



# Distributed Repository for Biomedical Applications

L. Corradi, I. Porro, A. Schenone, M. Fato

*University of Genoa*

*Dept. Computer Communication and System Sciences (DIST)*

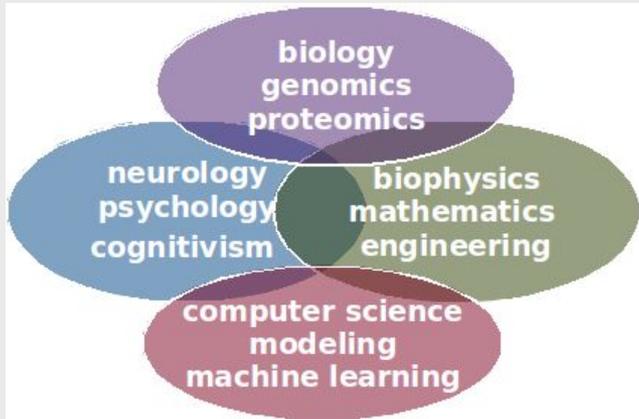
*BIOLAB*

*Contact: [ivan.porro@unige.it](mailto:ivan.porro@unige.it)*

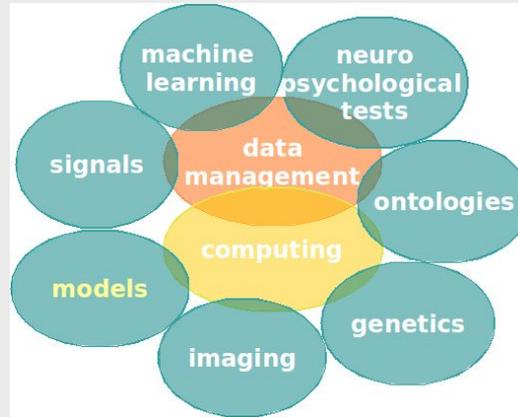
Conferenza GARR 2010  
Welcome to the Future Internet!  
Torino, 26-28 Ottobre 2010

# Multidisciplinary → issues

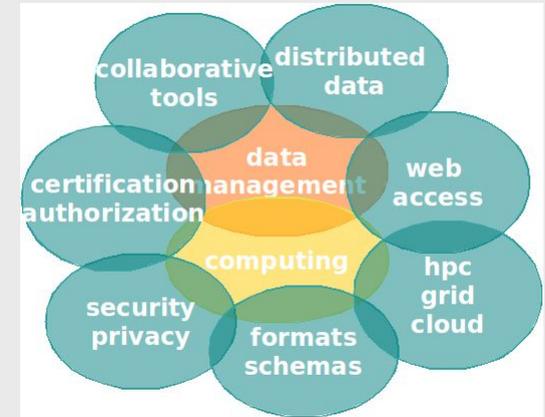
## Domain



## Data



## Technology



- Scientific paradigms
- Methods and algorithms
- Data analysis techniques

- Formats
- Schemas
- Procedures
- Semantics

- Methods
- Tools
- SW environments

# Support virtual, distributed experiments in neuroscience

**Multiscale approach**

**Multimodal analysis**

**Multicentric data**

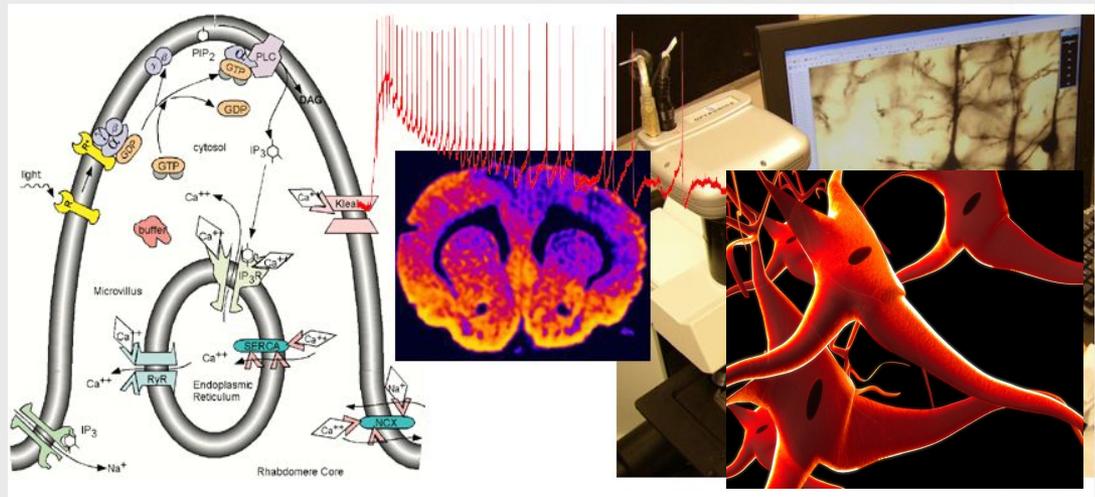
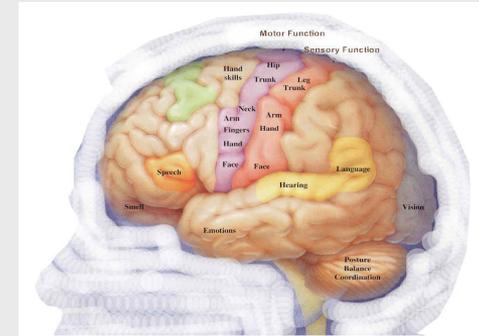
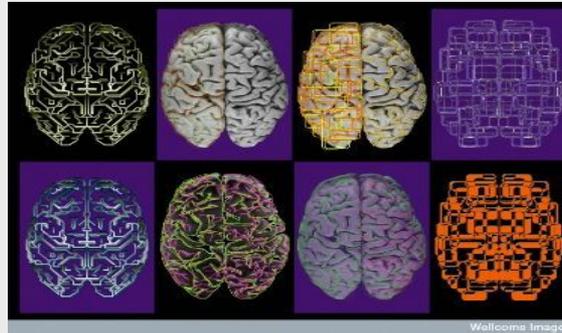
**Data integration**

