Protecting and Sharing of Semantically-Enabled, User-Orientated Electronic Laboratory Notebook Focusing on a Case Study in the e-Science Domain

Tahir Farooq, Richard Kavanagh, Peter Dew
School of Computing, University of Leeds,
Leeds, United Kingdom
tahir_farooq@hotmail.com, scsrek@leeds.ac.uk

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

***Abstract*— We discuss the addition to an existing Electronic Laboratory Notebook (ELN) system, a means to permit the sharing of modelling data. One advantage is that sharing of such data is a means of assisting the publication process. This is done by presenting the modelling data and the reasoning behind its creation. This sharing of data is managed in a user sensitive fashion by restricting the release of data based upon the role someone performs. Further sensitivity is shown by fine-grained access control, which permits only part of the ELN to be shown. The performance of the solution presented is reviewed via quantitative analysis that showed a reasonable degree of end-user acceptance of the proposed approach.**

Keywords-electronic laboratory notebook; sharing; privacy; fine-grained access.

I. INTRODUCTION

In science and engineering, many international communities of researchers employ complex computational models. Such communities often use paper-based laboratory notebooks [1]. Research has previously focused on encouraging scientist in these communities to use an Electronic Laboratory Notebook (ELN) to create, store and retrieve provenance data about modelling, as a means of providing consistency of recording provenance data. The ELN was specifically designed to capture and store high quality metadata for the modelling process and modeller's reasoning, whereas previously this provenance metadata was recorded in an *ad-hoc* and unstructured fashion. This is in contrast to ELNs for physical experiments [2], where meta-data is often captured in a structured fashion.

In this paper, we advance on this previous work by allowing users to fully share electronic records that meet the technical and scientific requirements of such communities [3]. We refer to this is a community ELN called ELN-PS (protection and sharing of ELN) within a distributed, multisite research environment.

Working with one such community, namely the Atmospheric Chemistry Community we aim to enhance sharing of the modeller's data and its associated meta-data for the betterment of the community. This community studies aspects of chemical reaction mechanisms that take place in the lower atmosphere (troposphere). This community relies upon a highly comprehensive database of chemical mechanisms to drive their modelling process. This database is known as the Master Chemical Mechanism (MCM)[4]. It acts as the benchmark for this community and as such records in the database are carefully evaluated. The MCM database describes the detailed gas phase tropospheric degradation chemistry of a series of Volatile Organic Compounds (VOCs). Acting as the benchmark for the community it has a wide variety of atmospheric science and policy applications where detail knowledge of chemical reactions is required. MCMv3.2 [3], for example, contains 6,700 species involved in 17,000 reactions. Members of this community are involved in ensuring that the last research is evaluated and where necessary updates are made to the relevant MCM entries. One aspect of a community based ELN is to support the MCM updating process. If reviewers are given detailed information about the modelling that has been performed in the community then they are better able to understand the simulation results presented and the reasoning behind them. This therefore makes the updating of this central database easier.

Simulation data and its associated meta-data is an important commodity which may also be used by modellers for supporting publications, in that if their reasoning and process to obtain results can be followed by reviews the results and publication can be reviewed more readily. This process however requires careful management of the access to the associated data so that it respects the publication and evaluating processes. In this paper, the following contributions are made:

- An architecture that permits ELNs to be shared in a fashion that respects the publication process, allowing for the reviewing of ELNs and controlling the sharing of ELNs within the community. This includes a means to protect ELNs in an end-to-end fashion between the modeller and reviewer.
- A mechanism for the sharing of part of an ELN. This allows for a particular series of simulations known as a trail to be shared. This means only the data relevant to the evaluation and publication process is shared and not all the work of a given modeller.
- Finally, an assessment of an impact of sharing data within the selected community is discussed.

These advances extend the previous work on the ELN for individual modellers which is fully reported in the earlier paper [5] and is summarised here in Section 2. A previous user evaluation [5] of this work showed that it vastly improved the efficiency of the modelling process, promoted

 ISBN: 978-1-61208-339-1

good practice and facilitate easy and transparent knowledge transfer. The ELN for the community is expected like its predecessor to be relevant to other communities such as those that use detailed chemical reaction mechanisms such as GRI-Mechanism [6] and fields such as astrochemistry.

The rest of this paper is structured as follows: Section 3 details the requirements for the sharing of ELNs and in particular the lifecycle of the release of ELN data. In Section 4 the platform is introduced that provides the provenance sharing. This architecture is known as ELN-Protection and Sharing (ELN-PS). A web-based implementation of this architecture is discussed in Section 5, which is then used to elicite feedback from members of the atmospheric community in Section 6 with a qualitative analysis of the ELN-PS system. In the last section, we conclude and present our future work.

