A Distributed repository implementation for molecular dynamics learning objects

Sergio Tasso[1], Simonetta Pallottelli[1], Antonio Laganà[2] and Marina Rui[3]

[1] Department of Mathematics and Computer Science, University of Perugia
via Vanvitelli, 1, I-06123 Perugia, Italy
{ sergio, simona }@unipg.it
[2] Department of Chemistry, University of Perugia
via Elce di Sotto, 8, I-06123 Perugia, Italy
lagana05@gmail.com
[3] Department of Chemistry and Industrial Chemistry, University of Genoa
via Dodecaneso, 31, I-16146 Genova, Italy
marina@chimica.unige.it

Abstract

The present report illustrates the implementation of a distributed repository tailored on the needs of research based education activities. The distributed repository architecture and the related software environment (tools and requirements) are described. A use case dealing with a course on Scattering and Molecular Dynamics and its applications to materials technologies is also discussed.

1. Introduction

Among the activities undertaken by the European Chemistry Thematic Network Association (ECTNA) [1] in order to enhance research-based education through its Virtual Education Community (VEC) are those on distributed repositories. The present report illustrates the implementation of a distributed repository approach for the Chemistry, Molecular and Materials Science and Technology (CMMST) field with particular concern for Scattering and Molecular Dynamics. This endeavor has been undertaken for the first time in the year 2009 within the "A virtual campus for chemistry and chemical engineering" workpackage of the EC2E2N project [2] and then continued as a regular activity of the VEC standing committee of ECTNA.

Along this direction the first step has consisted in singling out a common European syllabus as a result of a detailed analysis of the contents of the different national curricula. On the ground of such syllabus some guidelines for the production of the virtual campus e-learning material, the elaboration of the accreditation mechanisms of the higher education institutions, the development of the procedures needed for carrying out internet based self-evaluation sessions as well as various other auxiliary instruments were worked out. Further components of the virtual campus are still at their designing and prototyping phase as from a roadmap set by the VEC committee.

The main effort of our work was devoted to the assemblage of a distributed repository aimed at easily storing, identifying, localizing and reusing CMMST educational materials. The re-use of the educational material on a distributed repository bears the advantage of inducing a virtuous cycle when it is accompanied by the possibility of improving the used material. In this spirit, part of the teaching and learning materials developed by the VEC members has been packed into units (called Learning Objects or shortly LOs) of the distributed repository. These LOs not only represent a well-defined self-comprehensive topic of the syllabus but are also entirely multimedia and designed for networked usage.

The other important effort of our work was indeed the grafting of the LOs on the European Grid Infrastructure (EGI). The resulting product, called GLOREP (Grid Learning Object REPository) [3][4], is a management system of Learning objects based on a Client/server communication paradigm. A CMMST GLOREP application is discussed in detail in the present report by pinpointing the fact that the assembled LOs are entirely managed on line with the possibility of running related packages on the GRID for research based educational activities.

In section 2 some models of the GLOREP federation are described;
In section 3 a description of the GLOREP environment is made;
In section 4 a list the new tools developed is given;
In section 5 the GLOREP modules are illustrated;
In section 6 the GLOREP requirements are analysed;
In section 7 a molecular dynamics use case is discussed.

2. The GLOREP federation

GLOREP is a distributed management system of LOs that we have developed using a Client/server communication paradigm. It consists of the following components:

- A server bearing a CMS (Drupal [5]) offering the needed repository management activities at back-end-level (like backup, protection, access control) and providing, through a web portal, various services for clients (like up/download LOs, file management) at front-end level.
- A set of clients requiring the services offered by the server.
- A network (of the internet type as is EGI) allowing clients to use available facilities after authentication.
- A virtual Organization (like COMPCHEM) providing access to remote file systems where the LOs are stored.

