VIRT&L-COMM.10.2016.6

GEMS EXPRESSION OF INTEREST FOR THE EGI, EUDAT AND INDIGO-DATACLOUD H2020 PROJECT PROPOSAL EINFRA12 (A)

Antonio Laganà, Master-up s.r.l.,Perugia, Italy
Sergio Rampino, Scuola Normale di Pisa, Italy
Loriano Storchi, Università di Chieti, Chieti, Italy
Ernesto Garcia, Universidad del Pais Vasco, Vitoria, Spain
Cecilia Coletti, Università di Chieti, Chieti, Italy
Edward Blurock, Blurock Consulting, Sweden
Carles Bo, Institut Catala de Investigacio Quimica, Tarragona, Spain
Mirco Mariotti, Università degli studi di Perugia, Perugia, Italy
Giuseppe Vitillaro, ISTM-CNR, Perugia

Abstract: *In this paper we present a proposal to empower the Molecular science community with a high level of accuracy ab initio complex simulator of chemical processes through the establishing of an open and collaborative platform for the production, sharing, reusing, management and preservation of data. The proposal is articulated into the following subservices: define the grid of relevant molecular geometries and store them, perform ab initio calculations and store related results, fit the computed ab initio values to a suitable functional form, generate and store the related Potential energy routine, carry out quasi-classical dynamical calculations and store related outcomes, carry out quantum-classical dynamical calculations and store related outcomes, parse dynamical and kinetic data to fine-grained RDFs in the ChemConnect data repository and in the ioChem-BD repository and assemble a prototype inter-departimental cloud (CONCA-cloud).*
The proposal is designed as part of the activities to be committed by the Chemistry, Molecular, Materials Sciences and Technologies (CMMST) Virtual Research Community (VRC) within the joint EGI, EUDAT and INDIGO-DataCloud H2020 project for the EINFRA12 (a) call.

Service and activity descriptions

Describe the services that will be concerned by technical integration, consolidation and operations. Please provide a service description with a user point of view and make a distinction between final services and services components (enhancing parts of a service).

Thematic Service

Please provide as many service entries as necessary. Service is defined as: "the way to provide value to customers through bringing about results that they want to achieve".

SERVICE GEMS (General Molecular Simulator): *a proposal to empower the CMMST community by a higher accuracy and more complex simulator of chemical processes through the establishing of an open and collaborative platform for the production, sharing, reusing, management and preservation of data* (define an appropriate grid of relevant molecular geometries, perform ab initio electronic structure calculations, fit the computed potential energy values to a suitable functional form, reuse the routine by incorporating it into a dynamical code, calculate and store the resulting detailed dynamical quantities for reuse, statistically average dynamical outcomes to evaluate kinetic quantities to store and reuse for modeling complex chemical processes) for the benefit of researchers and innovators.

ISSN: 2279-8773

PROSUMERS: producers and users of ab initio and DFT data, of dynamical properties, of energy transformation, production and storage means for the Computational and Theoretical Chemistry division of EuChemS, for HEIs with highly qualified Departments in Molecular and Material science and technology.

SUBSERVICE n. 1a1: *defines the grid of relevant molecular and* ***stores them***

a. Present version based on the platform HERLA available at the Università degli studi di Perugia, Perugia, Italy in association with other European HEIs
b. Contact, report: <*Dr. Sergio Rampino, Scuola Normale Superiore, Pisa, Italy*>
c. TRL 8
d. Opening to outside world and scaling up access through EGI

SUBSERVICE n. 1b: *performs ab initio calculations and* ***stores related results***

a. Present version based on the platform available at the Università di Chieti, Chieti, Italy in association with other European HEIs
b. Contact, report: <*Dr. Loriano Storchi, Università di Chieti, Chieti, Italy*>
c. TRL 8
d. Opening to outside world and scaling up access through EGI

SUBSERVICE n. 1a2: *fits the computed ab initio values to a suitable functional form, generates and* ***stores the related Potential energy routine***

a. Present version based on the platform HERLA available at the Università degli studi di Perugia, Perugia, Italy in association with other European HEIs
b. Contact, report: <*Dr. Sergio Rampino, Scuola Normale Superiore, Pisa, Italy*>
c. TRL 8
d. Opening to outside world and scaling up access through EGI

