The MammoGrid project

presented by

Salvator Roberto Amendolia

CERN TT Group

DTI Global Watch Mission

CERN mission in distributed IT

CERN, 28 June 2004

What is Mammogrid?

- A project to build a:

    pan-European distributed Database of mammography images using GRID Technologies.

- **Aim:** To provide a demonstrator for use in epidemiological studies, quality control and validation of computer aided detection algorithms.

Mammogrid Consortium

- CERN (Proj. Management - Tech. Coord.)
  - Vitamib (France) - subcontractor Finance/Admin
- Mirada Solutions (UK) – Medical Image Analysis S/W
- University of Oxford (UK) – Medical Vision Laboratory
- University of Pisa (I) – Medical Physics section
- University of Sassari (I) – Maths & Physics Dept
- University West of England (UK) – Computing Research
- University of Cambridge (UK) – Addenbrookes Hospital
- University Hospital of Udine (I) – Inst of Diagnostic Imaging
  - Ospedale Valdese Torino (I) - Breast Screening Unit - subcontr.
  - Zybert Computing Ltd. (UK) - Subcontractor for GRIDserver

Times & $

- The project is funded by the European Commission (Framework Program 5)

- Start: Sep. 2002 - Duration: 36 months

- Cost: 2 M€

EU and Health Grids

- EU is putting a large emphasis on Health Care Grid applications

- MammoGrid is one of the IST FP5 projects specifically targeted at deploying Health Grids

- see also Workpackage 8 in EU-DataGrid and action NA4 in EGEE

Healthcare & IT

- Healthcare is being transformed by effective use of IT in hospitals, community & industry

- Increasingly powerful IT at affordable price

- Governments increasingly committing to electronic delivery of healthcare

- New ways to sense anatomy & physiology, support diagnosis, model genetic basis of disease

Opportunities

- **Teaching tools**
- **Diagnostic support**
  - Customised atlases
  - Tele-diagnosis
  - Data mining
- **Epidemiology**
- **Standardisation**
- **Image normalisation**
- **Ontology**
- **Quality control**

How can the Grid help?

- **High bandwidth of communication**
  - Images are huge
  - Clinical needs can't wait

- **Virtual databases**

- **Security**

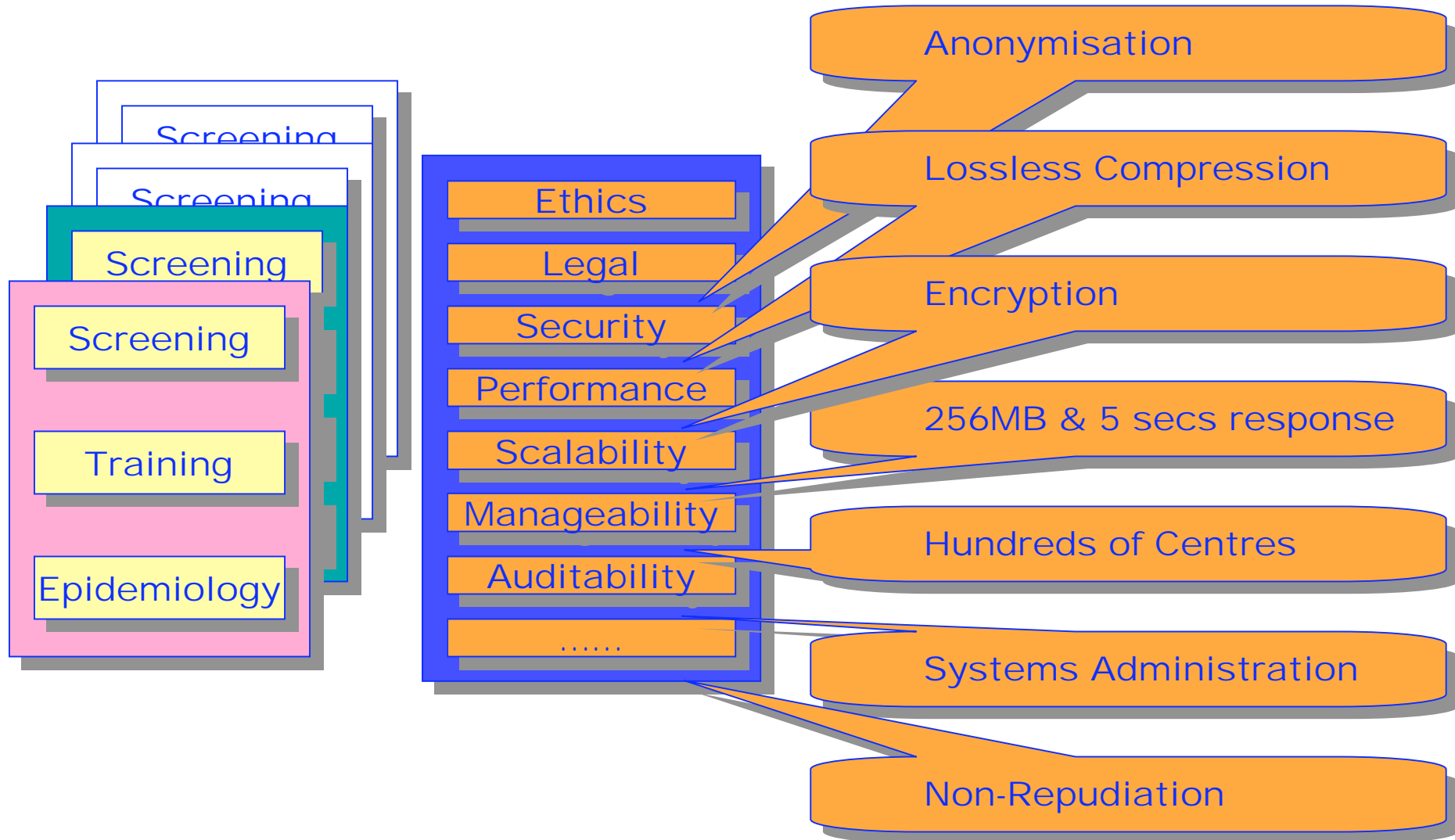- **Distributed computing**

Why a Mammography Database?

- Breast cancer is a huge problem:
  - 10% of women develop breast cancer,
  - 19% of cancer deaths are due to breast cancer,
  - 24% of all cancer cases are breast cancers,
  - there are 348,000 cases in EU & USA, 50,000 die every year

- Early diagnosis through mammography screening improves prognosis

- Quality control in acquisition, diagnosis and efficient data management is vital.

