Dipartimento di Informatica, Sistemistica e Telematica
Università degli Studi di Genova

biolab

NA4 Generic Applications Meeting

GRID node for biomedical applications: distributed image analysis for early diagnosis of Alzheimer Disease

Livia Torterolo

9 -11 January 2006 – Catania, Italy

Contents

- Introduction to SPM analysis for diagnosis in AD
- Development of SPM service through a web portal
- GRID implementation of SPM service
- Current Status and Plans

Partners

GRID node has been installed at BioLab, University of Genoa.

The first part of work has been done in collaboration with:

- *San Raffaele of Milano*
- *University of Milano – Bicocca*

GRID implementation has carried out by Bio-Lab team.

Contents

- Introduction to SPM analysis for diagnosis in AD
- Development of SPM service through a web portal
- GRID implementation of SPM service
- Current Status and Plans

Introduction to Alzheimer Disease

- AD is the most common form of dementia, accounting for more than half of all dementias in elderly people
- Clinically, AD is characterized by a progressive loss of cognitive abilities
- Memory loss is typically the earliest sign of AD

Analysis of PET images

Conventional Reporting: qualitative analysis of flux images

F18-FDG

Qualitative analysis of images doesn't reveal anomalies in the distribution of the tracer in cortical, subcortical and cerebellar regions.

The metabolic study is in the normal range.

Qualitative analysis of images reveals a reduction in the accumulation of the tracer in correspondence of the inferior parietal lobe, bilaterally.

Hypometabolism in these regions.

Tc99-ECD

Qualitative analysis of images doesn't reveal anomalies in the distribution of the tracer in cortical, subcortical and cerebellar regions.

The metabolic study is in the normal range.

Qualitative analysis of images reveals a reduction in the accumulation of the tracer in correspondence of the inferior parietal lobe, bilaterally.

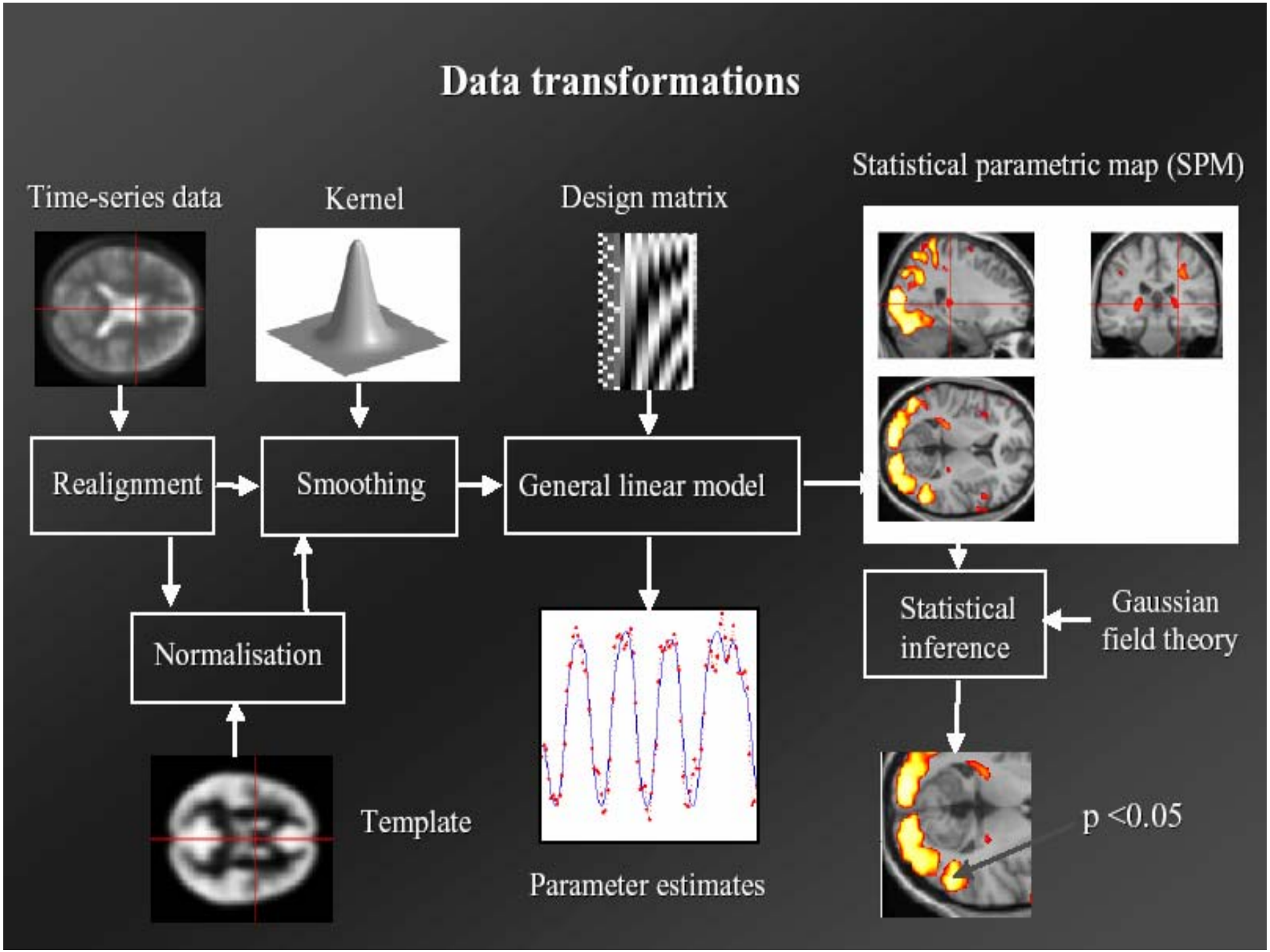
Parietal hypoperfusion.

