biowep: a workflow enactment portal for bioinformatics applications

Romano P.[1], Bartocci E.[2], Bertolini G.[3], De Paoli F.[3], Marra D.[1],
Mauri G.[3], Merelli E.[2] and Milanesi L.[4,5]

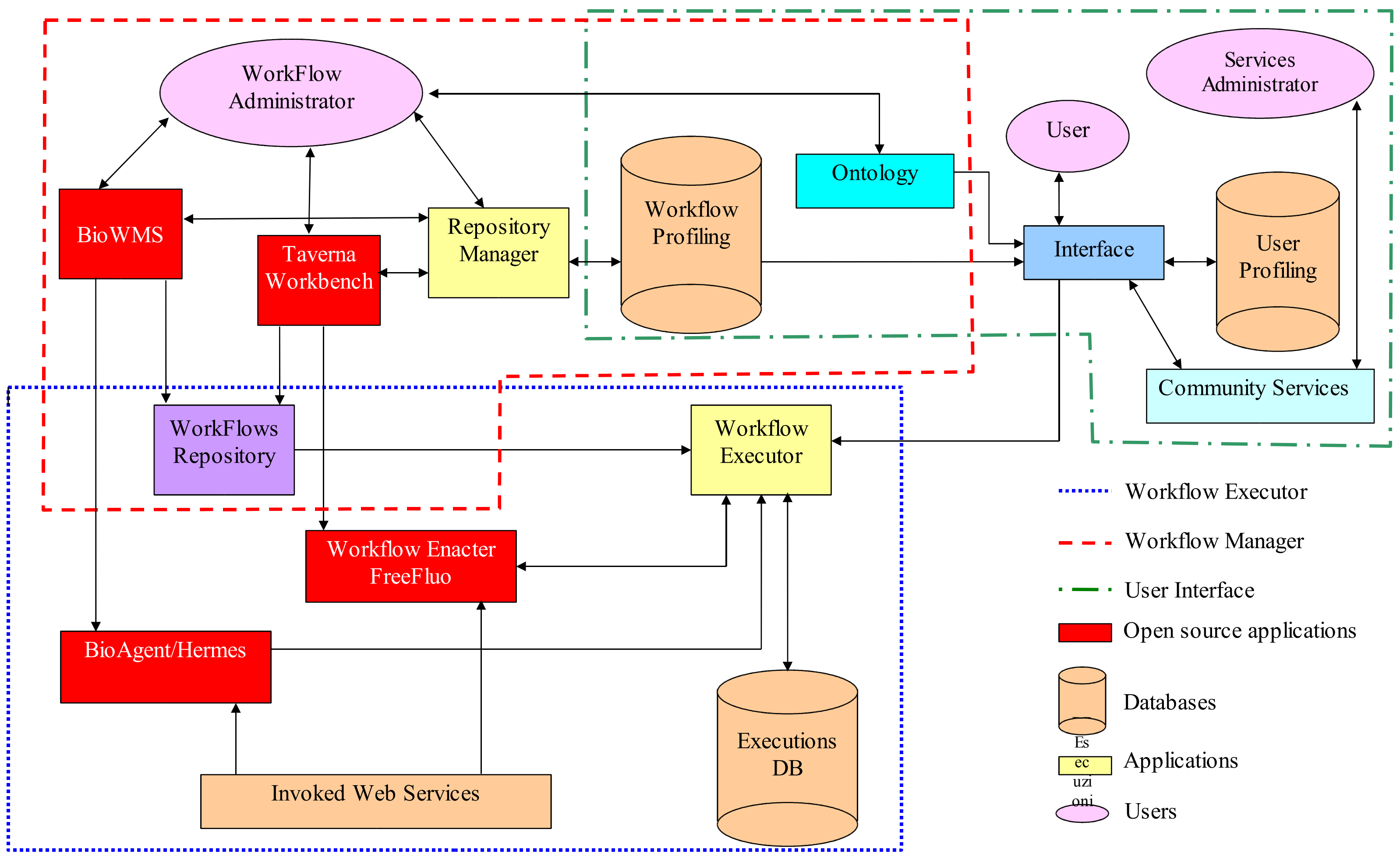
[1]National Cancer Research Institute, Genoa, [2]University of Camerino, Camerino (MC),
[3]University of Milan Bicocca, Milan, [4]National Research Council, Milan, [5]CILEA, Segrate (MI)

