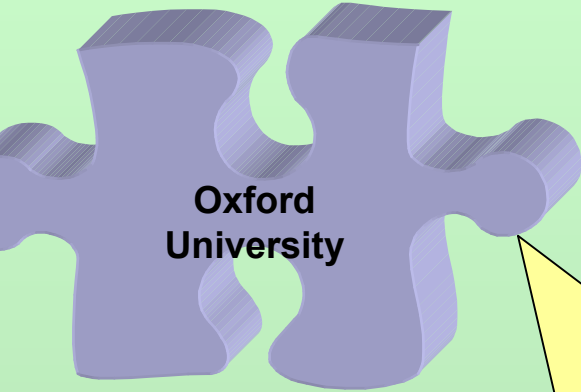
Meta-Data describes service definition

Udine Hospital:

- I store mammograms
- You can talk to me via SOAP or DICOM
- I can run algorithms (CAD) for you
- You need to use XXX authentication format (so that I can verify your site and country)
- I authenticate myself with XXX or YYY format
- You can subscribe to ...



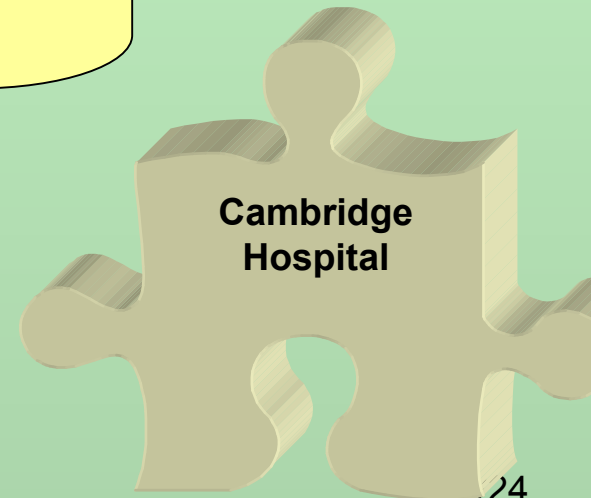
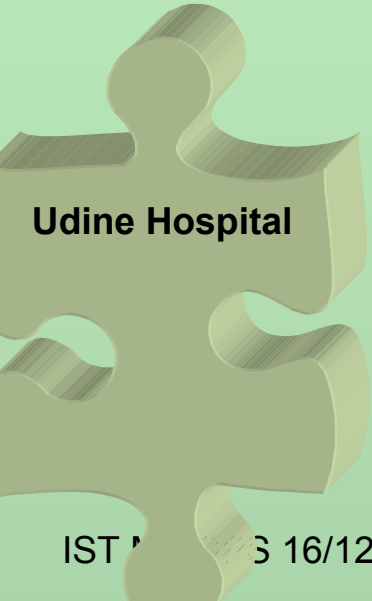
Meta-Data services and grid