In building G-LOREP the Drupal CMS was adopted because of its access and management easiness. Furthermore, Drupal does not require high-level programming skills and simplifies the handling of multimedia and heterogeneous material. It allows also the exchange of data information with various other CMSs. As a preliminary implementation, we adopted the centralized architecture of the Federated Repository shown in Fig.1. In this architecture, the elements of the Federated Repository cooperate to manage and fetch the LOs using an Index Registry inside a root member. The bottle-neck of such architecture is given by the root server that keeps track of content creation and modification enabling as well the search facilities on the rest of the federation.
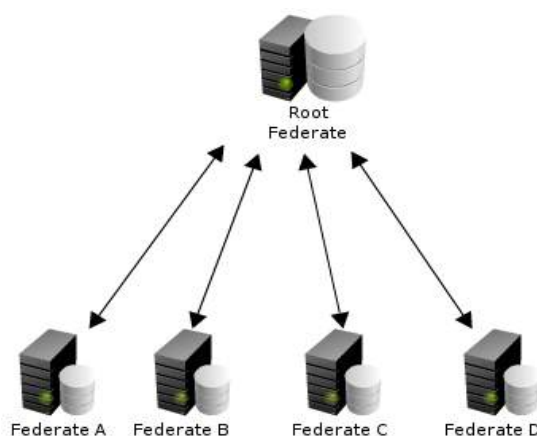

Fig.1 The centralized architecture

The second implementation of GLOREP adopted the peer distributed architecture (shown in Fig. 2) in which each node is both a client and a server. All servers are able to perform all operations with the same power and privilege. In this decentralized architecture, a malfunctioning or network loss would not prevent the federation from working because other federated members could provide a basic set of functions to users. A disadvantage of this type of architecture, however, is the complexity of the procedure to be adopted to keep homogeneous the knowledge basis of the federation on the servers.
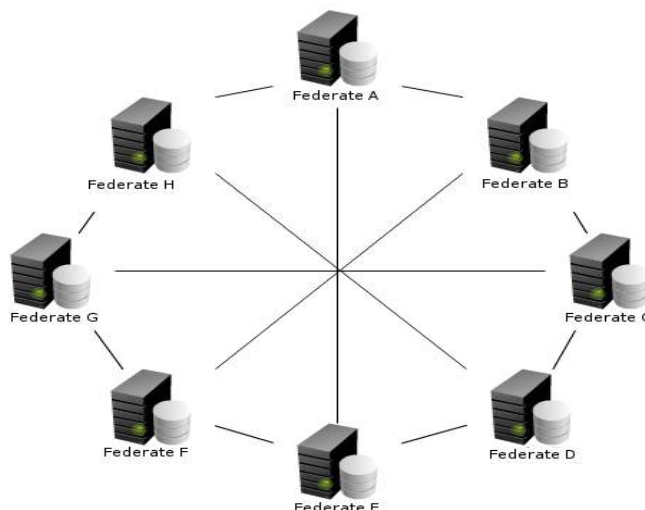
Fig.2 The peer architecture

Finally the most recently adopted G-LOREP architecture is the hybrid one illustrated in Fig.3 [6]. In this architecture no server has privileges or special abilities. A shared database accessible by all servers stores a map of the federation and an index of its content. In this way, the servers are peer entities, but they have a primary data source for external data. The database contains the identifiers and addresses of all servers and the information about their status. The database also contains the metadata of all the objects belonging to the federation. The shared database is directly accessed and update by all servers. When a new server joins the federation, it receives the information on how to access the shared database necessary to download the complete status of the federation and all LOs metadata.
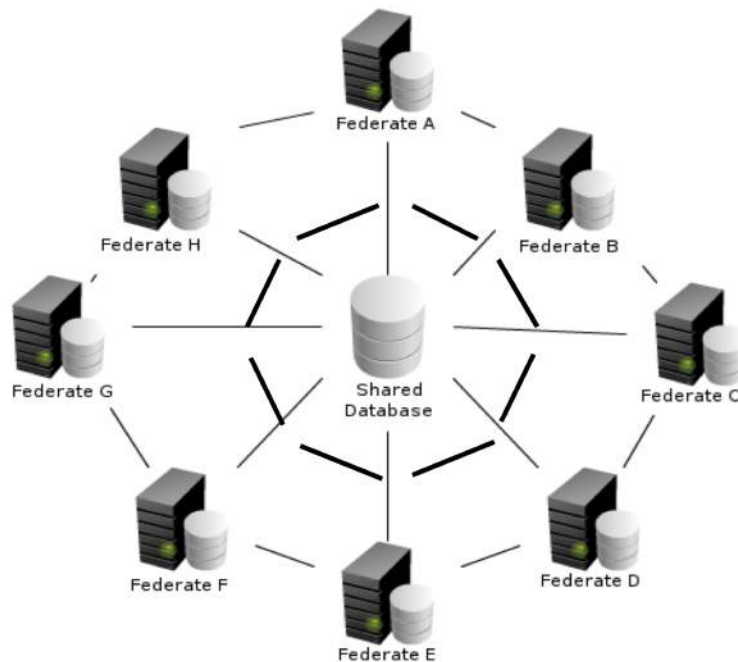

