Available online at www.sciencedirect.com
ScienceDirect
Procedia Technology 9 (2013) 215 – 225
Procedia Technology

CENTERIS 2013 - Conference on ENTERprise Information Systems / ProjMAN 2013 - International Conference on Project MANagement / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies

# A Semantic Peer-to-Peer Network for Service Composition in Scientific Domains

Tadeu Classe[a,]*, Regina Braga[a], Fernanda Campos[a], José Maria N. David[a]

[a]*Master Program in Computer Science, Federal University of Juiz de Fora, Juiz de Fora, Brazil*

**Abstract**

Changes are taking place in scientific research. In particular, there is a greater collaboration among groups of researchers, which leads to an increase use of information processing techniques, and a need to share results. Using peer-to-peer networks and semantic web, we create an access point to share distributed knowledge and scientific applications, where scientists can work with heterogeneous information in scientific communities according to their interests. This paper presents the e-ScienceNet, an architecture to creation of scientific communities in the context of semantic e-Science, facilitating the creation and sharing of scientific data and services. Beyond of presents the e-ScienceNet prototype.



*Keywords*: Semantic Web; Peer-to-Peer; Ontology; Service Composition; Web Services.

## 1. Introduction

In recent years, significant changes are occurring in scientific research nature. In particular, there is a greater collaboration among large groups of researchers, which leads to an increase use of information processing techniques, and a need to share results and observations among the participants of the process [1]. Using

* Corresponding author. Tel.: +55 0xx 32 2102-3387;
*E-mail address:* tadeu.classe@ice.ufjf.br, {regina.braga, fernanda.campos, jose.david}@ufjf.edu.br.


doi:10.1016/j.protcy.2013.12.024

technologies such as peer to peer networks [2] and semantic web [3], we create an access point to distributed knowledge bases, where scientists can work with heterogeneous information and create scientific communities according to their specialties and interests.

However, the use of computational resources in scientific research is not trivial, because, in general, scientists from areas others than computing have trouble dealing with low-level computer technology. Besides these difficulties, there is also the difficulty of finding software artifacts that perform specific tasks necessary for the experiment being carried out [4].

Some research addressing interactions in semantic web [5], [6], [7], [8], [9], [10] are fragmented and present an overview of the problem. Considering these technologies, this paper uses the semantic web to create a semantic description of the available scientific data and applications in a peer to peer network using ontologies. It provides a semantically rich description of data, facilitating the access of scientists to information and applications provided by other scientists and related to their interests and specialties.

This work has as its main objective the design and development of an architecture for supporting the management of scientific experiments in a peer to peer network. This architecture, named e-ScienceNet, is responsible for providing data and services in a scientific collaboration network, which are defined as public to all users (scientists) connected to the network, and, when necessary, defined as private to only those users previously registered. By using a generic data model, all information provided by a component may be shared with the other components into e-ScienceNet.

Considering the requirements of scientific services in order to execute a scientific experiment, this paper emphasizes the specification of two managers, named Composition Manager and Service Manager. These managers are essential in scientific context, since the composition of distributed scientific applications to carry out an experiment is important and often, scientists do not have adequate support to accomplish such a composition.

Composition Manager and Service Manager were specified in order to help scientists in this composition, preventing them from having to worry about technical issues. Thus, using the e-ScienceNet, peers can perform searches and create service flows (service compositions) according to their demand. Furthermore, these compositions can be shared with others, saving time in carrying out experiments, because they do not need to be developed from the beginning.

The paper is organized as follows: section 2 presents the theoretical background. In Section 3 some related works are analyzed. Section 4 presents the e-ScienceNet architecture. In section 5 the e-ScienceNet prototype is presented and a Proof of Concept is executed and Section 6 presents conclusions and future works.