SUBSERVICE n. 2a: *carries out quasi-classical dynamical calculations and* ***stores related outcomes****, statistically averages dynamical outcomes to evaluate and* ***stores kinetic quantities***

a. Present version based on the platform available at the University of the Basque Country, Vitoria, Spain in association with other European HEIs
b. Contact, report: <*Prof. Ernesto Garcia, Departamento de Quimica Fisica, Universidad del Pais Vasco, Vitoria, Spain*>
c. TRL 8
d. Opening to outside world and scaling up access through EGI

ISSN: 2279-8773

SUBSERVICE n. 2b: *carries out quantum-classical dynamical calculations and* ***stores related outcomes****, statistically averages dynamical outcomes to evaluate and* ***stores kinetic quantities***

a. Present version based on the platform available at the Università di Chieti, Chieti, Italy in association with other European HEIs
b. Contact, report: <*Prof. Cecilia Coletti, Università di Chieti, Chieti, Italy*>
c. TRL 8
d. Opening to outside world and scaling up access through EGI

SUBSERVICE n. 3a: *dynamical and kinetic data* ***is parsed to fine-grained*** *RDFs in* ***ChemConnect*** *(a data repository for combustion relevant applications) providing interconnectivity between data sets and advanced searching capabilities.*

a. Present version based on the platform available at the Edward Blurock SME site in association with other SMEs and European HEIs (COST Action CM1404: Smart Energy Carriers). Service at http://www.chemicalkinetics.info
b. Contact, report: <*Dr. Edward Blurock, Blurock Consulting AB*>
c. TRL 8
d. Sitting on the Google Cloud Platform with interfaces RDF, NoSQL (Google datastore), opening to outside world and scaling up access through EGI
e. Interest in dissemination of the service for access from cloud – e.g. with DataHub/FedDataPlatform
f. Access data from virtual machines running in the (EGI) cloud in various ways.

SUBSERVICE n. 3b: *the results of computational research projects of chemical relevance are* ***parsed, organized, visualized, analyzed and published*** *on the distributed* ***ioChem-BD repository.***

a. Present version based on the platform hosted at the Institute of Chemical Research of Catalonia (ICIQ) and instances running at URV, UdG, UAB, BSC (Spain) plus other European HEIs. Associated with COST Action CM1305. Service at ioChem-BD, http://www.iochem-bd.org
b. Contact, report: <*Prof. Carles Bo, Institut Catala de Investigacio Quimica, Tarragona, Spain*>
c. TRL 8
d. Opening to outside world and scaling up access through EGI. Currently developing Docker containers
e. Could be used for access from cloud DataHub/FedDataPlatform and operate an open instance on top of EGI cloud.

SUBSERVICE n.4: *a prototype inter-dipartimental cloud* **(CONCA-cloud)** *is assembled to the end of offering a platform for open and collaborative distributed computing between the local Department of Physics and Earth*

ISSN: 2279-8773

Science and the Department of Chemistry, Biology and Biotechnology of the Università degli Studi, Perugia (IT)

a. Contact, report: < *Dr. Mirco Mariotti, Department of Physics and Earth Science, Università degli studi di Perugia; Dr. Giuseppe Vitillaro, ISTM-CNR, Perugia* >
b. Link to the online: <>
c. TLR8
d. Opening to ouside world and scaling up access through EGI
e.
f. Service already implemented at the Physics and Earth Science of the Università degli studi di Perugia
g. Bring the service into the EGI cloud for different use
h. Possibility of generalizing this framework to other science domains