II. BACKGROUND

The work presented here extends our previous ELN for individual modellers [5]. The previous system is hence described here briefly in order to assist the understanding of this paper. The existing ELN is made up of three main components, namely: the Core ELN, the inline provenance node navigator (IPNav) and the notebook retrieval (see Figure 1). These components are described next:

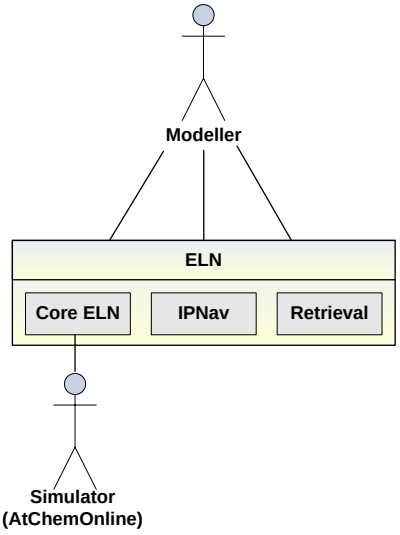

Figure 1. The ELN for Individual Modellers.

A. The Core ELN

This is principally responsible for executing simulation requests and recording all the parameters that go into the computational parametric modelling process. The modelling is performed by an external component called AtChem Online [7]. The core ELN records the output data from the AtChem modelling tool and links it to the provenance data, which indicates both the settings used as input and the user's original reasoning behind running the simulation. Simulations are performed iteratively and after each run the user is expected to change the parameters of the simulation, to further develop their model. The core records the step by step modelling process in a systematic and as far as possible, automated fashion. A feature of the core is that it supports the modeller in generating annotations to explain the rationale for making a parameter change at the time the change is introduced. Recording this reasoning at this point improves the quality of the annotations and their value to the modeller and other members of the scientific community once the notebook is shared. These annotations once made provide a narrative to the work of the modeller, giving complete coverage of their reasoning which includes both the successful and the unsuccessful formulations of the model. These annotations once combined with details of the modelling process provide the meta-data which we call the inline provenance of the model.

B. Inline Provenance Node Navigator (IPNav)

The IPNav [3] structures and displays to the end user the provenance as a graph/tree structure. It thus fully represents the inline provenance gathered as part of a series of successive interations of the model. It allows this provenance to be navigated and presents the modeller with the ability to compare different iterations of the model's development, using an inbuilt differencing tool. This viewer is particularly important for third party users of the ELN, whom of course did not develop the model and hence were not privy to the decisions and reasoning process of the original modeller.

C. Retrieval

The ELN retrieval function provides the ability to search and recover from the database past models which then allows the inline provenance node navigator to display the individual runs of the ELN. Its also allows the user to view both the experimental data and its provenance.

In addition to the advancements made with sharing, it should be noted since the previously reported version of the ELN was produced, an evaluation study identified that there was a need to reduce the time and complexity of setting up the ELN on a modeller's local computer. This was because it requires a number of third party software, namely: Python, Python Yet Another Markup Language (PyYAML), Natural Language Toolkit (NLTK), My Structured Query Language (MySQL), curl, diff, NetBeans and Java's Software Development Kit (SDK). This issue was resolved by using virtualisation which provided a prefabricated environment for the ELN.

III. REQUIREMENTS FOR ELN-PS

In Section 2, the ELN for individuals was discussed, including that of the generation of provenance meta-data. This provenance data that is generated assists the publication process and is a valuable resource for e-Scientists as it helps with: the repeatability of experiments, tracking experimental runs, managing the data generated, verifying experiment results and acts as a source of experimental insight [8]. The lifecycle and associated

 ISBN: 978-1-61208-339-1

requirements that govern the release of this provenance data are now to be discussed. The ELN lifecycle process is concerned with the management of the end-to-end provenance flow from the initial models creation to the use in the wider ELN community. This lifecycle, as shown in Figure 2, highlights the ELN protection and sharing requirements.

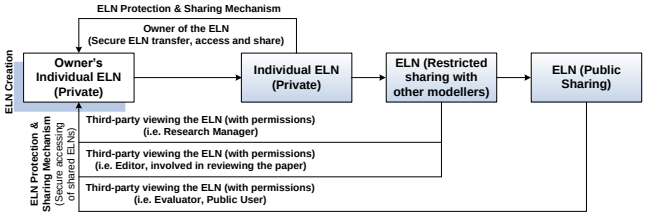

Figure 2. ELN Lifecycle Process.

Requirement 1: The principle of the ELN protection and sharing control is that the owner of the ELN (modeller) is required to share the personal ELN to the wider ELN community in a secure way. The protection and sharing of ELN thus has to be followed in a staged process. There is therefore three stages of release of an ELN in the wider ELN community: 1) "Private" so that only ELN owner has access 2) "Shared" enables ELN owner (modeller) to share personal ELN with other modellers 3) "Public", so that any community member (if ELN owner has allowed) can viewan ELN.

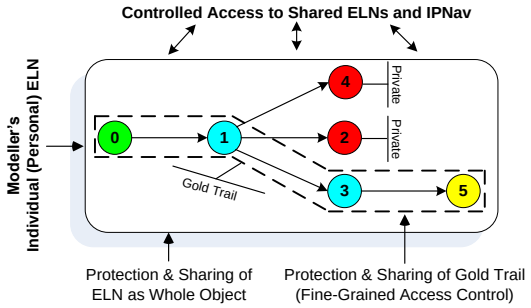

Figure 3. Fine-grained Access Control of ELN Trails.

Requirement 2: At some stage in the modelling process, community members may be interested in sharing only part of their personal ELN, i.e., access control to an ELN at a fine-grained is required. This is explored in Figure 3, where each simulation run is coded by colour indicating its position in the trail. For example green indicates the base run, red the dead ends and yellow highlights the gold/latest simulation run. Cyan means, the intermediate runs.