Fig.3 The adopted Architecture

The adoption of a shared database is crucial for the increase of the federation performances. However, it plays only a data storage passive role. If the database goes down, the federation-related changes are not forwarded until the shared database is up again and no new server is allowed to join the federation. Yet, the repositories continue to work because all the functionalities are available on the servers and most of the shared information is replicated inside them and each server works with both local and remote data.

An important feature of the hybrid platform is its scalability. Each server can, in fact, receive participation requests from new members wishing to join the federation. To cope with this type of requests the new server needs to be validated. If the federation is open, the wishing-to-join server will be asked to provide the federation authentication data and after its verification, the federation will accept the new server. When a new server joins the federation, it is provided with data necessary to connect to the shared database. Thanks to this, it will be able to get a complete picture of the status of the federation. Then it will execute a bootstrap procedure whose purpose is to download remote metadata and cache files to the new server and update the shared database and the other servers with the content provided by the newcomer.

3. The GLOREP environment

When all servers and the shared database contain the same knowledge basis of the federation, they are considered as synchronized (or shortly "synch"). An obvious goal of the software is to reach and maintain the synch status between the servers. In order to keep the federation in synch, every node operation (like creation, deletion, or modification) is then mirrored on the shared database and notified to all servers.

In order to manage the contents we introduced a new Drupal node type called Linkable Object.

To recognize a Linkable Object inside the federation we use a FUID (Federation Unique Identifier). When an object is created, the system sets its metadata containing FUID, title, description, optional Software Attachment (SA), etc. Then the linkable object metadata are sent to the shared database and finally to the federation. When a server receives a notification that a new Linkable Object has been created, it generates a new node representing the remote Linkable

Object.

In addition, a dependency management was introduced: from a general point of view, LOs can be not only simple slides and text or pdf files, but they can be more complex objects like 3d virtual worlds and uncommon file formats which might need special software tools to be visualized by the end user. Dependencies allow the creation of a relationship between the LO and a software attachment able to tie them together. In this way, the system will ensure that the user always downloads the required software attachments when requesting LOs from the repository.

The non-automatic tagging of a LO is a long, costly, and error prone process. For this reason the adoption of metadata standards and related automatic tagging procedures is extremely important. During the last few years, various open metadata standards have become popular (e.g. IMS [7], SCORM Dublin Core [8] and IEEE LOM [9]). After a careful analysis we opted for the IEEE Learning Object Metadata (IEEE LOM).