- A way to achieve reliability of screening and early diagnosis is through repositories of mammography data for research and training that contain sufficiently large statistical samples

The MammoGrid Challenge

- Building this repository is not trivial because:
  - Large numbers of exemplars required (==> enough abnormal cases to make significant statistical inferences).

  - To understand the variables (anatomy, genetics, diets, habits…) cases must be obtained from many geographically remote locations.

  - Data is large: 2 breasts x 2 views x 4K x 4K pixels x 2 bytes = 128Mbyte per patient per visit, 3M women per year UK, ~ 400 Terabytes in UK alone, (26M women in US), >100M mammograms on file.

  - Acquisition is highly variable, same image may look different depending on machine and parameters. How do you compare?

  - Patient privacy and data security is key.

  - Data need to be sorted efficiently according to demographics, acquisition information and image quantification parameters.

GRID

Challenges

Screening
Screening
Screening
Screening

Screening

Training

Epidemiology

Ethics

Legal

Security

Performance

Scalability

Manageability

Auditability

......

Anonymisation

Lossless Compression

Encryption

256MB & 5 secs response

Hundreds of Centres

Systems Administration

Non-Repudiation

MammoGrid Objectives

Evaluation of Grids technologies and of the requirements for Grid in a pan-European mammography database.

Implementation & deployment of the database and of a standardization system (enabling comparison of mammograms independently of scanner settings)

Development of software tools to automatically extract image information (to perform quality controls on the acquisition process and clinical studies in order to increase the performance of breast cancer screening programs)

Use & Exploitation of the system to provide a proof-of-principle for subsequent Europe-wide application

Project concentrates on applying emerging GRID technology rather than on developing it

High Energy Physics vs. MammoGrid

- MammoGrid heavily relies on technologies developed primarily in the field of High Energy Physics.
  - Similarities
    - Large number of big files
    - Files can be sensibly organized in directory tree
    - Need to replicate and move file copies between sites
    - Need to execute commands on the node which hosts data locally
  - Difficulties
    - Complexity of co-working in medical environment
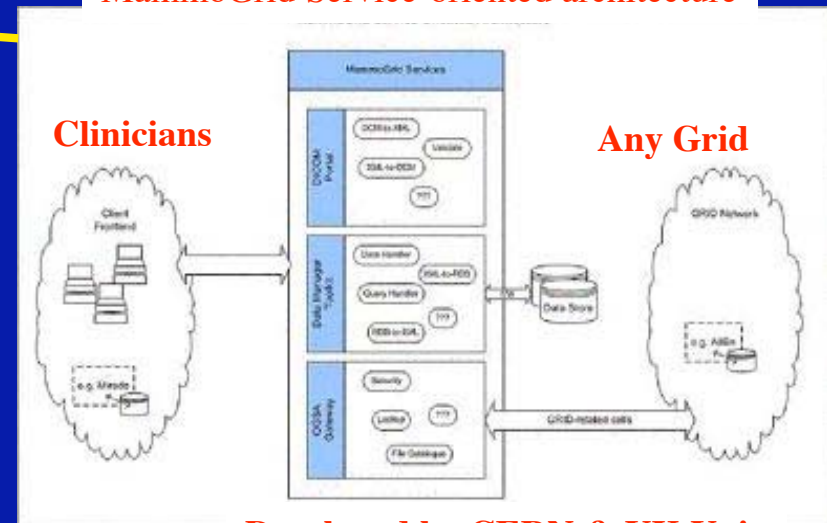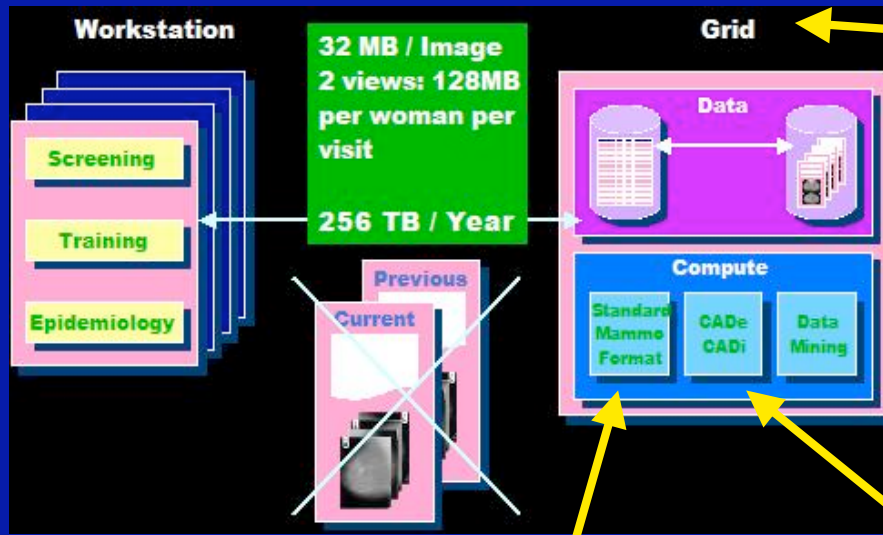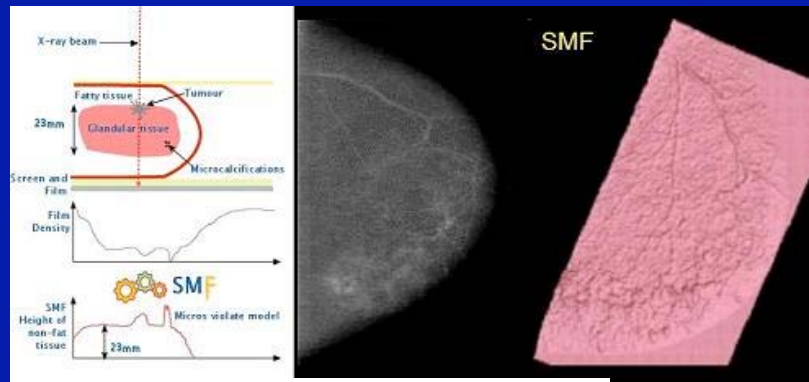   - Lack of trained IT personnel
  - Confidentiality !!!

