RESOURCE ENTITLEMENT MANAGEMENT SYSTEM

Mikael Linden
CSC – IT Center for Science, P.O. BOX 405, FI-02101 ESPOO, Finland
e-mail: mikael.linden@csc.fi
Tommi Nyrönen
CSC – IT Center for Science, P.O. BOX 405, FI-02101 ESPOO, Finland
e-mail: tommi.nyronen@csc.fi
Ilkka Lappalainen
The EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom
e-mail: ilkka@ebi.ac.uk

Paper type
Research paper

Abstract
The Resource Entitlement Management System (REMS) is an electronic tool that manages access rights to research resources, such as research datasets. Applicants can use their federated identity as authentication to the REMS, complete the data access application, commit to licence terms and submit the application to the appropriate resource. The REMS system circulates the application to the resource owner or denoted representative(s) in support of the data access granting process, and provides a reporting function for applications and granted data access rights.

This paper introduces the different aspects that manage access to research data and the requirements for the REMS system. The REMS has been piloted with the European Genome-phenome Archive (EGA), a service of the European Bioinformatics Institute (EBI), and with a complex data access committee of the Nordic Control Database (NCDB). Although originally developed for the life sciences as part of the ELIXIR project, the REMS system aims to be a discipline independent tool for access rights management.

Keywords
Authorisation, entitlement, research data, AAI, bioinformatics, infrastructure

1. Introduction

In the research and education networking community, cross-organisational identity management has been a subject of interest for years. The concept of an Authentication and Authorisation Infrastructure (AAI) was introduced in Switzerland in 2001 (Droz, 2001) and since then, 34 research and education identity federations have emerged to allow cross-organisational access to resources, relying mostly on the web single sign-on scenario of the SAML 2.0 technology (REFEDS, 2013).

The driving force behind identity federations has been federated authentication. An end user is able to use a single set of credentials (typically, username and password issued by home institution) to access resources in a remote security domain. In many scenarios, federated authentication has already provided significant benefits while the second "A" for cross-organisational authorisation has received less attention.

The management of research data has become a topical issue in contemporary science. One of the greatest challenges for research data management is posed by genome research (Smedley, 2010; Clarke, 2012; Church, 2010). Research projects produce research data, whose availability for secondary use reduces significantly the total costs and increases the scientific output of research projects. More and more data is created at accelerating speed. For instance, the efficiency (and subsequently cost) of sequence data generation has improved by six orders of magnitude from the early days of the Human Genome Project in 2000. Application of human genetic variation data in research requires that the research infrastructure is ready for controlling the use of data for the purposes it is meant and consented for. Whereas there are also other significant challenges, such as the discoverability of the research data, a key challenge is the access management of the datasets.

In general, open access to research data is desirable; when researchers have identified a dataset important for their study, they could simply download or access the data without any further complications (Clarke, 2012; Hrynaszkiewicz, 2012). However, there are various examples of ethical, legal, societal and intellectual property reasons (Cates, 2012; McGuire, 2011; Ostermann, 2012; Tabor, 2012) for controlling access to the research data. The examples below further clarify these reasons (commonly dubbed as ELSI) in the context of the life

sciences (Knoppers, 2012).

Ethical. Research data could be used for unethical purposes, such as, for political discrimination. If there is a risk of unethical research being carried out, it is a common practice that an ethical committee needs to screen the research proposal before access is granted to the researcher. On the other hand, it can be also unethical to deny access to a research dataset and make the researcher carry out a similar experiment again, if that research, for instance, requires sacrifice of the lives of test animals. **Societal** reasons to control data access are closely related yet distinct from ethical reasons as all societies do not share the same values. For example, societies may differ with regard to their public health system and insurance policies, the use of stem cells in research, minority groups (e.g. homosexuals and transgenders), abortion, or blood donation. Societal diversity will likely have a large impact on the use of genetic data in research.

Legal. Data protection laws protect the privacy of the data subjects if the datasets consist of personally identifiable information. In the life sciences, the datasets often consist of the human participants' health records and also their strong identifiers, such as the national identification numbers. The participants may consent their data for research purposes and the data can be anonymised/pseudonymised/aggregated, but this does not remove the legal obligation to manage access to the datasets.

Intellectual property. Different intellectual property rights, such as copyrights, may be associated with the data. The intellectual property rights limit access to the research data. For instance, the data owner may not want to open the data until all related patent applications are filed.

The REMS system can assist in access management when open access to datasets cannot be provided. Section 2 of this paper introduces previous work on access management. Section 3 presents general approaches to research data access management. Section 4 introduces the Resource Entitlement Management System (REMS), its requirements and section 4.3 its implementation. Section 6 presents a pilot carried out with the REMS system, section 7 identifies future work items and section 8 concludes the paper.

2. Previous Work

In the literature, authorisation is generally defined as a function f(S,O,A,e) which produces a "yes" or "no" answer to the question "Is subject S allowed to execute the action A to the object O given the effective environmental variables e". In the context of research data access, that translates to the question "Is the researcher S allowed to download (or execute a program that uses) the data O in a given environment (such as the network the researcher is in).