SPM (Statistical Parametric Mapping)

Functionality:

- *Image display*
- *Image segmentation*
- *Image realignment and co-registration*
- *Nonlinear spatial normalization*
- *Smoothing*
- *Statistical analysis – parameter estimation {GLM}*
- *Statistical analysis – interface {TGF}*
- *Results display (Graphical, tabular and image format)*

by Institute of Neurology at University College London


Data transformations
Time-series data
Kernel
Design matrix
Statistical parametric map (SPM)
Realignment
Smoothing
General linear model
Normalisation
Statistical inference
Gaussian field theory
Template
Parameter estimates
p <0.05

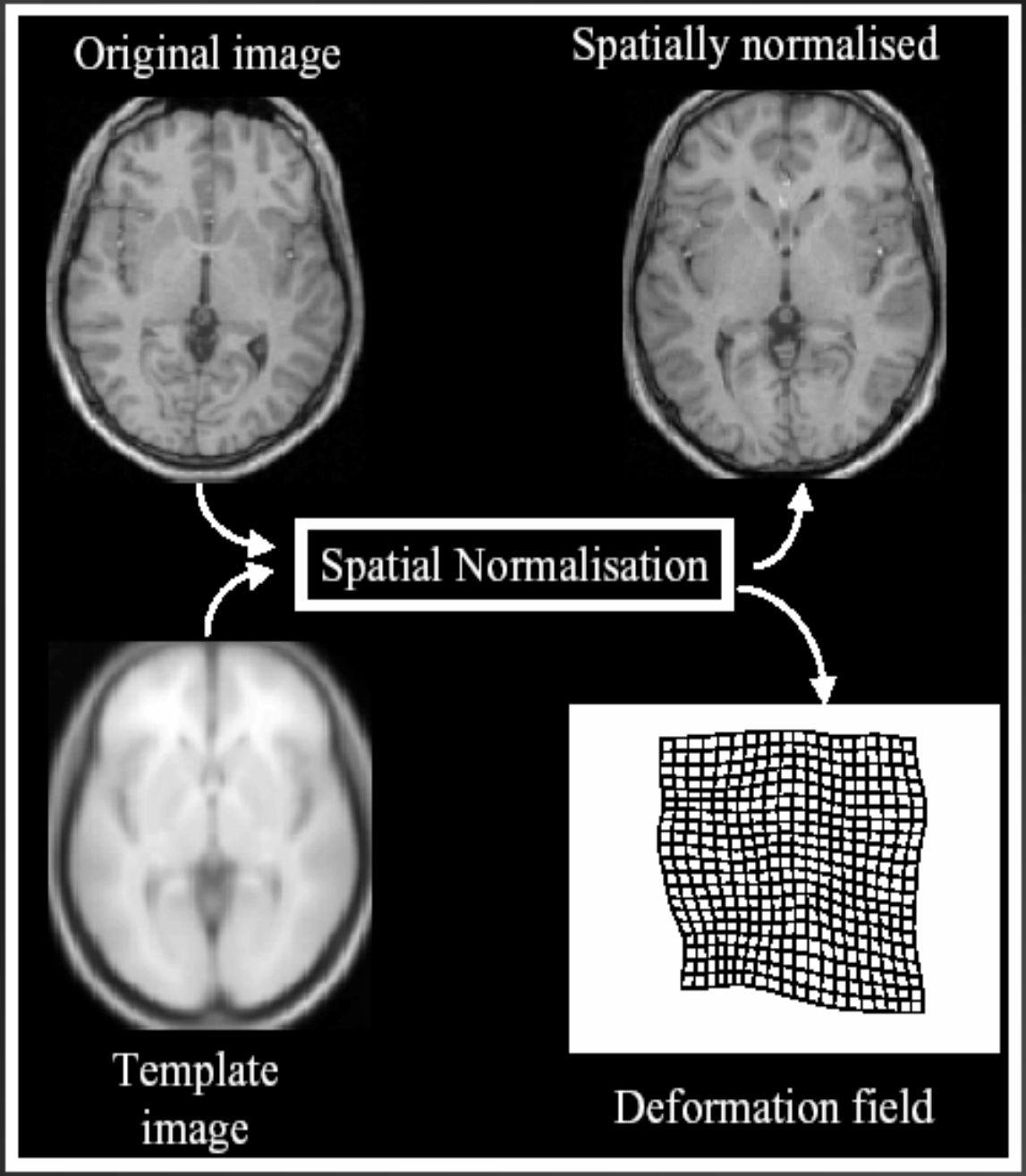
Spatial Normalisation

Determine the spatial transformation that minimises the sum of squared difference between an image and a linear combination of one or more templates.

Begins with an affine registration to match the size and position of the image.

Followed by a global non-linear warping to match the overall brain shape.

Uses a Bayesian framework to simultaneously maximise the smoothness of the warps.

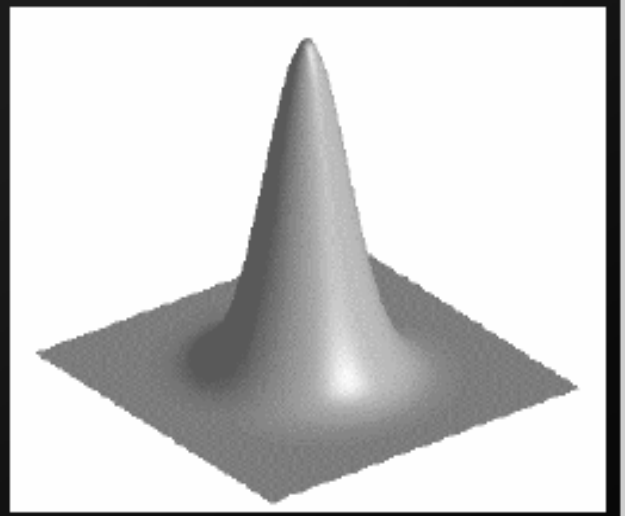


Smoothing

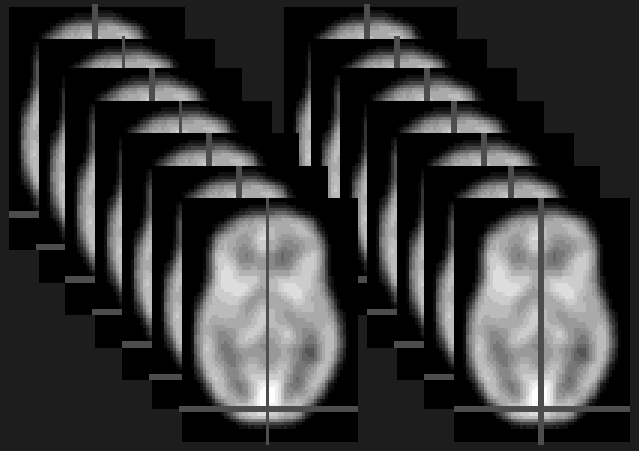
- Why Smooth?
  - Potentially increase signal to noise.
  - Inter-subject averaging.
  - Increase validity of SPM.
- In SPM, smoothing is a convolution with a Gaussian kernel.
- Kernel defined in terms of FWHM (full width at half maximum).