**Collaborative tools**

**Extensibility requests**

**Scalability issues**

**Complexity**



- **to fully support medical research** during neuroinformatics and bioinformatics multidisciplinary studies also providing integrated services and tools;
- **to handle multimodal/multiscale data and metadata**, enabling the injection in the repository of several different data types according to structured schemas;
- **to be highly extensible, designed in a very general way** in order to be able to fit different requirements coming from a *large variety of applications simply through a custom configuration*.

- Data models
  - **XCEDE** (XML-based Clinical and Experimental Data Exchange)
  - **FuGE** (The Functional Genomics Experiment)
- Ontologies
  - **OBI** (Ontology for Biomedical Investigations)
  - **Sequence Ontology**
- Data Grid
  - **gLite by EGEE** (SRM, DPM, ...)
  - **iRODS by Teragrid** (derived from **SRB** - Storage Resource Broker, form the same group at San Diego Supercomputing Center)

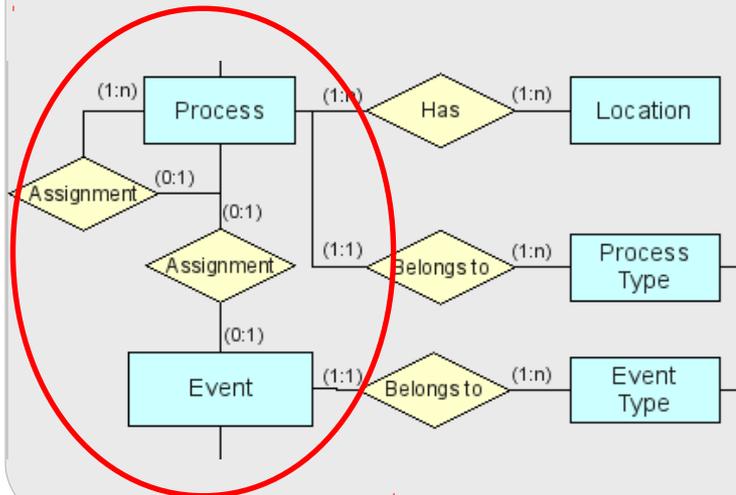
# Satisfy extensibility: Process-event model

- Designed to be highly flexible and easily extensible
- A multipurpose taxonomic schema composed by generic objects:

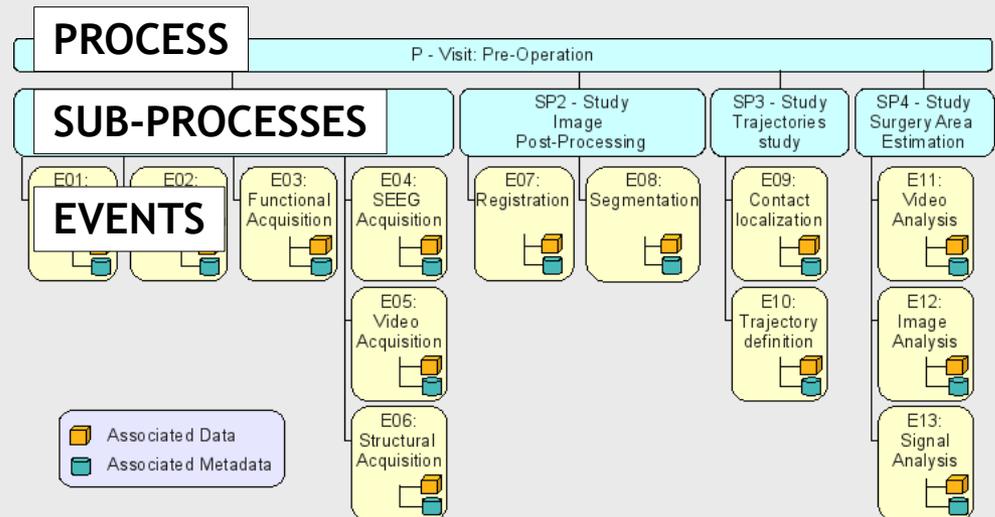
**Process:** structured sequence of processes and events