## 2. Theoretical Bases

Hendler [11] defines e-Science as concepts related to integration of computing in scientific research in various areas. A distributed architecture for e-Science using technologies such as peer to peer networks and semantic web, raises several requirements that must be considered in its planning [12], [13], such as: i) Storage: scientist should be able to store and process large volumes of data efficiently, regardless of their geographical location, ii) Property Management: the scientist should be able to maintain ownership of their data and services, providing its resources only after the other scientists accept the usage terms of the requested resource; iii) Transparency: the scientist should be able to discover, access and process data transparently, regardless of where this data is located; iv) Communities: the scientist should be able to create, maintain and dispose communities, whether restricted or not. To do this, they must create rules indicating the resources permissions; v) Security: scientist should be able to share their data securely. The use of encryption and authentication techniques is crucial to maintain the privacy of transmitted data; vi) Mobility: scientist should be able to access the available data from any device, including PCs and mobile devices; vii) Workflow: support the automation process, describing the scientific processes in a clear and computationally processable manner, thus creating a fully automated workflow; viii) Provenance: sufficient information should be stored at runtime, providing evidence of the concision of the generated data, allowing the reuse of results and enabling the reproduction of scientific experiments; ix) Notifications: scientists should receive notifications of new available data and services according to their interests; x) Decision Support: provide information and suggestions relevant to scientists according to their needs ; xi) Expansion: support the growth of infrastructure. The number of scientists, data and services should not affect or limit its use; xii) Component: consider the use of software components that can be customized to the changing needs of scientists over time.

A peer to peer (P2P) network is a completely decentralized network architecture that uses distributed resources of multiple computers. We name "nodes", the distributed computers that work to perform a common task, often sharing files on the Internet [14]. The peer to peer networks are by nature a distributed system without any hierarchical organization or centralized control over the data, where a node on the network can act as both a client and a server [2]. The design philosophy of peer to peer network systems is similar to e-Science domain, which is to provide its users the flexibility in cooperation among them [15]. In most peer to peer networks, the spread of requests to other nodes in the network is done randomly. By transmitting messages randomly to other nodes, there is a decrease in the efficiency of communication and a loss in accuracy of the result. Creating groups of nodes in a peer to peer network is a task known as overlay networks. These groups are aware of each other, thus forming a graph with the various groups and their respective nodes. If we use research areas and interests of scientists connected to the network to create these groups, we can prevent information from being propagated through the network randomly, sending the requests only to nodes that have semantic similarity with the areas of research and interest of the researcher. In this case, considering a peer to peer network, we can use ontologies [3] related to one or more network nodes.

The combination of P2P technology with the semantic web can bring several benefits to Science [10]. Considering the ontologies is it possible to perform different types of tasks, such as: routing between peers, search for resources, importing data from files and documents, common interest groups inference, mapping between peers and composition of scientific applications, among other features [8].

Considering the context of composition of scientific applications, with the increased use of web services in various areas, and the interest in developing scientific services, raises the necessity of composition of smaller services into complex services in order to carry out a scientific experiment. To help the composition of services, technologies such as scientific workflows and semantic web services should be considered [4].

## 3. Related Works

Research proposals have been presented in several papers considering the use of peer to peer systems together with the use of semantic web in scientific communities.

One of the classic works that can be highlighted is the Bibster tool [8], a semantic web system designed based on P2P networks. Such work has focused on the sharing of bibliographic information based on ontology mapping. These ontologies serve as a search resource for the system. Bianchini et al. [9] makes an approach using peer-to-peer networks with semantic web services based on a dynamic model of peers that provide similar services. In this proposal, each peer has a local knowledge structure, which includes registration and service publication units, and content-based ontologies.

The creation of a P2P approach for the discovery of web services, based on the use of semantic web technologies is also shown in the work of Gharzouli and Boufaida [16]. Its strategy is based on epidemic search algorithms, performing searches for services that are distributed across all peers. From these searches, their approach allows the creation of compositions. Di Modica et al. [17] proposes an architecture combining the P2P paradigm with Semantic Web technologies, grouping peers that have semantically related services. This work also relies on the use of JXTA technology [18], [19], which forms the basis of the network structure, allowing the creation of new peers and groups.

