



Distributed Repository for Biomedical Applications

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

University of Genoa

Dept. Computer Communication and System Sciences (DIST)

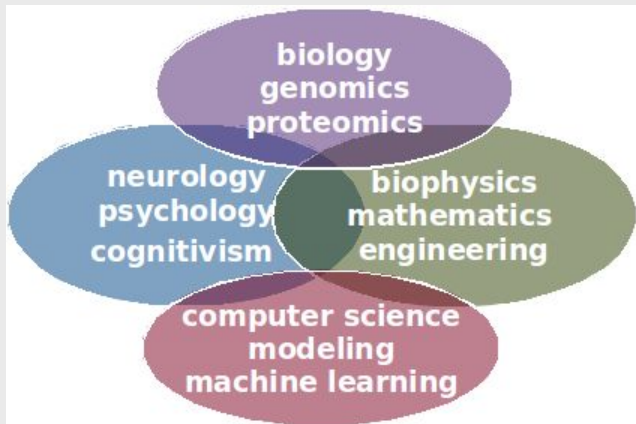
BIOLAB

Contact: ivan.porro@unige.it

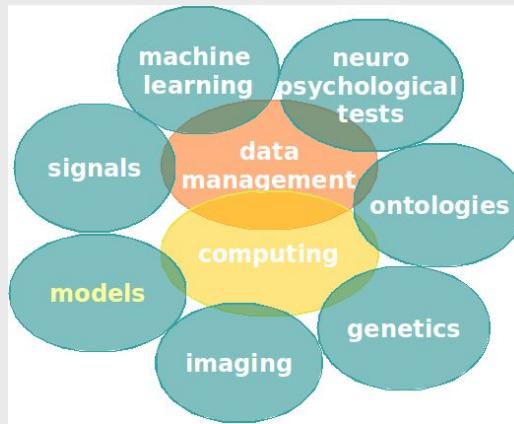
Conferenza GARR 2010
Welcome to the Future Internet!
Torino, 26-28 Ottobre 2010

Multidisciplinary → issues

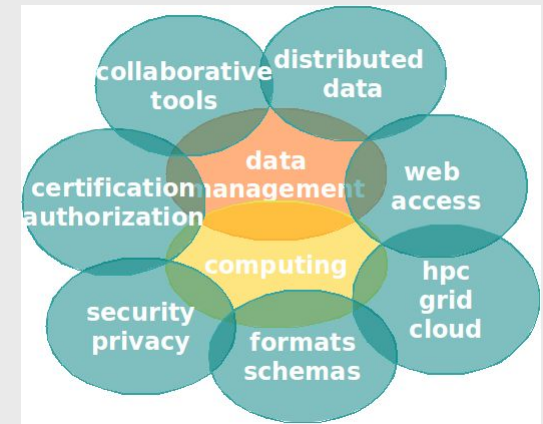
Domain



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

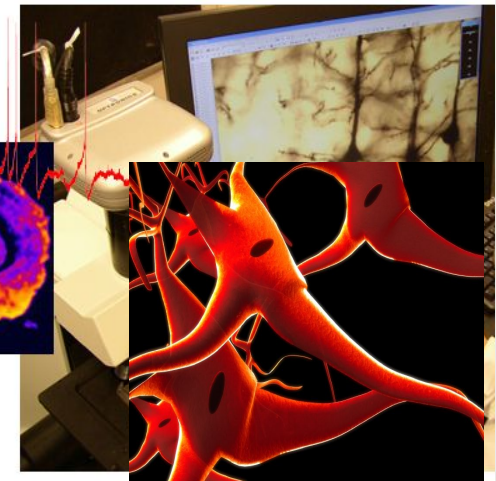
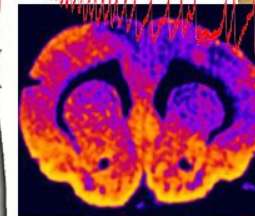
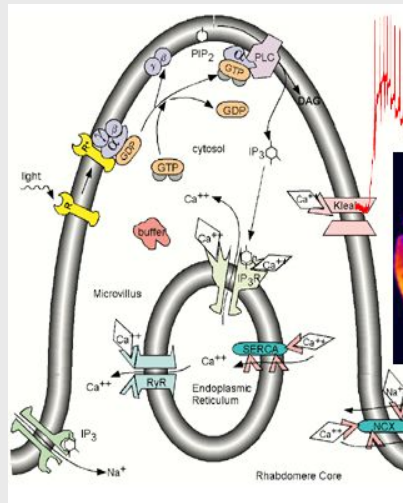
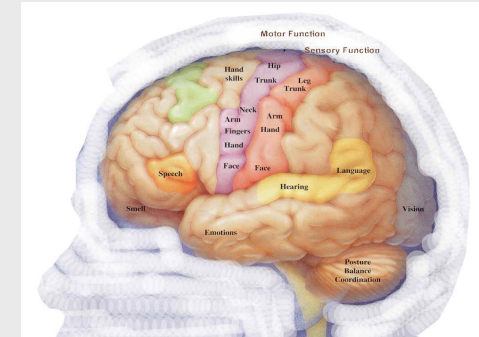
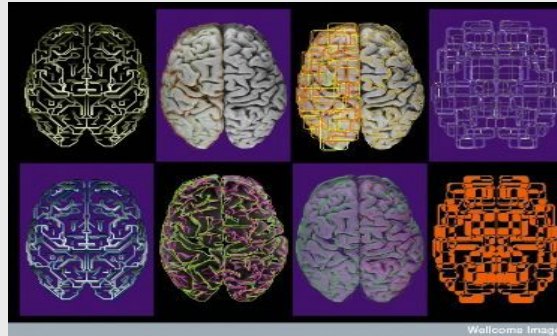
Data integration

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)

