Creating, Protecting and Sharing of Semantically-Enabled User-Orientated Electronic Laboratory Notebook in a Service Environment

Tahir Farooq
School of Computing, University of Leeds
Leeds, United Kingdom
tahir_farooq@hotmail.com

Zulkifly Mohd Zaki
Faculty of Science and Technology, Universiti Sains Islam Malaysia, Nilai, Malaysia
zulkifly@usim.edu.my

***Abstract*— We discuss in-addition to our previous paper, the creating, protecting and sharing of semantically-enabled, user-orientated Electronic Laboratory Notebook (ELN) in a service environment. The advantages are that ELN service manages the end-to-end provenance life cycle on behalf of the community modellers regardless of local ELN system which involves individual ELN support, ELN version control and hardware/software issues. The performance of the solution presented showed a reasonable degree of end-user acceptance of the proposed approach.**

Keywords-electronic laboratory notebook; collaborative software; sharing; fine-grained access; online; service environment.

I. INTRODUCTION

In our previous research [1], the Dynamic Role-Based Access Control (DRBAC) mechanism was discussed to protect and share the ELNs in a distributed, co-laboratory research environment. The ELN-DRBAC mechanism was used to allow community modellers to transfer locally created ELNs into a central repository where they can share their personal ELNs with other modellers in the community either as a whole ELN object or its elements (provenance trails) at fine-grained level. An evaluation was undertaken with the members of atmospheric chemistry community working on the EUROCHAMP-2 project [2]. The scientific goal of this community is to better understand the chemical processes (reactions) taking place in the lower atmosphere through the use of atmospheric simulation chambers. These processes can have significant impacts on both air quality and climate change.

In this paper, we extend our previous work to address the issues when creating ELNs in a local environment. The Electronic Laboratory Notebook - Creating, Protecting and Sharing (ELN-CPS) service provides an environment for the modellers to create, protect and share ELNs online without the use of local ELN system. This improves the consistency of conducting modelling experiments across the community. The local ELN system [3] was specifically designed to capture and retrieve high quality metadata concerning the modelling process together with modeller's reasoning. However, the analysis of the local user-orientated ELN system showed the following:

i) It was difficult to maintain the different versions of ELN modelling metadata; and
ii) The complexity of settings and configurations of the local ELN system requires technical assistance before modeller can start using it [3]. This may hinder the modeller from using the ELN.

The ELN-CPS service environment encapsulates the functionality of the local ELN system and ELN protecting and sharing mechanism. It consists of two main parts:

i) ELN protecting and sharing control which is based on dynamic role based access control; and
ii) ELN simulation service and associated sub services (i.e., ELN retrieval and IPNav services).

In this paper, the following contributions are made:

i) The existing user-orientated ELN system could be used across multiple platforms in a service environment. The ELN-CPS service is made online for creating, protecting and sharing of modelling metadata. It also addresses the need to set up the ELN system.
ii) The management of different versions of modelling metadata to ensure modellers can retrieve the relevant and correct metadata across different versions of simulations.
iii) The central archive mechanism of ELNs, particularly for group simulations are defined to avoid any disaster situation at local level.

This paper is structured as follows: Section 2 highlights the background of this research while Section 3 details the requirements for the ELN-CPS service. The ELN-CPS service architecture is discussed in Section 4 which is then used to elicite feedback from members of the atmospheric community in Section 5. Section 6 reviews the related work. In the last section, we conclude and present the future work.

II. BACKGROUND

The high level view of Electronic Laboratory Notebook Protecting and Sharing (ELN-PS) system is shown in Figure 1. The main function of the ELN-PS system is to enable the modellers to securely transfer and share their personal ELNs within or across the research laboratories. In other words, it is an interface to transfer local ELNs into the central ELN repository for sharing purposes.