**Event:** single step of the process

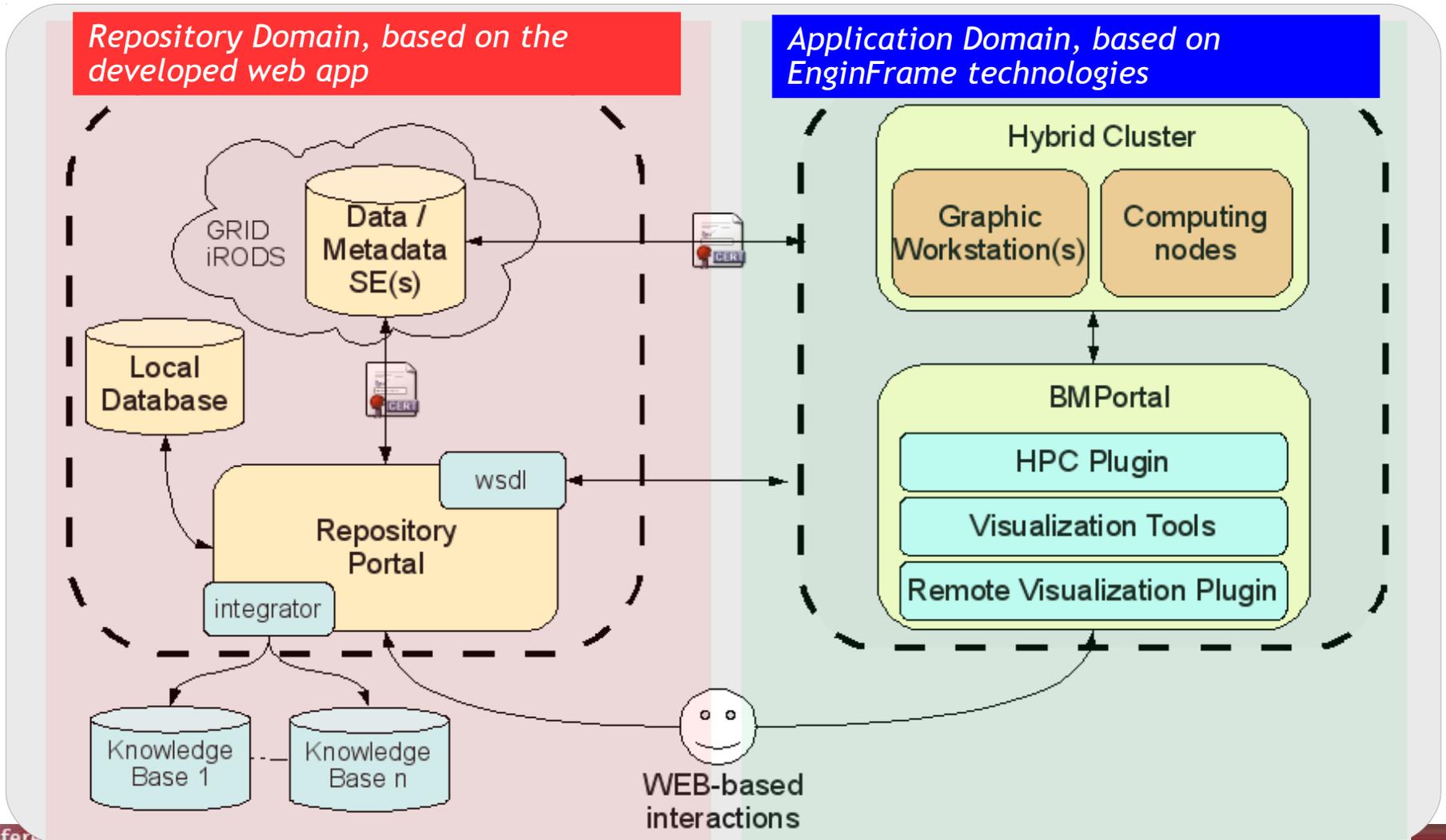
The model connect processes, events and data



*Example:* a pre-surgical assessment (process) is made by several data acquisition and processing events