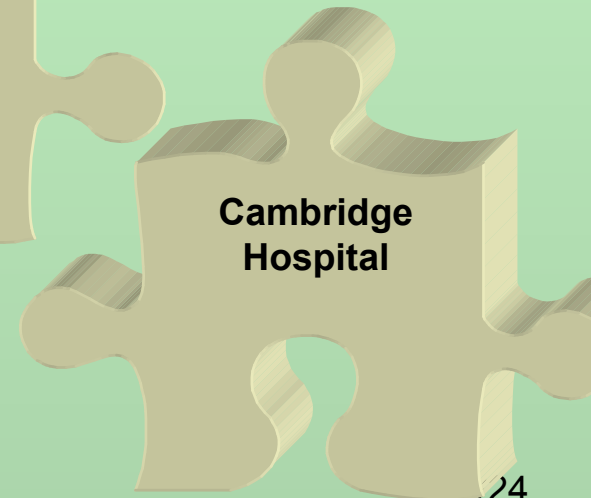
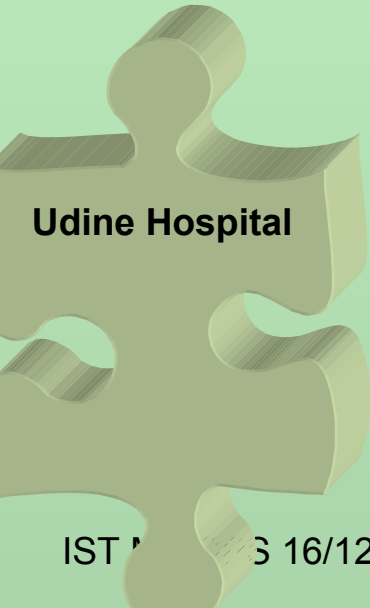
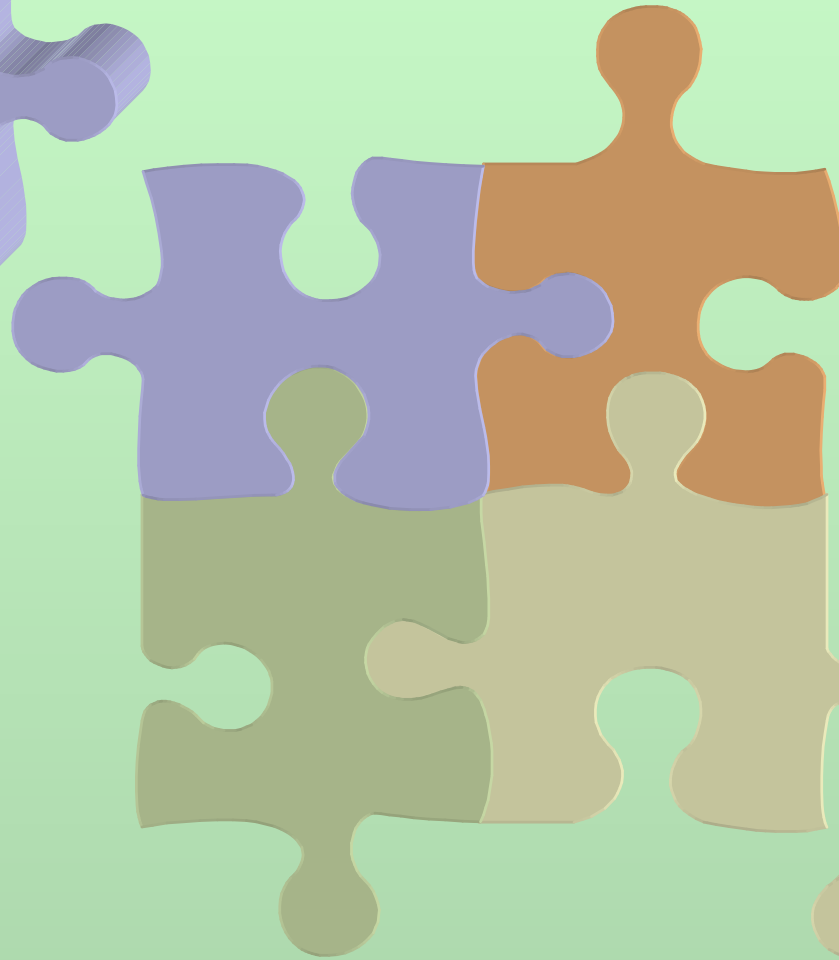
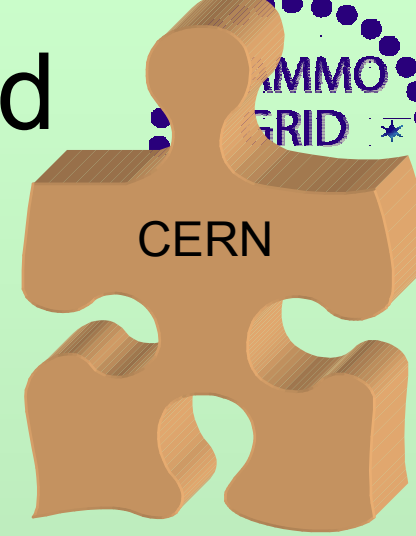
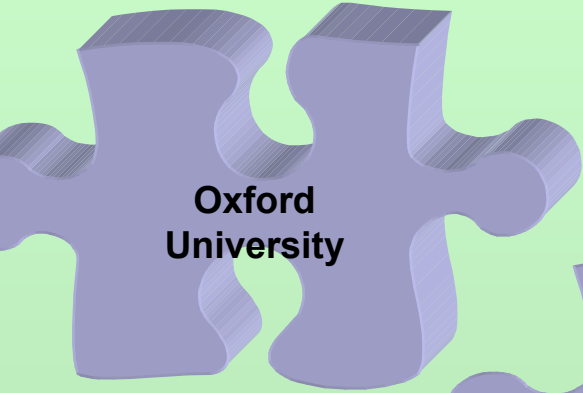
Meta-Data describes service definition