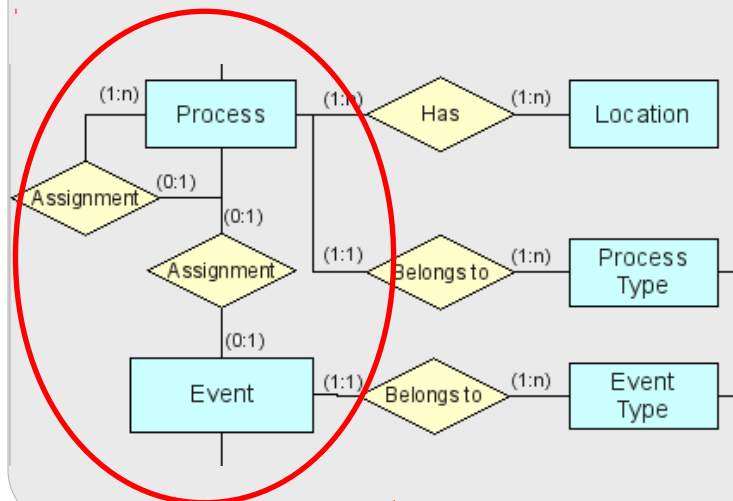
Satisfy extensibility: Process-event model

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

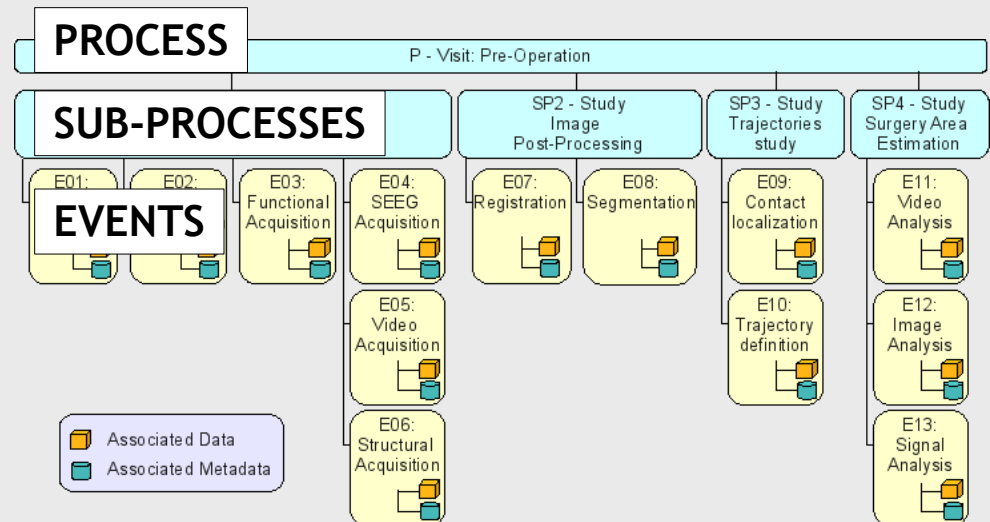
Process: structured sequence of processes and events

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





Admin / Power user

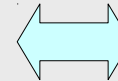
Data type creation form



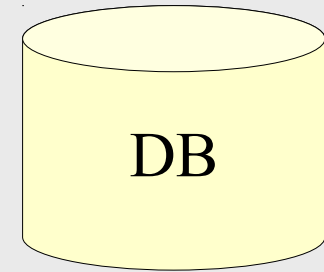
```

<ref uri="#xsi:type=boolean"/>
<attribute base="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://schemas.xmlsoap.org/soap/envelope/">
  <xs:base base="http://www.w3.org/2001/XMLSchema" />
</attribute>
<xs:element base="xs:string" type="xsd:string" required="1" minOccurs="1" maxOccurs="1" />
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>

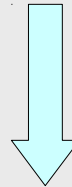
```



XML
+
Ad hoc structures



XSL Transformation



Medical user

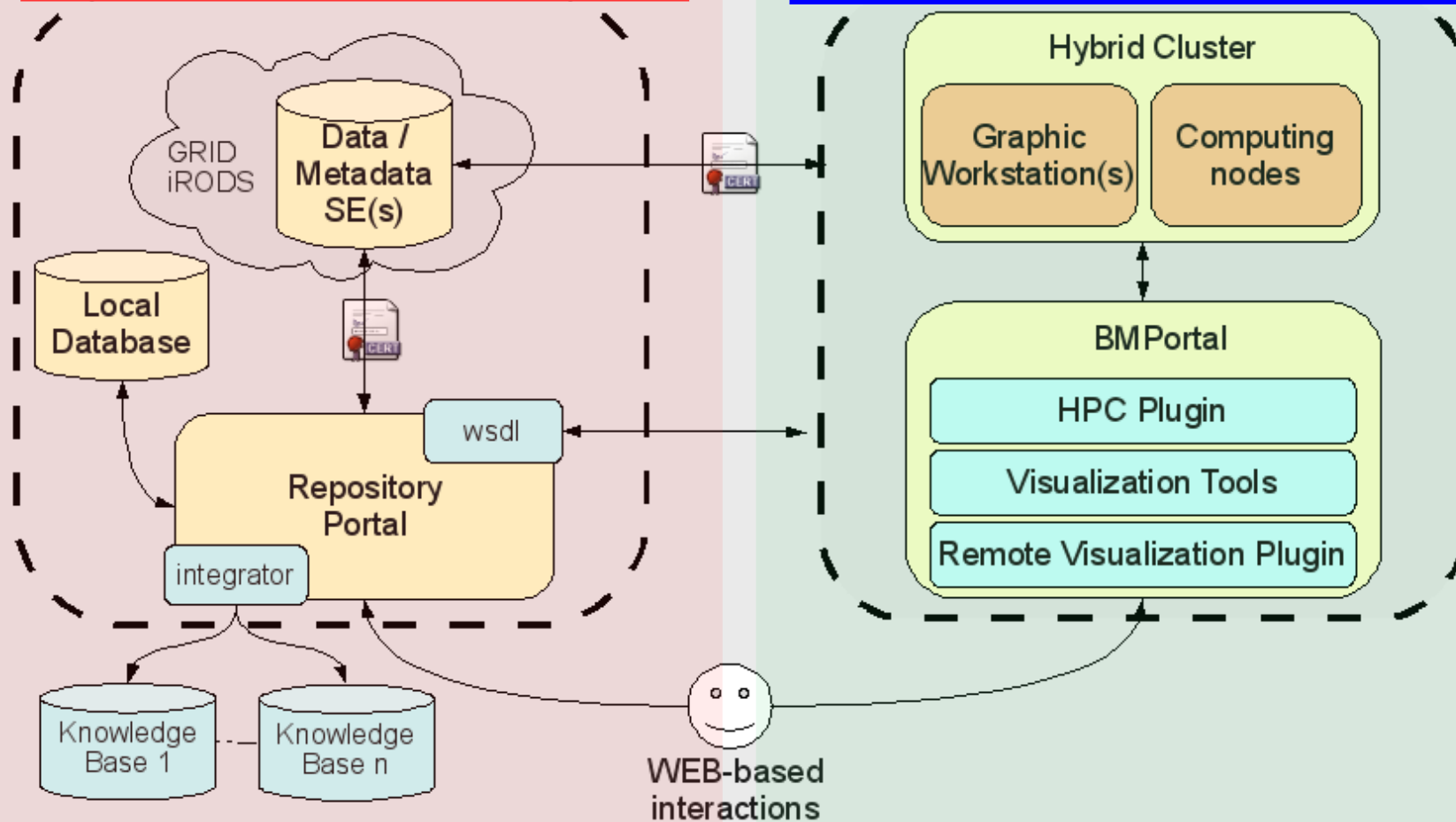
NO pre-existent
data types
Repository can store
anything (almost...)