MG: methods

- **The MG Information Infrastructure (MII) relies on two existing technologies:**
    - lightweight Grid-compliant software package AliEn (now moving to EGEE middleware)
    - reflective middleware, comprising high-level meta-data structures, based on CRISTAL
- **The two pilot applications in MG are:**
    - Pilot study 1: breast density measurement
    - Pilot study 2: CADe for Quality Control
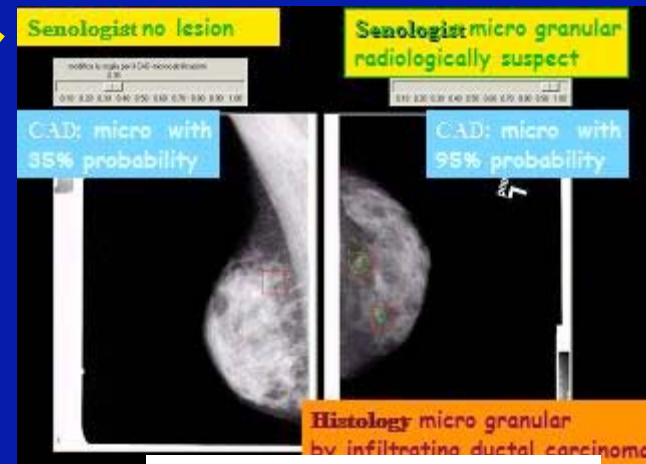
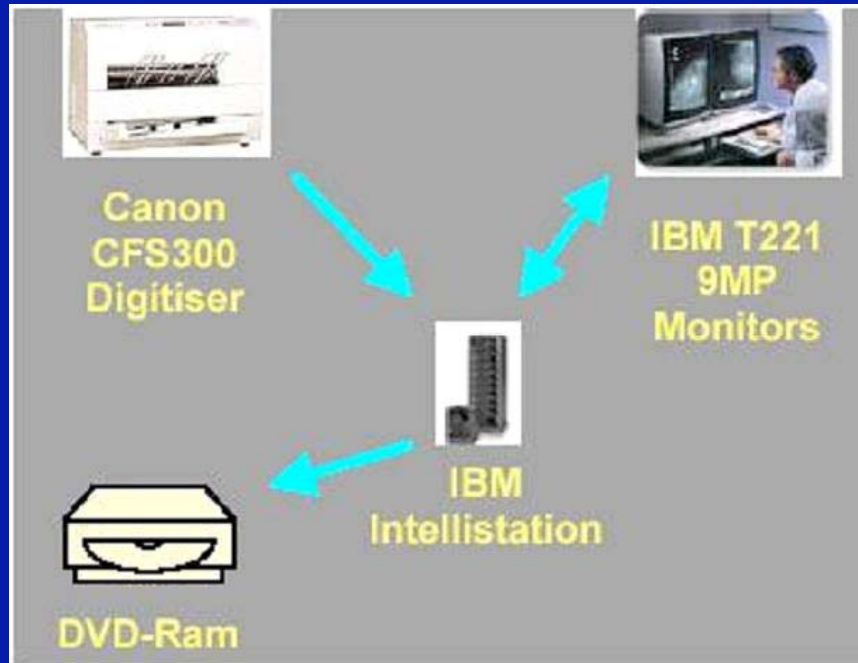# MammoGrid at a glance

MammoGrid Service-oriented architecture

Workstation — Grid

Screening
Training
Epidemiology

32 MB / Image
2 views: 128MB per woman per visit

256 TB / Year

Data

Previous
Current

Compute

Standard Mammo Format | CADe CADi | Data Mining

Clinicians — Any Grid

Developed by CERN & UK Univ

Developed by UK Univ & SME

Senologist no lesion
Senologist micro granular radiologically suspect
CAD: micro with 35% probability
CAD: micro with 95% probability
Histology micro granular by infiltrating ductal carcinoma

Developed by Italian Univ

Acquisition Station

- DAQ system successfully deployed by month 5
- Acquisition ongoing since month 6
- Basic components:
  - workstation, 2 high level screens
  - digitiser
- Statistics to date: 2000+ patients in each site

User requirements specifications

- Task: identify and specify functional and non-functional requirements
- Requirements engineering process adopted:
  - elicit the requirements of the MammoGrid,
  - analyze these requirements,
  - develop requirements definitions and specifications,
  - and validate these requirements

Software Engineers

(CERN & UWE)

Rational

Unified

Process Language

Radiologists &

Clinicians

(Udine & Cambridge)