Rationale and goals

The distribution over the Internet of biological information and the heterogeneity of software tools that are used in bioinformatics makes the adoption of new data integration and analysis network tools a necessity. New ICT tools, like Web Services (WS) and Workflow Management Systems (WMS), can support the creation and deployment of such systems. Many WS are already available, as well as tools for implementing them and some WMS have been proposed in bioinformatics, the Taverna Workbench from EBI being the most known. They assume that end users know which bioinformatics resources can be reached through a programatic interface and that they are skilled in programming and in building workflows. A portal enabling unskilled researchers to take profit from these new technologies is still missing. We present here a user-friendly web system that can support selection and execution of a set of predefined workflows, thus simplifying access to all end users.

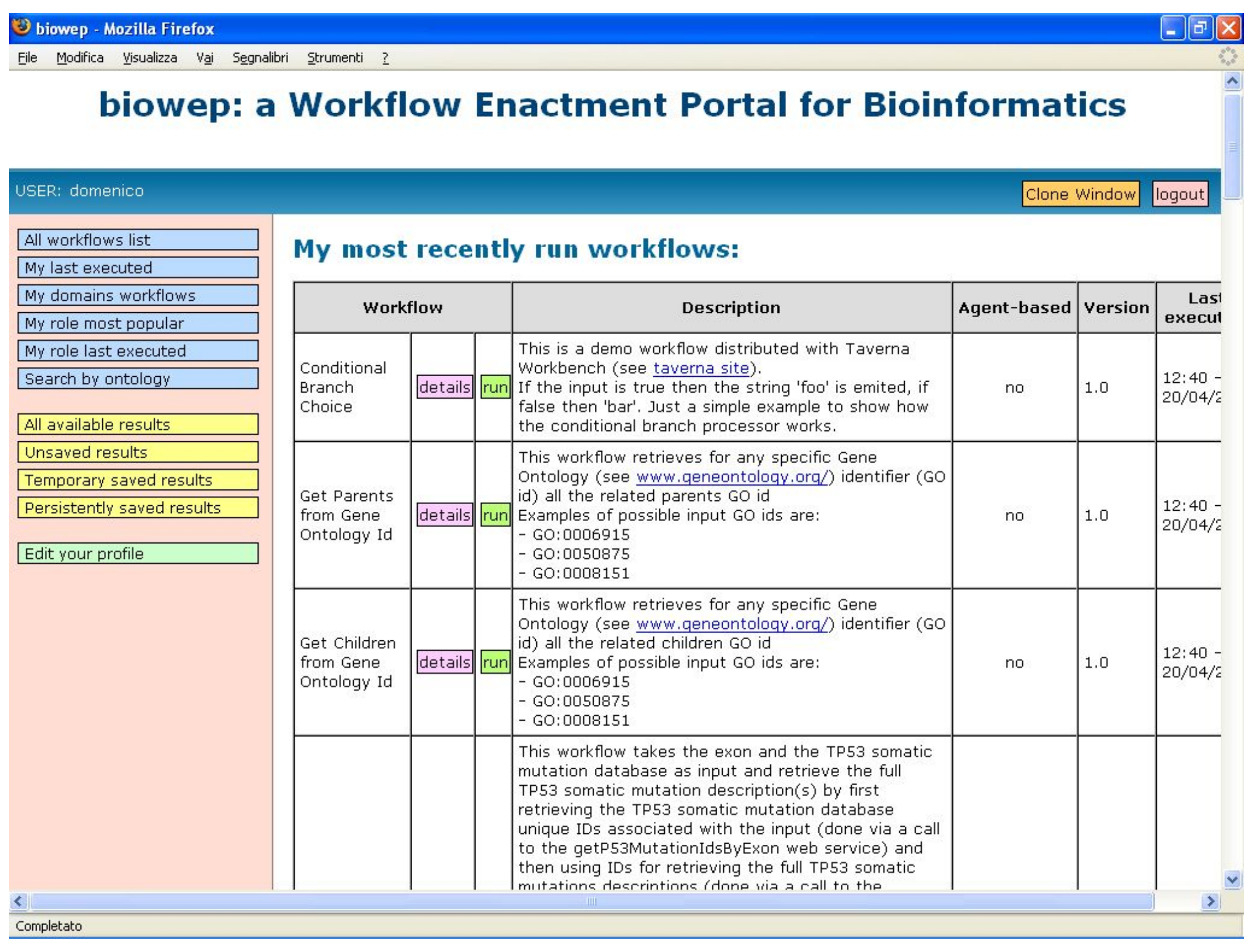
biowep architecture

