Cuarta Conferencia de Directores de Tecnología de Información, TICAL2014 Gestión de las TICs para la Investigación y la Colaboración, Cancún, del 26 al 28 de mayo de 2014

A CHAIN-REDS solution for accessing computational services

Roberto Barbera[a,b], Bruce Becker[c], Carla Carrubba[b], Giuseppina Inserra[b], Salma Jalife Villalón[d], Christos Kanellopoulos[e], Kostas Koumantaros[e], Rafael Mayo-García[f], Luis Núñez de Villavicencio[g], Ognjen Prnjat[e], Rita Ricceri[b], Manuel Rodriguez Pascual[f], Antonio Rubio-Montero[f], Federico Ruggieri[h,i]
(On behalf of the CHAIN-REDS project)

[a] Dpt. of Physics and Astronomy, University of Catania, Viale A. Doria 6, Catania 95125, Italy

[b] National Institute of Nuclear Physics-Catania, Via S. Sofia 64, Catania 95123, Italy
{roberto.barbera, carla.carrubba, giuseppina.inserra, rita.ricceri}@ct.infn.it

[c] Meraka Institute, Meiring Naudé Road, Pretoria, 0001, South Africa
bbecker@csir.co.za

d Corporación Universitaria para el Desarrollo de Internet, Parral No. 32 Col. Condesa
06140 Mexico D.F., Mexico
salmajalife@cudi.edu.mx

[e] Greek Research and Technology Network, 56 Mesogion Av., Athens, 11527, Greece
{skanct, kkoum, oprnjat}@admin.grnet.gr

[f] Centro de Investigaciones Energéticas Medioambientales y Tecnológicas,
Av. Complutense 40, Madrid, 28040, Spain
{rafael.mayo, manuel.rodriguez, antonio.rubio}@ciemat.es

[g] Escuela de Física, Universidad Industrial de Santander, Cra 27 calle 9,
Bucaramanga, Colombia
lnunez@uis.edu.co

[h] Consortium GARR, Via dei Tizii 6, Rome, 00185, Italy

[i] National Institute of Nuclear Physics-Roma Tre, Via Vasca Navale 84, Rome, 00146, Italy
federico.ruggieri@roma3.infn.it

Abstract. CHAIN-REDS started on December 2012 focused on promoting and supporting technological and scientific collaboration across different communities established in various continents. One of the main issues is the access to the offered computational services, which can be administrative services associated to Universities and Academic Networks, Data Repositories storing raw data and publications, or HPC and HTC infrastructures for scientific calculations just to mention a few. In this work, the solution proposed by the CHAIN-REDS project regarding these emerging necessities is presented.

Keywords: Global research; Open access; Data management; Standards; Identity provision; e-Infrastructure.

1 Introduction

CHAIN-REDS started on Dec, 1st 2012 focused on promoting and supporting technological and scientific collaboration across different communities established in various continents. Thus, it aims to facilitate these communities uptake and the final use of the services that e-Infrastructure offer them as Virtual Research Communities (VRCs), but also to single users.

To do so, it is essential to promote instruments and practices that can facilitate their inclusion in the community of users, i.e. the use of standards. Then, to build on the best practices currently adopted in Europe and other continents, and promote and facilitate interoperability among different e-Infrastructures and services is a must. In this way, it will be an asset to work out a step-by-step strategy for gradually developing the current technological possibilities in the regions targeted by CHAIN-REDS. Basic issues that should be addressed among others are access to administrative information by students and University staff (i.e. those closely related to a Faculty Secretariat, for example), access to monitoring and managerial tools to National Education and Research Networks (NREN) by their staff, access to HPC and HTC infrastructures by researchers or Data Access, Reproducibility and Trustworthiness (DART) regarding raw datasets and documents.

The project also aims to approach technological officers in charge of these issues, stakeholders and e-Science initiatives coordinators in order to build reliable and robust computational services suitable to real needs.

The first step is then to identify such best practices and approach the involved personnel in order to propose and even define a path towards a global computing ecosystem that will allow VRCs, research groups and even single researchers and students to access and efficiently use worldwide distributed resources (i.e. computing, storage, data, services, tools, applications).