Role-based access control (RBAC) has been a subject of research since the 1990's (e.g. Ferraiole, 1992; Ferraiole, 1995) and has become a mainstream approach to authorisation. In RBAC, a user is assigned to a role and a role is further assigned to an access rights. For instance, access to a service may be granted to all students or employees of a particular university using appropriately defined role. A special case of RBAC is group-based access control, where access rights for a particular researcher are coupled to membership to a group.

Grid computing frequently uses access management based on group membership (dubbed as virtual organisations or VOs). These groups are managed using Virtual Organisation Membership Tool (VOMS) (EMI, 2013). In SAML-based identity federations, similar group management elements exist in tools such as the SWITCH Group Management Tool (SWITCH, 2013), Internet2 COManage (Internet2, 2013), SUNET COIP (Nordunet, 2013) and the SURFconext of SURFnet (SURFnet, 2013). The VOOT research project also studied the management and sharing of the group membership information (GEANT, 2013).

Access control can further be divided into the functionality of a policy enforcement point (PEP) and policy decision point (PDP). The PDP evaluates the function f(S,O,A,e) and provides this result to the PEP that is used to enforce the access control. Standard languages and protocols, such as XACML (OASIS, 2013), have been developed for communication between these two components. Furthermore implementations such as Argus (SWITCH, 2013b) are already used by the National Grid Initiatives, a member of the European Grid Infrastructure.

3. Authorisation of Research Data

We have found that authorisation is typically granted based on role or group based access management, committing to appropriate licence terms or submission of a data access application.

3.1. Role or Group Membership

A researcher is authorised to access the data as a consequence of the assigned role or group membership. For example, access to a particular dataset will be granted to any researcher affiliated to a research institution that already has access to the dataset. Access may also be limited only to members of the research group within the institute carrying out a pre-defined experiment. Role or group membership may be managed by the home institute and it can be released as an attribute within the SAML 2.0 web single sign-on scenario, using standard attributes such as eduPersonAffiliation, organisationalUnit, or isMemberOf. For roles and groups managed by the research communities themselves, community-maintained attribute authorities can be used like the VOMS, discussed in section 2.

3.2. Commitment to the Licence Terms

Often, the researcher needs to commit to the dataset's licence terms before access to the data is granted. Depending on the discipline, the licence terms can also be dubbed as the Terms of Use, Code of Conduct or Data Access Agreement. Typically, the researchers must agree that they will use the datasets for the consented research purpose only, protect the data adequately and destroy it when the research project is completed.

Traditionally, as evidence of commitment to the licence terms, researcher will have to sign a paper document and send it to the party that owns or manages access to the datasets. However, in the European Union, the e-signature directive (EU, 1999) implies that under certain conditions electronic signatures can have the same legal effect as handwritten signatures. For instance, if an end user is authenticated reliably by an Identity Provider in an identity federation and the server log files show that he/she has pressed the "I approve these licence terms" button, it can be argued that this qualifies as equivalent to a handwritten signature. Ultimately, it is up to a court to evaluate if the evidence of the researcher's commitment to the licence terms is sufficient.

3.3. Submission of a Data Access Application

A widely adopted approach to grant access rights to a dataset requires a researcher to submit an application form that provides justification (e.g. an attached research proposal) for the request to access the dataset. The application is either accepted or rejected by the resource owner, or a denoted representative, such as a data access committee evaluating the quality of the intended research and compliance to the original informed consent.

Traditionally, the data access application is a printed and ink-signed document that is delivered in the postal mail. It may be also an electronic document sent as an e-mail attachment, but often an ink-signed copy of the application needs to be delivered as well. However, the application can be also converted into a fully electronic form. The common enterprise identity management software products already support electronic workflows that could be used to process electronic access right applications.

3.4. Discussion

The previous sections presented three approaches to manage access to research datasets. The approaches are not mutually exclusive and, in practice, it is typical that all three of them are used in parallel; the end user must be a researcher in good standing, commit to the applicable licence terms and present a written application to apply for access to the datasets. As discussed previously, it is possible to manage role or group memberships automatically or semi-automatically and it would be possible to implement a fully automatic system that grants access to the data based on compliance to the appropriate licence terms. However, the data access application based approach cannot be fully automated as each applicant within the application must be evaluated individually. It has been suggested that data access applications could be replaced by role or group membership models with additional committed licence terms (Hrynaszkiewicz, 2012). In practice, however, human data consented for research through controlled access mechanisms prevent such option (Knoppers, 2012).

The three models presented above could be combined by using a layered approach with increasing security and access control requirements. The first layer is fully open for data that can be discovered and shared publicly in the Internet. The second layer consists of aggregated information that in the genomic field are calculated over all samples in the dataset, for example allele frequencies for a particular genetic variation. The third layer consists of data that requires individual data access permissions to be set prior to data access. The original consent agreement will dictate the applied security model rather than the data type. The REMS software is designed to support data access decisions based on the licence term and data access application workflows.

4. Requirements for a Resource Entitlement Management System

The Resource Entitlement Management System (REMS) is a generic access management tool. It has been designed to support workflows that allow data access decisions to be made based on an electronic application

process, store data access permissions in a standardized format and provide interfaces to some of the most important services such as the SAML 2.0 for authentication or the European Genome-phenome Archive (EGA) for research consented data under controlled access management. Here we present the requirements of the REMS tool and the benefits it can bring to the access management of the research datasets.