The key component in ELN-PS system is the Electronic Laboratory Notebook Dynamic Role-Based Access Control (ELN-DRBAC) mechanism that manages the access control to transfer, protect and share ELNs. The use of ELN-DRBAC gives clear understanding to the people: what their responsibilities are and to whom they are giving access. The

 ISBN: 978-1-61208-389-6

typical NIST RBAC as discussed in research [1] is enhanced to DRBAC because a person in the community (with the right permissions) may need to share a whole ELN or a selection of it (provenance trails) at a fine-grained level.

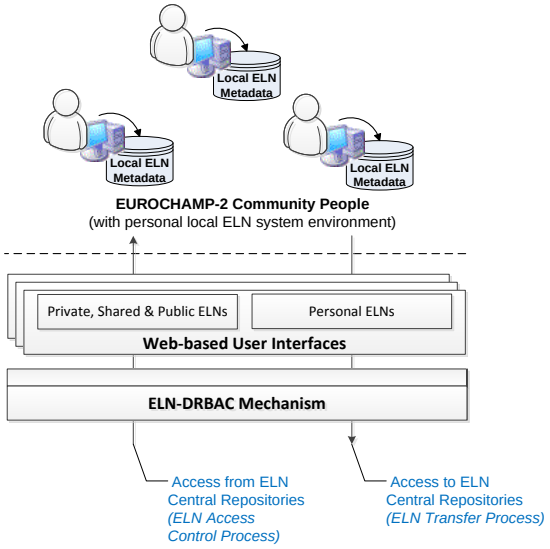

Figure 1. High level view of ELN-PS system.

In ELN-PS system mechanism, two dynamic allocations were introduced: a) persons to roles; and b) roles to permissions. The objective was to open the strict binding of three main components of the traditional RBAC system: users, roles and permissions to allow them to act dynamically within the RBAC framework. For example if a person with "modeller private" role wants to allow another person to download his/her personal ELNs, this mechanism gives the option to allocate "Download ELNs" permission to any person, independent of role allocation. This limited allocation of permission means that only those ELNs can be downloaded which are authorized by the ELN owner. Further a role may have different descriptions for different research groups. The "system manager" can activate and de-activate multiple permissions to any role according to the run time requirements of the domain people. In ELN-PS system, just like role sessions, the allocation of permissions to the roles is managed on the basis of permission sessions. So when a person uses ELN-PS system to protect and share ELNs, along with dynamic person-role sessions, the role-permission sessions are also created dynamically. This provides high level of dynamic inheritance in the ELN-PS system. Further detail about the ELN-PS system is referred to our previous research [1].

III. REQUIREMENTS FOR ELN-CPS SERVICE

In Section 2, the ELN-PS system is discussed to protect and share the generated ELNs locally in a co-laboratory research environment. The ELNs are valuable resources for e-Scientists as it helps with: the repeatability of experiments, tracking simulation runs, managing the data generated, verifying experiment results and acts as a source of simulation insight [4].

The ELN-CPS service environment is shown in Figure 2 which encapsulates the functionality of the local ELN system. There are two main parts: a) ELN protecting and sharing control (ELN-DRBAC mechanism); and b) ELN service and associated sub services (i.e., ELN retrieval and IPNav services).