This paper describes the solution proposed by the CHAIN-REDS project to achieve an easy global access to the services that are usually demanded by the scientific and academia community, which can be applied to a wide plethora of specific cases such as the access to student records, administrative tools, Libraries catalogues, Open Access Document Repositories, computational infrastructures or Data Repositories.

2 Identity provision

One of the major challenges in providing services is how the users access them, i.e. how they authenticate themselves and which roles are allowed to assume over that services. For many years, databases with information on users (username and password) have been provided by the service managers and more restricted solutions have been taken also into account such as that implemented to access Grid computing (personal certificates provided by an accredited Certification Authority). Nevertheless, those solutions have usually driven the users to a wide set of usernames and password pairs (with the difficulty of remembering all of them).

In the latest years, the concept of Identity provision has emerged as a valid solution. Furthermore, such a concept is of importance in the Academia, where every student,

professor or administrative staff has his/her username and password as he/she becomes part of an institution. Thus, being a University, an R&D Centre or an NREN accredited by an Identity Provider, a huge pool of services can be accessed by a single user with only a pair of associated credentials (username and password).

Such access has been successfully demonstrated in Grid computing, where no more personal certificates are needed and where now robot certificates can manage jobs for a long period of time, but it can be also applied to academic services for students such as those closely related to a Faculty Secretariat or to their academic record just to mention a few.

Moreover, there exists a close liaison with the access to advanced network services since the Identity provision is linked to the access by Eduroam [1].

It is worth mentioning the advances already made in Latin America in this respect. Thus, the RedCLARA Identity Provider Services has recently joined GrIDP federation, allowing the personnel belonging to the associated institutions to access computing services by means of the Science Gateway concept promoted by CHAIN-REDS (see below); such a success story has been possible thanks to the collaboration between the ELCIRA [2] and CHAIN-REDS projects.

By extending this Identity provision concept to other IT services offered by the Academia, a major step beyond will be made, which will redound in the involved personnel. Such a key issue is crucial for those little communities or even single users who do not belong to big initiatives, which have provided special access to their own services. Implementing Identity provision mechanisms to the academia IT services will allow everyone to access every service in an easy and transparent way.

CHAIN-REDS due to its close collaboration with eduGAIN [3] (the service developed within the GÉANT project [4] to interconnect identity federations around the world, simplifying access to content, services and resources for the global research and education community) is willing to provide the necessary support to the TICAL community, even more when it counts on a Catch-all Identity Provider for those people who are not part of an already accredited institution and are willing to access the services provided by CHAIN-REDS.

3 Science Gateways

The TeraGrid project defined a Science Gateway (SG) [5] as a community-developed set of tools, applications and data that is integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community. This way, it will be possible to abstract the final user from the technological complexity that would be underneath.

SGs have risen as an ideal option for using distributed environments. This is so because the technology underneath has provided four key points:

i. Easy access by the end users
ii. Unattended use of the available codes and/or applications
iii. Employment of standards for the technological implementation
iv. Interoperability and interaction of different infrastructures (middleware)

Thus, by means of SG, it is possible with an identity provision that it is usually provided by academic institutions or even social networks, to access a service (i), simply carry out an IT process by updating the input file (ii), clicking on a "Run" bottom, and retrieve the results as fast as possible since there exists a sufficient pool of resources even belonging to different infrastructure (iii and iv). As it can be easily inferred, this process can be done by a non-ICT skilled person.

SG relies on the adoption of robot-certificates for managing the jobs, so no additional personal certificates must be added to the authorization already provided by an Identity Federation. Then, once that the front-end is already implemented, computational frameworks are implemented and made available to the SG as independent linked modules. Doing so, it will be easier to integrate new tools, applications, and codes as part of the SG paradigm.

Because of all these advances, SG is supposed to provide a short term answer to interoperability when several different e-Infrastructures and services have to cooperate in order to fulfill the requirements of intercontinental communities and single users belonging to scientific, humanities and pure managerial sectors. SG paradigm is fostered by CHAIN-REDS; its version is available at the project website [6].