Medical Imagers

(Mirada, Oxford,

Pisa & Sassari)
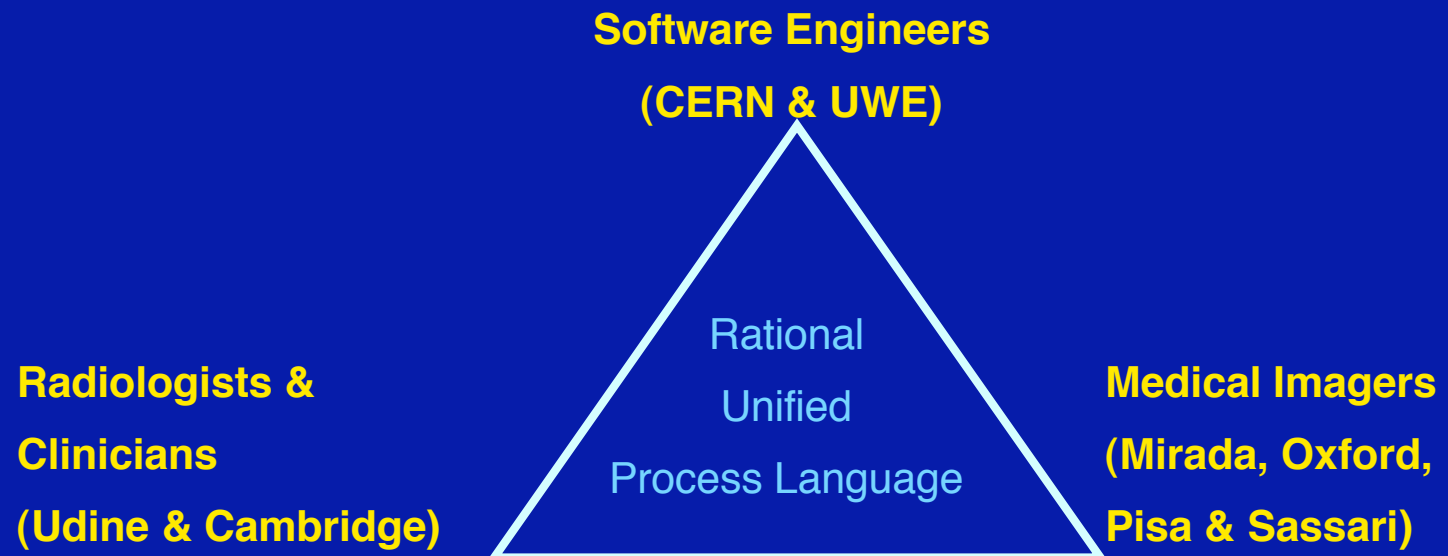
Example of CADe

modifica la soglia per il CAD microcalcificazioni
0.35
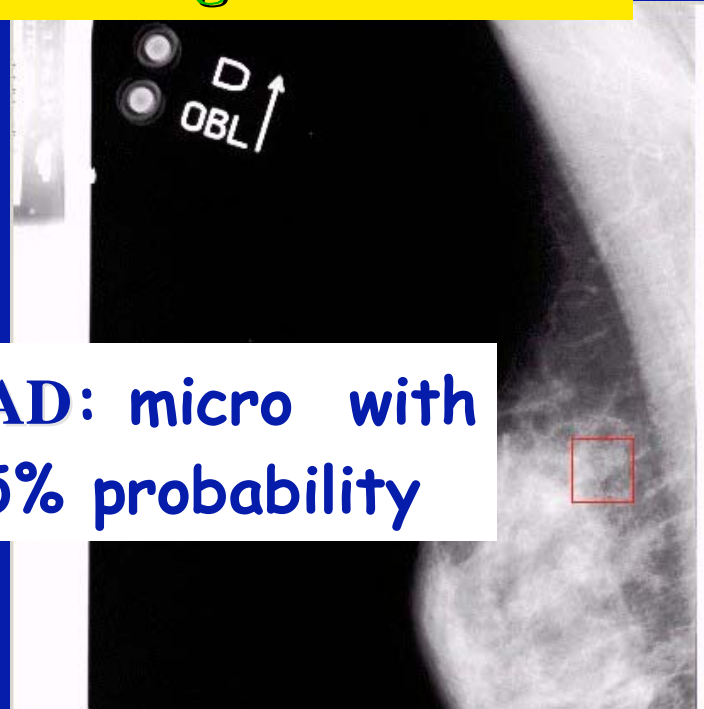
modifica la soglia per il CAD microcalcificazioni
0.95
0.80 0.90 1.00

Senologist: no lesion

Senologist: micro granular radiologically suspect

CAD: micro with 35% probability
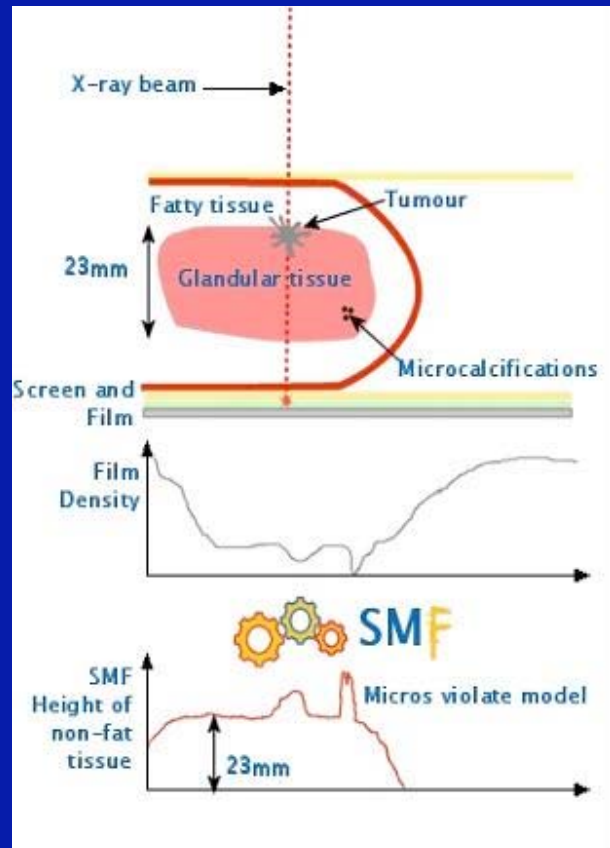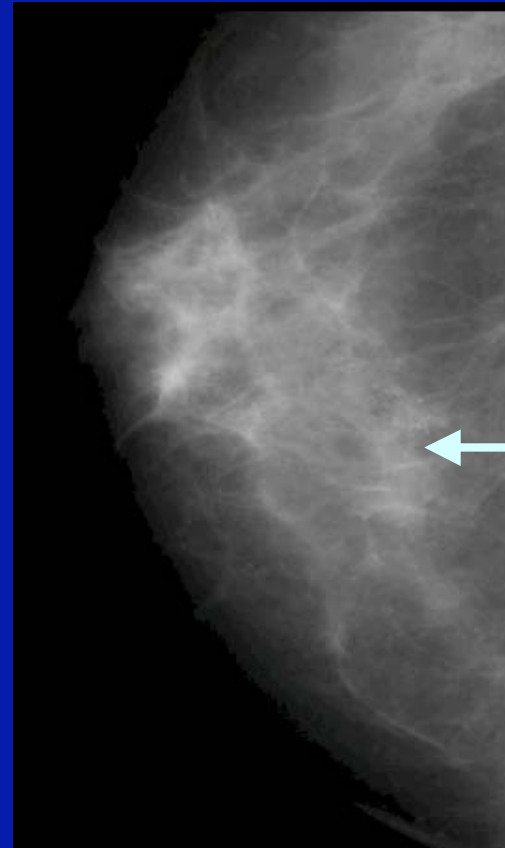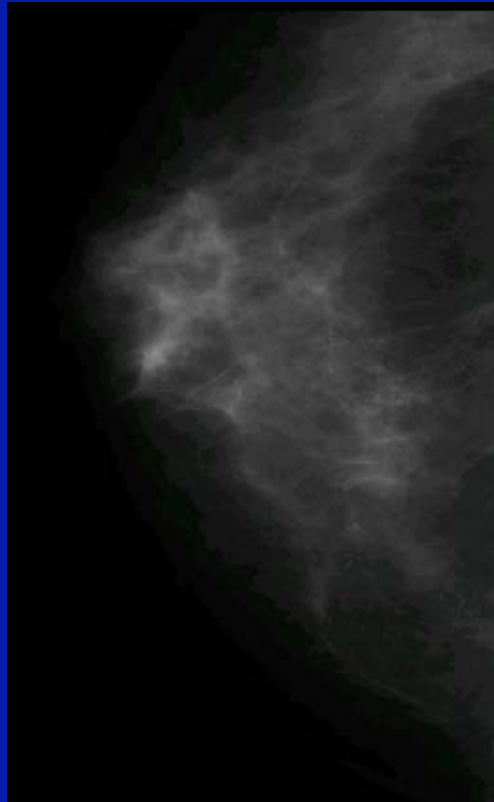
CAD: micro with 95% probability
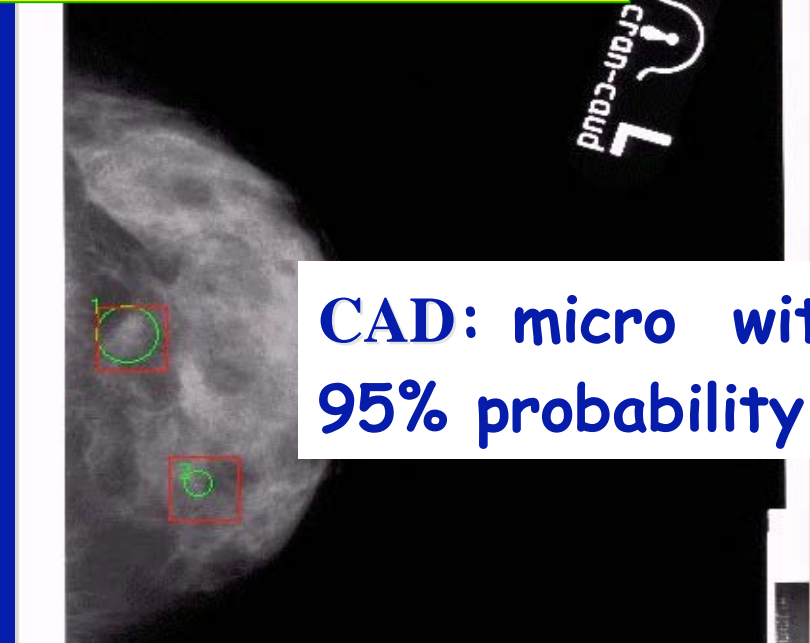
Histology: micro granular by infiltrating ductal carcinoma