Some of the features and ideas presented by these works contribute to the e-ScienceNet proposal, such as ontology mapping [8], the search for annotated services [9], [16] and the use of semantic groups of peers [17]. Our approach presents different aspects considering the works that were presented above. With e-ScienceNet, we can share all types of files and documents among peers. Information about these resources (name, extension, path, size, date, type, services descriptions (WSDL or other description language) etc.) are obtained and mapped to existing ontologies on each node (peer). The creation of semantic groups is made based on the research interests of each member of the network. Based on this semantic groups, the system performs the search considering the interest groups that the peer are included, finding adequate resources, including semantically annotated web services by the Service Manager (this manager will be explained latter). The search for web services enables the composition of services by the Composition Manager (this manager will be explained in the next section), and the result of this composition are mapped into ontologies and made available as reusable resources for further compositions.

## 4. E-ScienceNet Architecture

As pointed before, e-ScienceNet is a peer to peer semantic network that supports Science. The architecture of e-ScienceNet is shown in Fig 1, which encompasses some managers, each one with its specific features, in addition to the proposed use of semantic interest groups for peer grouping. The e-ScienceNet allows scientists that are geographically dispersed can be connected via nodes (peers) access, at the same time. This enables scientist to manage, share, analyze and discuss their works in a more efficient way.

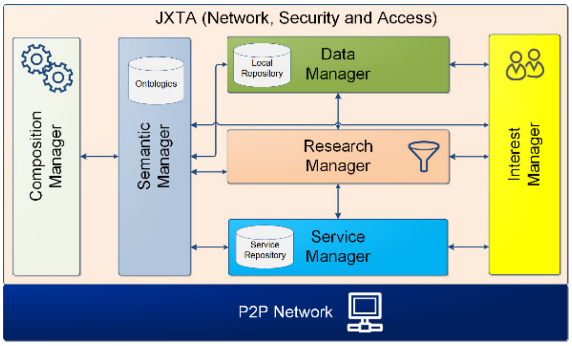


Fig. 1. e-ScienceNet architecture.

The main e-ScienceNet managers are detailed in Fig 1, and discussed below.

- Semantic Manager: it is considered one of the fundamental parts of the system. It is responsible for providing semantic support to other managers. According to Semantic Manager, the peers of e-ScienceNet must have an ontology, named PeerOntology (Fig 2), that describes the network structure and providing specific ontology restrictions related to ontology´s terms. Based on this ontology, the Semantic Manager stores and retrieves information about the peers, such as the groups to which a given peer belongs; resources, files and services related to the peer; resources associated to a given group, among others. All the semantic information passes through this manager, including also, domain ontologies, i.e., ontologies related to the domain of interest of researches.

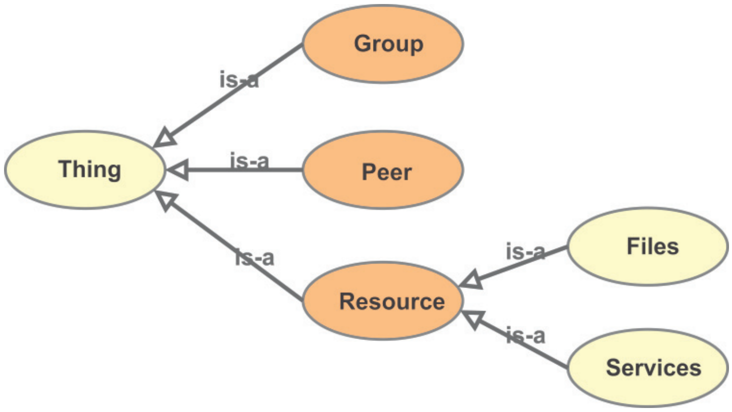


Fig. 2. PeerOntology: ontology found in all peers of e-ScienceNet.