- a Workflow Manager (WM), for the creation and annotation of workflows by using either the Taverna Workbench or the BioWMS. Input and output data, elaboration type and application domain of main steps of each workflow are annotated by using a classification of bioinformatics data and tasks.
- a User Interface (UI), for authentication and profiling of end users, selection of workflows from lists or by searching their annotation, management of results.
- a Workflow Executor (WE), for controlling the enactment of workflows that is carried out either by FreeFluo or by BioAgent/Hermes, a mobile agent-based middleware.



Results

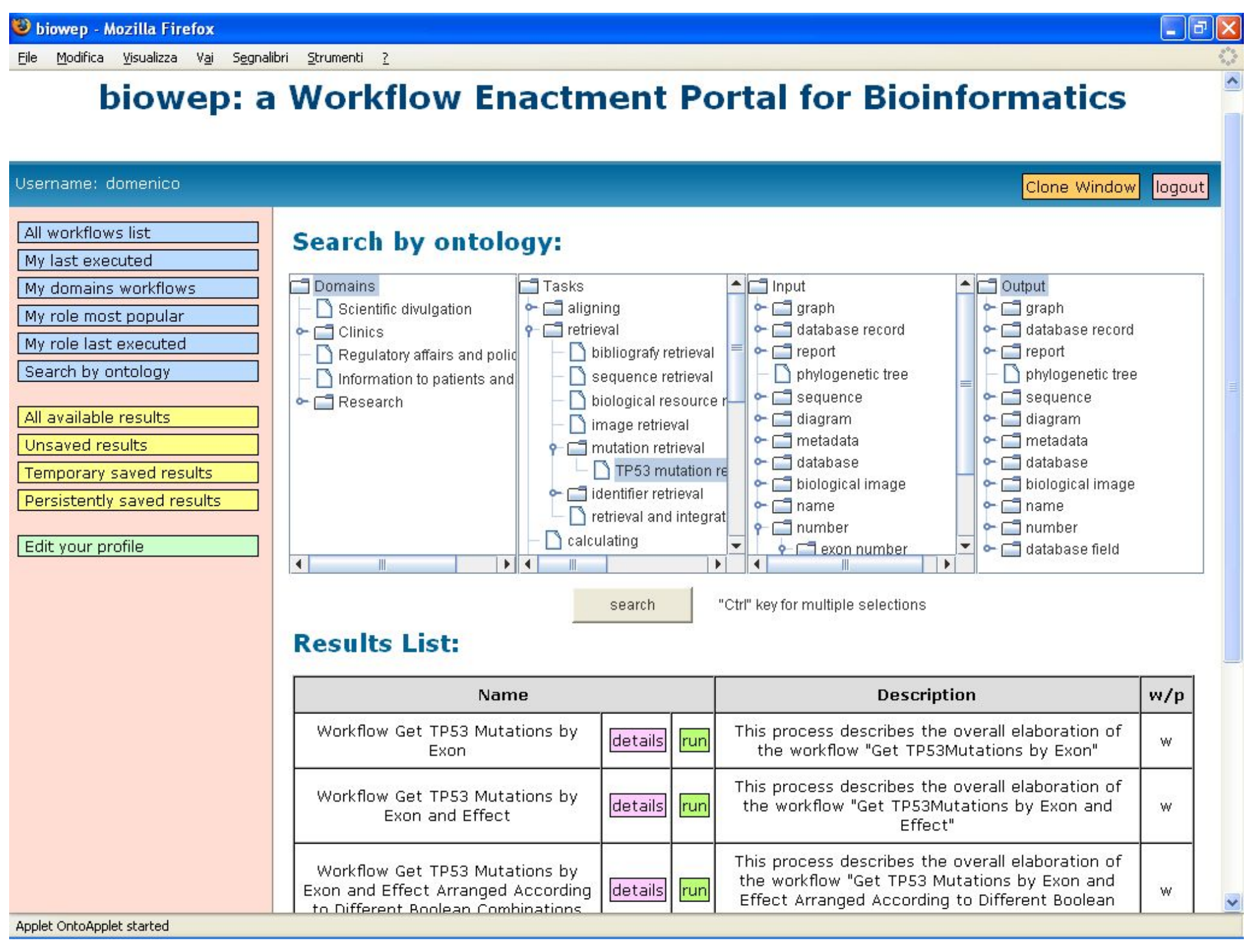
We designed a web based client application that allows for the selection and execution of a set of predefined, annotated workflows. A prototype system is available on-line for testing purposes. It includes workflows that are devoted to the retrieval of data from IARC TP53 Mutation Database and from CABRI biological resources catalogues. Some of them have been made available both in Taverna and in BioWMS formats.