SMF
Standard Mammographic Format

- Allows standardisation of mammograms acquisition across different sites

- Proprietary software

- Adapted to MG Grid environment
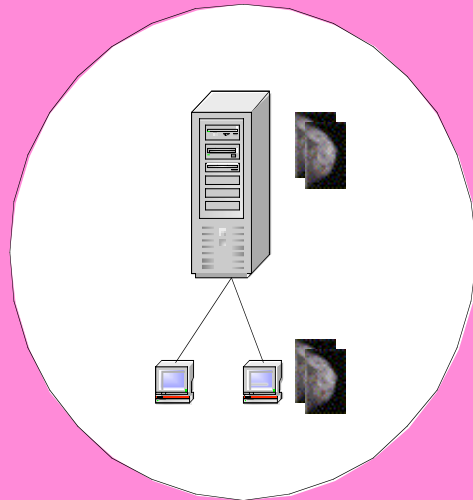
SMF operation

Challenge: imaging parameters
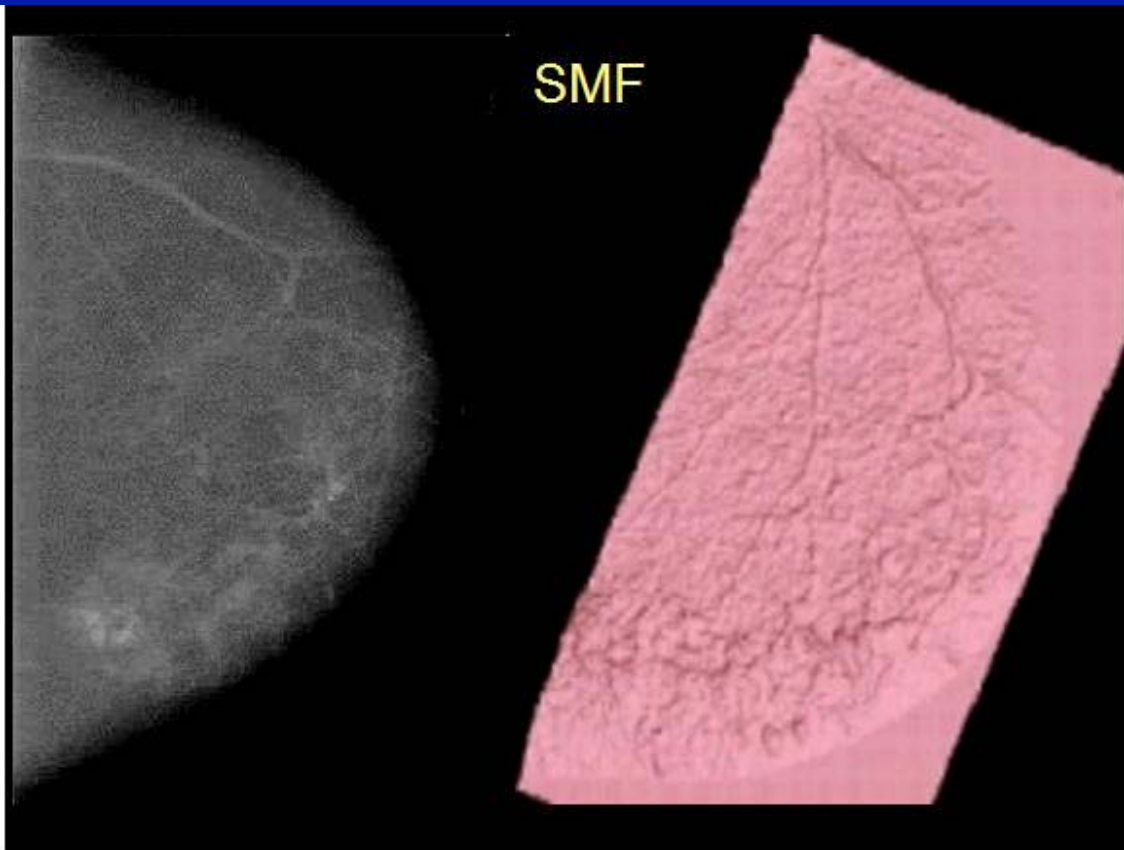
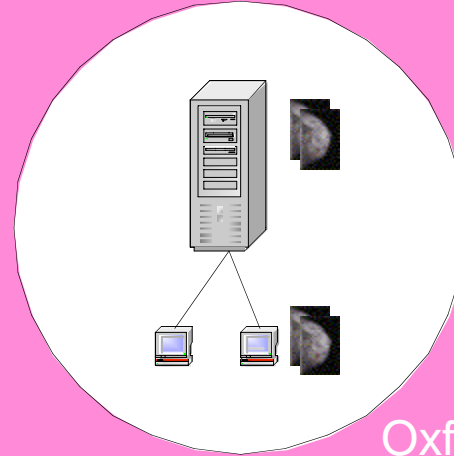This image difference corresponds to a change of just 40 mAs in exposure. Often this parameter is poorly controlled