- Data Manager: it is responsible for performing the physical access to data, files and information from the peers in e-ScienceNet. Data Manager allows the management of new files to be shared on the network. This manager also has the functionality of retrieving file information, such as file type, size, description, dates, among others. This information is sent to the Semantic Manager which performs the persistence in PeerOntology. This manager also participates as an information provider for the Research Manager.
- Service Manager: this manager is able to provide information about services available in peers. This module allows incorporating information about a service on the system and its data is processed by the Semantic Manager for the mapping and persistence in PeerOntology. As the Data Manager, the Service Manager provides information to the Research Manager.
- Research Manager: module responsible for the search among peers in e-ScienceNet. This manager receives information from the Data Manager and Service Manager, via Semantic Manager. This communication aims at obtaining data on available resources. The Research Manager performs its searches based on the configuration of the semantic interest groups that are associated with a peer. This manager also serves as the source of information for the Composition Manager, considering the service flow creation process. Its functionality can be seen in Fig 3 where three semantic groups can connect the requester peer (in red) to other peers, searching and getting results from their members (arrows).

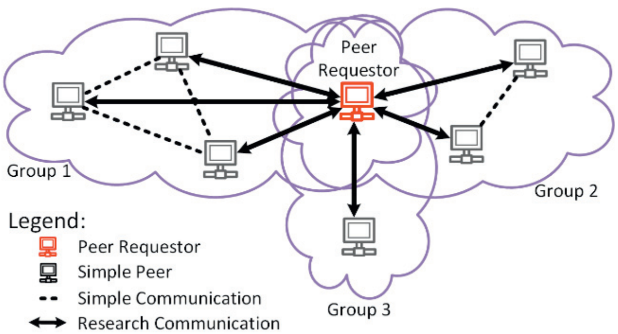


Fig. 3. Search through peers schema.

- Composition Manager: it is responsible for creating, publishing and sharing service compositions (services flows). Based on the search among e-ScienceNet peers and on information obtained from the search conducted by the Research Manager, it is possible the specification of flow of services in order to execute a given experiment. Observing the operation of this manager, sketched in Fig 4, it can be seen that the compositions created can be persisted, mapped into ontologies and made available for searching on e-ScienceNet. Thus, the Composition Manager enables the reuse and automation of scientific experiments, because a researcher does not need to create an experiment from scratch and can find a composition that already exists and can modify it to his/her needs. However, the main feature of the Composition Manager is the management and automation of composition of services for the specification of a scientific experiment, considering the semantic peers. As a result, it facilitates collaboration among scientists dispersed throughout the network.

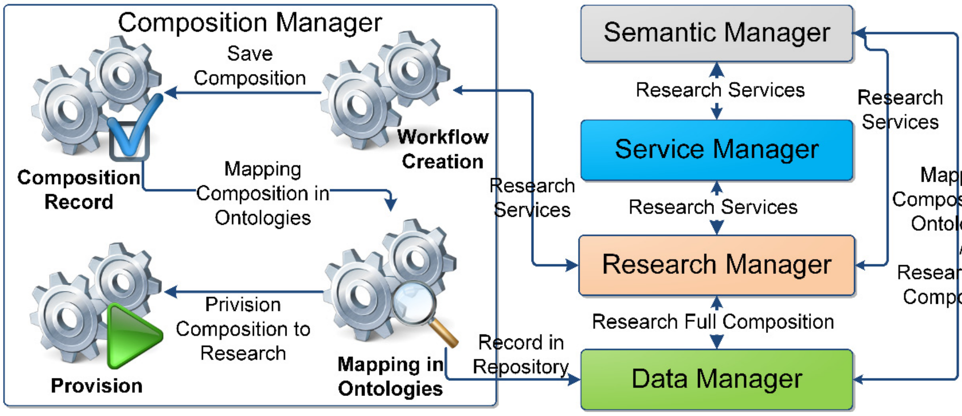


Fig. 4. Composition Manager operation.

- Interest Manager: module responsible for the configuration of the semantic interest group for each peer. An interest group is created based on a given domain ontology. A new peer when requesting its first connection to e-ScienceNet must select an interest group that will be connected. Initially, the peer has no knowledge about which groups are being provided by the system, then the Interest Manager sends a broadcast message to other peers connected to it, requesting information about which groups they belong to, considering that these peers can belong to any one of available groups in the network. These peers can query its local PeerOntology to obtain the groups. Each group is related to a given domain ontology and a domain ontology can be related to one or more groups. With this information, they return it to the new peer via an XML file. Thus, the new peer persists the obtained information in its own PeerOntology and select which of the available groups it wants to join. This process is detailed in Fig 5.