biowep: a Workflow Enactment Portal for Bioinformatics

My most recently run workflows:

Workflow			Description	Agent-based	Version	Last execut
Conditional Branch Choice	details	run	This is a demo workflow distributed with Taverna Workbench (see taverna site). If the input is true then the string 'foo' is emited, if false then 'bar'. Just a simple example to show how the conditional branch processor works.	no	1.0	12:40 - 20/04/2
Get Parents from Gene Ontology Id	details	run	This workflow retrieves for any specific Gene Ontology (see www.geneontology.org/) identifier (GO id) all the related parents GO id Examples of possible input GO ids are: - GO:0006915 - GO:0050875 - GO:0008151	no	1.0	12:40 - 20/04/2
Get Children from Gene Ontology Id	details	run	This workflow retrieves for any specific Gene Ontology (see www.geneontology.org/) identifier (GO id) all the related children GO id Examples of possible input GO ids are: - GO:0006915 - GO:0050875 - GO:0008151	no	1.0	12:40 - 20/04/2
			This workflow takes the exon and the TP53 somatic mutation database as input and retrieve the full TP53 somatic mutation description(s) by first retrieving the TP53 somatic mutation database unique IDs associated with the input (done via a call to the getP53MutationIdsByExon web service) and then using IDs for retrieving the full TP53 somatic mutations descriptions (done via a call to the			