Fine-grained access control, is the application of protection and sharing rules to control access to parts of an ELN's provenance trails. This ensures modellers have the flexibility to share certain parts of their ELN trails with others in the community. Therefore, by default, every navigation node is tagged as "private" and the modeller has the choice of applying access control permissions from a pool of accessibility options. One such option is to share the trail, with the node that represents the best experimental case as the final node, this is known simply as the "gold trail".

Requirement 3: During the release of an ELN to the community it will be required for many different people to be able to access the data. These people will have different roles, such as a researcher's supervisor, or a reviewer. It will be required to moderate the access to a given ELN based upon these roles.

Evaluation of Requirements: To assess the meeting of these requirements a qualitative user-orientated evaluation to assess the value of the ELN protection and sharing mechanism will be performed (see Section 6).

A. Scenario Cases

The requirements are drawn from the following scenario cases, which are derived from the working practices of the atmospheric chemistry project EUROCHAMP-2 [3], though remain generalisable to other communities with similar requirements. The scenario cases are divided into three main parts: a) the sharing of a whole ELN; b) sharing of ELN provenance trails at a fine-grained level; and c) management of ELNs in the central repository. These cases highlight the relevant characteristics and working procedures of modellers sharing ELNs.

1) Part-1:Sharing of a whole ELN

Helen is a modeller working in her local laboratory. After finishing simulating a toluene chamber experiment on her local computer, she transfers the first version of ELN (H-v1) into the community repository using ELN-PS system. During the transfer process, the default access of the ELN is set to private and its owner Helen. Hence, the ELN is neither visible or accessible to anyone other than Helen.

Helen shares the first version of ELN (H-v1) with her research manager Peter to get feedback. Helen allows Peter to access all the trails of ELN (H-v1), i.e., the whole ELN. Peter as a research manager examines all simulation runs of the ELN and suggest some updates in run 5 of the simulation. Helen takes his advice and transfers the second version of ELN (H-v2) into the repository and shares it with Peter.

2) Part-2: Sharing of ELN provenance trails at a fine-grained level

After examining the final ELN, Peter advises Helen to allow Mark to access gold trail of the ELN (H-v2) for review purposes as part of publishing a paper. Mark acting as an editor examines the ELN gold trail of the toluene chamber experiment. The results in the latest/gold trail helps him to make a positive recommendation to publish the paper.

After publishing the modeller's results, Helen marks the gold trail of ELN (H-v2) as public thus making it available for other community members.

 ISBN: 978-1-61208-339-1

3) Part-3: Management of ELNs in the central repository

Jill, another researcher working on toluene experiment recently joined the research group. She searches through the ELN-PS system to find related ELNs in the community repository. The search retrieves the shared ELN(s). For the previous retrieved ELN only the gold trail is displayed, because as a public user she is restricted to only viewing the published trail.

Helen now leaves the research group and her user status is blocked by Lindsey who is acting as systems manager. During the routine management searchers on the ELN repository, Andrew as a data manager finds that the owner of the toluene ELNs has left the group. Andrew follows the research group policy and allows Jill to access all toluene ELNs created by Helen thus allowing her to proceed with her research.

IV. ELN-PS SYSTEM ARCHITECTURE

An overview of ELN-PS system architecture is shown in Figure 4. A modeller with their own individual ELN is given the option to transfer it to the community ELN repository, via the use of an ELN transfer protocol. The provenance information for a particular simulation and its runs is transferred as Resource Description Framework (RDF) [1] metadata. These metadata contain the process provenance and associated annotations for simulation runs.

A modeller at the start of the transfer process uses the simulation retrieval function in the ELN for individuals. Once a simulation is chosen, the ELN performs an export of the simulation data and provenance. The "Transfer ELN" function of the ELN-PS system then allows the modeller to import the selected simulation and its associated runs into the community ELN repository. In reverse order, the "Download ELN" function of the ELN-PS system allows the modeller to download individual ELN from the community ELN repository.

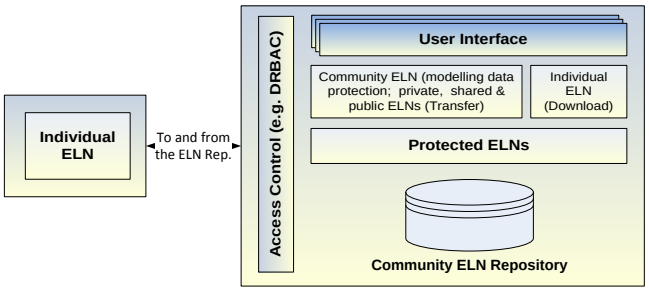

Figure 4. ELN-PS System Architecture.