The CHAIN.-REDS framework for SG is fully web-based and adopts official worldwide standards and protocols, through their most common implementations. It is built using the Liferay portal framework [7] and can be downloaded and installed by a virtual machine containing the development environment and examples of basic template portlets that can be customized to integrate specific applications.

On Feb 2014, the CHAIN-REDS SG offers a direct link to run 21 different applications [8] and its framework and technology have been already adopted by other initiatives such as agINFRA, DCH-RO or EarthServer just to mention a few. Nevertheless, what it is really worth mentioning is that new codes, applications and services can be easily included due to its modular design.

By the adoption of the SG paradigm, a University or academic network can provide in a unique interface a wide plethora of services to their users, which range from the use of scientific codes for research purposes to the access to IT administrative tools. Furthermore, small research groups are helped to be integrated to global research trends.

SG has been used by CHAIN-REDS to demonstrate final user based interoperability by seamlessly execute different applications on local clusters, Grid and Cloud [9] as part of the European Grid Initiative Technical Forum held in Madrid in September 2013.

4 The SCALAC initiative

Although the needs of advanced computing for Latin America and the Caribbean should be supported based on collaborative approaches, they are actually defined by a few research communities, whose inventory and needs analysis is usually specified in the individual acquisition of HPC and HTC facilities, instead of choosing collaborative systems. Given the costs for a country derived from the acquisition and

updating of these resources, it is likely that this trend will be reversed. Several examples have been in fact demonstrated by the CHAIN- REDS project in which initiatives and multinational projects have established a common framework for developing distributed applications to be shared by virtual communities living in Africa, Arab States, Asia, Europe, and Latin America and the Caribbean. Furthermore, by doing so, a collaborative research environment using highly specialized resources and networks of researchers distributed in different parts of the world has been intensified.

From this experience and with the CHAIN-REDS support, RedCLARA is promoting a collaborative scheme based on development prospects, taking into account the following aspects:

- Support for high-level scientific activity
- Support for academic and administrative-related activities
- Support for the transfer and use of advanced computing platforms for regional strategic projects.

The first aspect undertakes fundamental and applied research, which involves the use of specific HPC and HTC resources already integrated in national, multinational, regional or global projects. This scientific activity would be identified by scientific communities and their participation would be channeled through regional or global calls for national, multinational projects.

The second aspect considers support training at various levels (technical, specialized, scientific) and specific actions to be driven to the use of IT academic services.

The third aspect is linked to a direct support for multinational or regional strategic projects, such as projects supporting mainland (weather, early warning, health, safety) and intercontinental to global problems.

The Servicio de Computo Avanzado para Latinoamérica y el Caribe (SCALAC) integrates regional communities of researchers and technologists who combine their capabilities to offer advanced services to the region and to the rest of the computing world. Groups of researchers and technologists from each country provide personnel, experience and resources of computing and telecommunications networks to achieve this purpose.

SCALAC services are based on a virtual network of distributed resources already consolidated in the region constituting the e-infrastructure for the development of e-science projects. At the same time, the service foresees to provide services and pool of resources for academic administrative purposes (data storage, for example) and to promote university closely related initiatives such as La Referencia [10].

At the same time, SCALAC services can be accessed on a two-fold basis: SGs, where the CHAIN-REDS format is being set up in Mexico; and, directly by `ssh` commands.

5 Integration of data related capabilities

In the current days, there have been extraordinary advances in the network and computational capacities. Just to mention a few correlated ones, academic networks

have made available distributed infrastructure as grid or cloud and Infiniband links have deeply increased the parallel performances.