Gaussian convolution is separable

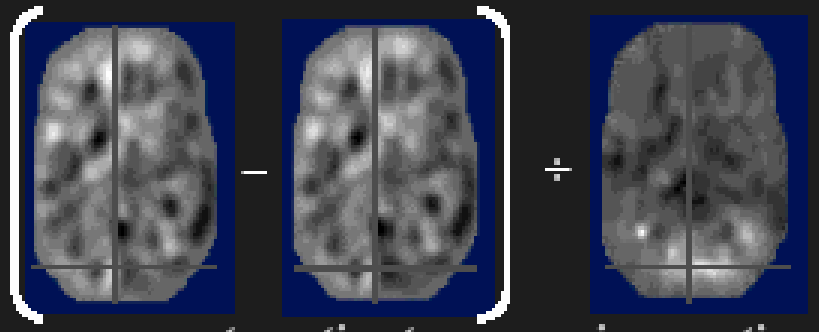
Gaussian smoothing kernel



Statistical Parametric Mapping…

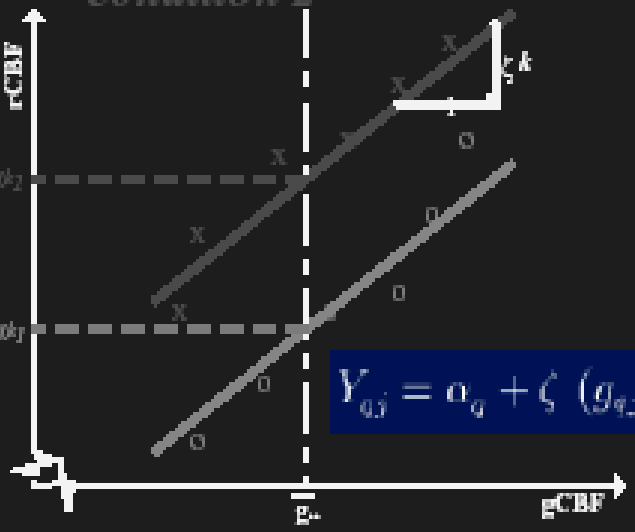


condition 1 *condition 2*



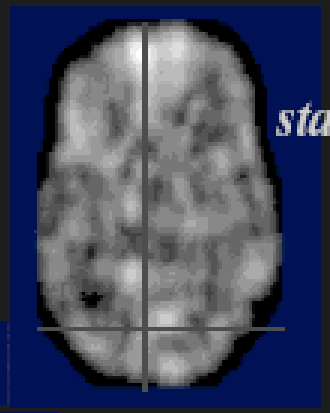
parameter estimate *variance estimate*

voxel by voxel modelling



$$Y_{qj} = a_q + \zeta\,(g_{qj} - \overline{g}_{\bullet\bullet}) + \epsilon_{qj}$$

$$\epsilon_{qj} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$



statistic image or SPM

Contents

- Introduction to SPM analysis for diagnosis in AD
- Development of SPM service through a web portal
- GRID implementation of SPM service
- Current Status and Plans

Why a portal?

SPM requirements:

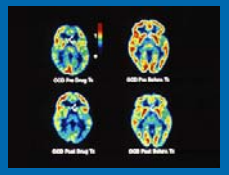
- the knowledge of functions used in the image analysis in order to provide the correct values of parameters and in order to understand the results
- a large set of images of normal patients to be used for comparison during the statistical parametric mapping. The accuracy of ipoperfusion maps is strictly related to the number of normal studies

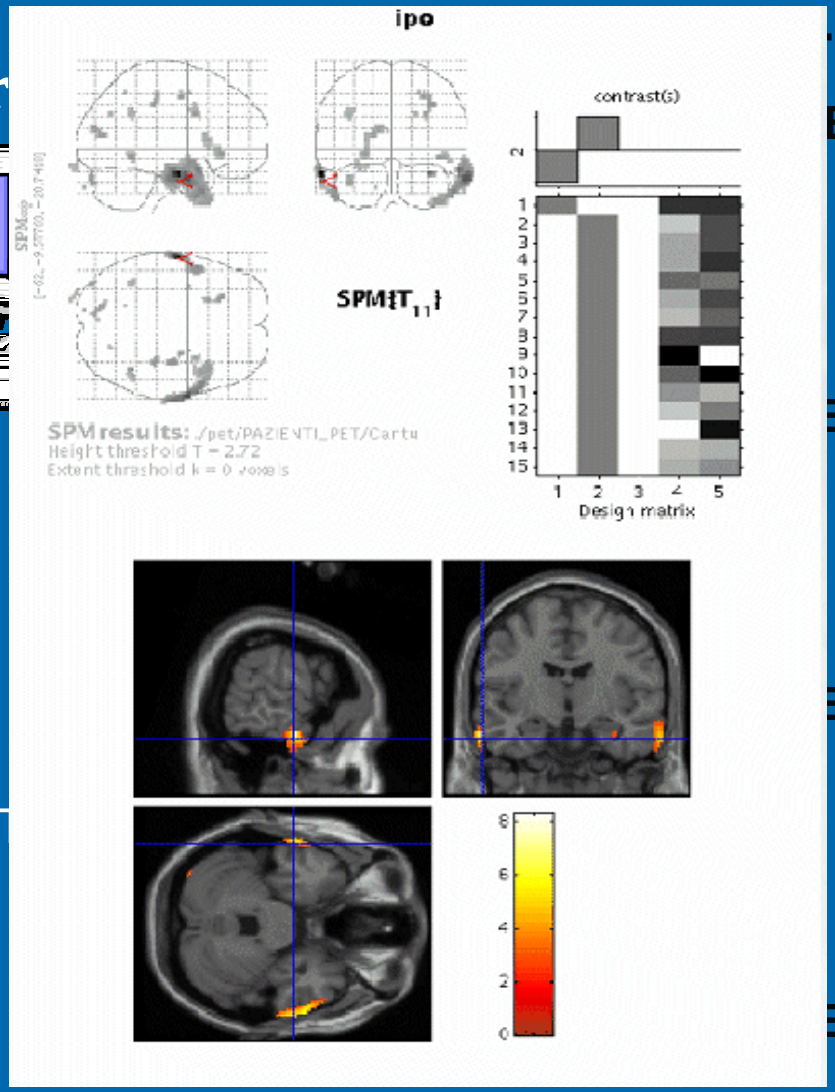
www.neuroinf.it

Web-based application

www.neur

IMAGE of PATHOLOGICAL SUBJECT
(PET or SPECT IMAGE)





of CONTROLS 1
ET, SPECT IMAGES)