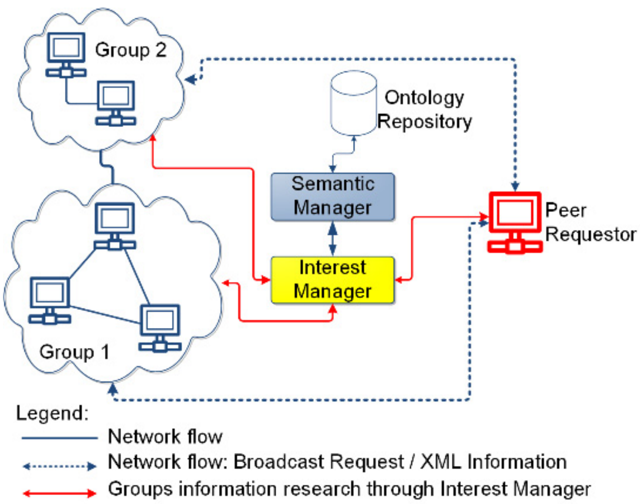


Fig. 5. Groups broadcast request.

### 4.1. E-ScienceNet Prototype

The Fig 6 shows the graphical interface of e-ScienceNet, where we can identify: the connection of a given peer in a particular interest group; peers that are currently connected to a given group, the information about resources that are being shared by members (peers), an interface to exchange messages between peers via chat, and the search interface that is connected to the Research Manager, in order to search for resources (data and services) on the network.

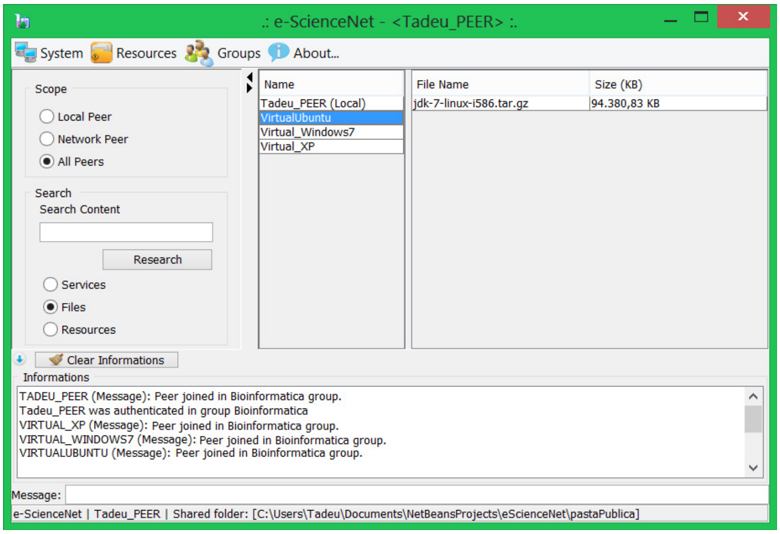


Fig. 6. e-ScienceNet interface.

The creation and exchange of interest groups (without ontologies) as well as the inclusion, sharing and visualization of files and their information, are already developed. These functionalities are presented in Fig 7.

Some of the technologies used in e-ScienceNet were (i) the Java programming language, (ii) the JXTA platform [20], which provides solutions for peer to peer network development, (iii) the OWL [21], for ontologies development and (iv) the JENA framework [22], enabling the use of semantic web, together with inference engines and XML query languages (SPARQL [23]). These technologies allowed the e-ScienceNet development according to P2P paradigm and use of semantic web.

### *4.2. Proof of Concept*

Aiming to present the feasibility of e-ScienceNet in scientific context, we present in this section an usage scenario, which serves as a proof of concept for the approach. Considering a group of Brazilian scientists which use e-ScienceNet to support the execution of their experiments, based on the creation of scientific service flows, results sharing, dissemination of knowledge and validation tests.