Microcalcification cluster – barely visible on the poor contrast image

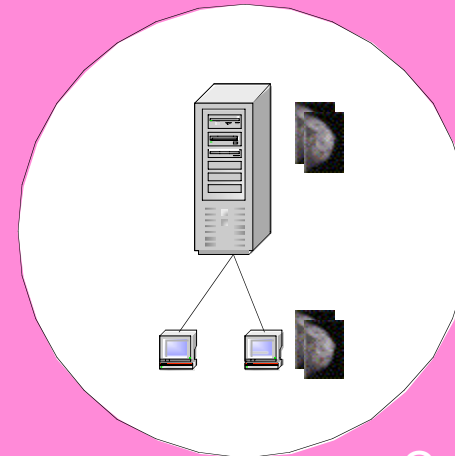The same breast; the exposure time on the right is shorter than that on the left.
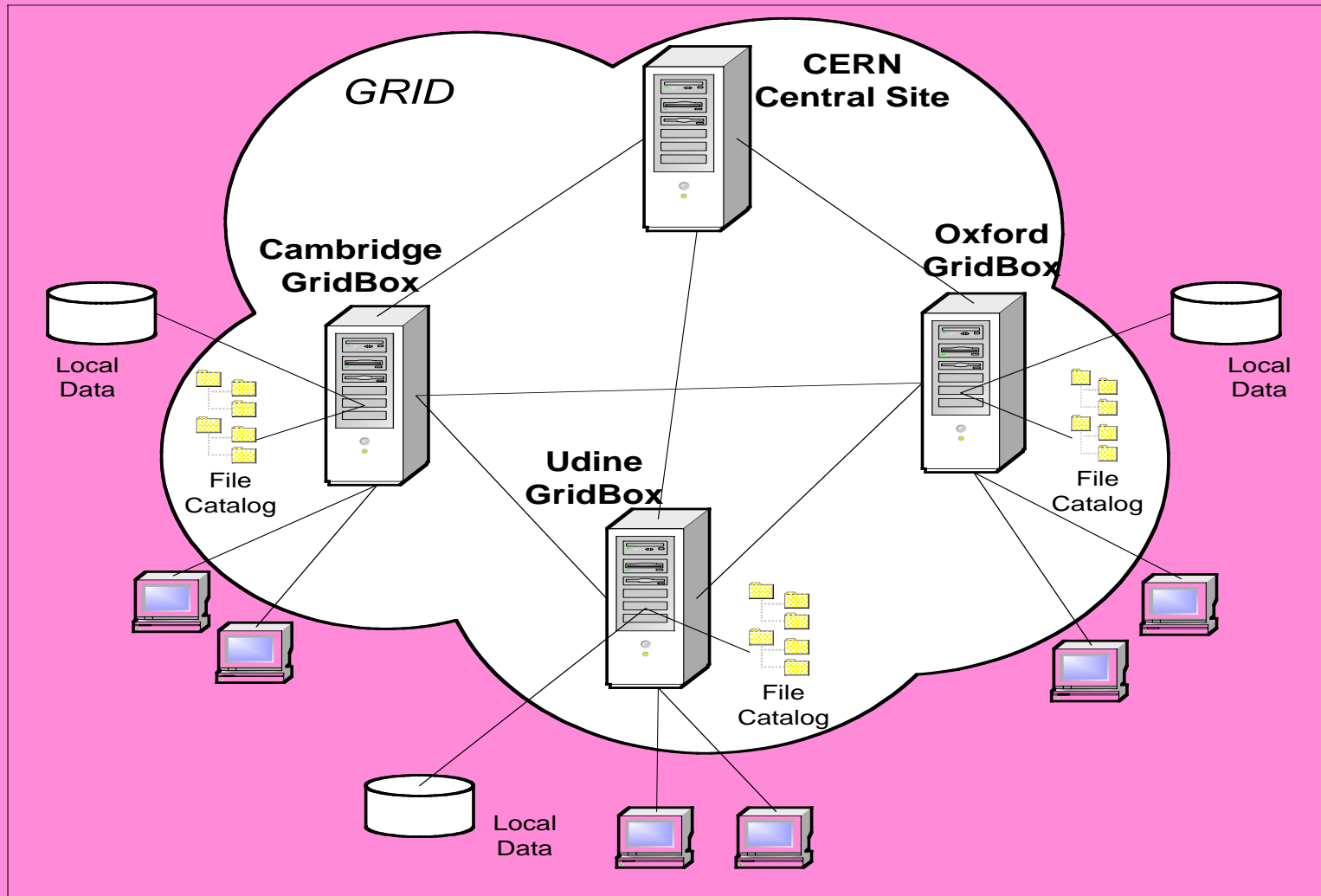
before MammoGrid...