The IEEE LOM is articulated into nine descriptive areas (categories) containing groups of attributes arranged in a tree structure and resulting in a total of 70 descriptive elements (see Fig. 4)
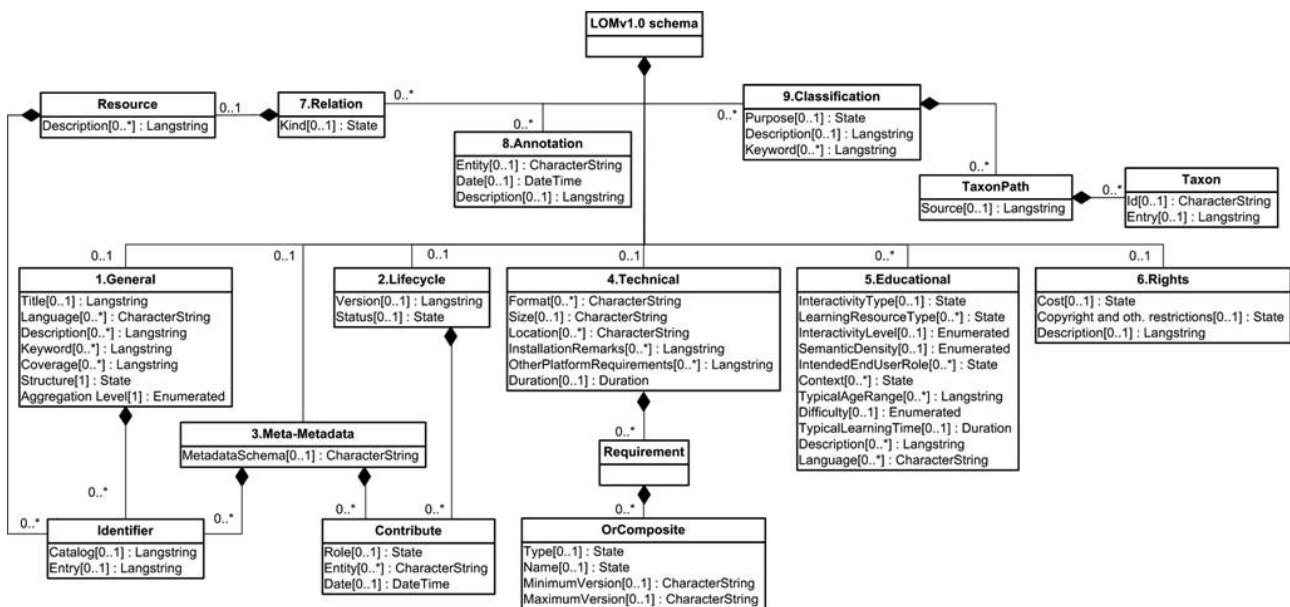

Fig.4 The IEEE LOM articulation

The peculiar feature of G-LOREP is that it adopts a distributed search that extends over all the federation repositories and operates through the use of keywords. The keyword control management is dealt with by organizing the keywords in Taxonomy resulting from a combination of specialized Thesauri. Each set of tree elements (a category) represents a scientific subject that can be differently evaluated depending on its tree level.

Categories are built out from the Dewey Decimal Classification (DDC) [10] schema for the following reasons:

DDC is an updated international standard;

DDC is a multi-discipline classification that covers all the relevant fields;

DDC allows associating a describing label to the subject numeric code;

DDC can define classes and subclasses on different specialization levels.

Currently the application supports the cataloguing of LO in the fields both of mathematics and chemistry [11]. The DDC structure for chemical subjects is shown below in fig. 5.