4.1. REMS Concept

Figure 1 represents the REMS concept. As a leader of a research group, a principal investigator (*the applicant*) identifies those resources to which all research group members needs to apply for access. The applicant then uses federated identity to log in to the REMS system and completes a data access application made available as an electronic form. The application typically consists of research proposal and a list of all those research group members (*members of the application*) who should have access to the data. The data access application process starts as the applicant submits the application (arrow 1 in the figure). Depending on the workflow configured for the resource, the REMS system may request the applicant and the members of the application to commit to the licence terms of the resource (2), and circulates (3) the application to the person(s), (such as, data access committee, DAC) responsible for granting access to the resources (*reviewers* and *approvers*)(4). The reviewers can only provide comments on the application for the approvers.

The reviewers and approvers receive an e-mail notification for a new data access application. The REMS supports federated identity authentication for both parties and a flexible workflow for evaluating, approving, rejecting or updating an application. For example, it is possible to configure the REMS so that the application has to be approved by two independent approvers to eliminate any human error. The REMS also provides a reporting functionality that allows data owners to review the history of all applications on their data, close active access permissions for any applicants and view statistics collected from the application process.

Once an application has been approved the REMS system signals that the data can be made available for the user (arrow 5). The user can, for instance, download the dataset to a local workstation or alternatively use it as the input for an executable application in a remote server. Therefore, REMS can be used as a Policy Administration Point (PAP) for the access control enforcement. In case of the European Genome-phenome Archive (EGA) the approved data access permissions are transferred from REMS to the EGA system in a secure manner.

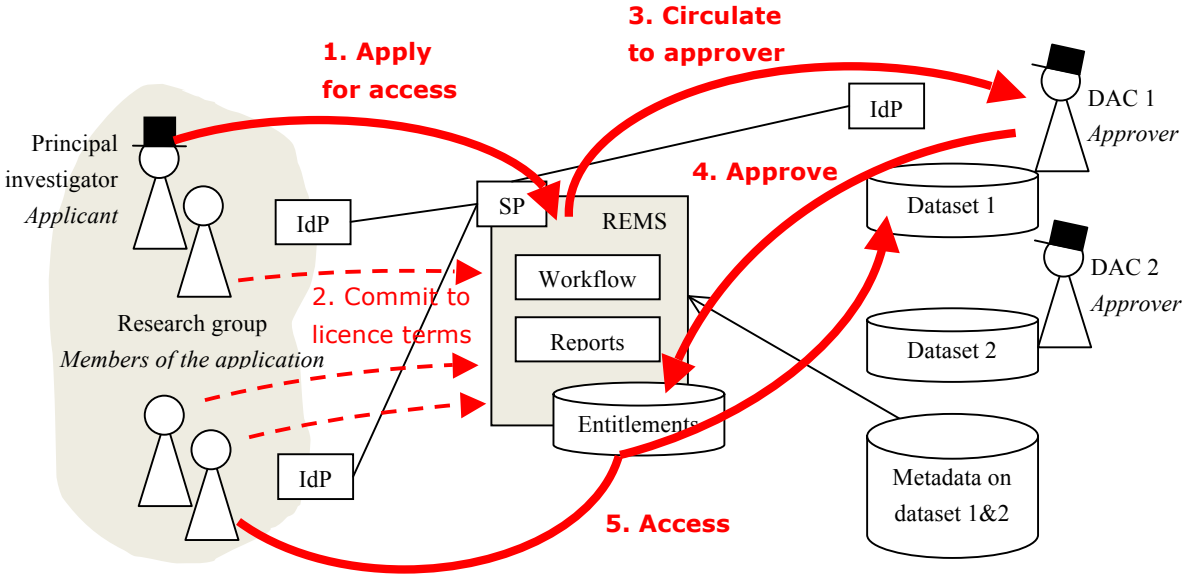

Figure 1. The REMS concept. A principal investigator applies for data access from a resource (1). All research group members must commit to the applicable licence terms (2). The REMS circulates application to the approver(s) (3) to facilitate data access decision making process (4). The REMS also provides an interface for the linked resources to query data access permissions that can be used together with federated identity to arrange access to the data (5). The IdP (identity provider) and SP (service provider) describe how federated identity is used in the REMS authentication process.

4.2. Requirements for the REMS system

This section presents the main requirements of the REMS system in detail.

4.2.1. The Application Process

The resources to which access is applied for can be anything (not just datasets, hence the more general word "resource") identified by a unique identifier (such as a URI or a persistent identifier, PID). The resources must be atomic i.e. from the access control perspective; they cannot be further split into smaller parts with more fine-grained access rights. The user who introduces a new resource to REMS becomes the owner of the resource.

A resource workflow supports both the application and data access approval process. The resource owner can configure one or more workflows for each resource to describe the process in detail. The REMS supports a single application process by allowing the same workflow to be configured to a number of resources. If the resources do not share the same workflow, the applicant must submit an application separately to each resource.

The workflow consists of the application form (see Figure 2), the resource's licence terms (if any) and definition of approvers and reviewers for the applications. It is also possible to have no approvers and reviewers, in which case the applicant just needs to approve the licence terms of the resource. In other words, in that setup, anyone who approves the licence terms gets access to the resource and the role of REMS is just to register people who start to use the resource and store their approval of the licence terms for the audit trail.