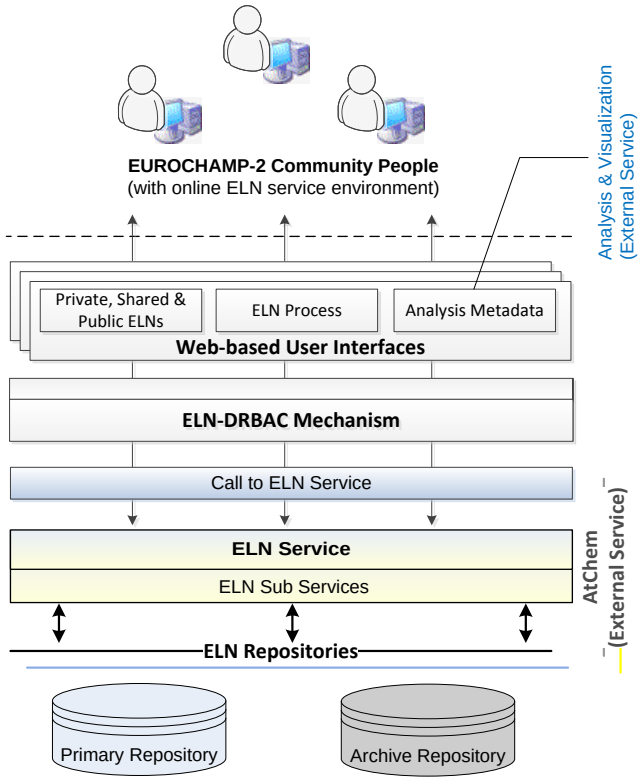

Figure 2. ELN-CPS service environment.

At this point, it is worth to note that the initial user requirements were captured in [1][2]. In this service, two types of service interfaces are required as given below.

Type 1 - Owner's active ELN: Owner of the ELN needs access via three functions: a) initial setup of a new ELN (or retrieve from the repository); b) the provenance capture for the next simulation step (this includes access to the inline provenance node navigator (IPNav) service to update current provenance trails); and c) transfer of ELN to the repository (after completing the simulation). Owner of the ELN can invite other members (with the necessary permissions) in the community to read and comments on the current set of the simulation steps. The comments will be recorded in the repository. For example, the research manager can view and comments on the state of modeller's personal ELN.

Type 2 - Third Party User: Retrieval of the ELN using the functions provided in the ELN-PS system. These are accessed through two services: Retrieve service (search

 ISBN: 978-1-61208-389-6

criteria and output of the list of matched ELNs); and ELN archives (ELN management, etc.).
These requirements arise from the Qualitative user-orientated evaluation was used to assess the meeting of the requirements and value of the ELN-CPS service. This is discussed in Section 5.

A. Scenario Cases

The scenario cases are used to develop the ELN-CPS service, which allows a modeller to create ELN simulations using Web-based interfaces, protected and shared by the ELN-PS system. The scenarios are derived from the working practices of the EUROCHAMP-2 community members, though remain generalisable to other communities with similar requirements.

Part-1: Creating ELN in the service environment

Helen is a modeller working in her laboratory. She is informed by the system manager that ELN system is now available in the service environment and she can access it through the Web-based interfaces. This service facility is provided to secure all ELNs centrally to avoid any disaster situation and to overcome the ongoing technical issues (i.e., ELN version control and the complexity of setting up and configurations) at the user's local ELN system.

Helen logs into the ELN-CPS service by activating "modeller private" role. She initiates a new version of simulation called "toluene chamber" simulation and shares the whole ELN with her "research manager". Helen understands that there is no need of transferring local ELN into the central repository because the ELN is already created online in the central repository and it is protected.

Part-2: Sharing of ELN in the service environment

Helen (modeller) and Alvin (research manager) exchange comments on a new simulation with different trails and finalize the results for evaluation. The golden trail is then shared with the publication editor for evaluation.

IV. ELN-CPS ARCHITECTURE

The ELN-CPS service architecture is designed with the vision to operate user-orientated ELN system in a distributed, co-laboratory research environment. The constraints of the existing ELN system architecture and framework [3] are also considered during the design phase. The ELN-CPS service architecture is designed based on the distributed computing model [5], as shown in Figure 3.