The access control layer in the ELN-PS system is built on the Dynamic Role Based Access Control (DRBAC) framework. DRBAC provides authorisation to the community ELN repository based on the assigned roles of users. The reason for using role based access control is that, it gives a clear understanding of responsibilities to each user. The roles are defined according to the job competency, authority and responsibility within the organisations to regulate access to the ELNs. In eScience communities like EUROCHAMP-2, the roles change dynamically (e.g., a person may at different time perform the role of a research manager and at other times the modeller role). Further in the wider ELN community, the research laboratories may need to define a custom description in the ELN access control mechanism. It is therefore important to dynamically allocate roles to the users and dynamically allocate permissions to the roles. Roles are used to embody the authority and responsibility of the main actors of the community in the system. The responsibilities of such roles therefore guide the need for access to the ELNs secured in the repository. The role based access control in ELN-PS is based upon the National Institute of Standards and Technology (NIST) Role Based Access Control (RBAC) model [9]. Further details on DRBAC can be found in the related work section of the paper. The role hierarchy for ELN-PS system is shown in Figure 5. These roles and their associated permissions are assigned dynamically to the community members as required so that they may perform different tasks such as: transfer, share and view ELNs.

The ELN-PS system role hierarchy is organised into three categories:

i) *Group*: Member of the wider ELN Community which has three sub-roles namely; Modeller, Research Manager and Editor;
ii) *Public*: Member of Public Community Group which has three roles namely Evaluator, Public User and Public Blogger;
iii) *Admin*: Administrators; has two roles namely; System Manager and Data Manager.

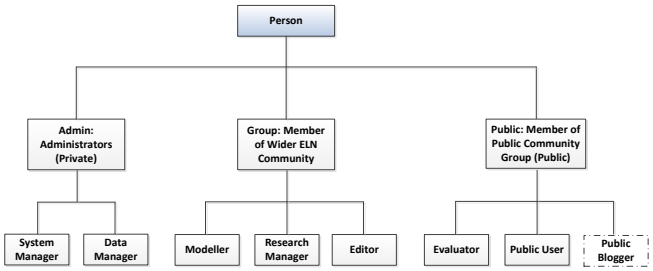

Figure 5. Role Hierarchy.

The modeller role, as shown in Figure 6, is further divided into three sub roles each of which deals with a different stage of release of the ELN data (see requirement 1):

i) *Modeller-Private role* can: a) transfer personal ELNs into the central repository; b) share the whole ELN; c) retrieve and view personal ELNs; d) download personal ELNs; and e) view/add comments.
ii) *Modeller-Public role* give permissions to the modeller to share and view gold/latest simulation trail of the ELN.

 ISBN: 978-1-61208-339-1

iii) *Modeller-Selective role* allows the ELN owner to share any simulation trail of the ELN with other modellers.

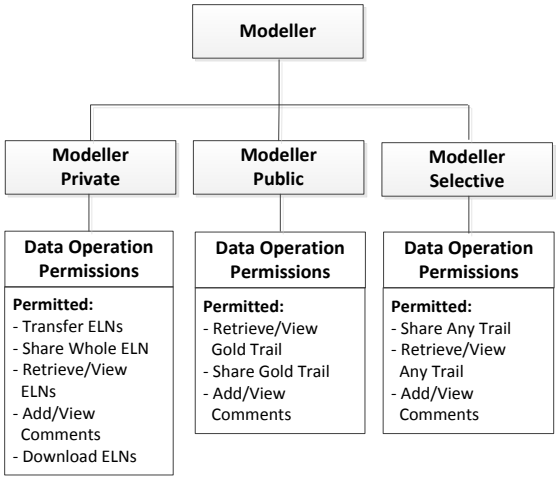

Figure 6. Modeller Role Properties.

The remaining roles are as follows: Research Manager allows a supervisor to view assigned researchers (modellers), view their shared ELNs and view/add comments. The Editor role allows the review process of a paper. It allows an editor to view the shared gold/latest trail and discuss it with fellow editors confidentially. The Public User role: can only view the final (publish) gold trail, provided it is shared by the ELN owner. The System and Data manager roles are associated with the management of the system including: archiving old ELNs, user management, roles management, role assignment, and role activation/de-activation.

The ELNs are transferred, shared and accessed through respective user interfaces of the ELN-PS system. The person-role and role-permission sessions are created dynamically within the system to open the static bindings of three main components of the traditional RBAC system: persons (users), roles and permissions. The access control layer in the ELN-PS system mainly addresses the authorisation process, which is based on the mapping of roles and permissions to ensure the person access to different services and functions. The user identification is done separately using a Form-based identification process [10]. The identification certifies the person credentials for the ELN-PS system. Figure 7 represents the internal view of person authorisation flow of the underlying security architecture. This is divided into two parts:

i) DRBAC Module
This module provides the allocation of community roles and associated permissions at run time for the entire session.

ii) ELN Access Control Module
The ELN access control functions and procedures to perform actions like Transfer ELNs, Share ELNs, View ELNs etc, which are provided in this module.

A person is authorised to access the community ELN repository according to the specific assigned roles. The authorisation process works with the generation of unique authorisation keys and security code for every person at run time. After the identification, when the access control request is received from the access control layer, the internal process of authorisation is started.

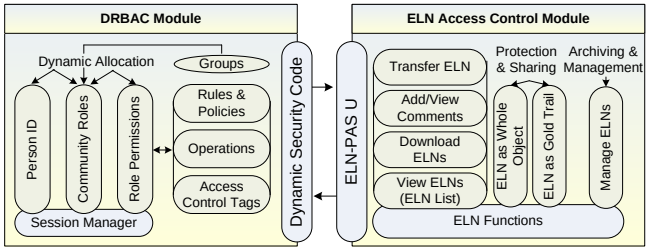

Figure 7. Authorisation Process in ELN-PS System.