Oxford University:

- I do not have storage
- I understand SOAP only
- I authenticate myself with XXX or YYY format



Meta-Data services and grid



Meta-Data services and grid

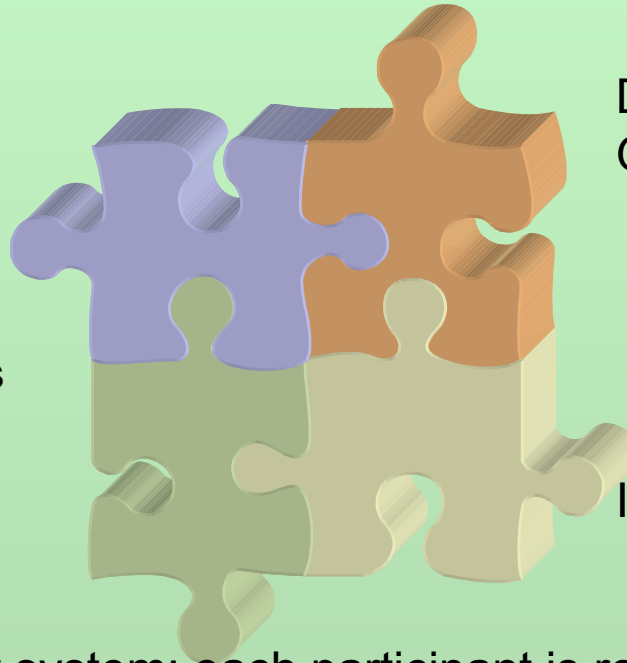


Grid metadata defines service orchestration mechanisms.

Follows OGSA specification

Implements OGSi:

- Service handles
- Lifetime management
- Represents agreements



Grid data services:

Data Virtualization

Grid Data Interfaces

- DataDescription
- DataAccess
- DataFactory
- DataManagement

Implemented by Data Services

- Flexible: plug-and-play system: each participant is responsible for managing its own metadata and can change its service description on-the-fly.
- Extensible: New sites can seamlessly join.
- Domain and Service ontology are defined independently

Outline



- The MammoGrid project
- Meta-Data and domain ontology (flexibility)
- Meta-Data, services and grid (openness)
- **Query negotiator**
- Outlook