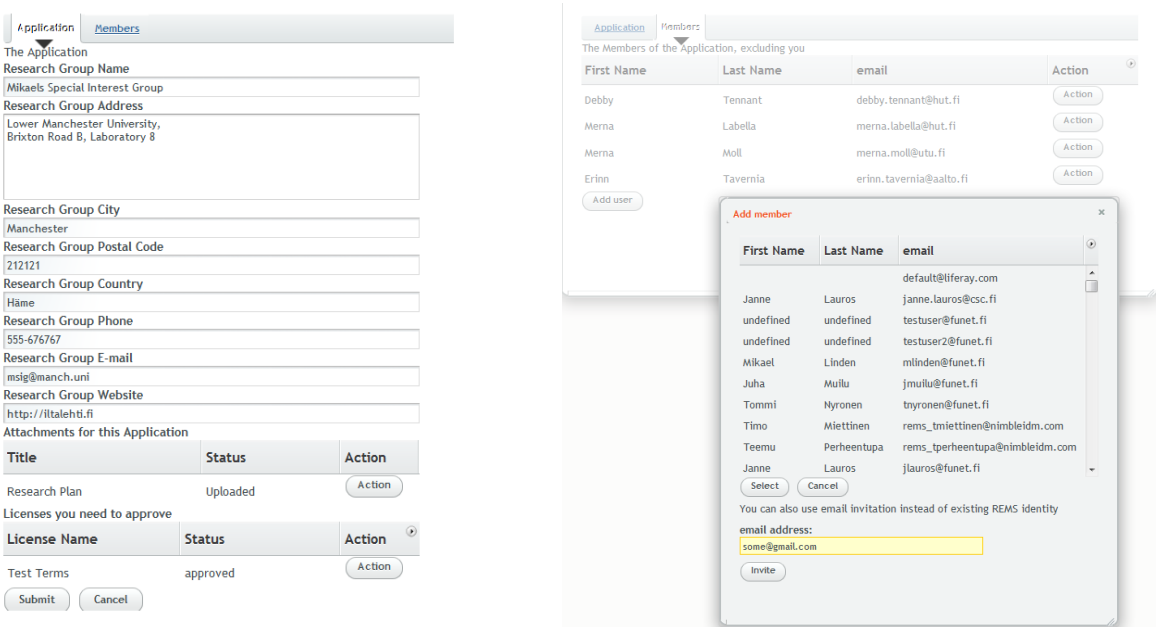

Figure 2. The applicant fills in the details of the application to an electronic form, and commits to the licence terms of the dataset (left). The applicant also invites the application members for whom equal access rights are applied (right). A screenshot from the REMS implementation.

In several disciplines, research is carried out by research groups and all members require access to the resources. A workflow can be configured to allow the applicant to attach a list of research group members to the application (see Figure 2). The applicant can either pick them from a list of existing REMS users or invite them by entering their e-mail addresses into the application. REMS invites all members of the application to approve the licence terms.

The approval process is typically simple; the application is delivered to the individual who owns the resource (see Figure 3). In the life sciences the applications are approved by data access committees who may have a secretary and a number of approvers (each with a single vote). To support multi-stage approval process the REMS provides workflows that can be configured to consist of several consecutive checkpoints that each application must pass to be approved. For example, the secretary can return an application for amendments if it does not include all the mandatory information. In the second round, a simple majority of the approvers is required to pass the application. At any stage approvers and reviewers (without vote) can register their comments to the application. These comments are visible to the approvers and reviewers.

During the requirements specification process, it was discussed if REMS should use group management service such as the VOOT protocols (GEANT, 2013). This would allow REMS to assign access rights to groups whereas the group access management and update processes could be done by an external service. However, this would limit the use of REMS for genomic projects where approvers must be able to assess each applicant or member of the applicant individually.

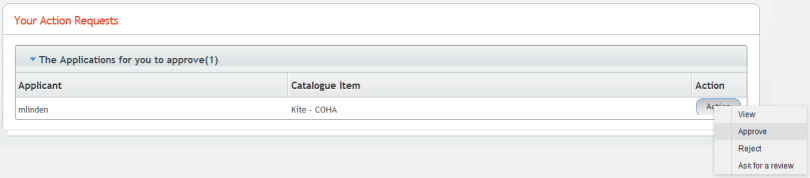

Figure 3. The application is circulated to the resource owner or a denoted representative for approval. A screenshot from the REMS implementation.

4.2.2. Application Update Process

Once approved, the applicant and the other members of the application can use the authorised resources. However, the conditions on active applications may change and these changes must be reflected on granted access rights.

The REMS includes application update process that allows applicant to update their active applications at any point. For example, as new members join the research team their rights to use the resource must be updated in the application. The applicant can also use existing application to apply access rights for new datasets. The update process triggers always a normal approval workflow in REMS with the exception of deleting existing approved members from the application.

The REMS approver for the application can remove the access rights for an entire application or separately from any member of the application at any time. The applicant has a responsibility to close the application as the research project is completed. However, in practice applicants tend not to close their applications. The REMS provides a timer function that can be configured for each application to remove automatically any unnecessary access rights. The timer is set when the application (or update) is approved. As the timer expires the applicant receives an e-mail notification to confirm that the research project is still on-going or alternatively to update any details on the application. Failure to update the application freezes data access from all members of the application.