# *DataGrid: why iRods?*

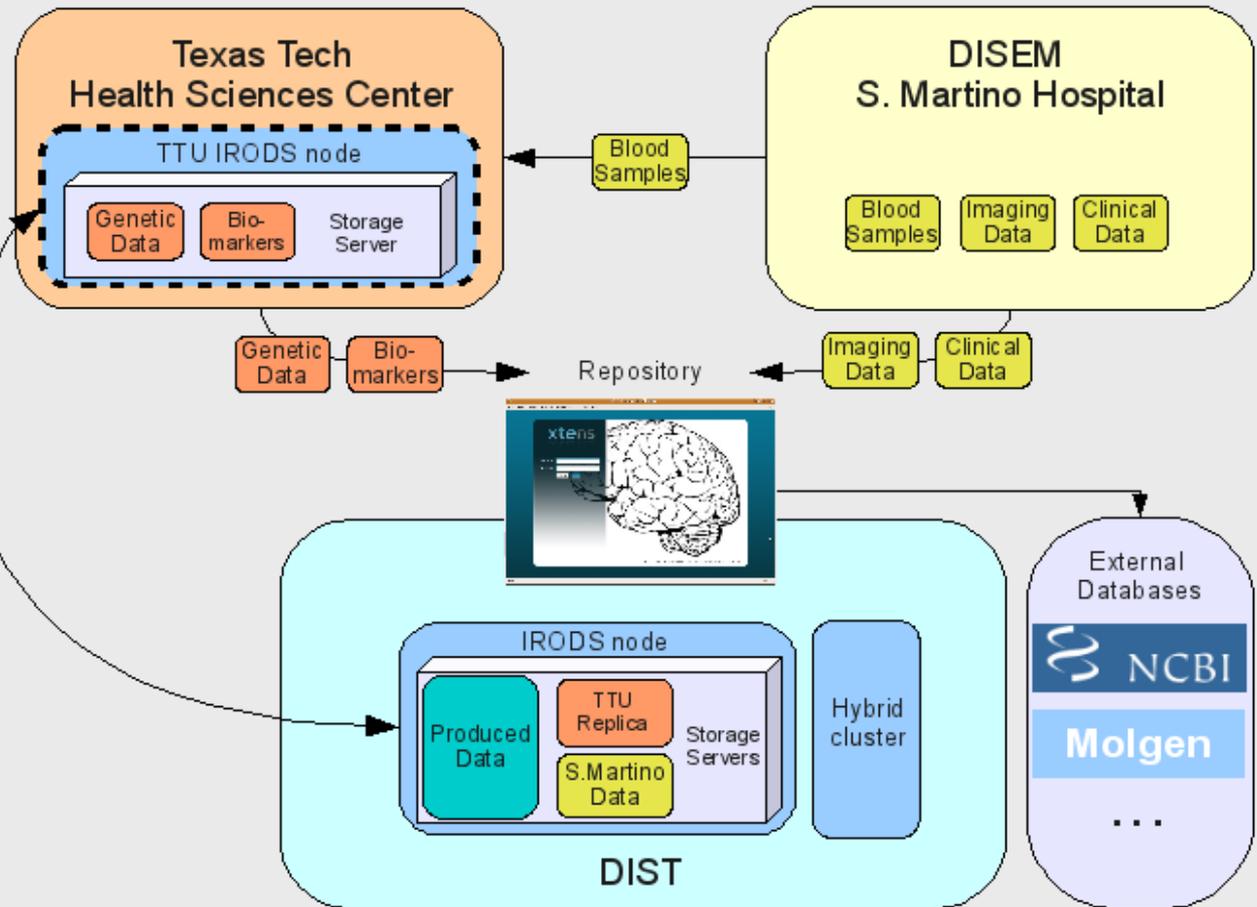
## **iRODS**

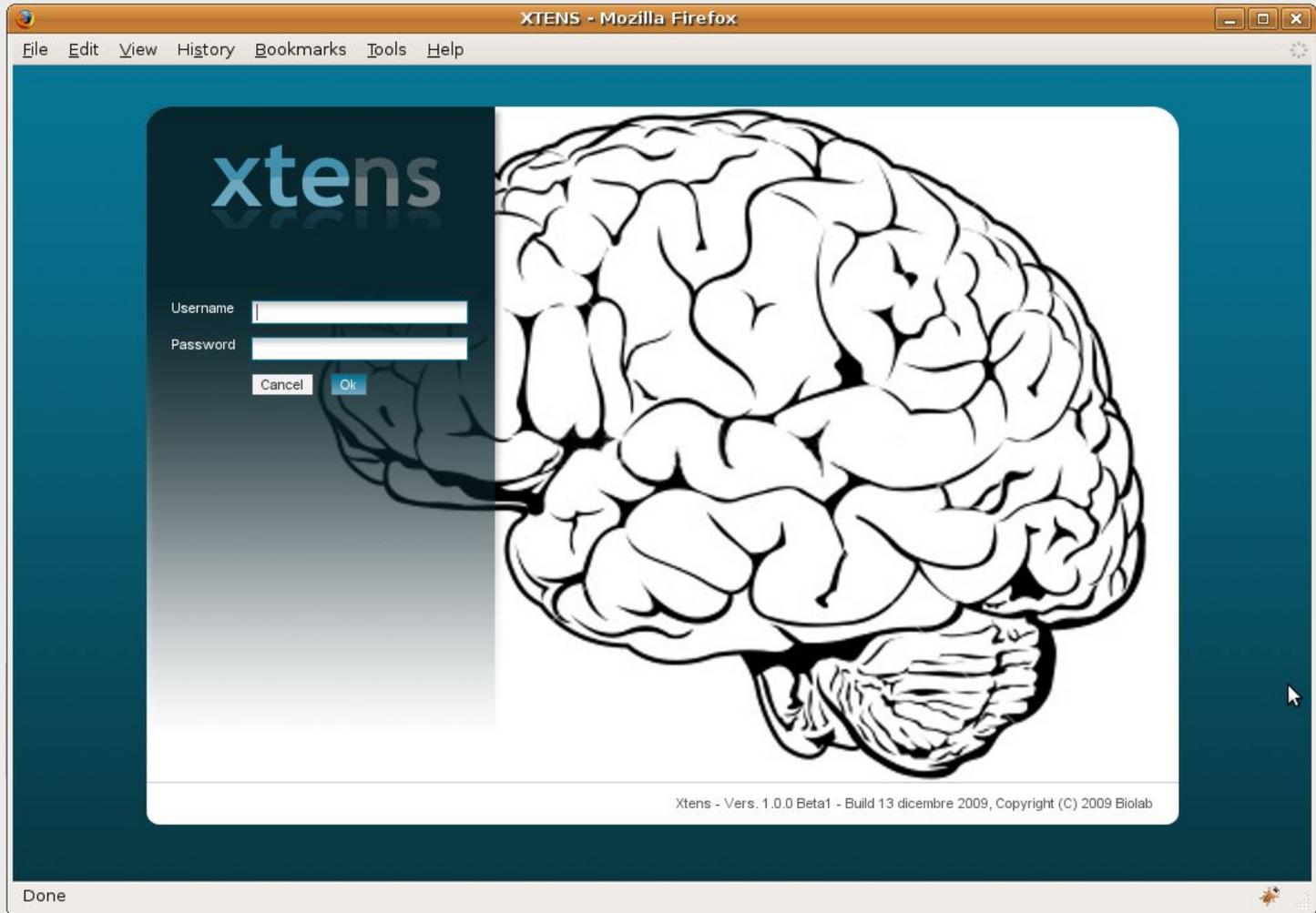
(the Integrated Rule-Oriented Data System ( <http://www.irods.org> )