Query Negotiator

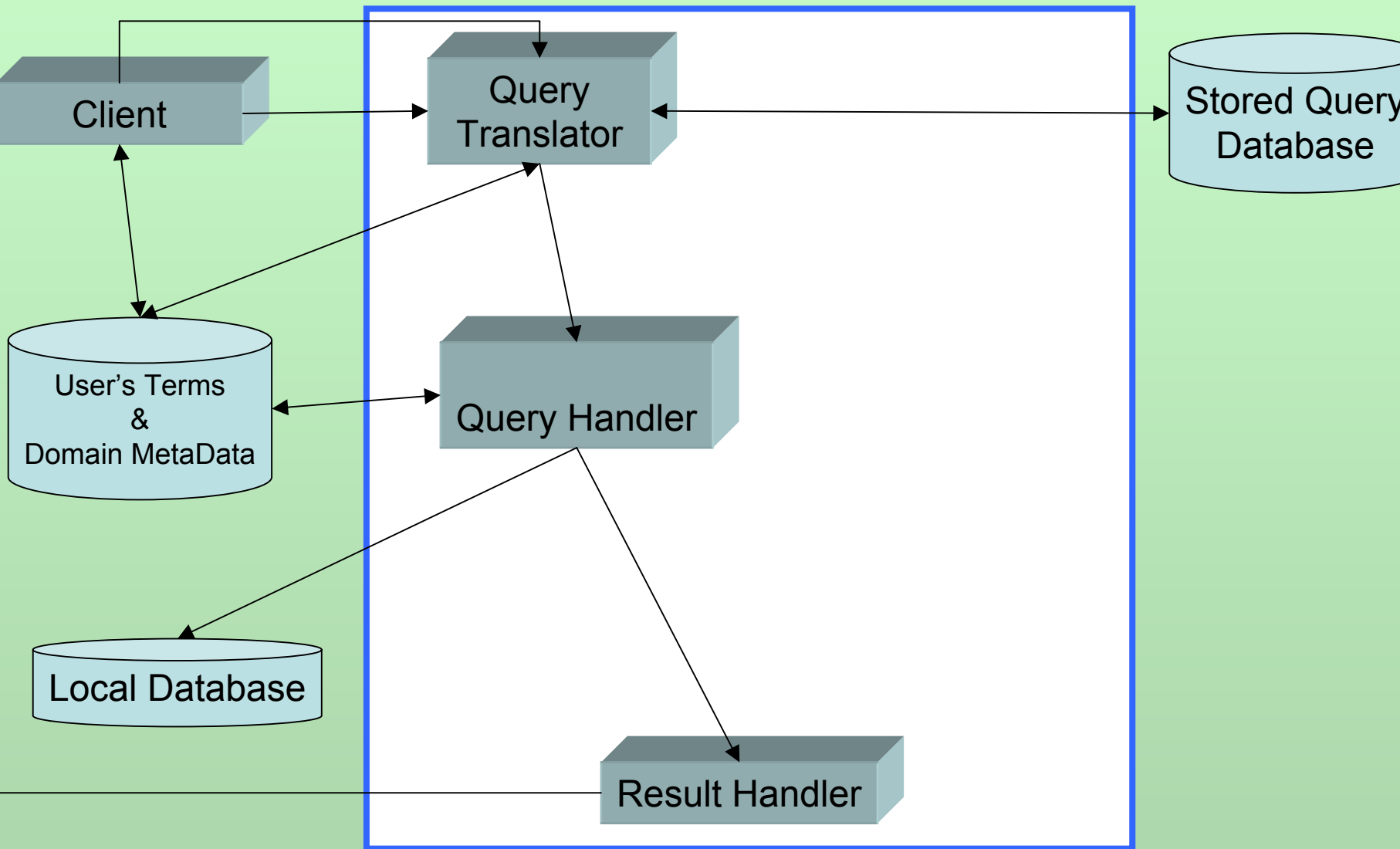
➤ Issues

- Data location, discovery
- Query caching
- Bandwidth, network, etc.
- Do I need to run an algorithm for this?