Automatically created form for **data input**

Architecture

Repository Domain, based on the developed web app

Application Domain, based on EnginFrame technologies



DataGrid: why iRods?

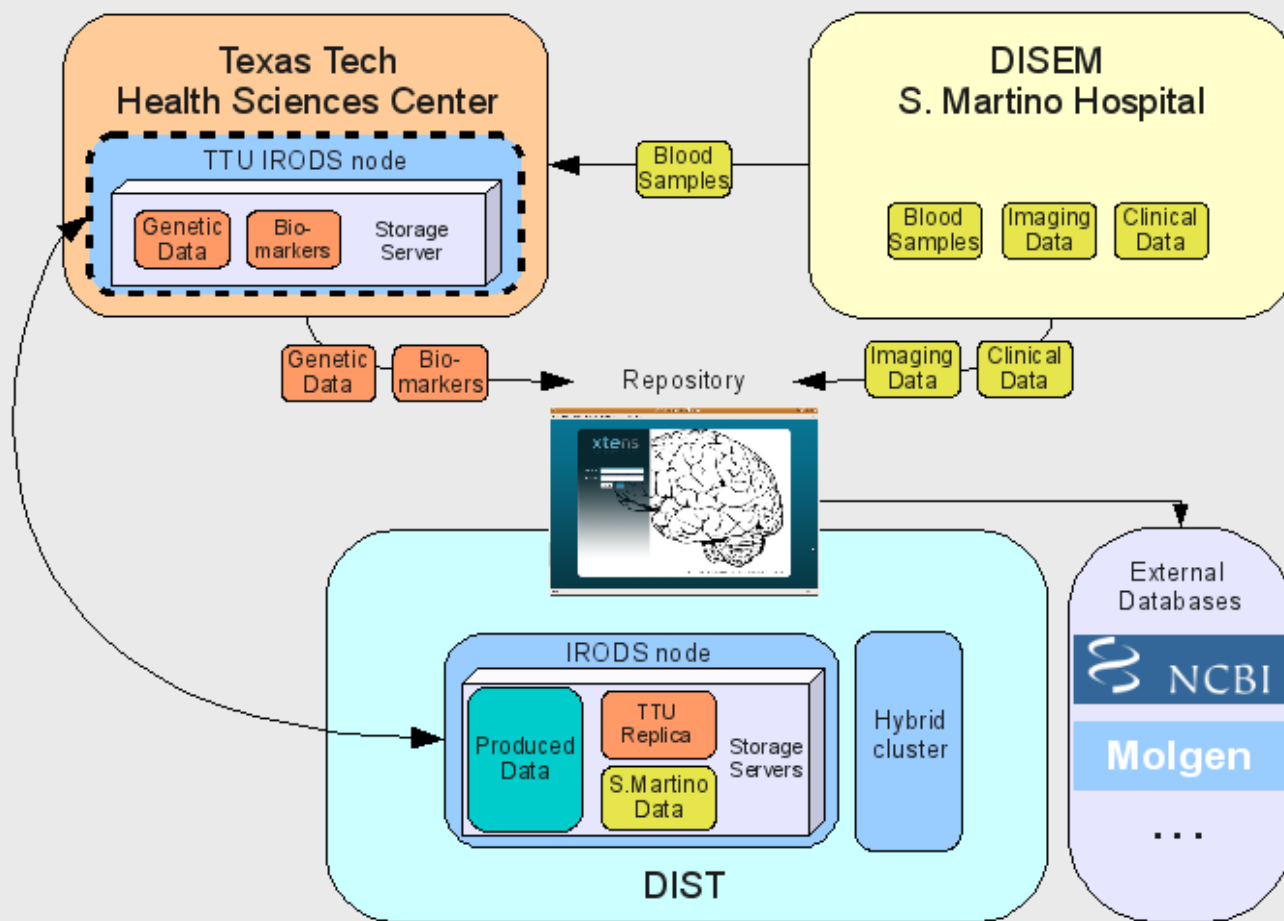
iRODS

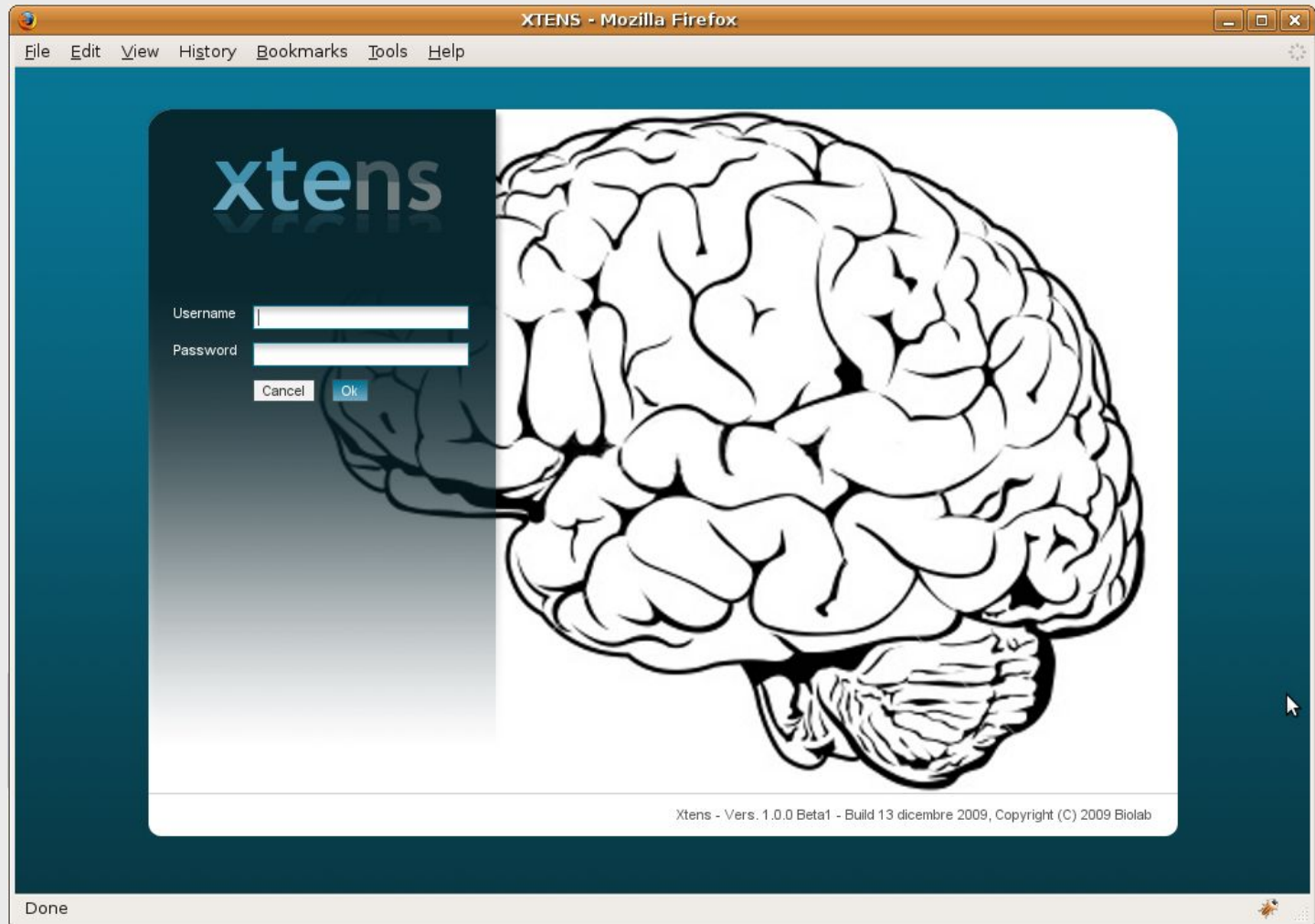
(the Integrated Rule-Oriented Data System (<http://www.irods.org>)

- 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

xtens User: admin

Home Configuration Actions Patients

Groups

Trovati 3 risultati:

#	Description	Actions
1	access control	[icon] [icon]
2	function list	[icon] [icon]
3	function-group page	[icon] [icon]
4	function list	[icon] [icon]

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	[icon] [icon]
<input type="checkbox"/>	/admin/function/	function list	[icon] [icon]
<input type="checkbox"/>	/admin/functionGroup/	function-group page	[icon] [icon]
<input type="checkbox"/>	/admin/functions/	function list	[icon] [icon]
<input type="checkbox"/>	/admin/group/		
<input type="checkbox"/>	/admin/groupFunction/		
<input type="checkbox"/>	/admin/groupOperator/		
<input type="checkbox"/>	/admin/groupPage/		
<input type="checkbox"/>	/admin/newdata/		

Associate with a group: admin [Associate] [New page]

← Groups

← Operators

Pages →

Functions →

xtens User: admin

Home Configuration Actions Patients

Operators

Surname: [input] All Groups [dropdown]
Name: [input] [Search]