Person, roles, permission and related identification keys are stored in the DRBAC database. If a person request authorisation two keys are generated. The first key contains the person and role identifications and the second key contains role and permission identifications. If a person is allocated multiple roles then multiple pairs of keys are generated for that particular person. The same mechanism is used when a role is allocated multiple permissions, i.e., the multiple keys are generated for that particular role. After successful processing of the unique authorisation keys, the secured information is forwarded for generation of a unique dynamic security code for every user. This security code is then combined with the "ELN-PAS U" (ELN Protection, Access and Sharing Unit) to access the ELN metadata. ELN-PAS U carries out the following operations:

i) Transfer ELNs: This allows modeller to select local ELNs using import function and transfer into the central repository.

ii) Share ELNs: Modeller can share personal ELNs as a whole or selected provenance trails with other community members like research manager, editor, evaluator etc.

iii) Access ELNs: This contains three types of access levels: a) "Private" so that only ELN owners have access on their personal ELNs; b) "Shared" enables modellers to share ELNs with other modellers and allows to view shared ELNs; and c) lastly "Public", so anybody can view public ELNs within or across the community.

iv) View/Add comments: This allows modeller to exchange comments privately with research manager or editor on a particular ELN or its provenance trail.

v) Archive and manage ELNs: This is for the data manager to archive and manage the: a) old ELNs; and b) ELNs of the modeller who left the research group.

 ISBN: 978-1-61208-339-1

The authorisation process in the ELN-PS system ensures that:

i) Roles are allocated dynamically (at run time). If a role is not assigned to a person, then the related authorisation key will not be added in the security code. So, a person cannot use that role. The security code stops working if a role is de-activated or blocked at any moment during program execution.

ii) Permissions are allocated dynamically (at run time) to the roles and could be activated or de-activated at any moment of the processing time.

iii) With the use of the dynamic security code, ELN-PAS unit works on a safe and protected mechanism. It prevents a private ELN becoming available automatically to any person in the community without the proper permissions being given by the modeller/data owner.

V. WEB BASED ELN-PS SYSTEM

In this section, we introduce the implementation of ELN-PS system that was created for the purpose of eliciting feedback from members of the EUROCHAMP-2 community. The architecture is presented as a Web based implementation and is shown in Figure 8.

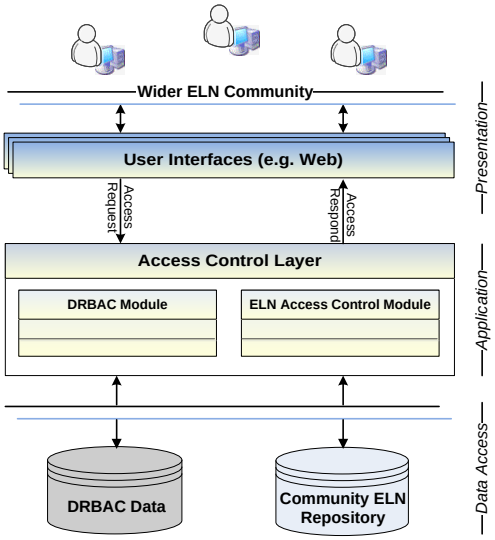

Figure 8. Implementation of ELN-PS System.

The authorisation process in the ELN-PS system was discussed previously in Section 4 so will not be repeated here. The implementation is based on 3-tier web architecture. It is coded in PHP, JavaScript and HTML and is hence reliant on a web browser to render the application executable [11]. The advantage of using a web based implementation is that it copes well with the distributed nature of the community in question. It presents the ability to update and maintain the web application without distributing and installing software on many different user computers. The geographically distributed environment and nature of users, requires the use of a centralised protection and sharing system that is accessible anywhere and is platform independent.

MySQL was used to implement the backend database. For the server-side scripting, PHP was used along with the semantic library for PHP ARC2 [12] in order to read the RDF data, associated with each ELN. A key aspect of the development was the Graphical User Interface (GUI), as even if the required functionality was met, if the GUI was hard to understand or unfriendly, then the program will ultimately be a failure. Interface design encompasses three distinct, but related constructs: usability, visualisation, and functionality [13]. A fourth component of accessibility has emerged as a critical factor in regards to the design of Web-based applications. The ELN-PS system thus uses a Cascading Style Sheet (CSS) for styling information. An example, rendering of the ELN-PS GUI showing the "Share ELN" interface is shown in Figure 9.

Figure 9. Share ELN Interface.

The "Share ELN" function allows a modeller to share individual ELNs as a whole object with other community members such as with their research manager. The "Add Comments" section allows modeller to exchange comments privately with their research manager or an editor. These comments are then recorded against the ELN as part of the life cycle information, these comment then may be retrieved at a later date.

VI. QUALITATIVE EVALUATION WITH END USERS

In this section, we perform a qualitative evaluation of the ELN-PS system. Qualitative evaluation captures descriptive data collected through the observations and interviews with end users and gives a voice to the participant's experiences [14]. It is used here as a mechanism for assessment on how well the ELN-PS performed. In this research, the goal of conducting qualitative evaluation with the end users was to determine the potential value, likely advantages and disadvantage of using:

i) The ELN-PS system;

 ISBN: 978-1-61208-339-1

ii) An DRBAC mechanism to protect and share ELNs;
iii) The concept of fine-grained access control to share only the selected ELN trails such as the gold simulation trail.