Each researcher (scientist) is connected to e-ScienceNet and can share on the network resources, services and documents, which are fundamental for the creation and execution of their research. In this configuration, each scientist is considered a peer of e-ScienceNet and also part of semantic interest groups related to their research goals. Documents and services made available by a peer can be searched by peers at the same interest group. Through these searches, one can specify service flows considering the available services and considering also the semantic compatibility between these services and their inputs and outputs. As an example, considering previous author's research, Da Silva et al. [4] and Costa et al. [24], it is possible to perform experiments involving genetic sequencing subdomain, related to Bioinformatics domain. More specifically, research related to nucleotides and DNA sequences. With this specific interest, a scientist can use the system to conduct research on peers that are inserted into groups that focus on this domain.

Based on the search results, scientist can select the services that are more adequate and, with the help of Composition Manager, can compose service flows to support his experiment execution. Fig 8 details the composition functionality.

Considering issues such as reuse and automation, e-ScienceNet allows that specified compositions are persisted in a local resource repository managed by the Data Manager. Additionally, these compositions are mapped to ontologies related to peer in order to make possible its reuse in further experiments. Therefore, a scientist can search for compositions that have already been specified, and then simply select one to reuse, changing it in order to be modified to his experiment. Thus, we consider that the use of e-ScienceNet can save time to scientists because it can offer reusable knowledge and resources according to their interest and research areas.