New operator

Click to insert a new operator

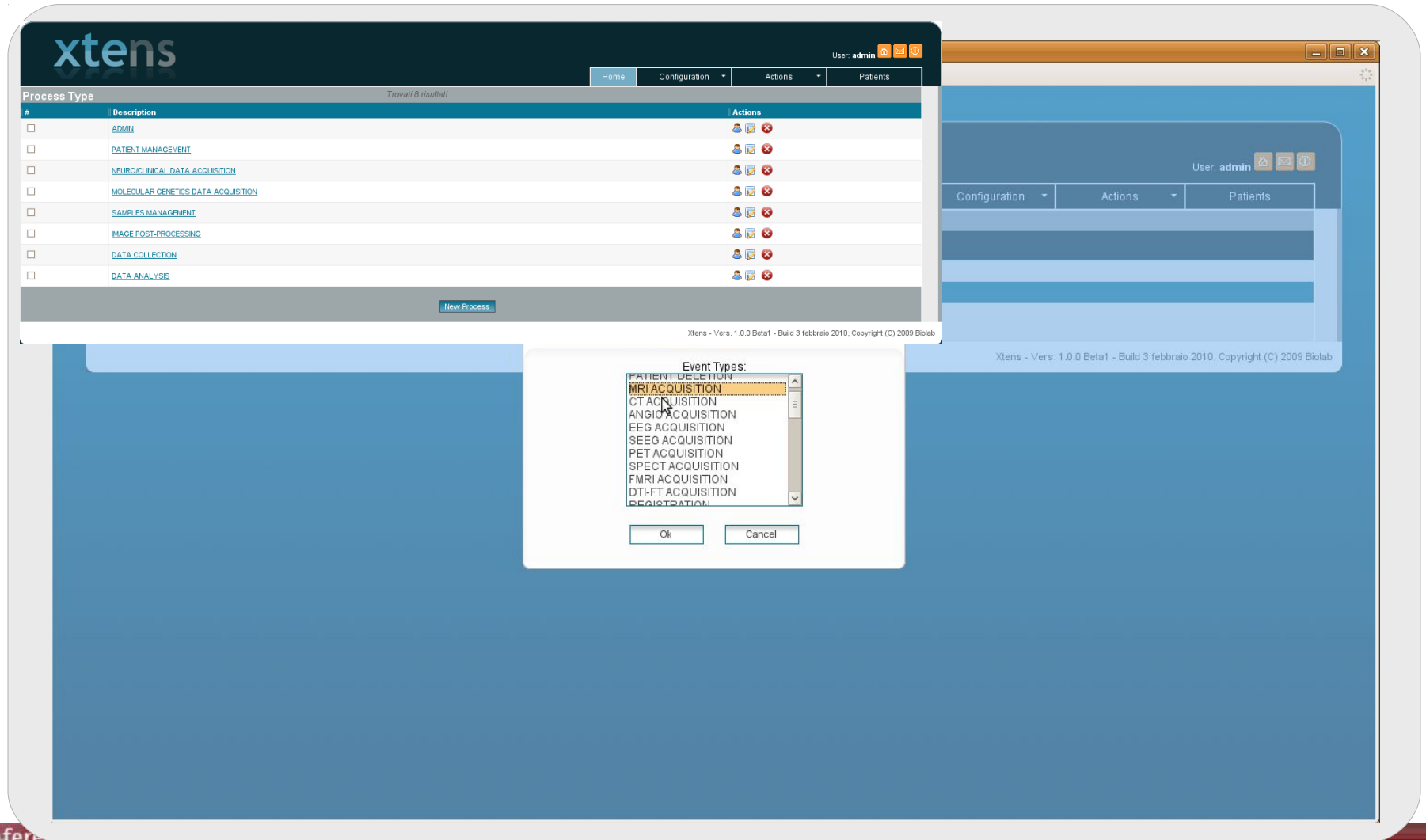
Functions

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

Associate with a group: admin [Associate] [New function]

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

Manage Processes / events



The screenshot displays the xtens web application interface. The main window shows a table of process types with columns for #, Description, and Actions. The table lists various processes such as ADMIN, PATIENT MANAGEMENT, and NEUROCLINICAL DATA ACQUISITION. A 'New Process' button is visible at the bottom of the table. A modal dialog titled 'Event Types:' is open, showing a list of event types including PATIENT DELETION, MRI ACQUISITION, CT ACQUISITION, and others. The dialog has 'Ok' and 'Cancel' buttons. The background shows a blurred view of the application's configuration and actions tabs.

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

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

xtens

User: admin

Home

Configuration

Actions

Patients

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
-------------------	-----------------	---------------------	---------------------------------

Metdata information

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

Link data
to processes

Ontology

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						
		Remove		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						
		Remove		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		

xtens User: admin Home Configuration Actions Patients

Insert Data

Data type selection

Select Data Type: MRI

Upload File

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

Upload one or more files

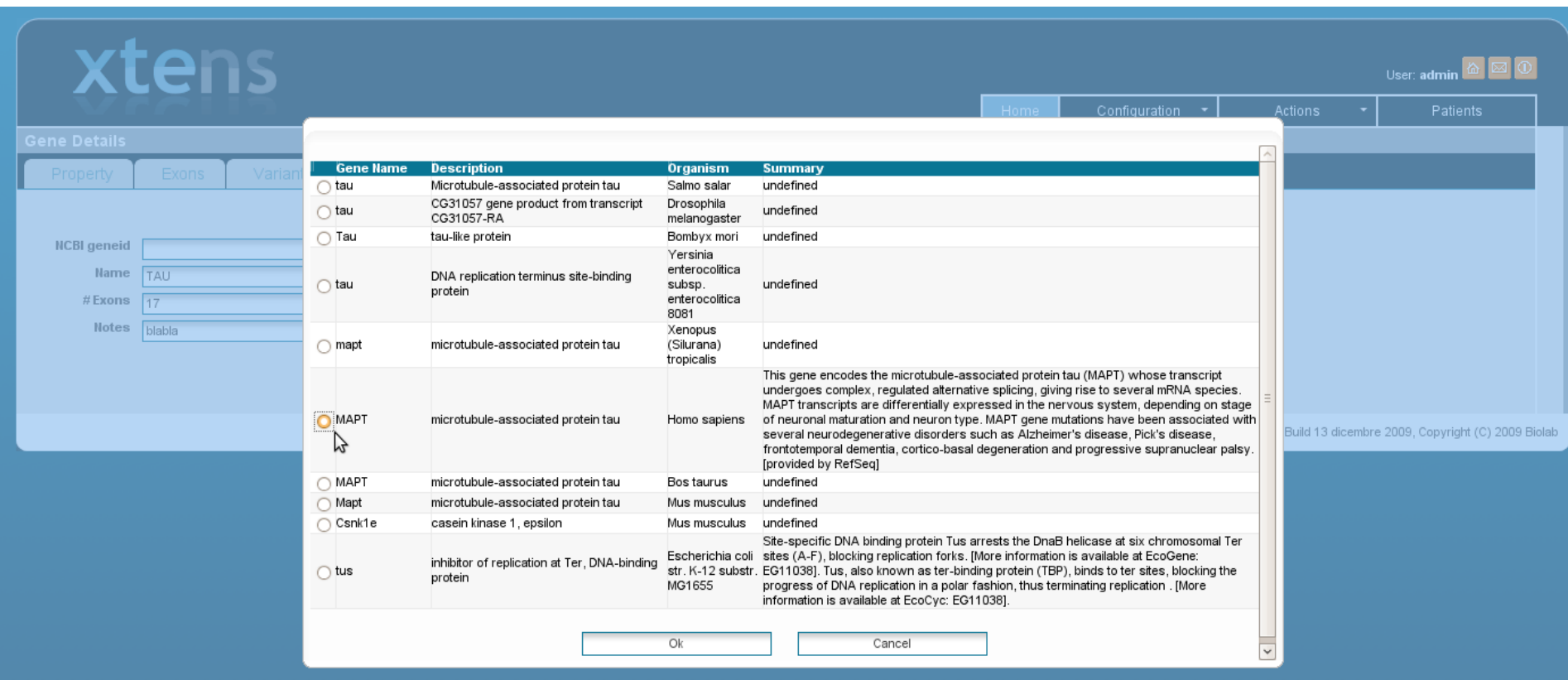
GENERIC INFORMATION

INSTITUTION_NAME	<input type="text"/>
MANUFACTURER	<input type="text"/>
MAGNETIC_FIELD_STRENGTH	<input type="text"/>
IMAGE_TYPE	<input type="text"/>