The evaluation plan included:

i) An introduction to the protection and sharing of ELN followed by questions/answers session;
ii) The demonstration of each scenario case in the ELN-system; and
iii) The collection of end user's feedback using a specific evaluation questionnaires, designed for this purpose.

A likert scale was used to assess the answer of each question in the evaluation process. Values were ranked from 1-5, 1 being very poor, 2 = poor, 3 = good, 4 = very good and 5 = excellent. After the demonstration of each scenario case, users provided feedback. The answers obtained from two members of the community are given in Table 1.

TABLE 1. ANSWERS FROM USER SURVEY.

Questions To Users	User	
	1	2
i) Do you understand the ELN protection and sharing process in the ELN-PS system?	3	4
ii) Do you value the DRBAC mechanism, adopted to protect and share the ELN?	4	3
iii) Do you think the role names, defined in the ELN-PS system give clear understanding to the people about their position?	3	4
iv) Are you satisfied with the privacy policy to protect and share ELN as whole object?	3	4
v) Do you understand the concept of fine-grained access control to share only a selected ELN trail (like gold simulation trail)?	3	4
vi) Do you see the value in giving an option to the ELN owner to restrict a third-party to view just the gold simulation trial?	4	4
vii) Do you think, it is good to provide extra functionality to share ELN trails other than gold trail?	4	3
viii) How you rate the design of the user interfaces? Is it clear and user friendly?	3	3

At the end of the evaluation, the recommendations and comments from the participants was recorded. A sample of their recommendations and comments are provided below:

A. Recommendations:

User 1: "Some adjustments to the design of the interface to improve usability".

User 2: "What I can see is that: a) the system should send an email auto notification to the person (e.g., supervisor) who will share the data with modeller, b) provides a variety of searching tips (e.g., search by date, by name of the simulation, search by range of date, search by EUROCHAMP chambers or search by collaborator partners), and c) includes the simulation trails (i.e., IPNav), so that the modeller can easily visualise the trails of the simulation runs".

B. Comments:

User 1:

1. "Confidence is the key to use".
2. "Facilitates remote supervision of student(s) / scientist(s) – gives the ELN unique educational aspects".
3. "Fine-grained access control is important as some users will want to share more/less of the ELN than others – again a key aspect of its usability".
4. "This is a key aspect if the community is going to really use this system as a primary scientific tool".
5. "Option to even "delete" ELN will make people feel safer".

User 2:

1. "Currently, I think the role names are sufficient unless if they changes from the users".
2. "This is a very good idea of protecting the provenance data".
3. "The user interfaces need to be improved".

The results were overall very encouraging, both participants rated the value of this protection and sharing mechanism between good and very good (i.e. 3 or 4 out of 5). They saw the value of being able to securely share the ELN as whole object and partly (i.e. at fine-grained level) with third parties (like research managers, editors, evaluators, etc.). These results can, therefore, be considered as an initial feedback before going into the larger community for further evaluation. The major concerns they have shown was the trust relationship with the third party (i.e. the data centre where the ELNs will be kept). User 1's comments in regards to confidence relates to the trust in the third party storing the data. We have presented a role-based access model surrounding the storage of ELNs but security must encompass the system as a whole and the end users need assurances that the service provider will maintain the relevant security around the ELNs stored.

VII. RELATED WORK

Access control is critical to information security and data protection. Within the Atmospheric Chemistry community, sharing of digital resources with a different degree of sensitivity is crucial as it ensures modelers are confident with the protection of their data. Details of the comparison of various access control models have been discussed in [15][16]. Based on the enhanced dynamic role based access control, this paper has introduced the ELN protection and sharing mechanism for secure access and sharing of ELNs from a central repository as whole objects or its elements (provenance trails) at fine grained level.

 ISBN: 978-1-61208-339-1

Generally, roles are defined as either static or dynamic. Static roles are normally based on a strictly defined association of users to the system that are established early and rarely change, dynamic roles ensures these associations are assessed at runtime as requests for access are made [17]. The NIST model [9] discusses RBAC and provides: a strict definition of RBAC sets and relations, while also defining a common vocabulary, setting the scope of the RBAC uniform features and introduces a functional specification providing administrative, review and system functions. It however does not incorporate a scalability attribute or permissions which deny access (i.e., negative permissions).

PERMIS [18] is a RBAC authorisation system that uses X.509 attribute certificates [19] to hold user's roles. Authorisation decisions are made through the PERMIS's access control decision engine based on the roles assigned to the user. PERMIS however does not define a mechanism for aggregating attributes from multiple authorities, where the user is known by different names at each authority. In AAA (Authentication, Authorisation and Auditing/accounting) [20], the RBAC framework is based on PERMIS which uses federated identity providers. Roles are used to identify users only to provide static access. The rest of the security is applied through the use of public and private keys. The concept of DRBAC is also not discussed in relation to sharing experiment metadata as is the case here.

In the eScience domain, CARMEN [21] and myExperiment [22] also discuss the data protection and sharing issues in a distributed environment. However, dynamic access control for sharing of metadata is limited or not discussed. Access control at a fine-grained level and varying descriptions of roles among different research groups is also not addressed.