540 – Chemistry & allied sciences
–541 – Physical Chemistry
--541.2 – Theoretical Chemistry
--541.3 – Miscellaneous Topics in Physical Chemistry
---541.33 – Surface Chemistry
---541.34 – Solutions Chemistry
---541.35 – Photochemistry
---541.36 – Thermochemistry & Thermodynamics
---541.37 – Electro Chemistry & Magneto Chemistry
---541.38 – Radio Chemistry (Nuclear Chemistry)
---541.39 – Chemical reactions
--541.7 – Optical Activity
–542 – Techniques, Procedures, Apparatus, Equipment & Materials
--542.1 – Laboratories
--542.2 – Containers and Accessory Equipment
--542.3 – Testing and Measuring
--542.4 – Heating and Distilling
--542.6 – Filtering and Dialysis
--542.7 – Gas Production, Processing, Measuring
--542.8 – Auxiliary Techniques and Procedures, Electrical and Electronic Equipment
---542.84 – Electrical Equipment
---542.85 – Chemistry Data Processing
–543 – Analytical chemistry
--543.1 – General Topics in Analytical Chemistry
--543.2 – Classical Methods
--543.4 – Electro Chemical Analysis
--543.5 – Optical Spectroscopy (Spectrum Analysis)
--543.6 – Non–Optical Spectroscopy
--543.8 – Chromatography
–546 – Inorganic Chemistry
--546.2 – Hydrogen and its Compounds
--546.3 – Metals, Metallic Compunds, Alloys
--546.4 – Group 3
--546.5 – Groups 4, 5, 6, 7
--546.6 – Groups 8, 9, 10, 11, 12, 13, 14
--546.7 – Groups 15, 16, 17, 18
--546.8 – Periodic Law and Periodic Table
–547 – Organic Chemistry
--547.1 – Physical And Theoretical Chemistry
--547.2 – Organic Chemical Reactions
--547.4 – Aliphatic Compounds
--547.5 – Cyclic Compounds
--547.6 – Aromatic Compounds
--547.7 – Macro–Molecules and Related Compounds
--547.8 – Other Organic Substances
–548 – Crystallography
--548.1 – Geometrical Crystallography
--548.3 – Chemical Crystallography
--548.5 – Crystallization and Crystal Growth
--548.7 – Mathematical Crystallography
--548.8 – Physical and Structural Crystallography
--548.9 – Optical Crystallography
–549 – Mineralogy
--549.1 – Determinative Mineralogy
--549.2 – Native Elements
--549.3 – Sulfides, Sulfosalts, Related Minerals
--549.4 – Halides
--549.5 – Oxides
--549.6 – Silicates
--549.7 – Other Minerals

Fig.5 The Dewey Decimal Classification for chemical subjects

4. New Modules

With respect to the Drupal standard configuration, we implemented the following new modules in order to manage the wide federation and its contents :

Collaborative Repositories (CollabRep)
- To manage the federation: it can be used to create, join and quit a federation, and it performs synchronization recovery measures in case of communication issues during updates.

Linkable Learning Object (LinkableObject)
- To manage the LO and SA: it enables the LO creation and upload it to the servers. It also allows managing LO permissions, defining who can create and who can view and download the learning material. It also defines how SAs are managed and downloaded together with the LO.

Search Linkable Learning Object (SearchLO)
- To manage a distributed search of LO on wide federation: it is a searching system of LO easy and intuitive. Users can use most of the metadata recommended by IEEE LOM to refine their search.

Taxonomy Assistant (TA2.0)
- To manage LO classification step: it analyses the related textual content entered by the user as LO title and description in order to help with the selection of the category better related to the LO [12] [13].

5. GLOREP requirements

The current requirements for GLOREP are:
- Operating System: Ubuntu server Linux distribution (or virtual machine for testing) ;
- DRUPAL version: Drupal 6.x;
- Disk space: minimum 60MB;
- Web server: Apache 1.3;
- Database: MySQL 5.0.15 (no PostgreSQL);
- PHP version: Php 5.2.

There are two downloadable installations of GLOREP [14]:
- Linux O.S. hosted
- Virtual machine hosted (see also the GLOREP Installation Guide [15]).

Campus tests have been performed between the Department of Mathematics and Informatics and the Department of Chemistry of the University of Perugia.
Remote tests have been performed between the University of Perugia and the University of Genoa in order to verify again all the requirements and fix those related to the networking of different campuses bearing different requirements.
GLOREP is now ready for a broader distribution among the cluster of the VEC universities and to pass GLOREP to the newest versions of Drupal.