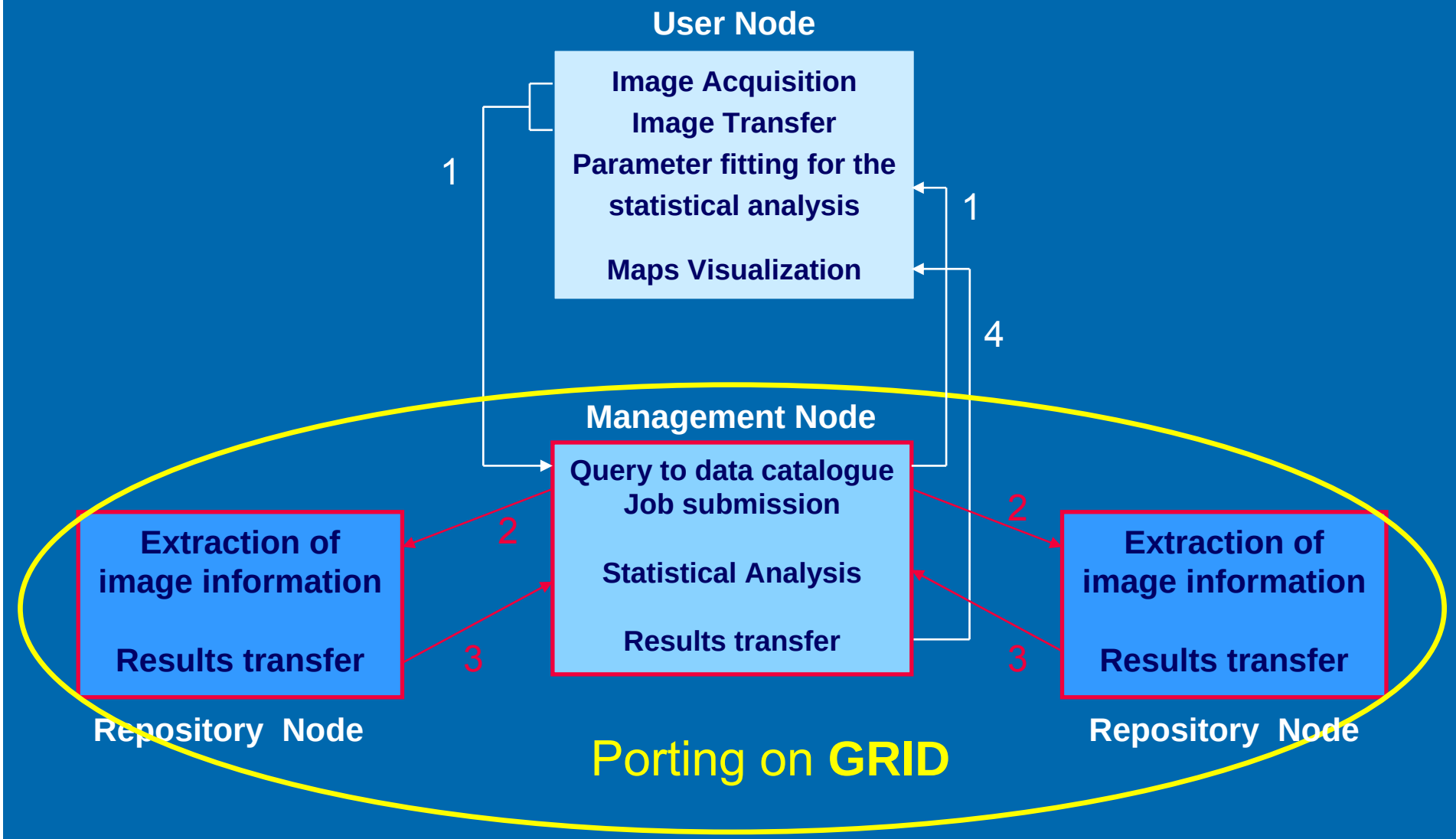
of CONTROLS 2
ET, SPECT IMAGES)

of CONTROLS 3
ET, SPECT IMAGES)

• • •

of CONTROLS n
ET, SPECT IMAGES)