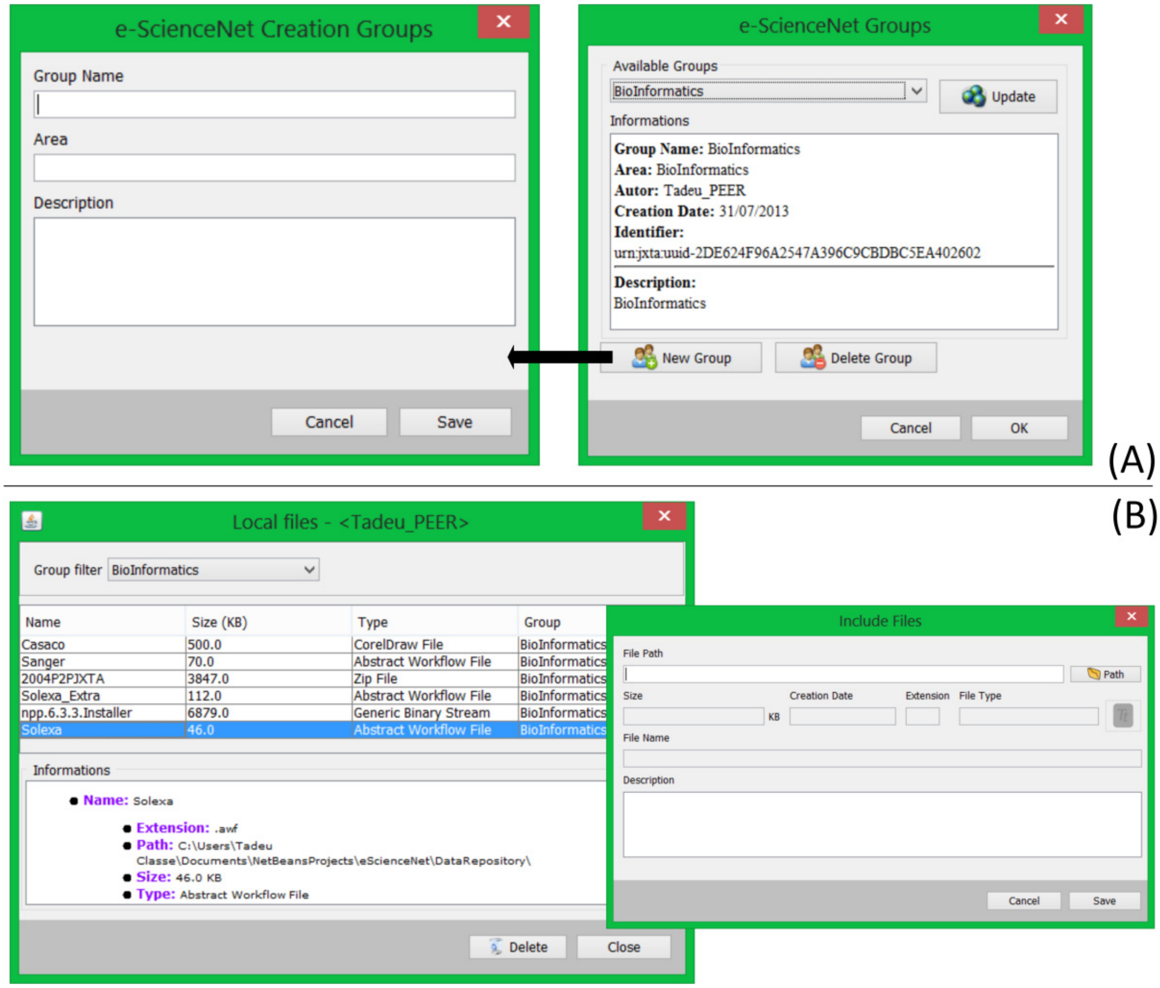


Fig. 7. (A) - Groups information and groups creation. (B) - Resource and file information.

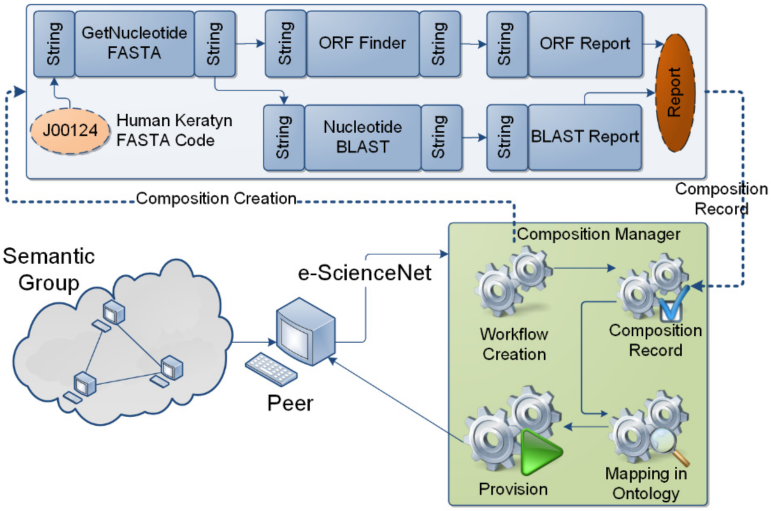


Fig. 8. Composition creation schema.

## 5. Final Remarks and Future Works

The development of scientific experiments is a complex process where different methods and services can and should be used together for their construction. It is believed that with the e-ScienceNet, compositions of services in order to execute experiments can be developed and shared, helping scientists and enabling reusability and automation. This composition of services is possible through the use of two managers: Composition Manager and Service Manager, which is aimed at facilitating the automation of service composition for conducting distributed scientific experiments.

Beyond the Composition and Service Managers, some other managers related to the semantic functionalities of e-ScienceNet were also detailed, with emphasis on the Semantic Manager and Interest Manager, which have functionalities that enable semantic discovery of services and creation of semantic groups in the network, facilitating the connection between scientists according to their research interests.

E-ScienceNet is still a prototype and some points still need to be improved and provided. It is necessary to evolve the Interest Manager, considering associations among domain ontologies and PeerOntology, allowing semantic inferences in order to improve resources selection for each new member. In this context, the use of a third ontology (MappingOntology) for mapping PeerOntology with domain ontologies is being discussed. Furthermore, new scientific services must be included in e-ScienceNet, so that they can be made available, queried by the Research Manager, and used by the Composition Manager.

## References

[1] Medjahed, B. and Bouguettaya, A., 2011. Service composition for the Semantic Web. Springer.
[2] Lua, E. K., Crowcroft, J., Pias, M., Sharma, R, Lim, S., 2005. A survey and comparison of peer-to-peer overlay network schemes. IEEE Communications Surveys and Tutorials, 7(2), pp. 72-93.
[3] Berners-Lee, T., Hendler, J., Lassila, O., 2001. The Semantic Web. Scientific American, 284(5), pp. 28-37.
[4] Da Silva, L. M., Braga, R., Campos, F., 2012. Composer-Science: A Semantic Service Based Framework for Workflow Composition in e-Science Projects. Information Science, 186(1), pp. 186-208.