ROWLBAC [23] explores the relationship between OWL and RBAC. It proposed two different approaches in the representation of roles i.e. roles as classes and roles as values; using a standard description logic reasoner. The role permissions defined are however limited. For example it explains the permissions on the basis of a Boolean function and does not discuss the access control at a fine grain level.

Smirnov et al. [24] present a RBAC model that is extended by adding a trust factor for a distributed environment. In this work, it gives every trust value for each user by introducing trust management to access control. It does not however, discuss the dynamic access control at fine grain level regarding metadata such as an ELN.

In [25], a model of the context-based access control for the information shared in a smart space is proposed. It uses open source Smart-M3 platform [25] and is built on the combination of the role based and attribute based access control models. Roles are assigned dynamically based on the participant's trust level. However, in this research, the co-ordination of roles in a hierarchy model and activation/de-activation of multiple roles are not discussed. Further the access control about the flow of data among different roles is not addressed. For example, like in the Atmospheric Chemistry domain, the experiment metadata is not accessible for Public role unless it is evaluated.

Carminati et al. [26][27] propose an access control system based on the Semantic Web technologies for social networks. It enables granting of access based on 'friendship" relation with the resource owner and on evaluation of the confidence level of the user. This works seems suited with the Atmospheric Chemistry community where modeller collaborated with other modeller in a laboratory or between other laboratories. However, in this research, we take it one step further providing access and share simulations metadata at a fine grained level.

VIII. CONCLUSIONS AND FUTURE WORK

The aim of this research was to study how to share modelling data and its provenance across a research community. The proposed architecture has been realised and in this instance tailored to the EUROCHAMP-2 community. It demonstrates the sharing of ELNs in a secure manner. It further shows how the DRBAC model allows for the protection of ELN provenance trails at a fine-grained level, thus ensuring that only data relevant to the evaluation and publication process is shared.

The qualitative evaluation demonstrated how a role based access system could be understood and accepted by a research community. In addition it showed how offering user's fine-grained access control over what they share elicits acceptance, especially when a community is sensitive to the sharing of a trail of important simulation runs.

This research can be considered to be an initial step in defining an access control model to protect and share ELNs within research communities. Future work will be to introduce the ELN-PS system to other eScience communities. The changes required are considered to only be need in the ELN system for individuals, so it can be tailored to a given community, leaving the RBAC system intact. In order to get more conclusive results on the value of the ELN-PS system, a larger set of ELN modellers is needed. However, before going into the larger community for further evaluation, the issue of establishing a trust relationship between end-users and service providers will need addressing. Only when a credible service provider such as the British Atmospheric Data Centre (BADC) [28] in EUROCHAMP-2 case, with robust plans for the safe storage of ELN data, will a community be willing to share their ELNs. Integration of technologies, such as Secure Socket Layer (SSL) Protocol [29] and encryption/decryption algorithms [30] into the ELN-PS system are also required to instill greater confidence from end-users.

Further, we aim to define the ELN as a service in the cloud. Cloud computing delivers the infrastructure, platform and software as services, which are made available by subscription in a pay per use model [31]. By defining the ELN service in the cloud, it could be managed on an on demand basis. Cloud computing would also offer a highly scalable solution which would be able to meet the ongoing

 ISBN: 978-1-61208-339-1

demands of several different ELN oriented research communities.

REFERENCES

[1] C. Martin, M. Haji, P. M. Dew, M. J. Pilling, and P. K. Jimack, "Semantically-enhanced model-experiment-evaluation processes (SeMEEPs) within the atmospheric chemistry community," Second International Provenance and Annotation Workshop, IPAW, Springer-Verlag, 2008, pp. 293-308.

[2] Taylor K., Essex J.W., Frey J.G., Mills H.R., Hughes G., and Zaluska E., "The semantic grid and chemistry: experiences with CombeChem," J. Web Semantics, 2006, 4 (2), 84-101, doi:10.1016/j.websem.2006.03.003.

[3] Zulkifly Mohd Zaki, Peter Dew, Mohammed H. Haji, Lydia MS Lau, Andrew Rickard, and Jennifer Young, "A user-orientated electronic laboratory notebook for retrieval and extraction of provenance information for EUROCHAMP-2," The 7th IEEE International Conference on e-Science, December 2011. pp. 371-378.

[4] Jenkin, M.E., S.M. Saunders, V. Wagner, and M.J. Pilling, "Protocol for the development of the master chemical mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds," Atmos. Chem. Phys., 2003, 181-193, doi:10.5194/acp-3-181-2003.

[5] Mohd Zaki Z, Dew PM, Lau LMS, Rickard AR, Young JC, Farooq T, et al., "Architecture design of a user-orientated electronic laboratory notebook: A case study within an atmospheric chemistry community," Future Generation Computer Systems, 2013, vol. 29, issue 8, pages 2182–2196.

[6] M. Frenklach, T. Bowman, and G. Smith, (n.d.), "GRI-mechanism," http://www.me.berkeley.edu/gri-mech/, [retrieved: 11th April, 2014].

[7] The University of Leeds. (n.d.), "AtChemOnline homepage," https://atchem.leeds.ac.uk/webapp/, [retrieved: 11th April, 2014].