Service overview	
Thematic Service name	*GEMS*
Service description	(1) High-level description of what the service does in terms of functionalities it provides to the end users *-defines the grid of relevant few body molecular geometries,* *-performs ab initio calculations and* ***stores related results****,* *-fits the computed values to a suitable functional form,* *-generates and* ***stores the related Potential energy routine****,* *-carries out dynamical calculations and* ***stores related outcomes****,* *-statistically averages dynamical outcomes to evaluate and* ***stores kinetic quantities*** *and presents them in a form suitable for further use* *- offers a synergistic platform based on complementary scientific interests and an interdepartmental OPENSTACK distributed compute technology* (2) Resources it enables access to *Necessary compute capabilities are expected to be provided by the project EGI, EUDAT and INDIGO-DataCloud e-infrastructures.* *Necessary additional cloud capabilities are expected to be provided by the project EGI, EUDAT and INDIGO-DataCloud e-infrastructures.* (3) Relevant documentation *Examples of the service capabilities from a user point of view are given in* - D. Skouteris, A. Costantini, A. Laganà, G. Sipos, A. Balasko, P. Kacsuk,

ISSN: 2279-8773

	Implementation of the ABC quantum mechanical reactive scattering program on the EGEE grid platform, Lecture Notes in Computer Science 5072, 1108-1121 (2008) - S. Rampino, D. Skouteris, A. Laganà, E. Garcia, A comparison of the isotope effect for the N + N_2 reaction calculated on two potential energy surfaces, Lecture Notes in Computer Science 5072, 1081-1093 (2008). - A. Laganà, N. Faginas Lago, S. Rampino, F. Huarte-Larranaga, E. Garcia, Thermal rate coefficients in collinear versus bent transition state reactions: the N+N_2 case study, Physica Scripta 78 (5), 58116-58125 (2008). - S. Rampino, A. Monari, S. Evangelisti, E. Rossi, K. Ruud, A. Laganà, A priori modeling of chemical reactions on a grid-based virtual laboratory, Cracow 09 Grid Workshop, 164-171 (2010) ISBN978-83-61433-01-09 - L. Storchi, F. Tarantelli, A. Lagana', Computing molecular energy surfaces on a grid, Lecture Notes in Computer Science 3980, 675 (2006). - L. Storchi, C. Manuali, O. Gervasi, G. Vitillaro, A. Lagana', F. Tarantelli, Linear algebra computation benchmarks on a model grid platform, Lecture Notes on Computer Science, 2658, 297, (2003). - C. Manuali, A. Laganà, A new collaborative Grid framework for SSCs Cracow 09 Grid Workshop, 188-195 (2010) ISBN978-83-61433-01-09 - A. Costantini, O. Gervasi, A. Laganà, Molecular distributed computing on the grid: Quantum reactive scattering and visualization tools, Cracow 09 Grid Workshop, 196-203 (2010) ISBN978-83-61433-01-09 - S. Evangelisti, S. Herres-Pawlis, F. Esposito, A. Laganà, G. Terstyanszky, A community for Open Access Molecular Science, VIRT&L-COMM.8.2015.8 http://www.hpc.unipg.it/ojs/index.php/virtlcomm/article/view/96. - C. Coletti, G.D. Billing, Quantum-classical calculation of cross sections and rate constants for the H2+CN->HCN+H reaction, J. Chem. Phys., 113 (2000), 11101-11108 - C. Coletti, G.D. Billing, Vibrational energy transfer in molecular oxygen collisions, Chem. Phys. Lett., 356 (2002), 14-22 S. Evangelisti, S. Herres-Pawlis, F. Esposito, A. Laganà, G. Terstyanszky, A community for Open Access molecular science, VIRT&L-COMM.8.2015.8 A. Laganà, The molecular science community for open science, VIRT&L-COMM.9.2016.6 G. Vitillaro, M. Mariotti, An interdepartmental Openstack platform for scientific applications, VIRT&L-COMM (submitted)
Service provider	*Master-up srl (SME), Italy (IT), University of Perugia, Perugia, Italy (IT), University of the Basque Country in Vitoria, Vitoria, Spain (ES), Institute of Chemical Research of Catalonia (ICIQ), Tarragona, Spain (ES), University of Chieti, Chieti (IT), Blurock Consulting AB (SME), Sweden (SE)*
Service catalogue	*The service is planned to become part of the EGI.eu catalogue by linking the different subservices.*
Value	*Provide either intermediate or final high level of accuracy data obtained from an ab initio study of the efficiency of a few body chemical process to the user or contribute to its production by running in a collaborative workflow part (or all) of the study for new systems leveraging on the e-infrastructure made available by EGI and on the expertise made available by other members of the community which are rewarded through a credit system for the work done on behalf of the community.* *Provide an interdepartmental cloud with a dedicated sharing of hardware (optical fiber connection, storage, compute units of the two departments) and software (Physical and Molecular science applications, educational services).*
Current TLR level1, acceptance	(1) service status in terms of completeness and maturity (including link to relevant documentation)