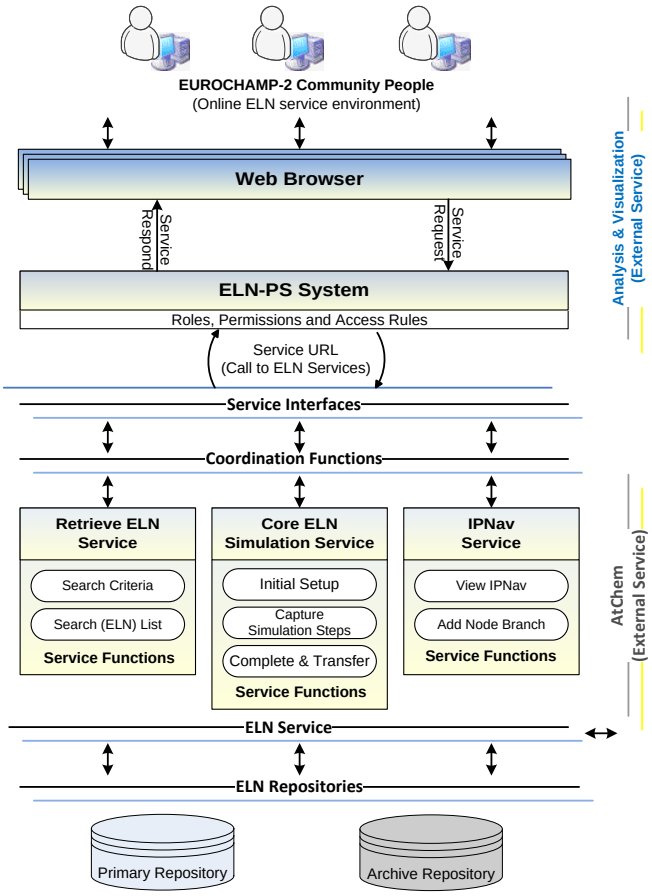

Figure 3. ELN-CPS service architecture.

The architecture grants the maximum configuration flexibility, because each function is mapped to a distinct physical (ELN development service) layer, which can be independently replicated. It presents the interrelation between different components of the ELN service, all accessible through the Web interfaces. In order to ensure loose coupling at the front-end of ELN-PS system, the proven Model-View-Controller (MVC) architecture [6] is embraced that efficiently handles session management for multiple Hypertext Transfer Protocol (HTTP) requests. The user interfaces as Java Server Faces (JSF) [7] does not contain any processing logic. These pages represent view part of MVC. Links in the user interface that requires processing logic to be executed submits the request that is mapped to an action. The action then selects and invokes the required processing logic. The processing logic is encapsulated in plain Java objects deployed as standard service.

The ELN-CPS service is developed using the JSF framework. All functions of the existing local ELN system are reused for this service. The service responses (i.e., send and receive) are managed according to the JSF framework. The ELN-CPS service keeps the detail of the service URL and invokes the service when the request for "Develop ELN" is generated from the modeller side. The sub services

 ISBN: 978-1-61208-389-6

are executed inside the main ELN service interface. JSF is a MVC framework that is capable of managing the front and middle tiers. To meet this goal, the design of the ELN-CPS service is based on the core JSF patterns as well as the industry standard development guidelines of scalability, flexibility and platform infrastructure. To serve the web pages, the Apache Tomcat web server is used which is one of the fastest and easiest to configure Java application servers [8]. The ELN-CPS service is implemented via the web. The modeller can access the service by using any web browser (e.g., google chrome, Mozilla Firefox, etc.) that they preferred. The core purpose of the scenario cases given in section 3 were to:

i) Create the ELN simulations in a Web-based environment. Existing user-orientated ELN system only allows the creation of ELNs in a local environment.
ii) Retrieve the ELN simulations using ELN service. It gives an advantage to the modellers to load the personal and shared ELNs into the online simulation interface for further simulations.
iii) Protecting the ELN-CPS service through ELN-PS system.

Figure 4 shows a service interface in the ELN-CPS service environment to create a new simulation for a modeller who has assigned "modeller private" role. The modeller can invoke the service using "Develop ELN" function, which generates a request call and returns the service interface to create new simulation in the Web browser.

Figure 4. Creating new simulation.

This interface contains the basic parameters like simulation name, simulation type, EUROCHAMP chambers, etc. After setting up the initial parameters, the submit function is used to proceed further in mechanism development. To analyze the model output, the modeller has the option to download input/output model files in .zip format and can use their own tools as an external service.