For any organisation hosting resources, it is important to be able to demonstrate to the funding institutes that their services are actively used by the researcher community. Publications based on the datasets is a good indicator. When updating, refreshing or closing an application, REMS will ask the researcher to fill in data on any publications that have used the approved resources.

4.2.3. Reporting on Access Rights

When a manual process is converted to an electronic workflow it will also allow automation of report generation. The most important report REMS provides is a full audit trail available to the resource owner. The audit trail is a chronological record of all committed actions on any of the registered applications. The resource owners may also request a report that describes specific access rights, throughput times or publications using the resource. The applicant and members of an application can also request reports on their access rights and on the history of the application (see Figure 4).

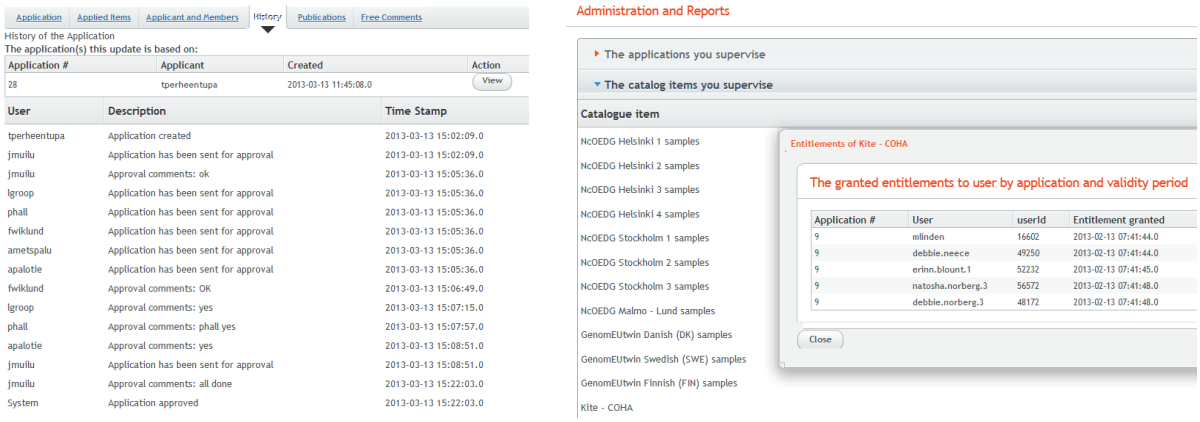

Figure 4. An approver can browse the history of an application (left) and have a report on all access rights granted to a resource (right). A screenshot from the REMS implementation.

4.2.4. External Interfaces

The REMS has multiple external interfaces. For the authentication of the end users, REMS implements a standard SAML 2.0 Service Provider functionality. All end users, including the applicants, other members of an application, approvers, reviewers and resource owners are authenticated against a SAML 2.0 Identity Provider. REMS retrieves the user's basic attributes (such as name, title, e-mail address, home organisation and his/her type of affiliation) from his/her home organisation. A shared unique identifier (such as eduPersonPrincipalName) is requested as well to be able to match the user identity between the REMS (PDP) and the external server enforcing the access control (PEP).

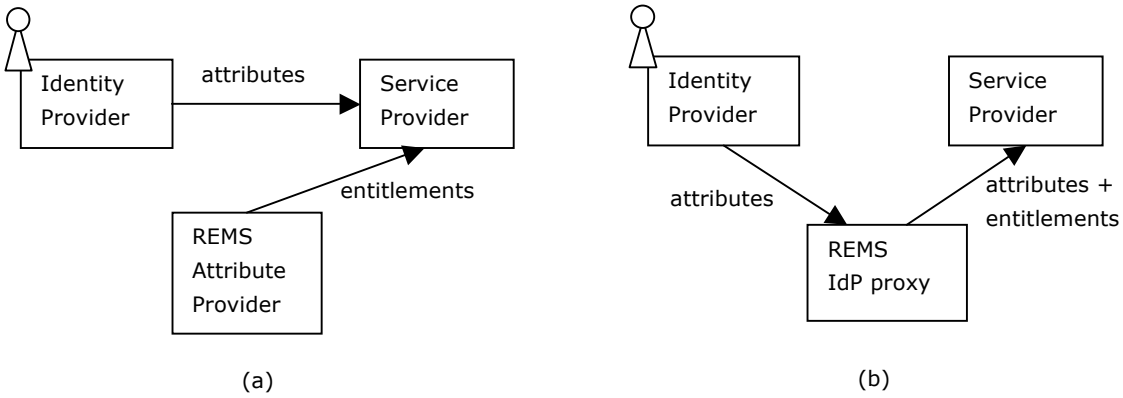

Figure 5. REMS can act as an attribute provider (a) or as an IdP proxy (b).

The goal is to facilitate a number of different ways that REMS can be integrated with the user attributes required by the Service Providers (SP). The relying parties (PEP) can use SAML protocols to rely on REMS (PDP) in access control. REMS can act as a SAML Attribute Provider that delivers the user's access rights in an eduPersonEntitlement attribute as a response to a SAML attribute query (see Figure 5 a). Alternatively, REMS can be configured to be a SAML proxy that resides in the SAML flow between the Identity and Service Providers and enriches the SAML assertions with eduPersonEntitlement attributes (see Figure 5 b).