[1] Technology Readiness Level: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

criteria and validation/verification results	*The various subservices are locally implemented and use their own compute infrastructure. During (and after) the life of the project it is planned to exploit the EGI infrastructure* (2) service acceptance criteria defined by customers and/or users (including e.g. aspects related to interoperability, availability, installability, performance, portability, recoverability, safety, scalability, usability) *EGI rules for the VRCs (and in particular those of the CMMST VRC) will be adopted* (3) results of validation and verification activities involving service providers and user communities. *EGI rules for the VRCs (and in particular those of the CMMST VRC) will be adopted by leveraging on the tools referenced in* C. Manuali, A. Laganà, A new collaborative Grid framework for SSCs Cracow 09 Grid Workshop, 188-195 (2010) ISBN978-83-61433-01-09
Access policy	*Policy-based: users are granted access to the service based on policies defined by the service provider(s) see* C. Manuali, A. Laganà, A new collaborative Grid framework for SSCs Cracow 09 Grid Workshop, 188-195 (2010) ISBN978-83-61433-01-09
Terms of use	Please provide a reference to the service terms of use, i.e. the rules which one must agree to abide by in order to use the service *EGI rules for the VRCs (and in particular those of the CMMST VRC) will be adopted*
User groups and scientific disciplines served	*The service is intended mainly for members of the CMMST VRC through the above specified access policy. The service is also open to non members and companies by making reference to https://mailman.egi.eu/mailman/listinfo/egi-pay-for-use*
Service business model	*Costs of the service subservices are proposed by their providers and approved by the community management board. Sharing of the incomes will occur through the credit mechanism implemented as described in* C. Manuali, A. Laganà, A new collaborative Grid framework for SSCs Cracow 09 Grid Workshop, 188-195 (2010) ISBN978-83-61433-01-09

Service architecture

Define the service by describing its components. A service is usually composed of different service components that enable or enhance the service. A service component is a logical part of a service that provides a function enabling or enhancing a service. Although a service component underlies one or more services, it usually does not create value for a customer alone and is therefore not a service by itself. Examples of service component are software, and services that are provided or could be provided by e-Infrastructures. For example:
- EGI https://www.egi.eu/services/
- EUDAT https://eudat.eu/services-support
- INDIGO https://www.indigo-datacloud.eu/service-component

Service components	Name of component	Functional description, applicable standards and needed resource capacity (if applicable) e.g. CPU Time, storage capacity etc.	Provider If already appointed

ioChem-BD CAS	Central authentication service (to be adapted to EGI services)	
ioChem-BD Create	Java webapp like GUI enables uploading, organizing, visualizing and analyzing files, and publishing datasets to Browse component. It provides a Linux shell client for massive uploads.	
	CPU Time: 24 cores/128 GB/ 100 concurrent users Storage capacity: 100 TB	
ioChem-BD Browse	Open-access digital repository. Modified version of DSpace CPU Time: 8 cores/16 GB Long term storage. 100 TB	
ioChem-BD Find	Central service. Content syndication of all instances. Provides tools for global searches. Data mining tools. Modified version of DSpace	
	CPU Time: 64 cores/128 GB Storage capacity: 500 TB	
	* 16 core Intel Xeon 2.40Ghz (120Gflops double precision) and the development environment mvapich/openmpi for parallel message passing applications, for both peer to peer infiniband and UTP copper gigabit Ethernet links. * 4 GPU Fermi NVIDIA S2050 (2TFlops double precision, 2GPU per node) and the CUDA development environment for applications able to exploit the high performance of the NVIDIA coprocessor. * virtualized data centre equipped with a 3COM 3848 switch, 48 UTP ethernet copper gigabit gates, a 3COM 2924-SFP switch, 24 UTP ethernet	