V. EVALUATION WITH END USERS

The qualitative evaluation is presented in this section. This evaluation was the extension of our previous evaluation work [1]. The goal was to determine the potential value, likely advantages and disadvantages of using the ELN-CPS service. The evaluation plan included the demonstration of the two scenario cases (mentioned in Section 3) developed for ELN-CPS service and the collection of end user's feedback using a specific evaluation questionnaires, designed for this purpose. Prior to this evaluation, the informal evaluations feedback were also captured during different presentations and discussions with the community members. However, due to geographical distributed locations of the community members and time constraint, we prefer to start formal evaluation process with two key evaluators having substantial experience of developing atmospheric chemistry models. Both of the evaluators regularly perform *in silico* experiments that make use of the Master Chemical Mechanism (MCM), so the evaluators could easily relate to and understand the ELN functionality they were presented with. The use of a small number of potential users, with close links to the software development team, to evaluate scientific software has also been applied successfully in the large e-Science projects, such as myGrid [9][10] and myExperiment [11][12]. By acting to meet the requirements of a small number of local, well known scientists (who acts as pioneers); whilst thinking about the requirements of the wider user community, a widely adopted software application can be developed.

A standard marking method using a likert scale was used to assess the answer of each question in the evaluation. Values were ranked from 1-5; 1 = poor - 5 = excellent. After the demonstration of each scenario case, the users were requested to answer the questionnaires and comments in the appropriate boxes. The answers obtained from the questions evaluated by two members of the community are given in

.

TABLE 1. ANSWERS FROM USER SURVEY.

Questions To Users		User	
		1	2
i)	Do you acknowledge the concept of ELN-CPS service?	4	4
ii)	Do you think the ELN-CPS service can overcome the maintenance issues, created in the local ELN system for the individual modeller?	3	4
iii)	For security and protection of ELNs, do you think it is good to create, store and maintain ELNs online?	2	3
iv)	Will you recommend the implementation of the ELN-CPS service in your laboratory?	3	3

At the end of the evaluation, the recommendations and comments from the participants were recorded. Samples of participant's recommendations and comments are provided below:

Recommendations and comments:

User 1:

i) "Community needs to trust the party who controls ELN online with their scientific legacy – how do you gain their trust?"
ii) An option to "back up" the private online ELN on their own private systems would promote the use of this service?"
iii) "Need third party to look after the online service (need guarantees), e.g., British Atmospheric Data Centre (BADC)".

User 2:

i) "Definitely the service can overcome the issue of maintenance. In local ELN, if you upgrade the functionality then you need to reinstall the ELN in every modeller's machine that use the local ELN. If using online, upgrading of the functionality would be easier".
ii) "If I can trust the online system, then I will use it. But you need to think about how to persuade people to trust the service".

The evaluation results were very encouraging and both participants acknowledge the value of this service. These results could be considered as an initial feedback before going into the larger community for further evaluation. The major concern in the evaluation was regarding the trust relationship, i.e., how community people will be agreed to create ELNs outside their personal system? For this, the participants have given the following suggestions:

i) To enhance the trust, there is a need to have an option for the end-users to store online ELN in their personal systems too;
ii) The service should allow the modeller to choose either to perform their simulation runs locally or via online;
iii) Need third party to look after the online service with their scientific legacy, e.g., British Atmospheric Data Centre (BADC) [19]. Guarantees should be signed by third party to ensure the security of the online ELN repositories.

VI. RELATED WORK

A service is an implementation of a well-defined logic functionality (in this research, it refers to the ELN service), and such services can then be consumed by clients (i.e., end-user application or another service) in different applications or processes [13]. Service Oriented Architecture (SOA) [20] is an architectural style for building software applications that use services available in a network such as the Web. SOA offers flexible approaches to distributed systems engineering including reuse of the existing applications.

Within the EUROCHAMP-2 community, the previously reported ELN [2] was based on a standalone local ELN to capture and retrieve the quality metadata performed by the modellers in a laboratory. However, there was an issue identified which was the need to reduce the complexity of setting up the ELN before it can be used [3]. Initially, virtualization [14] which provided a prefabricated environment for the ELN was implemented to address this issue; however, there is still a need for the ELN to maintain different versions of ELN metadata based on the role of the modellers. Two prominent projects related to this research are discussed. Although there are sheer number of work done related to this research, but we are only interested in the features aspects presented in those projects.