REMS itself has a simple catalogue user interface where an applicant can list all the resources the REMS system is aware of. However, REMS does not intend to solve the resource discovery problem but assumes that normally the applicants use some external catalogue service to locate the dataset(s) they want to apply for access to. Instead, REMS provides an interface which the external catalogue can use to trigger an application to one or more datasets in REMS. REMS then steps in and manages the application process for the end user as usual.

4.3. Benefits of REMS

The REMS is designed to make the application process easier for the applicants, reviewers and approvers by transforming a manual process into an electronic one. The REMS enhanced reporting tools improve application traceability, throughput times are expected to reduce and the system provides an audit trail of all committed actions. Finally, REMS is believed to be a discipline-independent access rights management tool for research

datasets. REMS has been developed with the requirements of biomedical data access control, but the tool can be used to control access to versatile resources. To get the benefits of scale, we assume applications of REMS in other research communities e.g. environmental sciences, linguistics or economics datasets that require access control.

5. REMS Implementation

ELIXIR is a pan-European research infrastructure for biological information. It unites Europe's leading life science organisations in managing and safeguarding the massive amounts of data being generated every day by publicly funded research. ELIXIR will provide the facilities necessary for life science researchers - from bench biologists to cheminformaticians - to make the most of our rapidly growing store of information about living systems.

Based on the requirements presented in section 4.2, the REMS system have been implemented in the ELIXIR EGA AAI pilot, a common project for CSC – IT Center for Science, the Institute for Molecular Medicine Finland and the European Bioinformatics Institute. The ELIXIR EGA AAI pilot has committed to release of the REMS implementation under an open source licence. As the proposed Finnish ELIXIR node, CSC – IT center for science is also prepared to host a REMS instance for the ELIXIR community.

The implementation is done on Java as a portlet compatible with the Liferay version 6.1.0 portal, using the Vaadin 6 web application framework. The screenshots in Figure 2, Figure 3 and Figure 4 are from that implementation. Also a screencast of REMS is available on the Internet[1].

Although developed for the life sciences as part of the ELIXIR project, the REMS implementation aims at being a discipline independent tool for managing access to scientific datasets. In Finland, also other disciplines, such as the humanities and social sciences, have shown interest in the use of the REMS system.

6. Pilot with the European Genome-phenome Archive (EGA)

A pilot on the REMS system has been carried out together with the life sciences community. This section shortly introduces the pilot.

6.1. Description of the EGA and the Nordic Control Database

The EGA (European Genome-phenome Archive[2]) is a service of the EBI (European Bioinformatics Institute). The EGA provides permanent archive and data dissemination services for all genetic and phenotypic data where the study participants have signed an informed consent agreement that authorises data release only for a specific research use or to bona fide researchers but where consent agreement prevents fully public data dissemination.

The EGA has implemented a distributed data access-granting policy. Data access is granted by a Data Access Committee (DAC) and not by the EGA. The DAC is typically formed from the same organisation that monitored the original data collection and analyses or a designate of this organisation. The applicants apply for data access directly from the DAC and all applicants within an application must also sign a Data Access Agreement that controls how the applicant must store, analyse and transfer the data once downloaded from the EGA system. The EGA stores the DAC approved data access decisions into an account created for each approved applicant. The account allows the data users to access the authorised data files from the EGA archive. At the beginning of the pilot project the EGA stored more than 300 dataset consisting of genetic and phenotypic information. The EGA co-operates with more than 70 different DACs to provide authorised access to these datasets. Most of the DACs do not have access to a software that would support electronic data access application process.

The NCDB (Nordic Control Database[3]) is a resource that stores information from the Danish, Estonian, Finnish or Swedish samples used as controls in a number of genome-wide association studies. The resource was created by the Nordic Center of Excellence in Disease Genetics and it has been deposited into the EGA. The current version of the NCDB consists of 6000 samples divided into 11 separate datasets. The data access applications are processed and approved by the NCDB DAC.

The NCDB was chosen to the pilot because the NCDB DAC has a relatively complex structure. The DAC consists of five voting members and a secretary who prepares and presents the applications to the voting

[1] http://www.youtube.com/watch?v=2pkIRjksH2g

[2] https://www.ebi.ac.uk/ega/

[3] http://www.nordicdb.org/

members. It was believed that a successful pilot with the NCDB DAC would prove the applicability of the REMS tool also in smaller setups.

6.2. Description of the Pilot

In the ELIXIR EGA AAI pilot, the data access application process of the NCDB DAC was described and translated into a REMS workflow. Instead of exchanging ink-signed data access agreements an applicant could complete and submit an electronic application to REMS, and the DAC could process and approve the application electronically. These data access permissions are transferred to EGA using a secure interface.

To enable REMS user authentication, the REMS system was also registered as a Service Provider to the Haka federation, the identity federation of the Finnish research and higher education, enabling an applicant from that community to use his/her home institution's credentials to access the REMS system.

Access control to consented data will be of major importance for the pan-European ELIXIR. The ELIXIR EGA AAI pilot is the first ELIXIR pilot focusing on the access management of the bioinformatics information related to sensitive data, but the topic has since received more attention e.g. in the BioMedBridges EU FP7 project[4]. The pilot was also recognised by the EIROforum Federated Identity Management for Research Collaboration group as one of the projects that will explore the requirements on federated identity management and increase interest in it among the bioinformatics community (Broeder, 2012).