- lightweight (13.3MB package...)
- Stable and flexible
- it handles metadata (gLite needs to use additional components);
- it permits to use heterogeneous storage resources;
- it permits to easily perform operations on the stored data and metadata (the “Rule Engine”);
- it doesn’t bind to use a specific infrastructure;
- it supports GSI (Grid Security Infrastructure) as an authentication method;
- Interesting Roadmap (DropBox like drag&drop, ...)

**Discover**  
Biomarkers,  
new diagnosis  
criteria

**Understand**  
development of  
mild cognitive  
impairment  
(MCI) to  
Alzheimer's  
disease (AD)





# Manage users, groups, permissions

**Groups**

Trovati 3 risultati.

#	Description	Actions

**Pages**

Trovati 43 risultati, visualizzo da 1 a 10. [Prima/Precedente] 1, 2, 3, 4, 5 [Successiva/Ultima]

#	uri	Description	Actions
<input type="checkbox"/>	/admin/accesses/	access control	[Icons]
<input type="checkbox"/>	/admin/function/	function list	[Icons]
<input type="checkbox"/>	/admin/functionGrup/	function-group page	[Icons]
<input type="checkbox"/>	/admin/functions/	function list	[Icons]
<input type="checkbox"/>	/admin/group/		
<input type="checkbox"/>	/admin/groups/		
<input type="checkbox"/>	/admin/grpFunction/		
<input type="checkbox"/>	/admin/grpOperator/		
<input type="checkbox"/>	/admin/grpPage/		
<input type="checkbox"/>	/admin/bevdata/		

Associate with a group:  [Associate](#) [New page](#)

**Operators**

**New operator**

Surname:

Name:

All Groups

[Search](#)

[Click to insert a new operator](#)

**Functions**

#	Description	Code	Actions
<input type="checkbox"/>	<a href="#">Edit operator authentication fields</a>	viewOperatorAuth	[Icons]
<input type="checkbox"/>	<a href="#">Edit operators</a>	editOperator	[Icons]
<input type="checkbox"/>	<a href="#">Edit patient personal data</a>	editPatient	[Icons]
<input type="checkbox"/>	<a href="#">Enable configuration menu</a>	viewConfMenu	[Icons]
<input type="checkbox"/>	<a href="#">Enable functions deletion</a>	deleteFunction	[Icons]
<input type="checkbox"/>	<a href="#">Insert new operator</a>	newOperator	[Icons]
<input type="checkbox"/>	<a href="#">Insert new patient</a>	newPatient	[Icons]

Associate with a group:  [Associate](#) [New function](#)

xtens - Vers. 1.0.0 Beta1 - Build 3 febbraio 2010, Copyright (C) 2009 Biolab

← Groups

← Operators

Pages →

Functions →

# Manage Processes / events

The screenshot displays the 'xtens' web application interface. At the top, there is a navigation bar with 'Home', 'Configuration', 'Actions', and 'Patients' tabs. The main content area shows a table of process types with columns for '#', 'Description', and 'Actions'. Below the table is a 'New Process' button. A modal dialog titled 'Event Types:' is open, showing a list of event types including 'PATIENT DELETION', 'MRI ACQUISITION', 'CT ACQUISITION', 'ANGIO ACQUISITION', 'EEG ACQUISITION', 'SEEG ACQUISITION', 'PET ACQUISITION', 'SPECT ACQUISITION', 'FMRI ACQUISITION', and 'DTI-FI ACQUISITION REGISTRATION'. The 'MRI ACQUISITION' option is selected. The dialog has 'Ok' and 'Cancel' buttons.

#	Description	Actions
<input type="checkbox"/>	ADMIN	[User] [Info] [X]
<input type="checkbox"/>	PATIENT MANAGEMENT	[User] [Info] [X]
<input type="checkbox"/>	NEUROCLINICAL DATA ACQUISITION	[User] [Info] [X]
<input type="checkbox"/>	MOLECULAR GENETICS DATA ACQUISITION	[User] [Info] [X]
<input type="checkbox"/>	SAMPLES MANAGEMENT	[User] [Info] [X]
<input type="checkbox"/>	IMAGE POST-PROCESSING	[User] [Info] [X]
<input type="checkbox"/>	DATA COLLECTION	[User] [Info] [X]
<input type="checkbox"/>	DATA ANALYSIS	[User] [Info] [X]

Event Types:

- PATIENT DELETION
- MRI ACQUISITION**
- CT ACQUISITION
- ANGIO ACQUISITION
- EEG ACQUISITION
- SEEG ACQUISITION
- PET ACQUISITION
- SPECT ACQUISITION
- FMRI ACQUISITION
- DTI-FI ACQUISITION REGISTRATION