Linked Environments for Atmospheric Discovery (LEAD) is a large scale project to address meteorology research challenges to analyze and predict the atmosphere. Scientists are provided with necessary tools such as automated search, selection and transfer of required data products between computing resources [15] to build forecast models using model generated data and manage necessary resources for executing the model [16]. This includes: a) replacing the manual data management tasks with automated data discovery; b) allow transfer of large scale data products between resources; and c) searching and access of the data via GUI interface and underlying ontology [15]. LEAD project is deployed via web portal. This work seems similar to our research where features such as data discovery and transfer between computing resources are present. However, the ELN-CPS provides access and share simulations metadata at a fine grained level according to the modeller's role.

Open Source Project for a Network Data Access Protocol (OPeNDAP) is a system that facilitates scientific data archival and exchange between researchers [17]. The main feature of OPeNDAP is that it allows access to the data from a multiple applications. OPeNDAP uses the client/server model that utilizes the browsers to submit/receive requests/respond to the servers. It also allows researchers to browse and request data to be translated into a specific format. Three data object types provided by OPeNDAP [18], i.e., Data Descriptor Structure (DDS), which describes the structure of the data set, Data Attribute Structure (DAS) semantic metadata which gives the attribute values of the fields described in the DDS and the actual data in a binary structure. The different between OPeNDAP and ELN-CPS is that in the latter, it allows access and sharing of scientific data between modellers through the inline provenance node navigator (IPNav) and modeller's role. This includes access and sharing of the whole ELN or a selection of it at a fine-grained level in the trail.

VII. CONCLUSION AND FUTURE WORK

The research presented in this paper was the continuation of our previous research work. The aim was to introduce the ELN creating process in the service environment,

protected and shared through the ELN-PS system. The ELN-CPS service was developed and implemented to allow community modellers to create ELNs online and stored in the central ELN repository. This improved the consistency in the use of the ELN across the community and the management of the updated versions. The qualitative evaluation illustrated, how ELN-CPS service could be understood and accepted by a research community.

Future work will be to evaluate the ELN-CPS service in the wider ELN community. In order to get more conclusive results on the value of the ELN-CPS service, a larger set of ELN modellers is needed for evaluation. However, before going into the wider community for further evaluation, the issue of establishing a trust relationship between end-users and service providers needs to be addressed with robust plans for the safe storage of ELN data. This include considering the feedback from the evaluation which are to: 1) provide an option for the modelers to store their online ELNs into their local personal systems; 2) allow the modelers to perform their simulation runs locally or via online; and 3) allow the scientific legacy such as British Atmospheric Data Centre (BADC) [19] to look after the online service to ensure the security of the online ELN repositories. However, these recommendations need further investigation before it can be implemented.

ACKNOWLEDGEMENT

The research work presented in this paper was carried out at the School of Computing, Leeds University, supervised by Professor Peter M. Dew.

REFERENCES

[1] Tahir Farooq, Richard Kavanagh, Peter Dew, and Zulkifly Mohd Zaki, "Protecting and Sharing of Semantically-Enabled, User-Orientated Electronic Laboratory Notebook Focusing on a case Study in the e-Science Domain," The Sixth International Conference on Future Computational Technologies and Applications, FUTURE COMPUTING 2014, May 2014, pp. 50-58.

[2] Zulkifly Mohd Zaki, Peter M Dew, Lydia LMS Lau, Andrew R. Rickard, Jenny C. Young, Tahir Farooq, Michael J Pilling, and Chris J. Martin, "Architecture design of a user-orientated electronic laboratory notebook: A case study within an atmospheric chemistry community," Future Generation Computer Systems, vol. 29, Oct. 2013, pp. 2182–2196.