The pilot was concluded in April 2013 by a successful delivery of the system for the NCDB. The researchers can now apply for access to the NCDB electronically, and the DAC has a fully paperless system for processing the applications. Due to the relatively small number of data access applications in the NCDB, the experiences on the actual application process have remained limited. As a next step, it is planned to extend the REMS usage to new EGA DACs.

7. Future work

Depending on a successful establishment of the ELIXIR consortium agreement among the ELIXIR member countries, CSC – IT Center for Science is prepared to deliver REMS as a service from the Finnish ELIXIR node to the other ELIXIR nodes including EMBL-EBI. CSC plans to offer the REMS service also locally in Finland for the national biobanking (BBMRI.fi) community. Similar discussion has also started with other disciplines, such as humanities and social sciences. It is likely that CSC as a Finnish IT Center for Science can gain economics of scale if it operates REMS nationally for several disciplines.

Currently, the REMS service is registered as a Service Provider to Haka federation, the Finnish identity federation for research and education. However, the EGA has end users around the Europe and beyond. At the time of writing, REMS is approved as a pilot for the Enabling Users task of the GN3plus project with the aim of further integrating REMS into eduGAIN, the interfederation service developed and operated by the GÉANT project and network.

The development of the REMS tool continues at the CSC. Currently, the workflow configuration is a manual process. The REMS deployment would accelerate with the possibility to auto-generate a workflow for a simple resource based on the resource's metadata (for instance, the owner of a resource becomes automatically the single approver of a data access application). Making REMS a Policy Administration Point (PAP) for the Argus authorisation service (SWITCH, 2013b) would also provide interesting integration opportunities for enforcement of the access rights managed by REMS.

8. Conclusions

REMS is a generic tool developed for access management to research resources, mainly focusing on research datasets. Researchers can use their federated identity to log in to the REMS tool, complete and submit their data access application and commit to the resource's licence terms. The application will be circulated to the resource owner or their representative, such as a data access committee, for approval. The REMS provides also reporting functionalities and integration interfaces necessary for the access control enforcement.

This paper presented the requirements of the REMS system. Although originally developed for the life sciences, REMS is a discipline independent tool and also other disciplines have shown interest in it. The REMS implementation will be made available on an open source licence.

[4] http://www.biomedbridges.org/

Acknowledgements

Acknowledgements to the REMS development team members Janne Lauros and Henri Mikkonen. Acknowledgements to the Data Access Committee of the Nordic Control Database, especially doctor Juha Muilu and professor Aarno Palotie and to Timo Miettinen and Teemu Perheentupa of the FIMM IT service team. Acknowledgments also to the EBI team, including Bren Vaughan, Pedro Albuquerque, Andrew Lyall, Ugis Sarkans and Justin Paschall and to other persons who have followed and contributed to the pilot, including Gergely Sipos and Valter Nordh.

CSC's work for REMS has been supported by the Ministry of Education and Culture of Finland and by Academy of Finland grants 271642 and 263164 to construct Biomedinfra, the Finnish consortium for ELIXIR, BBMRI and EATRIS ESFRI. The work presented in this paper is partly supported by the EGI-InSPIRE project (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe), co-funded by the European Commission (contract number: RI-261323).

References

Broeder, 2012

Broeder, D. Jones, B. Kelsey, D. Kershaw, P. Lüders, S. Lyall, A. Nyrönen, T. Wartel, R. Weyer, H., 2012. Federated Identity Management for Research Collaboration. CERN-OPEN-2012-006. <https://cdsweb.cern.ch/record/1442597>

Cates, 2012

Abortion policy and science: can controversy and evidence co-exist? Cates W Jr. J Public Health Policy. 2012 Aug;33(3):363-7. PMID: 22622482

Clarke, 2012

Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, Vaughan B, Preuss D, Leinonen R, Shumway M, Sherry S, Flicek P; 1000 Genomes Project Consortium. The 1000 Genomes Project: data management and community access. Nat Methods. 2012 Apr 27;9(5):459-62. doi: 10.1038/nmeth.1974. PubMed PMID: 22543379; PubMed Central PMCID: PMC3340611.

Church, 2010

Church DM, Lappalainen I, Sneddon TP, Hinton J, Maguire M, Lopez J, Garner J, Paschall J, DiCuccio M, Yaschenko E, Scherer SW, Feuk L, Flicek P. Public data archives for genomic structural variation. Nat Genet. 2010 Oct;42(10):813-4. doi: 10.1038/ng1010-813. PubMed PMID: 20877315; PubMed Central PMCID: PMC3169170.

Droz, 2001

Droz, S. Graf, C. Hassenstein, G. Heim, C. Lenggenhager, T. Meier, D. Monnard, J. Roy, A. Sonderegger, H. Tschudin, C. Concept for an Electronic Academic Community in Switzerland and the creation of a Common Authentication and Authorization Infrastructure (AAI) for the Swiss Higher Education System. Oct, 2001. <https://www.switch.ch/aai/docs/concept.pdf>

EMI, 2013

European Middleware Initiative. Virtual Organisation Membership Service. < http://www.eu-emi.eu/products/> [Accessed 18 April 2013]

EU, 1999

Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures.