Design of the application's structure



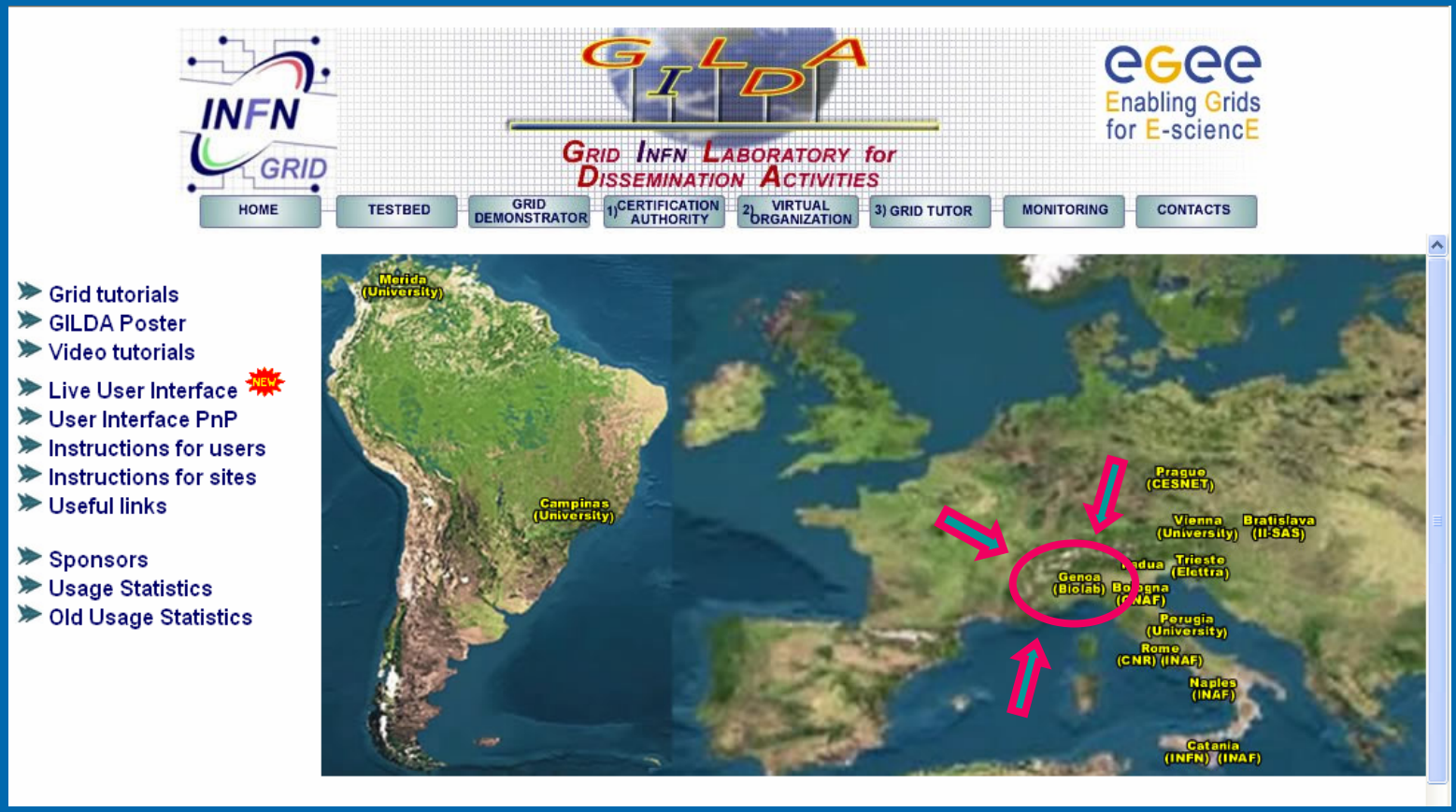
Contents

- Introduction to SPM analysis for diagnosis in AD
- Development of SPM service through a web portal
- GRID implementation of SPM service
- Current Status and Plans

SPM: Why GRID?

- Doctors would have an easy access to normal PET/SPECT databases without moving images from hospitals
- Waiting time for statistical parametric mapping depends on the number of normal images stored in every database ➡ It is very important to split this application into independent jobs and to run them at the same time to increase software performances

GILDA testbed



GILDA testbed: https://gilda.ct.infn.it/

LCG Installation

- Installation and Set up of the LCG Grid Environment v. 2.4.0.

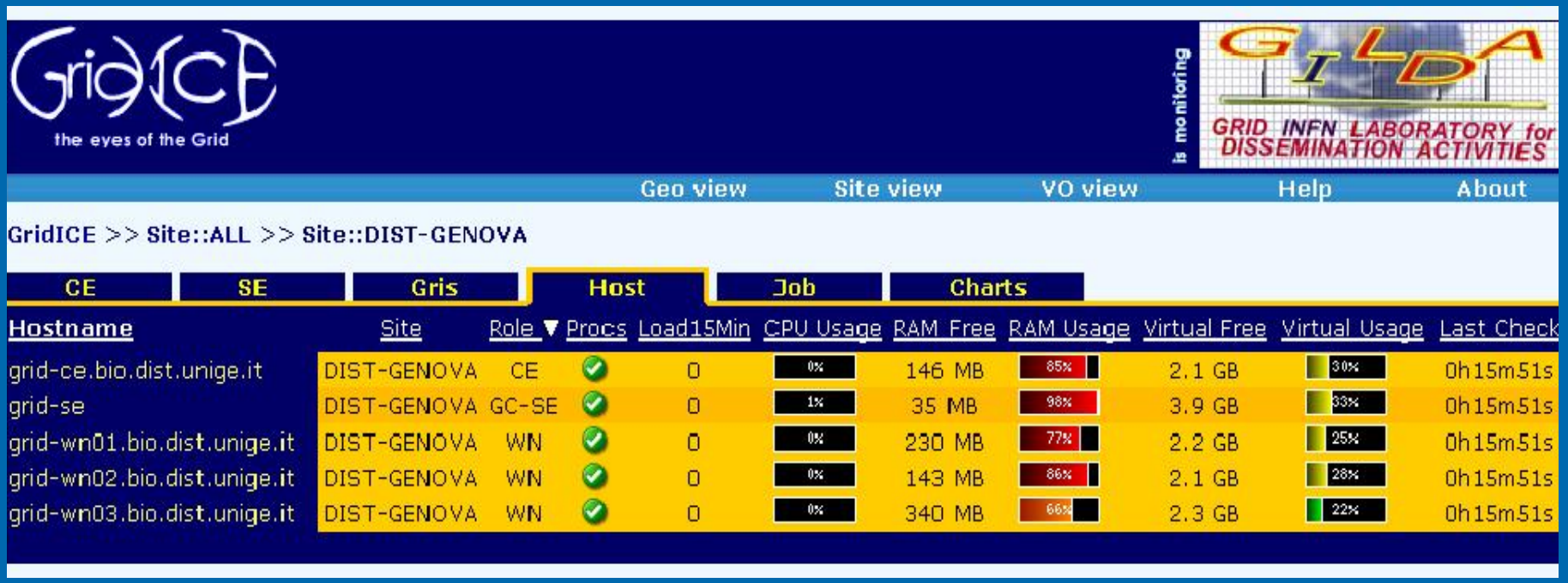
  6 machines are dedicated to LCG tests:

  1 CE, 1 SE, 3 WNs, 1 UI, 1 LCG install server

  → 12 CPUs, 7Giga (RAM), ≈ 400Giga of storage

- BioLab site is now on maintainance to upgrade from LCG 2.4.0 to LCG 2.6.0

DIST – GENOVA site



Hostname	Site	Role	Procs	Load15Min	CPU Usage	RAM Free	RAM Usage	Virtual Free	Virtual Usage	Last Check
grid-ce.bio.dist.unige.it	DIST-GENOVA	CE		0	0%	146 MB	85%	2.1 GB	30%	0h15m51s
grid-se	DIST-GENOVA	GC-SE		0	1%	35 MB	98%	3.9 GB	33%	0h15m51s
grid-wn01.bio.dist.unige.it	DIST-GENOVA	WN		0	0%	230 MB	77%	2.2 GB	25%	0h15m51s
grid-wn02.bio.dist.unige.it	DIST-GENOVA	WN		0	0%	143 MB	86%	2.1 GB	28%	0h15m51s
grid-wn03.bio.dist.unige.it	DIST-GENOVA	WN		0	0%	340 MB	66%	2.3 GB	22%	0h15m51s