GEOMETRICAL/RESOLUTION

VOXELRES_X	<input type="text"/>	MM
VOXELRES_Y	<input type="text"/>	MM
VOXELRES_Z	<input type="text"/>	MM
ROWS	<input type="text"/>	
COLUMNS	<input type="text"/>	

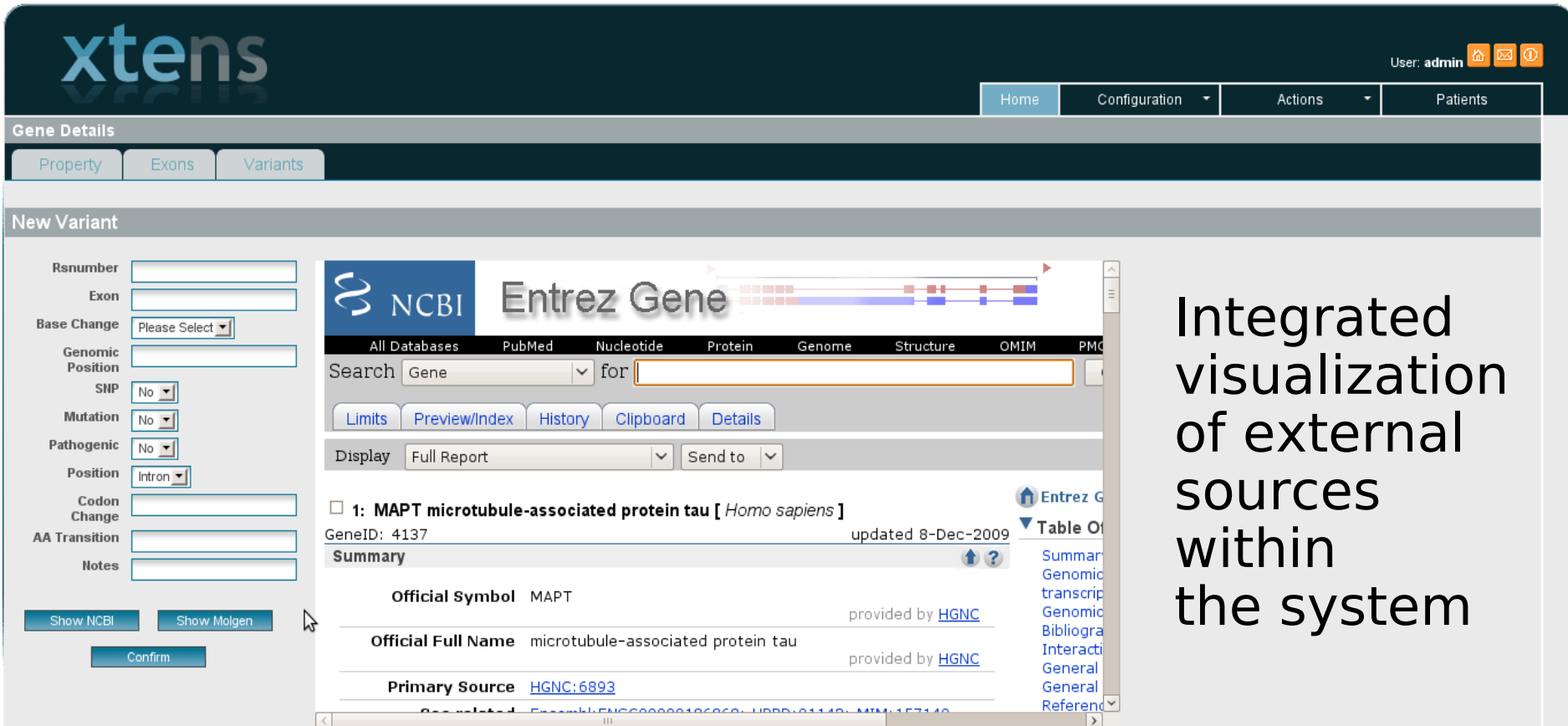
Insertion of metadata defined during data type creation



The screenshot shows the 'xtens' web application interface. A 'Gene Details' window is open, displaying a table of genes. The table has columns for Gene Name, Description, Organism, and Summary. The 'MAPT' gene is highlighted with a mouse cursor. The background shows the main application menu with 'Home', 'Configuration', 'Actions', and 'Patients' options. The user is logged in as 'admin'.

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].

Real time query NCBI to retrieve needed information



The screenshot displays the **xtens** web application interface. At the top, the **xtens** logo is on the left, and the user is logged in as **admin** on the right. A navigation bar includes links for **Home**, **Configuration**, **Actions**, and **Patients**.

The main content area is divided into two sections:

- Gene Details:** This section has tabs for **Property**, **Exons**, and **Variants**.
- New Variant:** This section contains a form for adding a new variant. The form includes fields for:
 - Rnumber** (text input)
 - Exon** (text input)
 - Base Change** (dropdown menu with "Please Select" as the current selection)
 - Genomic Position** (text input)
 - SNP** (dropdown menu with "No" as the current selection)
 - Mutation** (dropdown menu with "No" as the current selection)
 - Pathogenic** (dropdown menu with "No" as the current selection)
 - Position** (dropdown menu with "Intron" as the current selection)
 - Codon Change** (text input)
 - AA Transition** (text input)
 - Notes** (text input)

Below the form, there are buttons for **Show NCBI**, **Show Molgen**, and **Confirm**.

Overlaid on the right side of the interface is a window titled **Entrez Gene** from NCBI. It shows the search results for the gene **MAPT** (microtubule-associated protein tau) in *Homo sapiens*. The window includes a search bar, navigation tabs (All Databases, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC), and a list of search results. The first result is for **1: MAPT microtubule-associated protein tau [Homo sapiens]**, with GeneID: 4137 and updated 8-Dec-2009. The window also displays the official symbol (MAPT), official full name (microtubule-associated protein tau), and primary source (HGNC:6893).