[8] Greenwood, M., Goble, C.A., Stevens, R.D., Zhao, J., Addis, M, Marvin, et al., "Provenance of e-Science experiments experience from bioinformatics," The UK OST e-Science second All Hands Meeting (AHM'03), 2003, pp. 223-226.

[9] D. F. Ferraiolo, R. Sandhu, S. Gavrila, D. R. Kuhn, and R. Chandramoul, "Proposed NIST standard for role-based access control," ACM Transactions on Information and System Security, 2001, vol. 4, Issue 3, pages 224-274.

[10] Oracle. (2010). "The Java EE 5 tutorial," http://docs.oracle.com/javaee/5/tutorial/doc/bncbe.html, [retrieved: 11th April, 2014].

[11] M. Miller, "Cloud computing: web-based applications that change the way you work and collaborate online," A Book by Que Publishing, 2009.

[12] Annon, "Easy RDF and SPARQL for LAMP systems," https://github.com/semsol/arc2/wiki, [retrieved: 11th April, 2014].

[13] Information Resource Management Association, "Instructional Design: Concepts, Methodologies, Tools and Applications," IGI Global, 2011, DOI: 10.4018/978-1-60960-503-2.

[14] Breier, J. and Hudec, L., "New approach in information system security evaluation," IEEE first AESS European Conference on Satellite Telecommunications (ESTEL), 2012, pages 1-6, IEEE, DOI: 10.1109/ESTEL.2012.6400145.

[15] S. M. Hasani and N. Modiri., "Criteria specifications for the comparison and evaluation of access control models," International Journal of Computer Network and Information Security(IJCNIS), 2013, vol. 5, pp. 19-29.

[16] P. N. Mahalle, B. Anggorojati, N. R. Prasad, and R. Prasad, "Identity authentication and capability based access control (IACAC) for the internet of things," Journal of Cyber Security and Mobility, 2013, no. 4, pp. 309-348.

[17] J. Odell, H.V.D. Parunak, S. Brueckner, and J. Sauter, "Changing roles: dynamic role assignment," Journal of Object Technology, ETH Zurich, 2003, pp. 77–86

[18] Sacha Brostoff, M. Angela Sasse, David Chadwick, James Cunningham, Uche Mbanaso, and Sassa Otenko, "R-What? development of a role-based access control (RBAC) policy-writing tool for e-Scientists," Software: Practice and Experience, 2005, pp. 835-856.

[19] International Telecommunication Union (ITU), "Information technology - open systems interconnection - The Directory: Public key and attributes certificate frameworks," ITU-T Recommendations X.509, March 2000.

[20] R. O. Sinnott, A. J. Stell, and J. Watt, "Advanced security infrastructures for grid education," 10th World Multi-conference on Systemics, Cybernetics and Informatics, (WMSCI 2006), 2006, pages 182-196.

[21] Martyn Fletcher, Bojian Liang, Leslie Smith, Alastair Knowles, Tom Jackson, Mark Jessop, et al., "Neural network based pattern matching and spike detection tools and services in the CARMEN neuroinformatics project," Neural Networks, Special Issue on Neuroinformatics, 2008, vol. 21, Issue 8, pp. 1076-1084.

[22] De Roure, D. and Goble, C., "myExperiment: A web 2.0 virtual research environment for research using computation and services," In: Workshop On Integrating Digital Library Content with Computational Tools and Services at JCDL, 2009.

[23] T. Finin, A. Joshi, L. Kagal, J. Niu, R. Sandhu, W. Winsborough, et al., "ROWLBAC: Representing role based access control in OWL," SACMAT 08, 2008, pages 73-82.

[24] L. Zhao, S. Liu, J. Li, and H Xu, "A dynamic access control model based on trust," 2nd Conference on Environmental Science and Information Application Technology, 2010, vol. 1, pages 548-551.

[25] A. Smirnov, A. Kashevnik, N. Shilov, and N. Teslya, "Context-based access control model for smart space," 5th International Conference on Cyber Conflict (CyCon), 2013, pages 1-15.

[26] B. Carminati, E. Ferrari, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "A semantic web based framework for social network access control," Proc. of the 14th ACM symp. on Access control models and technologies, 2009, pp. 177-186.

[27] B. Carminati, E. Ferrari, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Semantic web-based social network access control," Comp. & Security, March–May 2011, vol. 30, issues 2–3, pp. 108–115.

[28] British Atmospheric Data Centre (2013), "Homepage," http://badc.nerc.ac.uk/, [retrieved: 11th April, 2014].

[29] Chou, W, "Inside SSL: the secure sockets layer protocol," IEEE Computer Society, 2002, ITPro, 47-52.

[30] C. Lu and S. Tseng, "Integrated design of AES (Advanced Encryption Standard) encrypter and decrypter," Proceedings of The IEEE International Conference on Application-Specific Systems, Architectures and Processors, 2002, pp. 277–285.

[31] R. Buyya, R. Ranjan, and R. N. Calheiros, "InterCloud: utility-oriented federation of cloud computing environments for Scaling of application services," 10th international conference on Algorithms and Architectures for Parallel Processing (ICA3PP), 2010, vol. part I, pages 13-31.

 ISBN: 978-1-61208-339-1