		copper gigabit gates, 2 IBM 2005-16B SAN switch, 16 16Gbits gates, 2 IBM TotalStorage DS4700 of 4.2Tb, 4Gbits SAN, 2Gbits fiber channel disks, IBM TotalStorage DS4100 of 4.6Tb, 2Gbits SAN, SATA disks, IBM TotalStorage 3573 Tape Library, IBM BladeCenter E (Type 8677), 5 IBM x346 providing the computational cluster with a network environment, a distributed storage and flexible back ups easy to adapt. .	

Service integration with generic e-Infrastructures
If applicable. Define the proposed technical integration/service enhancement activities for this service that are proposed to be funded in the project. For example software integration activities that concern general e-Infrastructure capabilities, like those provided by EGI, EUDAT, INDIGO-DataCloud and other software.

Integration activity and concerned service components	*Integration activities are needed for:* - *data repository for shared knowledge* - *distributed compute power both for HPC and HTC*
Overall necessary effort (Person Months) and timeline	*50 PMs distributed over 3 years.*
List of requested service components	*Services needed are for identity provisioning, authentication, authorization* *Resources are needed for specific compute needs* *Software developed for distributed infrastructures will be needed*

Infrastructure integration
This activity addresses challenge 2 of EINFRA-12 (A): "seamless operation of highly scalable and agile data and computing platforms and services dedicated to analytics including hardware and software components, database, compilers, analytics software, supported to easy user entry points for the community of users". By filling this section the applicant commits to make the service discoverable in a central service catalogue and available for access by international user communities.

Description of infrastructure integration activities relevant to the proposed thematic service	*Infrastructure integration activities will be needed to integrate compute, storage, etc. to the end of supporting GEMS subservices. Work will be provided partly (50%) as in-kind contribution to support wider usage and exploitation.*

ISSN: 2279-8773

(to be planned in the project)	

Training
to develop human capital and generate innovation by fostering adoption by new user communities. Training activities requested in this project must be specific to this call and to the thematic services in scope in your expression of interest. They must not duplicate training activities already funded by other initiatives and projects.

Description of training activities relevant to the proposed thematic service (to be planned in the project)	*Training activities for the users related to the general usage and exploitation of the GEMS services and subservices are planned as part of the project. Work will be provided partly (50%) as in-kind contribution to support wider usage and exploitation.*

Relevance to EINFRA-12 (A) challenges

EINFRA-12 (A) challenge	*Specify your contribution to the challenges highlighted by the e-INFRA-12 (a) call, providing whenever possible concrete examples and key performance indicators.*
1. The operation of a federated European data and distributed computing infrastructure for research and education communities will optimise the access to IT equipment and services	*Users can specialize their contributions to the field of their competences while exploiting the competences of the other members for the areas in which they are least competent*
2. All European researchers and educators are in equal footing to access essential resources	*Access is on equal footing for users while producers can have preferential use thanks to the credits accumulated for the activities provided on behalf of the community*
3. Partnerships with industrial and private partners	*Partnership is encouraged not only from researchers but also from industrial and private partners*
4. Train people in research and academic organisations	*Training is included in the services and subservices offered and are rewarded with credits*
5.	
6. More scientific communities will use storage and computing infrastructures with state-of-the-art services	*The service is aimed at providing a higher quality seamless use of the activities also to members of other communities (especially if scientific as is targeted in section 3)*
7. The open nature of	*This policy is encouraged by a credit reward policy*

ISSN: 2279-8773

the infrastructure will allow scientists, educators and students to improve the service quality	
8. Increase the incentives for scientific discovery and collaboration across disciplinary and geographical boundaries. It will further develop the European economic innovation capacity and provide stability to the e-infrastructure.	*The advanced nature of the services and subservices offered is aimed at fostering high quality innovation*

ISSN: 2279-8773