Integrated
visualization
of external
sources
within
the system

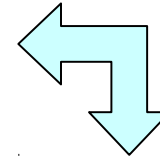
Sample List

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		
<input checked="" type="radio"/>	001-A729-04	X2523	BLOOD	09/03/2009	10/02/2009		
<input type="radio"/>	001-A755-09	X2528	BLOOD				11/09/2009
<input type="radio"/>	001-A802-02	X2529	BLOOD				



Details

Project: GENOVA

Sample Notes:

Fridge Location: ROOM 2

Temperature: -80 °C

Shelf:

Metal Rack:

Floor:

Box:

Box: BX789

001-A729-04 - Hole n°42

User: admin

Home Configuration Actions Patients

New Sample

Type:

Type	Fedex
BLOOD	<input type="text"/>
RAIN TISSUE	<input type="text"/>
ERUM	<input type="text"/>
VA	<input type="text"/>

ct:

le Notes:

e Location: ROOM 2

erature: -80 °C

Shelf:

Metal Rack:

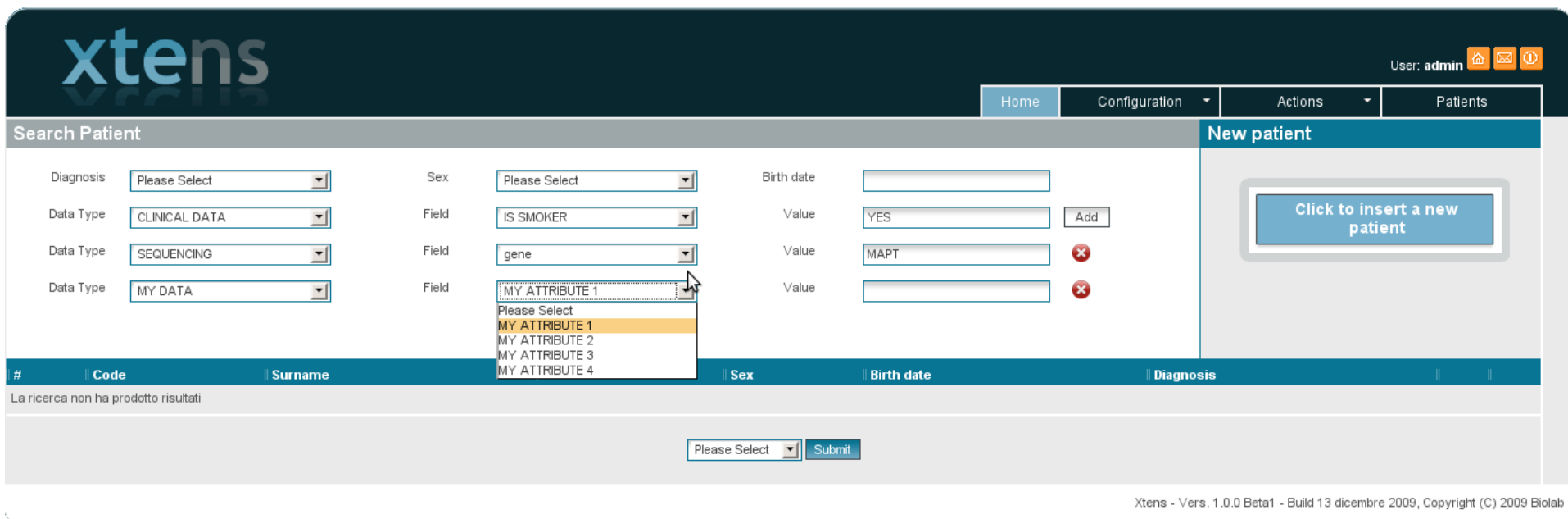
Floor:

Box:

Box: BX789

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


Smart search engine



The screenshot shows the 'xtens' web application interface. At the top, there's a navigation bar with 'Home', 'Configuration', 'Actions', and 'Patients' tabs. The user is logged in as 'admin'. Below the navigation bar, there's a 'Search Patient' section with several dropdown menus for 'Diagnosis', 'Data Type', 'Sex', 'Field', and 'Birth date'. The 'Field' dropdown is currently open, showing options like 'MY ATTRIBUTE 1', 'MY ATTRIBUTE 2', 'MY ATTRIBUTE 3', and 'MY ATTRIBUTE 4'. To the right of the search fields, there's a 'New patient' section with a button that says 'Click to insert a new patient'. Below the search fields, there's a table with columns for '#', 'Code', 'Surname', 'Sex', 'Birth date', and 'Diagnosis'. The table is currently empty, and a message below it says 'La ricerca non ha prodotto risultati'. At the bottom of the search section, there's a 'Please Select' dropdown and a 'Submit' button. The footer of the application shows '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


User: admin

[Home](#)
[Configuration](#)
[Actions](#)
[Patients](#)

Search Patient

Diagnosis

Please Select

Sex

Please Select

Birth date

Data Type

Please Select

Field

Please Select

Value

Add

Search

New patient

Click to insert a new patient

Trovati 6 risultati.

#	Code	Surname	Name	Sex	Birth date	Diagnosis	
	CNTGLI37L4	C	G	F	02/07/1937	AL ZHEIMER	Clinical Data Molecular Genetics Samples Associate Sample History
	CRTLCU81D	C	M	M	02/07/1937	AL ZHEIMER	Clinical Data Molecular Genetics Samples Associate Sample History
	PGGMRZ85H	P	G	M	02/06/1985	NON AMNESTIC MCI	Clinical Data Molecular Genetics Samples Associate Sample History
	RSSGRG69D	R	G	M	09/04/1969	CONTROL	Clinical Data Molecular Genetics Samples Associate Sample History
	RSSMRA81D	R	M	M	14/04/1981	AMNESTIC MCI	Clinical Data Molecular Genetics Samples Associate Sample History
	RZZSMN77L	R	S	F	02/07/1977	NON AMNESTIC MCI	Clinical Data Molecular Genetics Samples Associate Sample History