Ferraiole, 1992

Ferraiole, D. Kuhn, R. Role-Based Access Control. Proceedings of the 19th National Computer Security Conference, 1992

Ferraiole, 1995

Ferraiole, D. Cugini, J., Kuhn, R. Role-Based Access Control: features and Motivations. Proceedings of the 11th Conference in Computer Security Applications, 1995

GEANT, 2013

GÉANT. Identity and Trust Technologies for GÉANT services. <http://www.geant.net/Innovation/Research_Programmes/Pages/Identity_and_Trust_Technologies_for_G%C3%89ANT_Services.aspx> [Accessed 7 Apr 2013]

Hrynaszkiewicz, 2012

Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. Hrynaszkiewicz I, Cockerill MJ.BMC Res Notes. 2012 Sep 7;5:494. doi: 10.1186/1756-0500-5-494.

Internet2, 2013

Internet2. COmanage, Collaborative Organization Management, <http://www.internet2.edu/comanage/> [Accessed 7 April 2013]

Knoppers, 2012

Sampling populations of humans across the world: ELSI issues. Knoppers BM, Zawati MH, Kirby ES. Annu Rev Genomics Hum Genet. 2012;13:395-413. doi: 10.1146/annurev-genom-090711-163834. Epub 2012 Mar 8. Review. PMID: 22404491

McGuire, 2011

To share or not to share: a randomized trial of consent for data sharing in genome research. McGuire AL, Oliver JM, Slashinski MJ, Graves JL, Wang T, Kelly PA, Fisher W, Lau CC, Goss J, Okcu M, Treadwell-Deering D, Goldman AM, Noebels JL, Hilsenbeck SG. Genet Med. 2011 Nov;13(11):948-55. doi: 10.1097/GIM.0b013e3182227589.

Nordunet, 2013

NORDUnet. SUNET COIP. <https://portal.nordu.net/display/COIP> [Accessed 7 April 2013]

OASIS, 2013

OASIS. eXtensible Access Control Markup Language (XACML) Version 3.0. January, 2013.

Ostermann, 2012

Efficiency of the Austrian disease management program for diabetes mellitus type 2: a historic cohort study based on health insurance provider's routine data. Ostermann H, Hoess V, Mueller M. BMC Public Health. 2012 Jun 29;12:490. doi: 10.1186/1471-2458-12-490

REFEDS, 2013

REFEDS Federation Survey. TERENA. <https://refeds.terena.org/index.php/Federations> [Accessed 3 March 2013]

Smedley, 2010

Smedley D, Schofield P, Chen CK, Aidinis V, Ainali C, Bard J, Balling R, Birney E, Blake A, Bongcam-Rudloff E, Brookes AJ, Cesareni G, Chandras C, Eppig J, Flicek P, Gkoutos G, Greenaway S, Gruenberger M, Hériché JK, Lyall A, Mallon AM, Muddyman D, Reisinger F, Ringwald M, Rosenthal N, Schughart K, Swertz M, Thorisson GA, Zouberakis M, Hancock JM. Finding and sharing: new approaches to registries of databases and services for the biomedical sciences. Database (Oxford). 2010 Jul 6;2010:baq014. doi: 10.1093/database/baq014. PubMed PMID: 20627863; PubMed Central PMCID: PMC2911849.

SURFnet, 2013

SURFnet . SURFconext. <http://www.surfnet.nl/surfconext_en> [Accessed 7 April 2013]

SWITCH, 2013

SWITCH. Group Management Tool. <http://www.switch.ch/aai/support/tools/gmt.html> [Accessed 7 April 2013]

SWITCH, 2013b

SWITCH. Argus Authorisation Service. <http://www.switch.ch/grid/argus/index.html> [Accessed 7 April 2013]

Tabor, 2012

Informed consent for whole genome sequencing: a qualitative analysis of participant expectations and perceptions of risks, benefits, and harms. Tabor HK, Stock J, Brazg T, McMillin MJ, Dent KM, Yu JH, Shendure J, Bamshad MJ. Am J Med Genet A. 2012 Jun;158A(6):1310-9. doi: 10.1002/ajmg.a.35328. Epub 2012 Apr 24. PMID: 22532433

Biographies

Mr Mikael Linden received his doctoral degree from Tampere University of Technology in information security in 2009. He works at CSC – the IT Center for Science Ltd. and has been operating the Haka federation of the Finnish research and higher education since 2005 and has chaired the subtask for the policy development in the eduGAIN task (SA3 T3) of the GN3 project. He has consulted several research communities in authentication and authorization related issues.

Tommi Nyrönen, PhD works at CSC – the IT Center for Science Ltd. and is the head of the Finnish ELIXIR node. In Finland ELIXIR is part of the Biomedinfra consortium with BBMRI and EATRIS. After receiving Ph.D. in biochemistry (biocomputing) he has worked over 10 years to make ICT services for science. TN is an adjunct professor in computational drug design in the University of Helsinki, external member in Biocenter Finland bioinformatics infrastructure network, and a member of the board of the National Graduate School of Structural and Informational Biology - ISB.

Ilkka Lappalainen, PhD, works at the European Bioinformatics Institute (EBI) as a project leader for Variation Archives that also includes the European Genome-phenome Archive (EGA). After receiving his PhD in biochemistry he has worked as a scientist in Cambridge University and since 2007 as part of the EBI providing IT services that support the science.