6. The Scattering and Molecular Dynamics use case

As already mentioned, the work reported here has been stimulated by the research based education activities fostered by the interaction of ECTNA with COMPCHEM [16]. COMPCHEM is a Virtual Organization (VO) that operates on the European Grid Infrastructure (EGI) [17] with the aim of building a Virtual Research Community (VRC) in the field of the Chemistry, Molecular & Materials Sciences and Technologies (CMMST) [18]. This has stimulated the establishing of the already mentioned VEC aimed at developing software tools and applications for research based education [19]. The project has been elaborated by COMPCHEM through a specific Virtual Team in cooperation with other VOs (GAUSSIAN [20], IBERGRID [21]) and projects (MoSGrid [22] and Lifescale [23]) of EGI and has been submitted for participation to the CC Division of EUCHeMS [24].
The focus of the project is the exploitation of CC infrastructures and applications developed for research to the end of designing and assembling Learning Objects (LO) for research based education activities. This has prompted the development of LOs requiring the running of multi-scale simulations for a wide range of innovative fields, including chemical engineering, biochemistry, chemometrics, 'omic'-sciences (Metabolomics, Proteomics, Genomics and Transcriptomics), forensic chemistry, medicinal chemistry, food chemistry, energy production and storage, new materials, space technologies and others with definitely much more computing power and less operational constraints

than in the past.

A typical set of computational applications relevant to this purpose encompasses packages and programs devoted to:

- ab initio calculation of the electronic structure and assemblage of ab initio-based potential energy surfaces;

- integration of quantum and/or classical equations of motion;

- definition of additional treatments necessary to calculate the value of measurable quantities.

These packages and programs were gathered together and implemented in a coordinated workflow as a realistic simulator (see GEMS the Grid Empowered Molecular Simulator described in Ref. [25]) that serves the purpose of both innovative molecular science research and higher education especially when ported on the computing grid that involves about 30,000 CPUs and utilizes 1.5 PBytes of disk storage provided by more than 25 European research institutes and universities. The availability of such resources enabled the VO to implement some components of the simulator to their maximum level of throughput. As a result, they have been used to carry out massive computational campaigns simulating the molecular components of some innovative technologies, such as the gas clathrate hydrates formation for energy storage [26], the flux of ions through trans-membrane cellular micropores [27], the heat transfer around spacecraft in planetary re-entry [28] and the virtual measurement of scattered products in molecular beams apparatuses [29]. This type of calculations has been integrated by proper graphical treatments.

Furthermore CC is, indeed, one of the EchemTest libraries implemented at Master level (level 4) [30]. An important part of such library is the one devoted to Molecular dynamics. The related Molecular Dynamics LOs we consider here are articulated in three blocks: the two-, the few- and the many-body ones. In these LO blocks, the physical, mathematical and computational schemes necessary to handle the related problems are presented.

In the two-body LO block the different types of coordinates used to describe the problem, the full dimensional classical mechanics treatment of a two-body collision, the reduction of the dimensionality of the treatment, the analytical solutions for some model central field cases are discussed. The distinction between bound and scattering problems is made. The corresponding numerical procedures as well as their quantum equivalent to cope with the cases of classical mechanics failure are also explained.

In the few-body (for illustrative purposes only three and four are presented in detail) LO block, again, the different types of coordinates used to describe the problem, the separation between nuclear and electronic coordinates, the calculation and the representation of electronic structure, the case for the different reactive and non reactive channels, the full dimensional classical and quantum mechanics treatments are discussed.