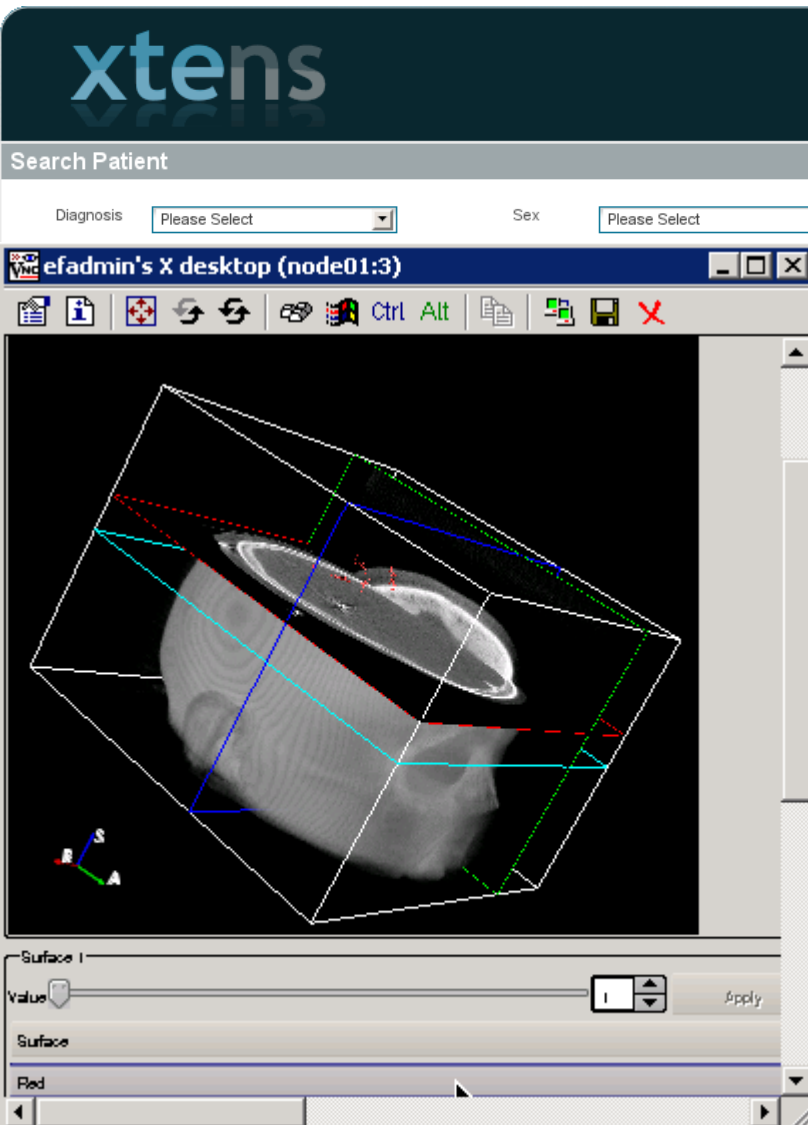
Clinical Details

Clinical Data	NeuroPsychological Tests	Data List																		
<h4>Clinical Data</h4> <table> <thead> <tr> <th>Field</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>USE OF ALCOHOL</td> <td>NO</td> </tr> <tr> <td>IS SMOKER</td> <td>YES</td> </tr> <tr> <td>IS SMOKER DATE</td> <td></td> </tr> <tr> <td>EX SMOKER</td> <td>YES</td> </tr> <tr> <td>DIABETES</td> <td>YES</td> </tr> <tr> <td>HIGH BLOOD PRESSURE</td> <td>YES</td> </tr> <tr> <td>HIGH BLOOD PRESSURE DATE</td> <td></td> </tr> <tr> <td>HIGH CHOLESTEROL</td> <td>NOT DECLARED</td> </tr> </tbody> </table>			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
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	MAPT	Known Variants	Found Variants																														
<h4>Haplotype</h4> <table> <thead> <tr> <th>ACE Haplotype</th> <th>MAPT Haplotype</th> <th>APOe</th> </tr> </thead> <tbody> <tr> <td>rs16832753 - C>G</td> <td>H1/H1</td> <td>E2/e3</td> </tr> </tbody> </table>				ACE Haplotype	MAPT Haplotype	APOe	rs16832753 - C>G	H1/H1	E2/e3																								
ACE Haplotype	MAPT Haplotype	APOe																															
rs16832753 - C>G	H1/H1	E2/e3																															
<h4>Known Variants</h4> <table> <thead> <tr> <th>Exon</th> <th>Rsnumber</th> <th>Gen Pos</th> <th>B Change</th> <th>SNP</th> <th>Mutation</th> <th>Pathogenic</th> <th>Position</th> <th>C Change</th> <th>AA Trans</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>rs63750529</td> <td>g.75859G</td> <td>G>A</td> <td>NO</td> <td>YES</td> <td>NO</td> <td>EXON</td> <td>From ACG to ACA</td> <td>-</td> </tr> <tr> <td>1</td> <td>rs63750959</td> <td>g.72930G</td> <td>G>T</td> <td>NO</td> <td>NO</td> <td>NO</td> <td>EXON</td> <td>From CGC to CTC</td> <td>-</td> </tr> </tbody> </table>				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	-
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	-																								
<h4>Found Variants</h4> <table> <thead> <tr> <th>Exon</th> <th>Rsnumber</th> <th>Gen Pos</th> <th>B Change</th> <th>SNP</th> <th>Mutation</th> <th>Pathogenic</th> <th>Position</th> <th>C Change</th> <th>AA Trans</th> </tr> </thead> <tbody> <tr> <td>5</td> <td>-</td> <td>g.92434C</td> <td>C>G</td> <td>NO</td> <td>YES</td> <td>NO</td> <td>EXON</td> <td>-</td> <td>-</td> </tr> </tbody> </table>				Exon	Rsnumber	Gen Pos	B Change	SNP	Mutation	Pathogenic	Position	C Change	AA Trans	5	-	g.92434C	C>G	NO	YES	NO	EXON	-	-										
Exon	Rsnumber	Gen Pos	B Change	SNP	Mutation	Pathogenic	Position	C Change	AA Trans																								
5	-	g.92434C	C>G	NO	YES	NO	EXON	-	-																								

Browse results and look at both clinical and genetic data



“in line” visualization for simple data

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