# Creation of new data type

## New Data Type Creation

**Data information**

Name	MRI	File Upload	YES
Description	MRI DATA		

**Process-Event Association**

Event Type	MRI ACQUISITION	Process Type	NEURO/CLINICAL DATA ACQUISITION
------------	-----------------	--------------	---------------------------------

**Metadata information**

Version	01.00.00	Ontology	Ontology for Biomedical Investigations (OBI)
---------	----------	----------	--

**Group 0**

Name	GENERIC INFORMATION		
------	---------------------	--	--

**Attribute**

Type	STRING	<input type="checkbox"/> Required <input type="checkbox"/> Is List <input type="checkbox"/> Has Unit	<input type="checkbox"/> Enable XTENS Connection	GENE
Name	INSTITUTION_NAME		Unit Value	
Value				
	<input type="button" value="Remove"/> <input type="button" value="Add"/>			<input type="button" value="Remove"/> <input type="button" value="Add"/>

**Attribute**

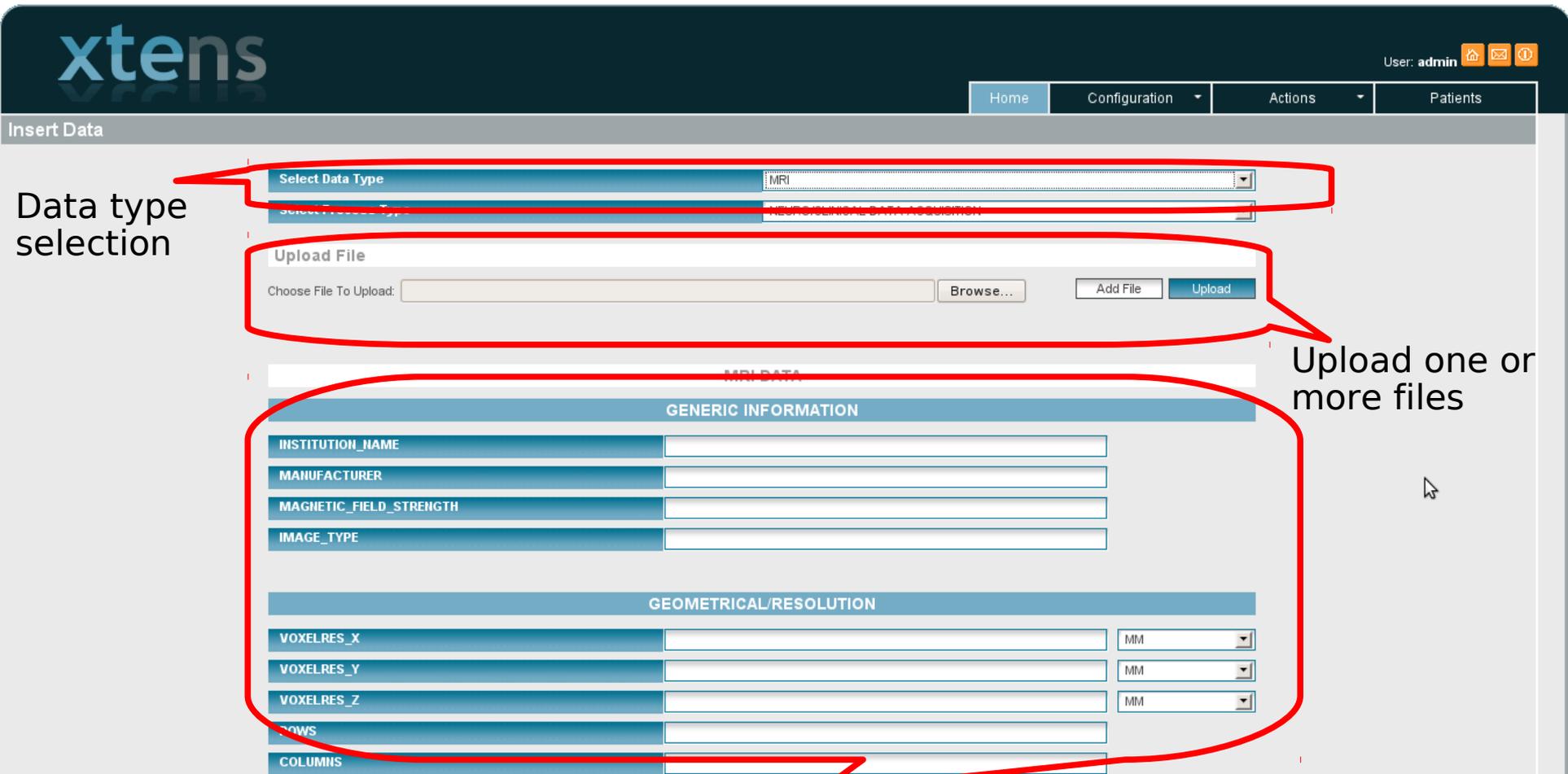
Type	STRING	<input type="checkbox"/> Required <input type="checkbox"/> Is List <input type="checkbox"/> Has Unit	<input type="checkbox"/> Enable XTENS Connection	GENE
Name	MANUFACTURER		Unit Value	
Value				
	<input type="button" value="Remove"/> <input type="button" value="Add"/>			<input type="button" value="Remove"/> <input type="button" value="Add"/>