LCG node: Technical characteristics (I)

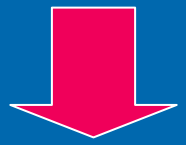
- is powered by an autonomous power line and a uninterruptible power supply with a capacity of 3000 VA is provided
- the UPS is big enough to maintain online the node for about 30 - 45 ‘
- storage element has a redundant power supply and redundant hot-swap cooling system (fan)
- storage is provided by 4 U320 scsi 10.000 rpm disks (about 400 GByte) and a raid 5 controller with 128 mb r/w onboard cache

LCG node: Technical characteristics(II)

- network is provided by a 24 port Gigabit switch and each node has a dual gigabit network adapter (only one channel is used by now)
- node is connected to Internet through the local department LAN
- 1Gbit/sec connection to the GARR POP in Genoa
- the University of Genoa is connected to the Milan Internet Exchange (Internet backbone access point) with a 155 Mbps line

Installation phase

- with this type of hardware (brand HP Proliant DL 140 and DL 380 server) we didn't have any troubles



- node’s installation has been performed without problems
- very good system administration support from staff in Catania

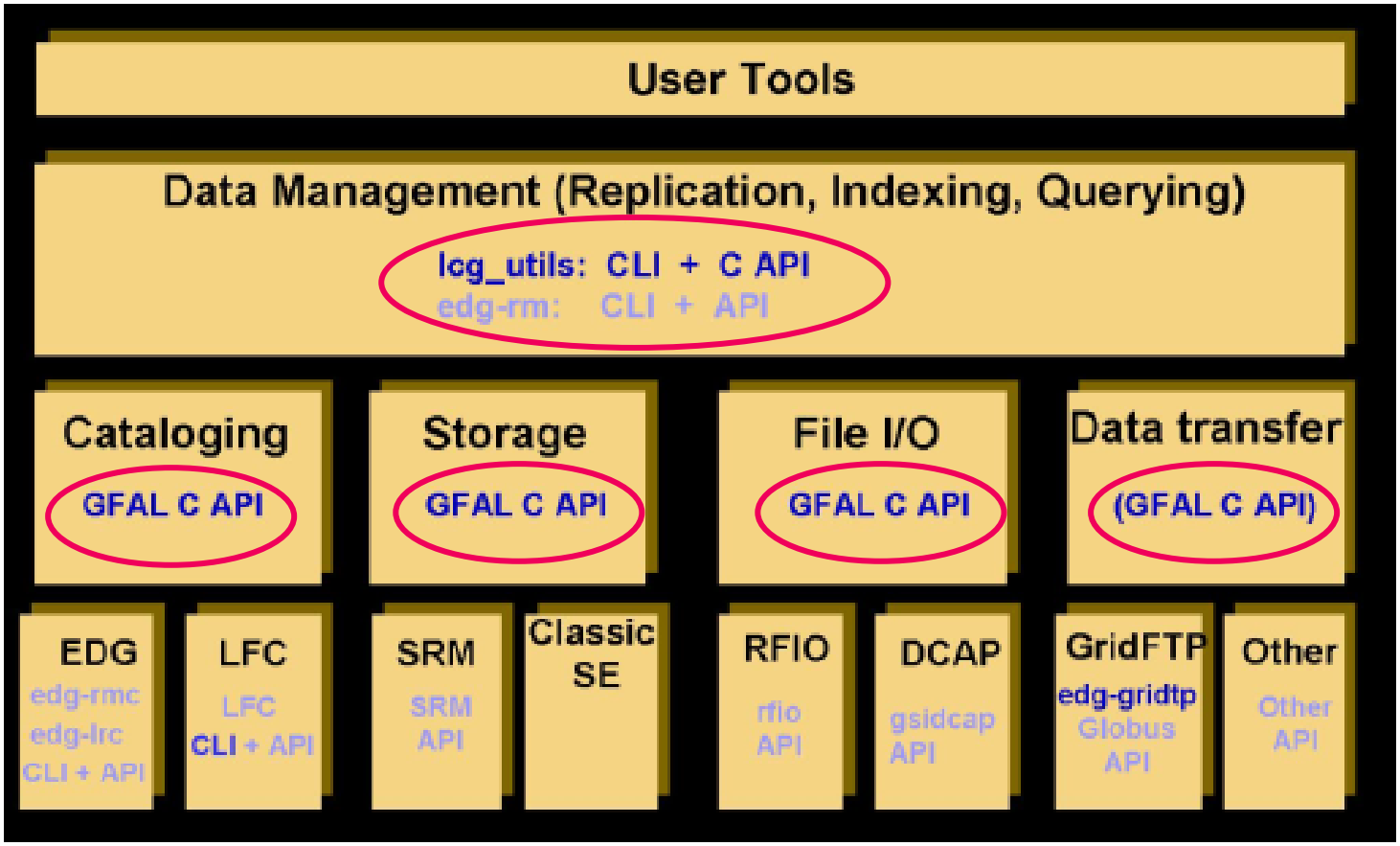
Development phase: Migration of the application to LCG

Main objectives:

- to distribute PET/SPECT images on different Storage Elements available on the GRID and register data - with related metadata - on LCG File Catalog (LFC).

- to access images from User Interface using logical file names (LFN) without copying them on Worker Nodes.

Data Management and File Access



LCG_utils

- **CLI:**
  - lcg-* commands as *lcg-cp, lcg-cr, lcg-del, lcg-la, lcg-lr, ecc….*
- **C API:**
  - int **lcg_cp** (char *src_file, char *dest_file, char *vo, int nbstreams, char * conf_file, int insecure, int insecure);
  - int **lcg_del** (char *file, int aflag, char *se, char *vo, char *conf_file, int insecure, int verbose);

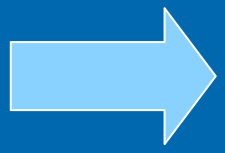
    ...

    ...