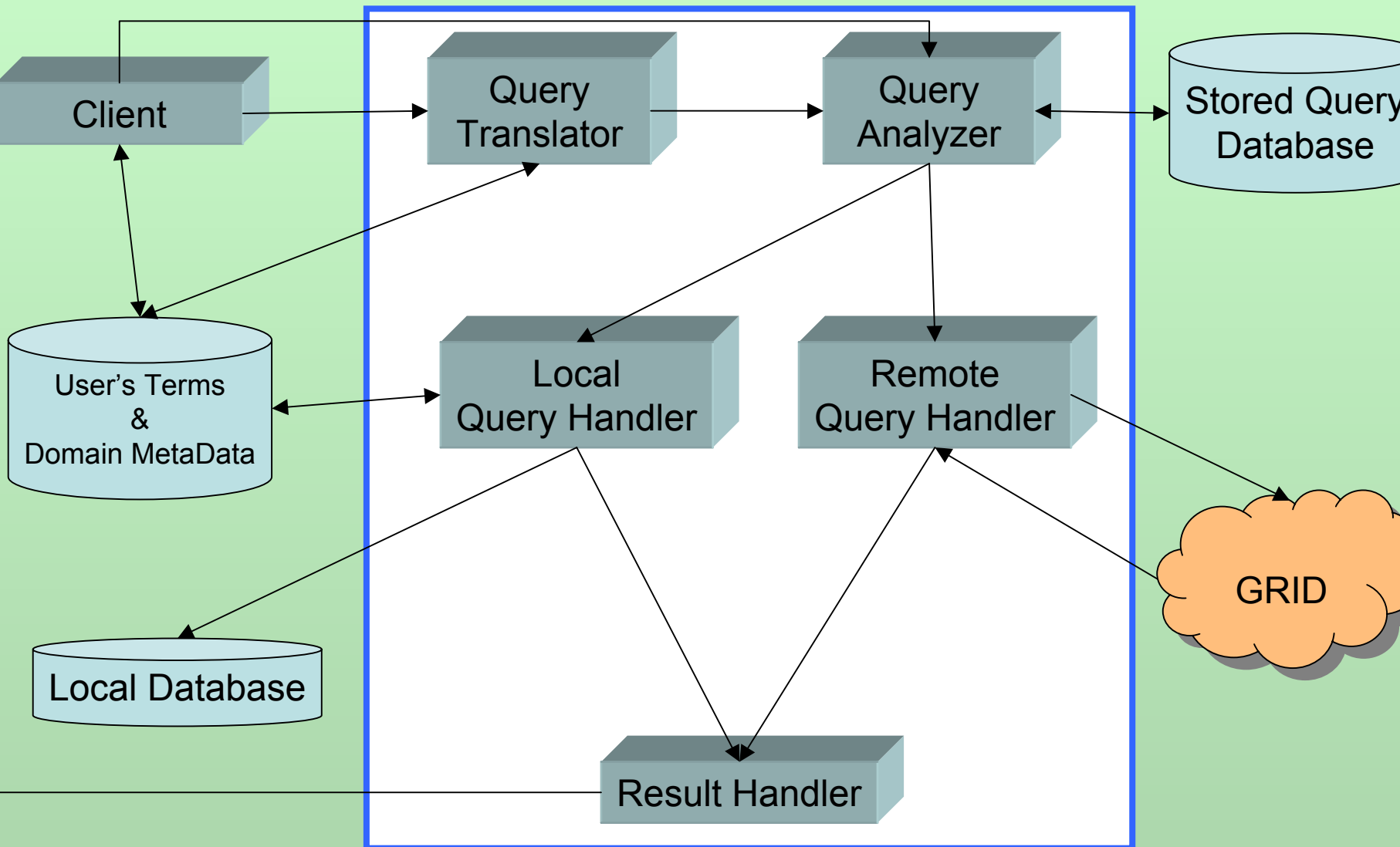


➤ What belongs to the negotiator?