Nevertheless, a major new challenge has arisen due to the huge amount of computational calculations and services that have been made: the management of data. Both in academic and scientific fields, the stored data have dramatically increased and, even more, their use is demanded by more and more people. This fact can be easily showed: the number of Data Repositories (DRs) and Open Access Document Repositories (OADRs) and the volume of data they store have largely increased in the latest years. As a consequence, it is necessary that the data will be easily used as well; furthermore, when these data concern datasets and publications. Otherwise, advances made on middleware interoperability will be meaningless since the computational resources will not be properly exploited. In this regard, CHAIN-REDS is promoting interoperability as a main objective and a worldwide demo has been recently shown in September 2013 covering different regions worldwide [9,11]. To this, the proper data curation must be also taken into account.

CHAIN-REDS has also identified this issue and is proposing a solution through the Data Access, Reproducibility and Trustworthiness (DART) challenge. DART is conceived in a way that can be adapted to scientific, academic and administrative purposes interchangeably. This is so because is based on technological standards, profits from Identity providers to be accessed, allows the integration of any kind of toll or application, and adopts the PID assignment for the data transfer. As it has been aforementioned, it can be seamlessly applied to either a bioinformatics study, to the production of a network statistical data report or to a university secretariat procedure.

To achieve DART, several tools have been implemented by the project. The Knowledge Base (KB) [12] provides information about the deployment of e-Infrastructure related topics per country and even about specific Distributed Computing Infrastructures (DCIs) by means of a Site or a Table view. During 2013, the project has been working on extending the CHAIN KB with information related to data infrastructure. To do so, it has collected both issues and best practices and has surveyed the involved regions in order to discover data repositories. The reason for that is to promote data sharing across different e-Infrastructure and continents widening the scope of the existing CHAIN KB to Data Infrastructure and to finally provide proof-of principle use-cases for data sharing across the continents by the end of CHAIN-REDS in May 2015.

Thus, the main developed tool has been the Semantic Search Engine (SSE) [13], which allows users to search for any specific term and find the datasets related to it in all the repositories already linked to the KB, but also find potential relationship in terms of authors, subjects and publishers with other repositories.

Both CHAIN-REDS KB and SSE can work with any kind of data already stored under some metadata standards, i.e. the stored raw data must not be changed, only some metadata labels must be added in order to make them searchable

5.1 4.1 Standards promoted

After a deep analysis on the best practices that on-going data initiatives were carrying out, the following standards were selected by the project for pursuing DART:

- OAI-PMH [14] for metadata retrieval.
- Dublin Core [15] as metadata schema.
- SPARQL [16] for semantic web search.
- XML [17] as potential standard for the interchange of data represented as a set of tables.

To those, Persistent Identifiers (PID) [18] must be added as a tool to know where and how data and metadata are stored. Such a circumstance is achieved by assigning an identifier to a digital object, i.e. as d.o.i works for articles.

With this selection in mind, a work plan for improving the CHAIN-REDS KB by adding new functionalities was executed. Such a work plan is roughly described by the following steps:

Step1. Integration of the KLIOS [19] services inside the KB in order to extend its functionalities.

Step2. Dynamically include in the KB Data Repositories (DRs) and Open Access Documents Repositories (OADRs) worldwide by using the already defined standards. The information related to these repositories should be presented to the user in the same way as the already available DCIs one.

Step3. Take this integration as a proof-of-principle for demonstrating the work carried out. Repositories containing documents (articles, proceedings, books, etc.) were selected in principle for testing and depurating the KB functionalities with the aim of improving its characteristics to other fields and data (see items below).

Step4. Provide the KB with the following capabilities, so further extraction and exploitation of raw data by any user could be performed: semantic web enrichment; semantic search engine; and, a tool for extracting the data associated to the repositories.

Step5. Start working with some specific communities on a strategy for demonstrating DART, i.e. the access by a user of data already stored in order to extract them and employ them as input in a scientific application for reproducing and/or extending the results of a given research. The new data and, may be, the new paper will be lately stored on the Data Infrastructure and will be easily found by the people belonging to the same domain.

To the date, February 2014, Step1 to Step4 have been already achieved and some prototypes already exist with regard to Step5.