[3] Z. Mohd Zaki, P. M. Dew, M. H. Haji, L.M.S. Lau, A. Rickard, and J. Young, "User-orientated electronic laboratory notebook for retrieval and extraction of provenance information for EUROCHAMP-2," Pro. IEEE Seventh International Conference on eScience (ESCIENCE 11), Dec. 2011, pp. 371-378.

[4] M. Greenwood, C. Goble, R. Stevens, J. Zhao, M. Addis, D. Marvin, L. Moreau, and T. Oinn, "Provenance of e-Science Experiments - experience from Bioinformatics," The UK OST e-Science second All Hands Meeting (AHM'03), pp. 223-226, [In UK e-Science All Hands Meeting, Nottingham, UK, Sep. 02-04, 2003].

[5] Systems and software engineering - Architecture description, International Standard ISO/IEC/ IEEE 42010, IEEE, 2011.

[6] Marty Hall, Integrating Servlets and JSP: The Model View Controller (MVC) Architecture, Prentice Hall and Sun Microsystems Press, 2010.

[7] R. Hightower, JSF life cycle, conversion, validation, and phase listeners Skill Level: Introductory, IBM Corporation, 2008.

[8] Xue Liu, Heo, J, and Lui Sha, "Modeling 3-tiered Web applications," 13th IEEE International Symposium, Sept. 2005, pp. 307-310, DOI:10.1109/MASCOTS.2005.40.

[9] D. Hull, K. Wolstencroft, R. Stevens, C. Goble, M. Pocock, P. Li, and T. Oinn, "Taverna: a tool for building and running workflows of services," Nucleic Acids Res, vol. 34 (Web Server issue), pp. 729-732, 2006.

[10] T. Oinn et al., "Taverna: lessons in creating a workflow environment for the life sciences," Concurrency and Computation: Practice & Experience - Workflow in Grid Systems, vol. 18, Aug. 2006, pp. 1067-1100.

[11] D. De Roure, and C. Goble, "myExperiment: A Web 2.0 Virtual Research Environment," International Workshop on Virtual Research Environments and Collaborative Work Environments, Edinburgh, UK, ePrint ID:263961, May 2007.

[12] Carole Anne Goble, and David Charles De Roure, "myExperiment: social networking for workflow-using e-scientists," In: Proceedings of the 2nd workshop on Workflows in support of large-scale science, June 2007, pp. 1-2, [ACM: Monterey, California, USA].

[13] Carlos Granell, Laura Díaz, and Michael Gould, "Service-oriented applications for environmental models: Reusable geospatial services", Environmental Modelling & Software, vol. 25 Issue 2, Feb. 2010, pp. 182-198.

[14] VMWare, "Virtualisation," http://www.vmware.com/virtualization, (retrieved: 15 February, 2015).

[15] Beth Plale, Rahul Ramachandran, and S. Tanner, "Data Management Support for Adaptive Analysis and Prediction of the Atmosphere in LEAD," 22nd Conference on Interactive Information Processing Systems for Meteorology,Oceanography, and Hydrology (IIPS), 2006.

[16] Plale, B., K. Brewster, C. Mattocks, A. Bhangale, E. C. Withana, C. Herath, F. Terkhorn, and K. Chandrasekar, "Weather Forecast Data from The D2I-Vortex2 project," Bloomington, Indiana: Data to Insight Center, June 2012.

[17] P. Cornillon, J. Gallagher, and T. Sgouros, "OPeNDAP: Accessing data in a distributed, heterogeneous environment," Data Science Journal, vol. 2, Oct. 2003, pp. 164-174.

[18] Galip Aydin, Service Oriented Architecture for Geographic Information Systems Supporting Real Time Data Grids, PhD. Thesis, Indiana University, 2007.

[19] British Atmospheric Data Centre, Natural Environment Research Council, http://badc.nerc.ac.uk/home, (retrieved: 15 February, 2015).

[20] T. Erl, "Service-oriented Architecture: Concepts, Technology, and Design", Upper Saddle River: Prentice Hall. ISBN 0-13-185858-0, 2005.

 ISBN: 978-1-61208-389-6