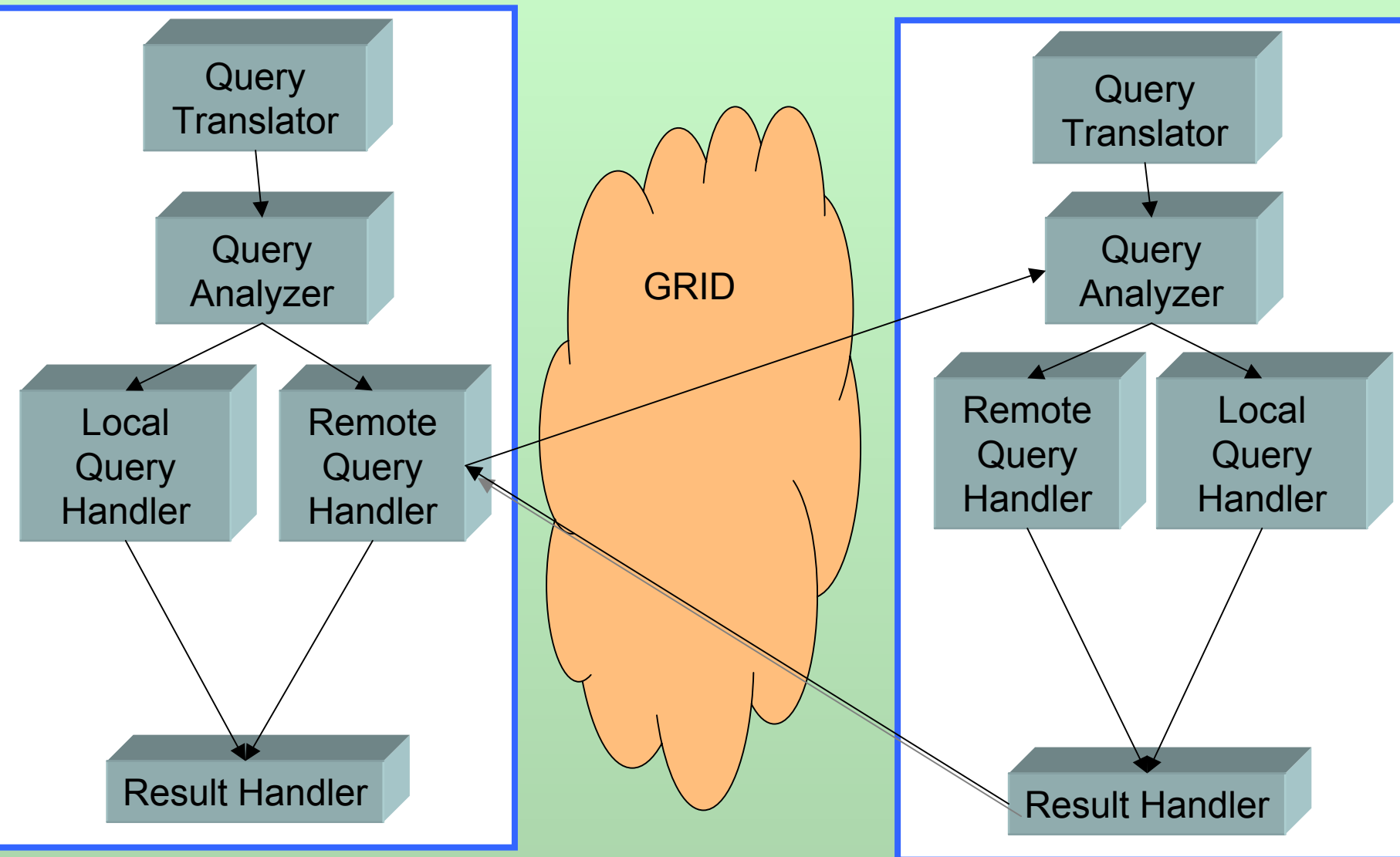
Query Negotiator



Query Negotiator



Query Negotiator



Outline

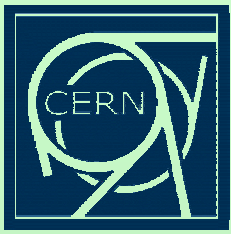


- The MammoGrid project
- Meta-Data and domain ontology (flexibility)
- Meta-Data, services and grid (openness)
- Query negotiator
- **Outlook**

Outlook



- Within the Mammogrid project we focus on three distinct areas where active metadata management is useful
- These provide key features such as flexibility, extensibility and efficiency
- The three meta-data management facilities work in an orchestrated fashion to define the overall characteristics of the Mammogrid system.
- Later it will be interesting to define interfacing two grid-enabled projects (Mammogrid and eDiamond?)
- To that end it might turn out that there is room for a further meta-layer which defines (and unifies) our three (and possibly more) meta-data concepts.



Thank You

Tamas.Hauer@cern.ch