5.2 The CHAIN-REDS Knowledge Base

In principle, the KB was implemented considering e-Infrastructure as an environment where research resources (hardware, software and content) can be readily shared and accessed where necessary to promote better and more effective research. Then, such environments integrate hard-, soft- and middleware components, networks, data repositories, and all sorts of support enabling virtual research collaborations with a final goal: to allow scientists across the world to do better (and faster) research using DCIs, independently of where they are and of the paradigm(s) adopted to achieve their goals. Thus, to better fulfill this long term milestone, the use of standards is more than an asset and, at the same time, a step forward to achieve sustainability.

The first release of the CHAIN KB presented integrated dynamically updated information about DCIs. Information about Regional and National Research and Education Networks, National Grid Initiatives, Certification Authorities, Identity Federation Providers, Regional Operation Centers, Grid sites and Applications (and already running on a Science Gateway) was available in both Country and Table views. All this work was performed bearing in mind the concepts mentioned in the previous paragraph.

But even when this on-line service was a clear step forward in harmonizing the different regional infrastructure information, new capabilities should be incorporated. Thus, it was decided to work with article and papers repositories on a first stage. Once standards (Std1-Std4) were identified to easily gather and access both OADRs and DRs, a demonstrator was built with them to visualize and access the repositories by means of both geo- and tab-views (as it was previously made for DCIs). Such a demonstrator was implemented with the advances carried out within the Knowledge Linking and sharIng in research dOmainS (KLIOS) project [19]. KLIOS is based on the interconnection and the integration of scientific resources through a grid of meta-data network and provide the following services: metadata harvesting; semantic enrichment; and, linked data semantic search.

Basically, the new KB capability is composed of a multi-layer structure where two harvesters running on either Grid or Cloud search for OAI-PMH end-points from OADRs and DRs. Above them, a semantic web-enrichment layer is used to act as a previous step before the linked-data search engine, which is on the top. The process of the metadata harvesters is as follows:

- Get the address of each repository publishing an OAI-PMH standard endpoint
- Retrieve, using the OAI-PMH repository address, the related Dublin Core encoded metadata in XML format
- Get the records from the XML files and, using the Apache Jena API, transform the metadata in RDF format
- Save the RDF files into a Virtuoso triple store according to an OWL-compliant ontology built using Protégé

CHAIN-REDS KB reports of 2,481 OADRs and 596 DRs (Feb 2014). From these repositories, the following information is depicted: the country where the data is stored; the name of the repository; the scientific domain it belongs to; and, the organization is maintaining it. All this information can be also geographically displayed on a world map. OADRs counts on DRIVER [20], OpenAIRE [21] and OpenDOAR [22] and currently refers to more than 30 million documents and; regarding DRs, they include Databib [23], DataCite [24] and La Referencia [10].

As it has been aforementioned, as a legacy from CHAIN, it has 86 entries in its DCIs Table view, which reports on the country where the DCI is settled, the regional network, the National Research Education Network (NREN) and the National Grid Initiative (NGI) it belongs to, the Certification Authority (CA) and the Identity Federation (IdF) it relies on for accessing it, the Regional Operation Centre (ROC) it is connected to and the sites it counts on. As in the previous case, the same information can be showed searching on a world map.

5.3 The Semantic Search Engine

The next action was to correlate OADRs and DRs to create linked data and discover new knowledge through semantic enrichment of metadata. This is possible by now by using the SSE [13]. Thus, a user can now access de "Linked Data" tab in the CHAIN-REDS KB, select (if desired) which language will use for its search, enter (if desired) a free keyword or some from a list of them and final make the semantic search.

The results are displayed on the screen on a two-fold possibility: as a list of them or as a graph where filters for visualization can be applied in order to detect semantic links. Doing so, new information can be obtained from the graphs, i.e. new knowledge through semantic enrichment of metadata. On Feb 2014, the CHAIN-REDS KB counts on datasets storing articles and papers; then, by performing a semantic search with the 'Linked Data' tool, authors who work on a scientific topic, articles related to the same kind of research or, investigations carried out with the same kind of facility or experimental setup, for example, can be easily discovered.