Oxford

Udine

Cambridge

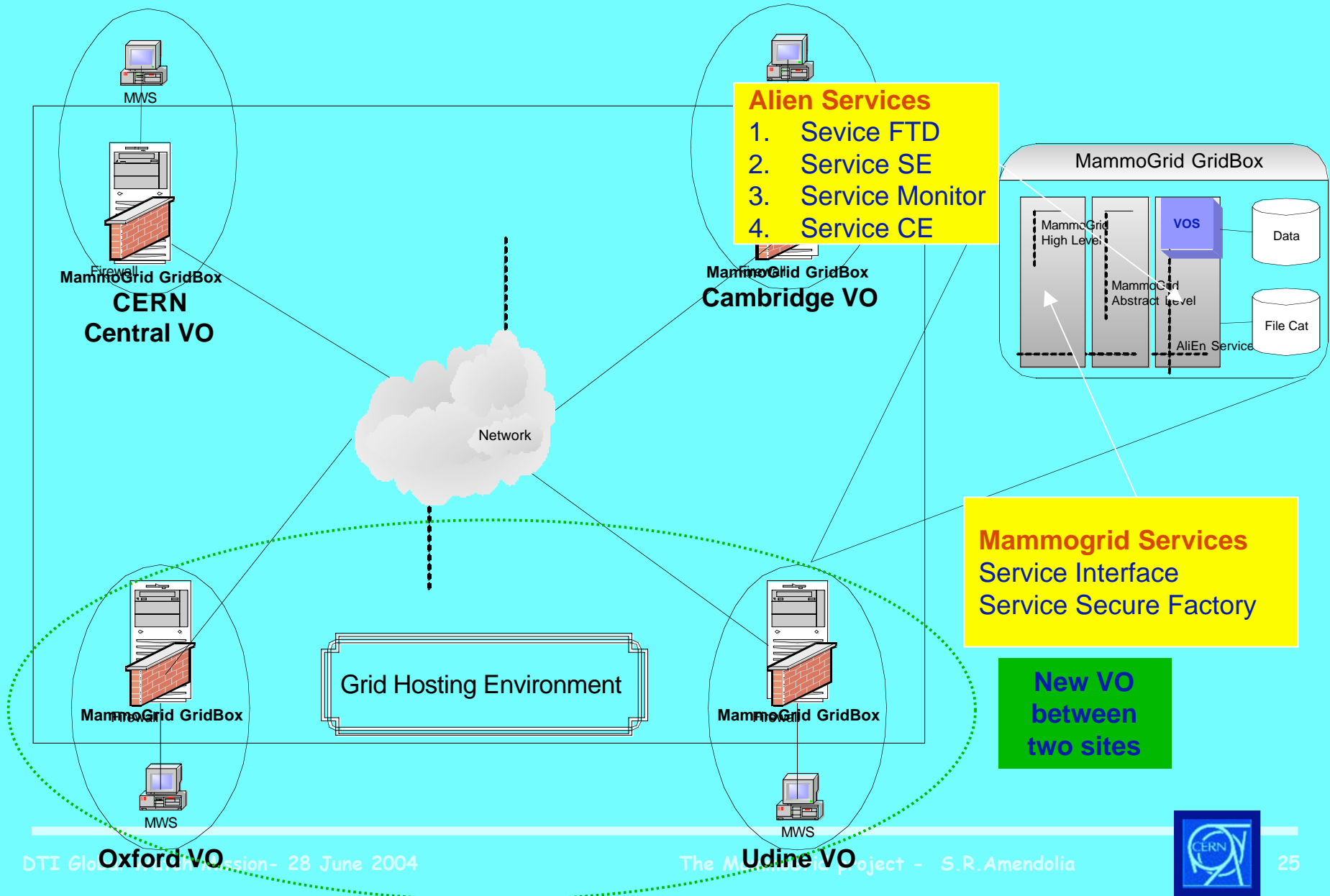**How to enable cooperation?**

# MammoGrid Infrastructure

Gridbox

- Built by UK sub-contracting firm

- All units delivered

- Successfully tested

- Hosts Grid middleware and Medical Services

Security in terms of VOs

Mammogrid Service Layers

- Medical imaging (MI) service layer:
  - Generic set of services for handling image-related data (parsing, transforming, storing etc);
  - Specific imaging services (image analysis, query services etc.)

- Grid-aware service layer:
  - Mediators between MI services and underlying Grid middleware;
  - Examples: storing/retrieving files on the Grid, Distributed Query Service, Remote Handler, Result Handler.
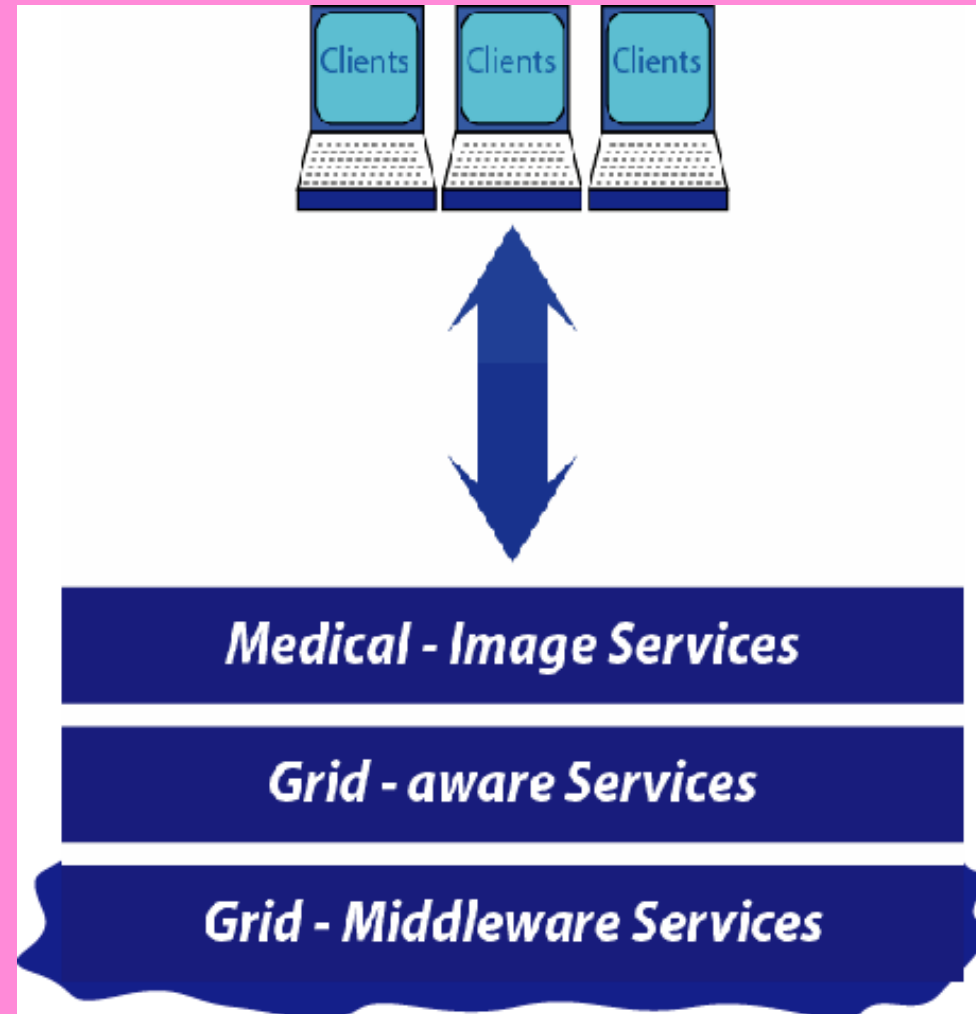
Prototype 2 Architecture