Used to store application's data and images on distributed storage elements and register them in LFC catalog

Grid File Access Library (GFAL)

GFAL provides calls for catalog interaction, storage management and file access and can be very handy when an application requires access to some part of a big Grid file but does not want to copy the whole file locally



we used GFAL API to access to distributed images without copying them locally

GFAL: File I/O API

Examples:

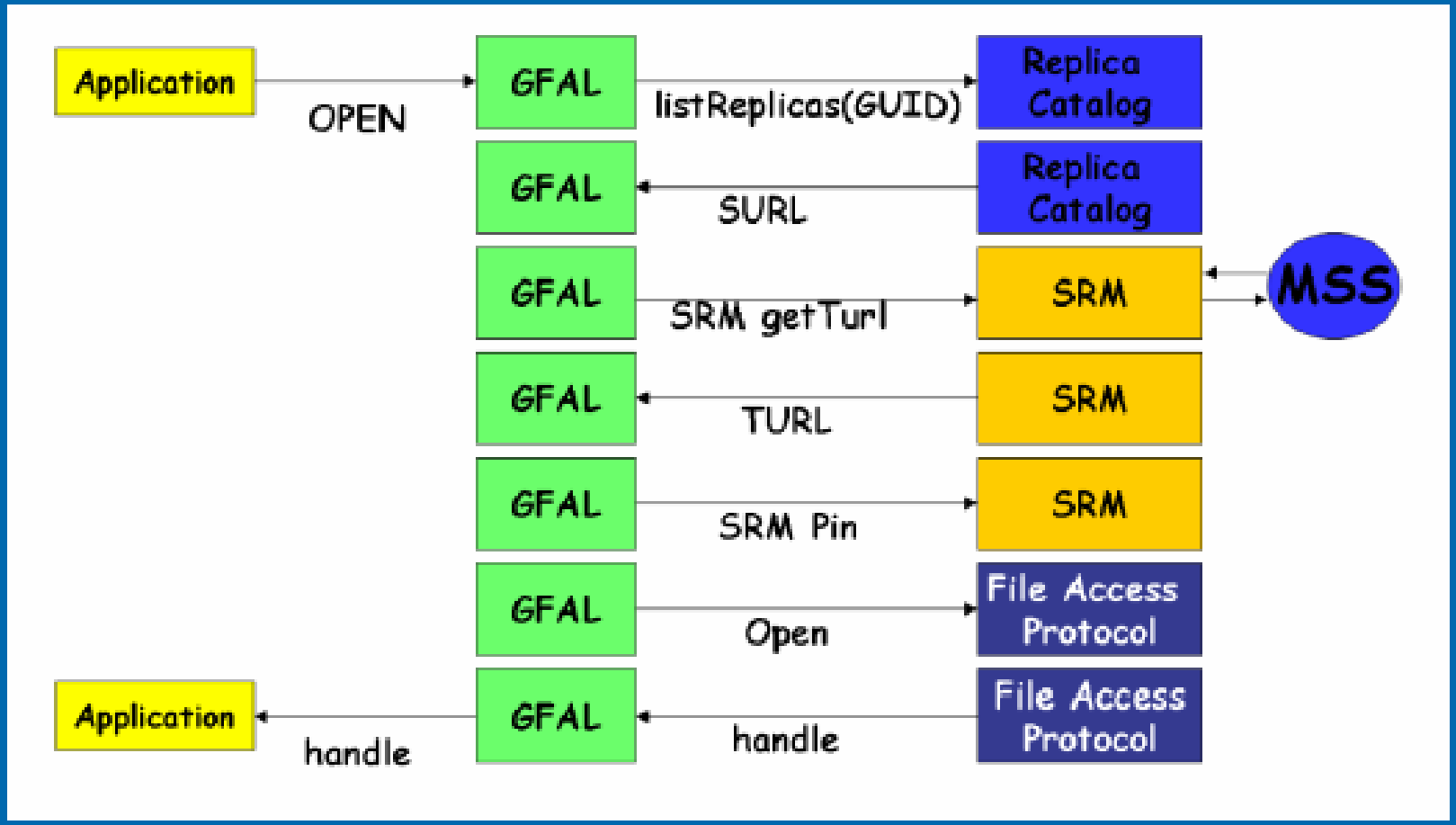
- *int* **gfal_open** *(const char * filename, int flags, mode_t mode);*
- *ssize_t* **gfal_read** *(int fd, void *buf, size_t size);*
- *int* **gfal_close** *(int fd);*

…..

GFAL accepts GUIDs, LFNs, SURLs and TURLs as file names, and, in the first two cases, it tries to find the closest replica of the file.

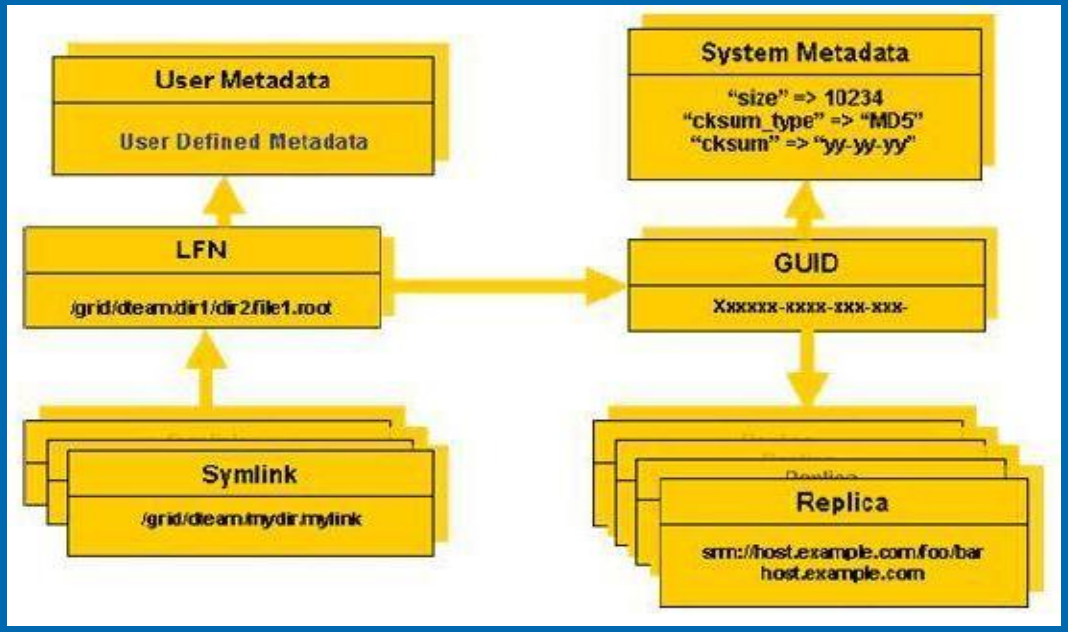
the application code has been modified to access images using their logical file names