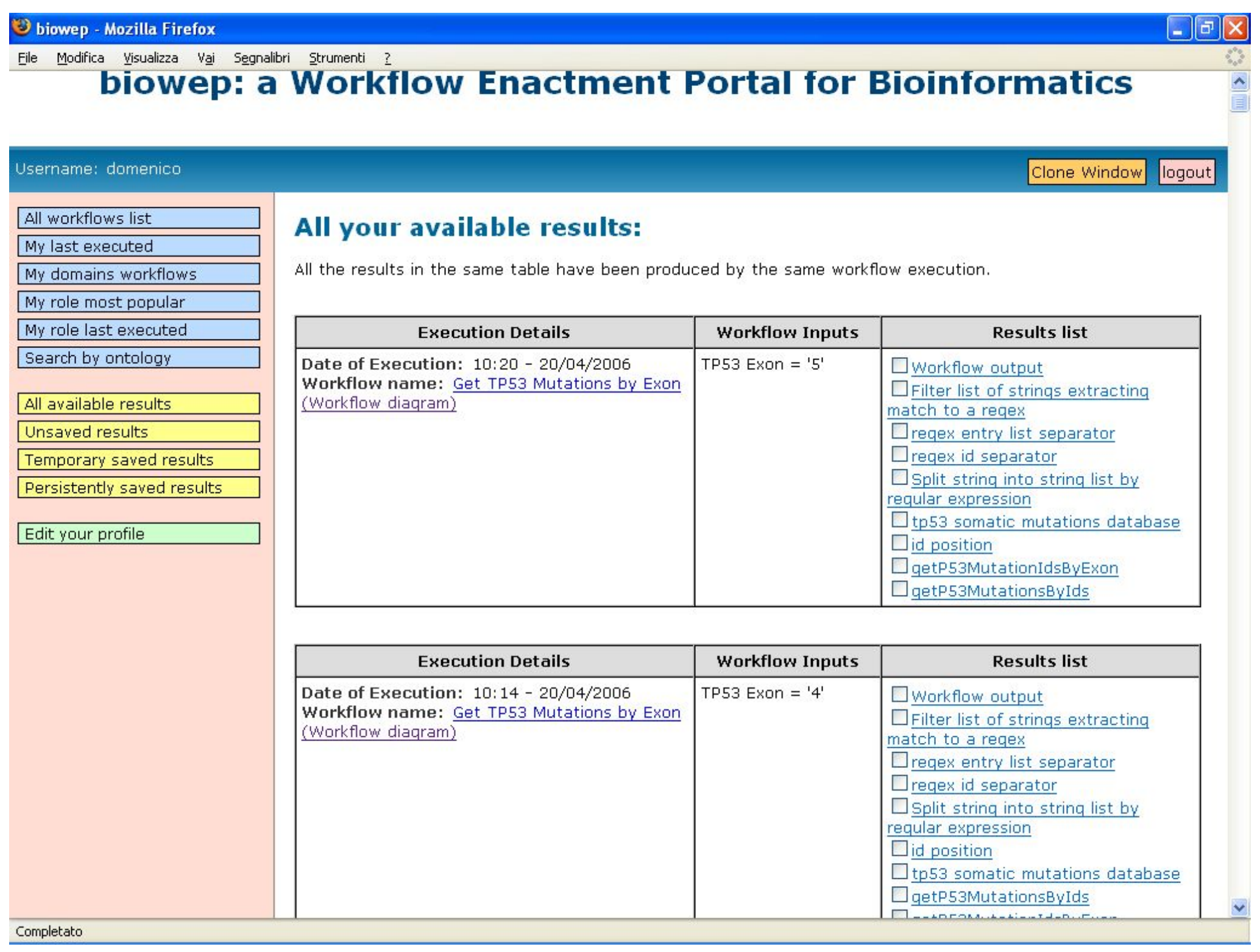


biowep: a Workflow Enactment Portal for Bioinformatics

Search by ontology:

Results List:

Name			Description	w/p
Workflow Get TP53 Mutations by Exon	details	run	This process describes the overall elaboration of the workflow "Get TP53Mutations by Exon"	w
Workflow Get TP53 Mutations by Exon and Effect	details	run	This process describes the overall elaboration of the workflow "Get TP53Mutations by Exon and Effect"	w
Workflow Get TP53 Mutations by Exon and Effect Arranged According to Different Boolean Combinations	details	run	This process describes the overall elaboration of the workflow "Get TP53 Mutations by Exon and Effect Arranged According to Different Boolean	w



biowep: a Workflow Enactment Portal for Bioinformatics

All your available results:

All the results in the same table have been produced by the same workflow execution.

Execution Details	Workflow Inputs	Results list
Date of Execution: 10:20 - 20/04/2006 **Workflow name:** Get TP53 Mutations by Exon (Workflow diagram)	TP53 Exon = '5'	☐ Workflow output ☐ Filter list of strings extracting match to a regex ☐ regex entry list separator ☐ regex id separator ☐ Split string into string list by regular expression ☐ tp53 somatic mutations database ☐ id position ☐ getP53MutationIdsByExon ☐ getP53MutationsByIds

Execution Details	Workflow Inputs	Results list
Date of Execution: 10:14 - 20/04/2006 **Workflow name:** Get TP53 Mutations by Exon (Workflow diagram)	TP53 Exon = '4'	☐ Workflow output ☐ Filter list of strings extracting match to a regex ☐ regex entry list separator ☐ regex id separator ☐ Split string into string list by regular expression ☐ id position ☐ tp53 somatic mutations database ☐ getP53MutationsByIds

Conclusions

The development and implementation of WS allowing the access to an exhaustive set of biomedical databases and analysis software and the creation of effective workflows through widely distributed WMS can significantly improve automation of in-silico analysis. biowep is available for interested researchers as a reference portal. They are invited to submit their workflows for insertion in the workflow repository.

Supplementary Information

Availability: Our system is partially based on open source. biowep is itself available under the GNU Lesser General Public Licence (LGPL). biowep is further being developed in the sphere of the Laboratory of Interdisciplinary Technologies in Bioinformatics.

Contacts: Paolo Romano (paolo.romano@istge.it)

Tools:

Biowep: http://bioinformatics.istge.it:8080/biowep/
Taverna Workbench: http://taverna.sourceforge.net/
BioWMS: http://litbio.unicam.it:8080/biowms/ (demo under testing)
BioAgent: http://www.bioagent.net/
Hermes: http://hermes.cs.unicam.it/

Acknowledgements
This work was partially supported by the Italian Ministry of Education, University and Research MIUR, projects "Oncology over Internet (O2I)" and "Laboratory of Interdisciplinary Technologies in Bioinformatics (LITBIO)".