MammoGrid Service-Oriented Architecture

any Grid
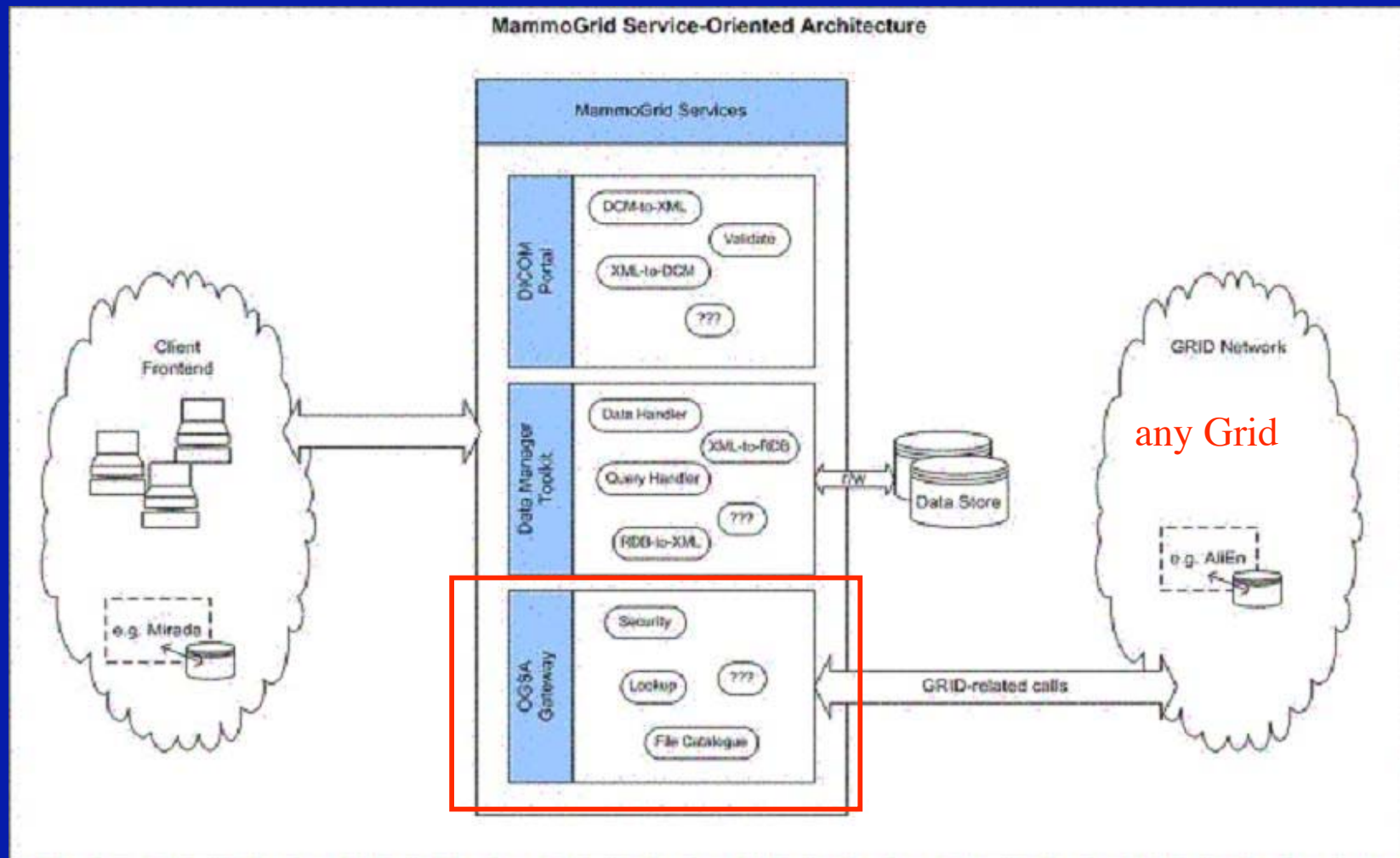
Projects

- **MammoGrid**
  - EU FP5
  - CERN + Univ. + Company in UK and Italy
- **eDiamond**
  - UK country wide
- **GP-CALMA**
  - Italy country wide

Synergy between projects

- Teaching & CPD   eDiamond
- Tele-diagnosis   eDiamond   GP-CALMA
- Quality control   MammoGrid   eDiamond
- Epidemiology   All
- Algorithm development:

  data mining   Mammogrid   eDiamond
- CADe development   MammoGrid   GP-CALMA

Generalising from Mammogrid & al.

- The architectural infrastructure is not limited to mammograms or even cancer

- The only specialisation to mammography is image normalisation SMF, and the knowledge embodied in programs about masses, breast dense tissue and microcalcifications

- Medical image analysis is moving beyond diagnosis to monitoring disease progression and therapy, e.g.

  - molecular medicine

  - software-enhanced pharmaceuticals (eg Parkinson's)

Transferring technology

- MammoGrid @ CERN is funded/supported by the Technology Transfer Group

- Extending Grids to societal applications at large helps in raising sensitivity (and funding...) from governments

- Fall-out will benefit basic research

- Many opportunities for IT related business

Exploitation

- ## Mirada
  - Establish SMF as a standard for breast density measurement.
  - Establish SMF as a standard for mammogram data exchange.
  - Prototype SMF based review workstation for CADiagnosis.

- ## CERN/UWE
  - Study the spinning-out of GRID/database technologies to address needs in healthcare (e.g. licensing agreements for Cristal software)

- ## Oxford
  - Develop patentable technologies for medical image analysis products.
  - Transfer of technology agreement with Mirada.

- ## Pisa/Sassari
  - Develop patentable technologies for medical image analysis products.

Final word

- Clinicians and other health professionals want to **use** the Grid & they profoundly wish to remain ignorant about how it works

- But come it will, and use it, they will