A GFAL call


Application
OPEN
GFAL
listReplicas(GUID)
Replica Catalog
GFAL
SURL
Replica Catalog
GFAL
SRM getTurl
SRM
MSS
GFAL
TURL
SRM
GFAL
SRM Pin
SRM
GFAL
Open
File Access Protocol
Application
handle
GFAL
handle
File Access Protocol

LCG File Catalog (LFC)

Functionalities:



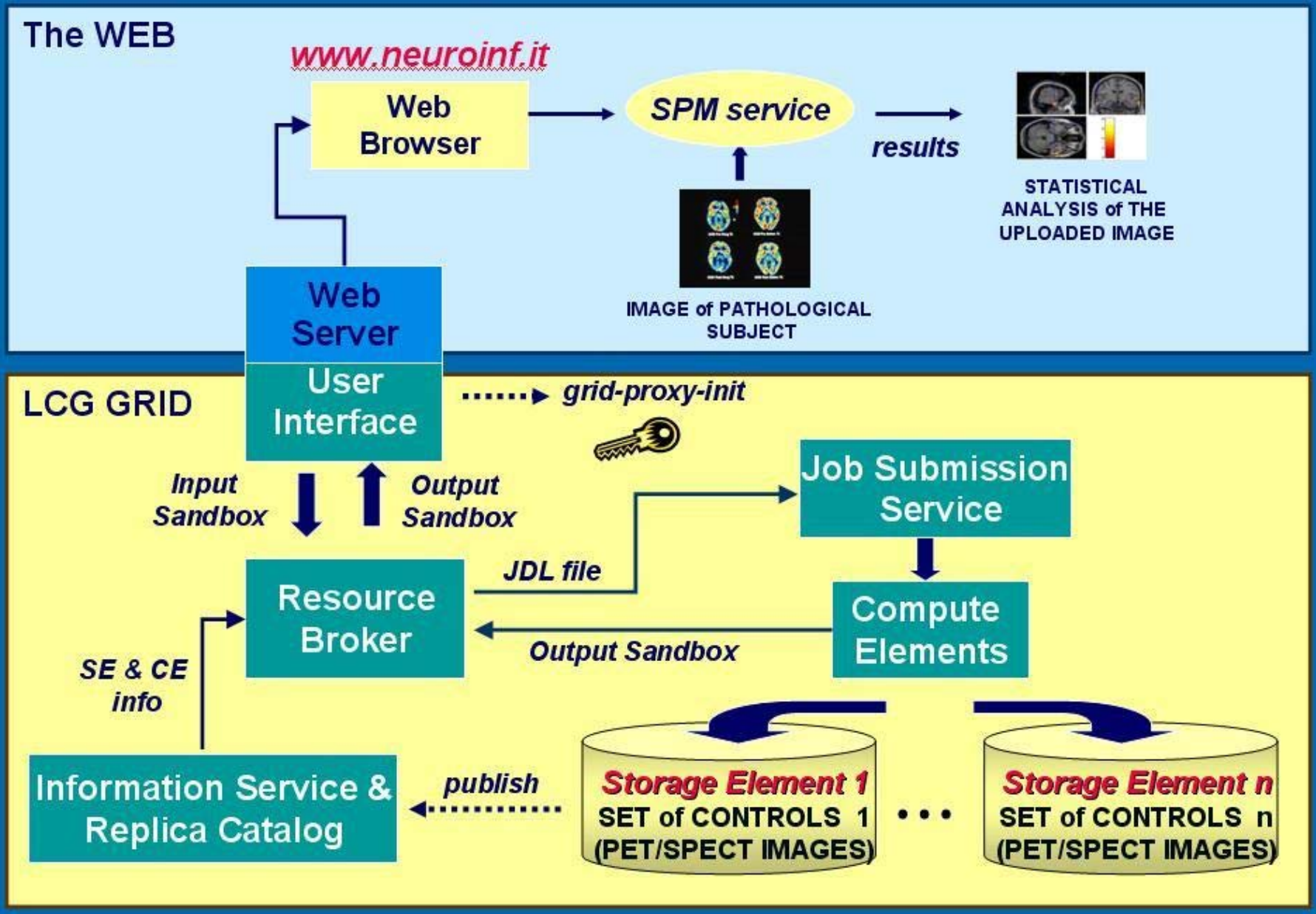
- Use of lcg_utils
- Use of GFAL calls
- Use of GSI certification
- Access to grid file in SEs from "anywhere"
- Several replicas of files in different sites
- Copy of data from/to local file system to GRID

Re-Design of the application

With reference to different LCG tools mentioned above, application code has been modified and structured in the following way:

- Registration and storage of data files (PET/SPECT images) on SEs available using lcg_utils.
- Development of a C program with GFAL C API in order to access distributed images using their LFNs and to extract some information necessary to SPM analysis without copying them locally.
- Job Submission: creation of a JDL file to submit the executable (and not the images) with GFAL call to the GRID.
- Statistical Analysis: running of SPM analysis from results obtained from job submission. Statistical analysis is performed outside GRID environment.

Final GRID structure



Problems

- If you want to access a file on a classic SE using GFAL from your UI, you won't be able to do this from the UI (because of insecure RFIO that needs uid correspondance), but you should be able to do it from a WN in the same site (there is indeed such correspondance between uids in WNs and in SEs).
  So the options are:
  1. Copy the file with gridftp to your UI, then access it locally
  2. Send your application (that contains the GFAL calls) in a job to a CE in the same site as the SE where the file you want to access sits.

- Support for Metadata using LFC catalog:
  - There is an only field for metadata
  - No way to do metadata queries

Contents

- Introduction to SPM analysis for diagnosis in AD
- Development of SPM service through a web portal
- GRID implementation of SPM service
- Current Status and Plans

Current Status and Plans

- Integration of GRID application with neuroinformatics web interface
- To find a better way to manage metadata
- Integrate the application in GENIUS
- To value the possibility of parallelize the statistical analysis

Thanks for your attention!