The technological description of how this process can be made follows. The first pillar is the harvester procedure. Then, each Resource Description Framework (RDF) file retrieved and saved in a Virtuoso-enabled triple store is mapped onto a Virtuoso Graph that contains the ontology expressly developed for the search engine. The ontology, built using Dublin Core and FOAF standards, consists of:

- Classes that describe the general concepts of the domain: Resource, Author, Organization, Repository and Dataset (where Resource is a given open access document);
- Object properties that describe the relationships among the ontology classes; the ontology developed for the service described in this paper has several specific properties such as `hasAuthor` (i.e., the relation between Resources and Authors) and `hasDataSet` (i.e., the relation between Resources and Datasets)
- Data properties (or attributes) that contain the characteristics or classes' parameters.

The highest-level, component is the Search Engine itself. Using it, visitors can either enter a keyword and submit a SPARQL query to the Virtuoso triple store or select a language and get, on the left side of the page, the list of subjects available in that language with the indication, between parentheses, of the number of records available for that particular subject. The results of a given query are listed in a summary view directly displayed on the webpage. For each record found, the title, the author(s) and a short description of the corresponding resource are provided. Clicking on the "More Info" link, visitors can access the detailed view of the resource.

In the "Dataset information" panel users get the link to the open access document and, if existing, to the corresponding dataset. Clicking on the "Graphs" tab, which appears at the top of the summary view, users can select one or more of the resources found and get a graphic view of the semantic connections among Authors, Subjects and Publishers. In this way, if new links appear, connecting different resources, users can infer new relations among resources, thus discovering new knowledge.

A programmable use of the CHAIN-REDS SSE is also possible due to the development of a RESTful API that has been created on purpose; now, it is possible to get and/or re-use the many millions of open access resources contained in the

CHAIN-REDS KB and stored in a Virtuoso RDF-compliant database by calling the Semantic Search Engine from a common website or even mobile application [25].

5.4 The CHAIN-REDS tools extension

In the previous subsections the technological base of the CHAIN-REDS tools has been briefly described. Their joint use form a framework that can be adopted by many disciplines and interests since they are built upon standards and can be easily customized to both scientific and administrative purposes. In addition, their modularity makes them ideal candidates to be coupled or integrated into other services.

Thus, now it is possible to perform either single or parallel semantic searching [26]. By passing the mouse over the "Semantic Search" link of the CHAIN-REDS webpage, any user can see a sub-menu with several items; the first two are:

- Single: the usual semantic search service described above; and,
- Parallel: the new parallel semantic search service that allow users to search in parallel (i.e., at the same time) across the millions of resources contained in the CHAIN-REDS Knowledge Base and in the ENGAGE Platform [27].

Parallel semantic search engines have been made available also in the SGs of some (collaborating) projects, enhancing and extending in this way the solutions proposed by CHAIN-REDS. This parallel semantic search can be found at:

- agINFRA, here the user can search in parallel across the millions of resources contained in the CHAIN-REDS Knowledge Base and in the OpenAgris [28] repository; and,
- DCH-RP, here the user can search in parallel across the tens of millions of resources contained in the CHAIN-REDS Knowledge Base and in the Europeana [29], Cultura Italia [30] and Isidore [31] repositories.

Another two extensions have been the programmable use of the Semantic Search Engine by a RESTful API and the implementation of the engine on a mobile app for both Apple store and Google Play. Any user can access them also by passing the mouse over the "Semantic Search" link of the CHAIN-REDS webpage.

As it can be easily inferred, now that the technological skeleton is implemented, it will be very easy to furthermore extend the impact of the CHAIN-REDS tools by customizing them to different services. For example, any student or administrative staff from any university can access to specific information by using his/her mobile. Based on the current Semantic Search Engine app, he/she can access the system and search for some records or data and also find with connected information is also available by the semantic enrichment capability. A typical case would be to access the service for knowing a qualification mark and, at the same time, know if the same Professor has already published other qualifications belonging to other subjects, information of interest for both students (in order to know if they have pass an exam) and Secretariat personnel (for knowing if any Professor has already finished his/her duties on time for closing an academic year). Thinking in NRENs, parallel search can be done for specific parameters in the different academic networks belonging to RedCLARA.