In the many-body LO block, the force field representation of the interaction is discussed and the theoretical and computational problems arising when dealing with a large number of particles are illustrated. Related statistical treatments and observables are analysed.
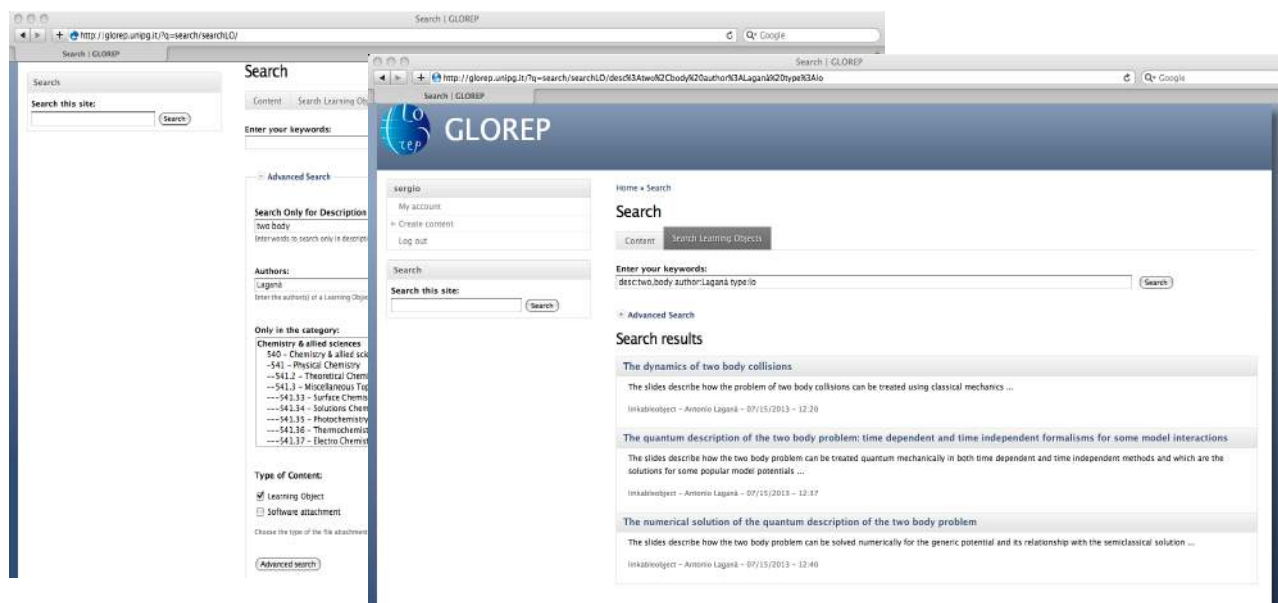

Fig.6 Searching and retrieving LO in GLOREP

The above sketched LOs are useful to illustrate realistic simulations relevant to some modern technologies. The first of them considered here is the aerospace one like the three body reaction $N + N_2$. In the related LOs the contours of the related potential energy surface are drawn, trajectories and wave packets integrated on them are illustrated, vector correlations between initial and final states, detailed state to state probabilities, cross sections, state specific rate

coefficients and more averaged quantities are computed and compared. From the educational point of view this is useful to illustrate the dynamical effects of some features of the potential energy surface like the formation of a stable aggregate at the barrier to reaction.

A different LO related to larger systems, is the gas clathrate hydrates (ice-like solid inclusion compounds resulting from the trapping of gas molecules within a lattice-like cage of water molecules) one [26]. Methane (CH_4) and carbon dioxide (CO_2) have molecular sizes suited to form such hydrates. By using the simulator components, it has been possible to assemble a new ab initio-based force field and run massive Molecular Dynamics calculations showing how the SDS surfactant is able to cage CH_4 molecules in water while CO_2 is not. From the educational point of view this is useful to illustrate how the ionic end of an additive interacts with the gas and how its tail, under proper circumstances, is able to push surrounding water molecules to gage around the gas molecule.

 The last LO considered here illustrates the ion channelling through a carbon nanotube [27] that helps to understand the behaviour of the pores of cell membranes. Placed on the membrane of the cell, the nanotubes allow particular ions to pass across the membrane. They can be used to explain how cellular pores open and close access to inside a cell and show how the determining mechanism is the water ion dressing and undressing.

7. Conclusions and perspectives

The implementation of a distributed repository tailored on the needs of research based education activities has been found not only to be useful for teaching and learning chemical concepts which require often strong multimedia support but also to be tools exploiting the evolution of Grid computing technologies. The distributed repository architecture and the related software environment (tools and requirements) described here show in fact the suitability of such approach for a continuous improvement of the implemented Learning Objects independently from their first publication. The Learning Objects designed for a course on Scattering and Molecular Dynamics and its applications to materials technologies show clearly that the advantages of the tools designed for the present version of GLOREP.