[5] Bussler, C., 2001. “The Role of B2B Protocols in Inter-Enterprise Process Execution”. In Proceedings of Workshop on Technologies for e-Services (TES 2001) (in cooperation with VLDB2001). Rome, Italy.

[6] Cingil, I. and Dogac, A., 2001. An architecture for supply chain integration and automation on the internet. Distributed and parallel databases, 10(1), pp. 59-102.

[7] Shim, S. S. Y., Pendyala, V. S., Sundaram, M., Gao, J. Z., 2000. Business-to-business E-Commerce Frameworks, IEEE Computer, 33(10), pp. 40-47.

[8] Haase, P., Schnizler, B., Broekstra, J., Ehrig, M., Van Harmelen, F., Menken, M.; Mika, P., Plechawski, M., Pyszlak, P., Siebes, R., Staab, S., Tempich, C., 2004. “Bibster - a semantics-based bibliographic peer-to-peer system”. In III International Semantic Web Conference (ISWC 2004). Hiroshima: Japan.

[9] Bianchini, D., De Antonellis, V., Melchiori, M., 2010. P2P-SDSD: On-the-fly Service-based Collaboration in Distributed Systems. International Journal of Metadata, Semantics and Ontologies, 5(3), pp. 222-237.

[10] Di Modica, G., Tomarchio, O., Vita, L., 2011. A p2p based architecture for semantic web service discovery. International Journal of Software Engineering and Knowledge Engineering, 21(7), pp. 1013-1035.

[11] Hendler, J., 2003. Science and the Semantic Web. Science, 299(5606), pp. 520-521.

[12] De Roure, D., Jennings, N., Shadbolt, N., 2003. The Semantic Grid: A Future e-Sciencee Infrastructure. Concurrency and Computation: Practice and Experience.

[13] De Roure, D., Jennings, N., Shadbolt, N., 2001. Research agenda for the semantic grid: a future e-Science infrastructure. Report commissioned for EPSRC/DTI Core e-Science Programme.

[14] Tanenbaum, A. S. and Van Steen, M., 2006. *“Distributed systems: principles and paradigms”*. Second ed. Upper Saddle River: Pearson Prentice Hall.

[15] Stuckenschmidt, H., Van Harmelen, F., Siberski, W., Staab, S., 2006. “Peer-to-peer and semantic web”. In Semantic Web and Peer to Peer. Berlin: Springer.

[16] Gharzouli, M. and Boufaida, M., 2011. PM4SWS: A P2P Model for Semantic Web Services Discovery and Composition. Journal of Advances in Information Technology, 2(1), pp. 15-26.

[17] Di Modica, G., Tomarchio, O. Vita, L., 2011. Resource and Service Discovery in SOAs: A P2P Oriented Semantic Approach. International Journal of Applied Mathematics and Computer Science, 21(2), pp. 285-294.

[18] Wilson, B. J., 2002. *“JXTA”*. Indianapolis: New Riders Publishing.

[19] Gradecki, J. D., 2002. *“Mastering JXTA: Building Java Peer-to-peer Applications”*. Indianapolis: Wiley Publishing.

[20] JXTA, 2013. JXTA(TM) Community Project [Online]. Available at: <http://java.net/projects/jxta/> [Accessed 3 April 2013].

[21] Web Ontology Language (OWL), 2013. W3C Semantic Web [Online]. Available at: <http://www.w3.org/2004/OWL/> [Accessed 3 April 2013].

[22] Apache JENA, 2013. Apache Jena [Online]. Available at: <http://jena.apache.org/> [Accessed 3 April 2013].

[23] DuChame, B. 2011. *“Learning SPARQL”*. Sebastopol: O'Realy Media.

[24] Costa, G. C. B., Braga, R., David, J. M. N., 2012. “Connecting Feature Models and Ontologies in Software Product Lines”. In XIV International Conference on Enterprise Information Systems (ICEIS 2012). Wroclaw, Poland.