**Attribute**

Type	FLOAT	<input type="checkbox"/> Required <input type="checkbox"/> Is List <input type="checkbox"/> Has Unit	<input type="checkbox"/> Enable XTENS Connection	GENE
Name	MAGNETIC_FIELD_STRENGTH		Unit Value	
Value				
	<input type="button" value="Remove"/> <input type="button" value="Add"/>			<input type="button" value="Remove"/> <input type="button" value="Add"/>

Link data to processes

Ontology



xtens

User: admin

Home Configuration Actions Patients

Insert Data

Select Data Type MRI

Upload File

Choose File To Upload:  Browse... Add File Upload

MRI DATA

GENERIC INFORMATION

INSTITUTION\_NAME

MANUFACTURER

MAGNETIC\_FIELD\_STRENGTH

IMAGE\_TYPE

GEOMETRICAL/RESOLUTION

VOXELRES\_X  MM

VOXELRES\_Y  MM

VOXELRES\_Z  MM

ROWS

COLUMNS

Data type selection

Upload one or more files

Insertion of metadata defined during data type creation

Gene Details

Property Exons Variants

NCBI geneid:

Name: TAU

# Exons: 17

Notes: blabla

Gene Name	Description	Organism	Summary
<input type="radio"/> tau	Microtubule-associated protein tau	Salmo salar	undefined
<input type="radio"/> tau	CG31057 gene product from transcript CG31057-RA	Drosophila melanogaster	undefined
<input type="radio"/> Tau	tau-like protein	Bombyx mori	undefined
<input type="radio"/> tau	DNA replication terminus site-binding protein	Yersinia enterocolitica subsp. enterocolitica 8081	undefined
<input type="radio"/> mapt	microtubule-associated protein tau	Xenopus (Silurana) tropicalis	undefined
<input checked="" type="radio"/> MAPT	microtubule-associated protein tau	Homo sapiens	This gene encodes the microtubule-associated protein tau (MAPT) whose transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species. MAPT transcripts are differentially expressed in the nervous system, depending on stage of neuronal maturation and neuron type. MAPT gene mutations have been associated with several neurodegenerative disorders such as Alzheimer's disease, Pick's disease, frontotemporal dementia, cortico-basal degeneration and progressive supranuclear palsy. [provided by RefSeq]
<input type="radio"/> MAPT	microtubule-associated protein tau	Bos taurus	undefined
<input type="radio"/> Mapt	microtubule-associated protein tau	Mus musculus	undefined
<input type="radio"/> Csnk1e	casein kinase 1, epsilon	Mus musculus	undefined
<input type="radio"/> tus	inhibitor of replication at Ter, DNA-binding protein	Escherichia coli str. K-12 substr. MG1655	Site-specific DNA binding protein Tus arrests the DnaB helicase at six chromosomal Ter sites (A-F), blocking replication forks. [More information is available at EcoGene: EG11038]. Tus, also known as ter-binding protein (TBP), binds to ter sites, blocking the progress of DNA replication in a polar fashion, thus terminating replication. [More information is available at EcoCyc: EG11038].

Ok Cancel

Build 13 dicembre 2009, Copyright (C) 2009 Biolab

Real time query NCBI to retrieve needed information

# Support everyday tasks

The screenshot displays the 'xtens' web application interface. At the top, the 'xtens' logo is on the left, and the user 'admin' is logged in on the right. A navigation bar includes 'Home', 'Configuration', 'Actions', and 'Patients'. Below this is a 'Gene Details' section with tabs for 'Property', 'Exons', and 'Variants'. The main area is titled 'New Variant' and contains a form with fields for 'Rnumber', 'Exon', 'Base Change', 'Genomic Position', 'SNP', 'Mutation', 'Pathogenic', 'Position', 'Codon Change', 'AA Transition', and 'Notes'. To the right of the form is an integrated 'Entrez Gene' window for the gene 'MAPT microtubule-associated protein tau [Homo sapiens]'. The Entrez window shows the gene's official symbol, full name, and primary source (HGNC:6893). A sidebar on the right of the Entrez window lists various data types like 'Summary', 'Genomic', 'transcript', etc. At the bottom of the screenshot, the version information 'Xtens - Vers. 1.0.0 Beta1 - Build 13 dicembre 2009, Copyright (C) 2009 Biolab' is visible.

Integrated visualization of external sources within the system

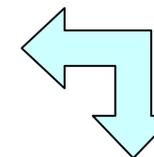
**Sample List**

**Search Parameters**

Code:  Sample Type:  Project:

Name:

#	Code	Name	Type	Shipment	Receipt	Storage	Drawing	
<input type="radio"/>	001-A729-01	X2520	BLOOD	09/16/2009	09/24/2009			<a href="#">Edit</a>
<input checked="" type="radio"/>	001-A729-04	X2523	BLOOD	09/03/2009	10/02/2009			<a href="#">Edit</a>
<input type="radio"/>	001-A755-09	X2528	BLOOD				11/09/2009	<a href="#">Edit</a>
<input type="radio"/>	001-A802-02	X2529	BLOOD					<a href="#">Edit</a>



**Details**

Project: GENOVA  
 Sample Notes: \_\_\_\_\_  
 Fridge Location: ROOM 2  
 Temperature: -80 °C

Shelf

Metal Rack

Box: BX789

001-A729-04 - Hole n°42

User: admin

Home Configuration Actions Patients

**New Sample**

Type:

type	Fedex	
BLOOD		<a href="#">Edit Patient details</a>
RAIN TISSUE		<a href="#">Edit Patient details</a>
ERUM		<a href="#">Edit Patient details</a>
VA		<a href="#">Edit Patient details</a>

ct:  
 ile Notes:  
 e Location: ROOM 2  
 erature: -80 °C

Shelf

Metal Rack

Box: BX789

Find bio-samples by location, FedEx number, patient code, ...