References

[1]   http://ectn-assoc.cpe.fr/network/ec2e2n/workpackage.htm
[2]   http://www.ec2e2n.net/
[3]   http://glorep.unipg.it/
[4]   S. Pallottelli, S. Tasso, N. Pannacci, A. Costantini, N. Faginas Lago, Distributed and Collaborative Learning Objects Repositories on Grid Networks, Lecture Notes in Computer Science 6019, 29-40 (2010);
[5]   https://drupal.org/
[6]   S. Tasso, S. Pallottelli, R. Bastianini, A. Laganà, Federation of distributed and collaborative Repositories and its application on Science Learning objects, Lecture Notes Computer Science 6784, 466-478 (2011)
[7]   http://www.imsglobal.org
[8]   http://www.adlnet.org/scorm/scorm-2004-4th/
[9]   http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
[10]   M. Dewey, Dewey Decimal Classification and relative index, 23th edition, OCLC, Dublin, Ohio (2011)
[11]   http://www.math.unipd.it/~biblio/msc-cdd/
[12]   S.Tasso, S. Pallottelli, R. Ferroni, R. Bastianini, A. Laganà, Taxonomy Management in a Federation of Distributed Repository: A Chemistry Use Case, Lecture Notes in Computer Science 7333, 358-370 (2012)
[13]   S. Tasso, S. Pallottelli, G. Ciavi, R. Bastianini, A. Laganà: An Efficient Taxonomy Assistant for a Federation of Science Distributed Repositories: A Chemistry Use Case, Computational Science and Its Applications, Lecture Notes in Computer Science 7971, 96-109 (2013)
[14]   http://glorep.unipg.it/installer/
[15]   http://glorep.unipg.it/installer/glorep_installation_2_0.pdf
[16]   https://www3.compchem.unipg.it/compchem/
[17]   http://www.egi.eu/
[18]   https://wiki.egi.eu/wiki/Towards_a_CMMST_VRC
[19]   http://ectn-assoc.cpe.fr/news/letter/2012/03/1302_201203_eChemist-VEC.pdf
[20]   http://www.gaussian.com/
[21]   http://www.ibergrid.eu/
[22]   http://www.sci-bus.eu/mosgrid, https://mosgrid.de/portal
[23]   www.scalalife.eu/
[24]   http://www.euchems.eu/divisions/computational-chemistry.html
[25]   A. Costantini, O. Gervasi, C. Manuali, N. Faginas Lago, S. Rampino, A. Laganà, COMPCHEM: progress towards GEMS a Grid Empowered Molecular Simulator and beyond, Journal of Grid Computing, 8(4), 571-586 (2010)
[26]   M. Albertí, A. Costantini, A. Laganà, F. Pirani, Are micelles needed to form methane hydrates in sodium dodecyl

sulfate solutions? J. Phys. Chem. B, 116 (14), 4220–4227 (2012); DOI: 10.1021/jp301124z

[27] D. Skouteris, A. Laganà, Study of the passage of an H^+ ion along a carbon nanotube using quantum wavepacket dynamics, Lecture Notes in Computer Science (ISSN:0302-9743) 3980, 757-762 (2006).

[28] M. Bartolomei, F. Pirani, A. Laganà, A. Lombardi, A full dimensional grid empowered simulation of the CO_2 + CO_2 processes, J. Comp. Chem. (ISSN:0192-8651), 33, 1806–1819 (2012)

[29] A. Laganà, E. Garcia, A. Paladini, P. Casavecchia, N. Balucani, The last mile of molecular reaction dynamics virtual experiments: the case of the OH (N=1-10) + CO (j=0-3) -> H + CO_2 reaction, Faraday Discussion of Chem. Soc. 157, 415-436 (2012)

[30] http://www.emtccm.org/