Of course, there is a wide plethora of cases where the access to data can be managed by the CHAIN-REDS tools. In what follows, an example devoted to scientific production of results is described

5.5 The DART concept

CHAIN-REDS is working on Data Accessibility, Reproducibility and Trustworthiness (DART). The first step was to achieve interoperability from the final user point of view. Thus, CHAIN project successfully performed an interoperability demo on Sep 2012 during the EGI Technical Forum that showed how jobs could seamlessly run on Grid and HPC infrastructure independently of the middleware they were using. This way, any user accessed the CHAIN Science Gateway [6] by means of his/her Identity Federation credentials and, on a user-friendly interface, submitted a job. In an unattended way, the system was able to perform a job of 9 different applications on 8 different computing infrastructure ruled by 7 different middleware and, of course, retrieve back the results.

On Sep 2013, and also as part of EGI Technical Forum, the demo 'Managing and using interoperable DCIs through a standard-based Science Gateway' was also successfully performed [9]. It was planned for either demonstrate interoperability (by allowing a scientist to seamlessly run applications on HPC machines, Grids and Clouds) and interoperation (by allowing a cloud-tenant of a real or virtual organization to seamlessly and easily manage Cloud resources pledged by providers owning/operating infrastructures based on different middleware stacks). Again, and logging into the user-friendly CHAIN-REDS Science Gateway using his/her federated credentials, the user could select an application from a menu and transparently execute it on HPC machines, Grids and Clouds. It is also worth mentioning that the fractions of executions on the three different platforms could be adjusted to simulate the need to "boost" the resources in case of temporary peaks of activity.

The next action is to add data management capabilities to this demo according to the strategy described in the previous Sections. The vision is that of a researcher of a given scientific domain finding publications he/she is interested in and being automatically redirected to the data used to produce those papers and to the applications used to produce those data. Or, alternatively and simply, access raw data of interest to be lately used as input in applications.

He/she can then access that data and the corresponding application in order to reproduce and extend the results of a given study. The new data (and the new paper if any) are stored on the Data Infrastructure and can be easily found by the people belonging to the same domain making possible to start the cycle again.

By now, all these actions can be carried out by using the CHAIN-EDS tools, i.e., the Science Gateway, the Knowledge Base and the Semantic Search Engine. The requirements that are needed and not directly managed by CHAIN-REDS are related to intellectual properties issues and unique identifiers (PID) referring to papers, data and applications. Nevertheless, CHAIN-REDS is supporting the assignment of PID to digital objects by means of the service provided by the partner GRNET [32].

CHAIN-REDS has started working on this DART challenge for providing proof-of principle use-cases for data sharing across continents. The first prototype has been

successfully tested. It counts on a couple of datasets stored in ZENODO [33] and DataCite [24] and applications that could make use of them. Thus, datasets related to molecular cross sections hosted by the Max-Planck Institute [34] can be downloaded from ZENODO and data on genes to be compared across species hosted by Ensembl [35] can be downloaded from DataCite. The former can be used to obtain molar absorption coefficients by using an application devoted by CIEMAT already included in a Science Gateway portlet called Molon [36] and the latter can be taken as input for several applications, like for example jModelTest2 [37,38], to obtain models of nucleotide substitution.

These calculations can be performed via the CHAIN-REDS Science Gateway and the final output can be stored on data repositories included in the Knowledge Base, searchable by the Semantic Search Engine, and identified by a PID, so the cycle can be started again.

This proof-of-concept test is going to be shown and proposed to the CHIAN-REDS collaborative communities for their own use with their own datasets and applications. These communities come from the Agriculture, e-Government, Earth Science, Cultural Heritage and Astroparticle domains. This concept is also being promoted in the CHAIN-REDS targeted regions in order to find success stories; Latin America has identified an astrophysics community and conversations are on the way. Nevertheless, as it has been explained before, DART challenge can be also applied to academic and network management scenarios and necessities seamlessly.