The screenshot shows the 'xtens' web application interface. At the top left is the 'xtens' logo. On the right, the user is identified as 'admin' with icons for home, mail, and help. A navigation menu includes 'Home', 'Configuration', 'Actions', and 'Patients'. The main content area is divided into two sections: 'Search Patient' and 'New patient'. The 'Search Patient' section contains several dropdown menus for 'Diagnosis', 'Data Type', 'Sex', 'Field', and 'Birth date'. The 'Field' dropdown is open, showing options like 'MY ATTRIBUTE 1', 'Please Select', 'MY ATTRIBUTE 2', 'MY ATTRIBUTE 3', and 'MY ATTRIBUTE 4'. Below the search form is a table with columns for '#', 'Code', 'Surname', 'Sex', 'Birth date', and 'Diagnosis'. The table is currently empty, with the message 'La ricerca non ha prodotto risultati' (The search did not produce results). A 'Submit' button is located at the bottom of the search form. The 'New patient' section features a large button that says 'Click to insert a new patient'. At the bottom right, the version information is displayed: 'Xtens - Vers. 1.0.0 Beta1 - Build 13 dicembre 2009, Copyright (C) 2009 Biolab'.

- Query data by integrating:
- data types
  - data types attribute values
  - diagnosis and patients attributes

# Scenario 4: Overview on patient

xtens

User: admin

Home Configuration Actions Patients

Search Patient

Diagnosis:  Sex:  Birth date:

Data Type:  Field:  Value:

New patient

Trovati 6 risultati.

#	Code	Surname	Name	Sex	Birth date	Diagnosis						
	CNTGLI37L4	C	G	F	02/07/1937	AL ZHEIMER	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>
	CRTLUCU81D	C	G	M	02/07/1937	AL ZHEIMER	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>
	PGGMRZ85H	C	G	M	06/06/1985	NON AMNESTIC MCI	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>
	RSSGRG69D	R	G	M	09/04/1969	CONTROL	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>
	RSSMRA81D	R	M	M	14/04/1981	AMNESTIC MCI	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>
	RZZSMN77L	R	S	F	02/07/1977	NON AMNESTIC MCI	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>

Browse results and look at both clinical and genetic data

Clinical Details

Clinical Data	NeuroPsychological Tests	Data List
<b>Clinical Data</b>		
Field	Value	
USE OF ALCOHOL	NO	
IS SMOKER	YES	
IS SMOKER DATE		
EX SMOKER	YES	
DIABETES	YES	
HIGH BLOOD PRESSURE	YES	
HIGH BLOOD PRESSURE DATE		
HIGH CHOLESTEROL	NOT DECLARED	

Molecular Genetics

Haplotype:  Known Variants: Found Variants

**Haplotype**

ACE Haplotype	MAPT Haplotype	APOe
rs16832753 - C>G	H1/H1	E2/e3

**Known Variants**

Exon	Rsnumber	Gen Pos	B Change	SNP	Mutation	Pathogenic	Position	C Change	AA Trans
2	rs63750529	g.75859G	G>A	NO	YES	NO	EXON	From ACG to ACA	-
1	rs63750959	g.72930G	G>T	NO	NO	NO	EXON	From CGC to CTC	-

**Found Variants**

Exon	Rsnumber	Gen Pos	B Change	SNP	Mutation	Pathogenic	Position	C Change	AA Trans
5	-	g.92434C	C>G	NO	YES	NO	EXON	-	-

**xtens**

Search Patient

Diagnosis:  Sex:

efadmin's X desktop (node01:3)

Surface 1

Value:  Apply

Surface

Red

Code	Surname
CNTLCU7L4D866V	CONTE
CPTLCU8D4898E	CORTI
POQMFZ3SH4886P	POGGI
PSQSPG68D9086P	ROSSI
PSQSPAR1D14869Q	ROSSI
RZ2SMW7L4D866L	RUZZENENTI

Clinical Details

Clinical Data	NeuroPsychological Test
DATA 1	
DATA 2	
DATA 3	
DATA 4	
DATA 5	

Name

img\_0124.jpg  
img\_0120.jpg  
img\_0125.jpg

File name: img\_0124.jpg

close or Esc Key

“in line” visualization for simple data

04/1981	AMNESTIC MCI	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>
07/1977	NON AMNESTIC MCI	<a href="#">Clinical Data</a>	<a href="#">Molecular Genetics</a>	<a href="#">Samples</a>	<a href="#">Associate Sample</a>	<a href="#">History</a>	<input type="checkbox"/>

**Molecular Genetics**

Haplotype: MAPT Known Variants Found Variants

Haplotype

ACE Haplotype || MAPT Haplotype || APOe

- Remote Desktop run visualization applications for complex data
- Advanced OpenGL enabled viewer available in EnginFrame
- Exploit remote visualization clusters



Thank you!

Q & A

[ivan.porro@unige.it](mailto:ivan.porro@unige.it)