At the same time, DART has been proposed to the Workflows Working Group of the EUDAT initiative [39] and now CHAN-REDS representatives are part of it.

6 Conclusions

CHAIN-REDS has as a major goal to propose a model for accessing and managing data. To achieve such an objective, several tools based on standards have been implemented and access to these services through identity providers has been relied on. Now, any user can perform a whole cycle of searching data and documents, retrieve the raw data, use them as an input of a service, and obtain final results susceptible of being stored under the same standards format.

For the specific case of Latin America, CHAIN-REDS is closing collaborating with RedCLARA. Thus, it is supporting the SCALAC service and integrating major repositories as La Referencia into the project Knowledge Base. Also, success stories are being aware of the CHAIN-REDS developments and will benefit from them.

In principle, scientific cases have been identified, but all the services promoted by CHAIN-REDS can be adopted by the Academia and the NRENs. In this sense, CHAIN-REDS is much interested in collaborating with this kind of institutions to support them in the use of its services.

Acknowledgements

This work has been co-funded by the CHAIN-REDS project (Grant Agreement 306819) under the European Commission Seventh Framework Programme.

References

1. Eduroam, https://www.eduroam.org/
2. ELCIRA, http://www.elcira.eu/
3. eduGAIN, http://www.geant.net/service/eduGAIN/Pages/home.aspx
4. GÉANT, http://www.geant.net/
5. N. Wilkins-Diehr et al. TeraGrid Science Gateways and Their Impact on Science. Birck and NCN Publications. Paper 434 (2008)
6. CHAIN-REDS Science Gateway, http://science-gateway.chain-project.eu/
7. Liferay, http://www.liferay.com/
8. CHAIN-REDS applications, http://www.chain-project.eu/applications
9. CHAIN-REDS interoperability demo, http://science-gateway.chain-project.eu/demo-status
10. La Referencia, http://lareferencia.redclara.net/rfr/
11. CHAIN-REDS D4.2, available at http://www.chain-project.eu/deliverables
12. CHAIN-REDS Knowledge Base, http://www.chain-project.eu/knowledge-base
13. CHAIN-REDS Semantic Search Engine, http://www.chain-project.eu/linked-data
14. OAI-PMH, http://www.openarchives.org/pmh/
15. Dublin Core, http://dublincore.org
16. SPARQL, http://www.w3.org/2001/sw/wiki/SPARQL
17. XML, http://www.w3.org/XML/
18. PID, http://www.pidconsortium.eu/
19. KLIOS, http://klios.ct.infn.it
20. The DRIVER initiative, http://www.driver-repository.eu/
21. The OpenAIRE initiative, https://www.openaire.eu/
22. The OpenDOAR initiative, http://www.opendoar.org/
23. The Databib initiative, http://databib.org/
24. The DataCite initiative, http://www.datacite.org/
25. The programmable use of the Semantic Search Engine, http://www.chain-project.eu/semantic-search-api
26. The parallel search in the CHAIN-REDS KB, http://www.chain-project.eu/parallel-semantic-search
27. ENGAGE, http://www.engagedata.eu/
28. The OpenAgris repository, http://aims.fao.org/openagris
29. The Europeana repository, http://www.europeana.eu/
30. The Cultura Italia repository, http://www.culturaitalia.it/
31. The Isidore repository, http://www.rechercheisidore.fr/
32. GRNET PID Service, http://epic.grnet.gr/
33. The ZENODO repository, http://zenodo.org/
34. The MPI-Mainz UV/VIS Spectral Atlas, http://satellite.mpic.de/spectral_atlas
35. Ensembl Gene Trees and Homologues, http://www.ensembl.org/info/website/tutorials/compara.html
36. Molon portlet, http://science-gateway.chain-project.eu/molon
37. M. Loureiro et al. Grid selection of models of nucleotide substitution. Studies in Health Technology and Informatics 159, 244-248 (2010)
38. jModelTest2 portlet, http://science-gateway.chain-project.eu/jmt
39. EUDAT, http://www.eudat.eu/workflows