

Towards a Translational Medical Research Ecosystem

Matthew Smith, Marian Harbach
Leibniz Universität Hannover
Distributed Computing & Security Group, L3S
Hannover, Germany
smith,harbach@dcsec.uni-hannover.de

Susanne Mertins
Harvard Medical School, Boston
Laboratory for Translational Research,
University Duisburg-Essen
Institute for Anatomy,
Susanne_Mertins@hms.harvard.edu

Andrew Lewis
Griffith University
School of ICT
Brisbane, Australia
a.lewis@griffith.edu.au

Lyn Griffiths
Griffith University
Genomics Research Centre
Brisbane, Australia
l.griffiths@griffith.edu.au

Abstract—In this paper we introduce a novel design for a translational medical research ecosystem. Translational medical research is an emerging field of work, which aims to bridge the gap between basic medical science research and clinical research/patient care. We analyze the key challenges of digital ecosystems for translational research, based on real world scenarios posed by the Lab for Translational Research at the Harvard Medical School and the Genomics Research Centre of the Griffith University, and show how traditional IT approaches fail to fulfill these challenges. We then introduce our design for a translational research ecosystem. Several key contributions are made: A novel approach to managing ad-hoc research ecosystems is introduced; a new security approach for translational research is proposed which allows each participating site to retain control over its data and define its own policies to ensure legal and ethical compliance; and a design for a novel interactive access control framework which allows users to easily share data, while adhering to their organization's policies is presented.

Digital Ecosystem; Translational Research; Mind Mesh; Access Control; Access Negotiation; Legal Compliance;

I. INTRODUCTION

Translational medical research [1] is an emerging field of work with the aim of bridging the gap between basic medical science research (bench) and clinical research/patient care (bed). The main focus of the translational research paradigm is the more efficient translation of basic science research discoveries into clinical science/patient care (bench- to bed-side). The return route (bed- to bench-side) describes the translation process in which bench-side researchers gain access to direct feedback, clinical data and impulses for new research. While the National Institute of Health (NIH) funds billions in basic science research, less than 10% of promising biomedical discoveries were established in clinical practice within 20 years [2]. Thus, the main motivation for funding in this area is to improve the bench to bedside aspect of translational research. Accordingly, the NIH identified translational research as a vital

part of their roadmap and endorsed the foundation of the Clinical and Translational Science Awards (CTSA) in 2005. These grants were created to stimulate progress from scientific innovation to health improvement and by 2012 will have an estimated annual budget of \$500 million [3] [4]. The main target of this funding is the creation of methodologies and collaborations to improve bench to bedside translation and is focused on the human and organizational aspects of translational research, since it is these human aspects that have been identified as the main hindrance in the translational process [5]. In particular, the difficulties of bridging the different operational procedures of bench and bedside organizations and the often dynamic, trans-discipline, trans-organizational nature of collaborations are an important aspect of this problem. Interestingly, IT solutions used in support of translational medical research (e.g. [6] [7] [8]) do not address this area but mainly focus on creating a centralized system for data integration, access and search in a particular problem domain. Thus, while they offer invaluable support to medical researchers within a single domain who have a clear translational path, they currently do not offer support for the entire spectrum of translational research [9]. We argue that one main reason for this shortcoming is the paradigm choice of the system, i.e. that centralized and domain specific IT infrastructure paradigms are used to realize the systems. This works well for certain highly organized areas of translational research, such as cancer research, but does not cope well with the more ad-hoc and dynamic requirements of general translational research, as we will demonstrate with the requirements from the Lab for Translational Research at the Harvard Medical School and the Genomics Research Centre of the Griffith University.

In this paper, we argue that a digital ecosystem paradigm better suits translational research needs and present a novel design for a transnational, translational research ecosystem to facilitate ad-hoc bench- to bed-side and bed- to bench-side research collaborations. Several key contributions are made –

This work is partially funded by the EU-FP7 M-Eco: Medical Ecosystems Project Grant No.247829, ARC Translating Health Discovery Super Science funding, the NHMRC, an Australian Govt International Science Linkages Grant, and the BMBF D-Grid Project DG12.

we identify why traditional approaches to IT infrastructures fall short of fulfilling the needs of translational researchers and deduce the key issues which a digital ecosystem based approach must fulfill; we present a novel design for a decentralized support infrastructure for the management of translational research efforts which suits the ad-hoc nature of these research efforts better than current non-ecosystem approaches; we present a new flexible security system, which can adapt to the different security requirements of bench- and bed-side research environments; and, in keeping with the digital ecosystem paradigm, each participating organization retains control over the security mechanisms and policies guarding their data and meta-data.

II. SCENARIO

In the following two example scenarios of translational research collaborations are sketched.

A. *Bedside to Bench*

There are two scenarios of main interest in the bed- to bench-side route: The validation of research hypotheses based on clinical trials/experience and conducting clinical trials during bench-side research. Gathering existing clinical data or locating suitable candidates for a clinical trial are the issues to be dealt with here. Both cases involve the sharing of bed-side data with bench-side researchers. Currently, bed-side data is mainly shared manually. Below, two example cases are presented:

Trial support: If a bench-side researcher knows of clinical data of relevance to her work, she needs to apply for ethical approval from the bench-side organization to request the data from the bed-side researcher. A committee grants this ethical approval on a case-by-case basis. Once approval is granted the data can be requested manually, i.e. a letter requesting the data is sent. The bed-side researcher must then apply for and get ethical approval from the bed-side organization. This also needs to be approved by a committee on a case-by-case basis. When the final ethical approval is granted, the bed-side researcher must manually prepare the data for the translational sharing process (e.g. anonymization, pseudonymization) and deliver it to the bench-side researcher (usually through print-outs or DVDs). This process typically takes around one year.

Hypothesis Control: While all efforts are made only to produce effective treatments out of basic research finding, it is nonetheless vital to check if hypothesis of bench side results achieve the expected results in clinical trials and practice. As an example: a link between endothelial progenitor cells in the bone marrow (BM-EPCs) and tumor vascularization was discovered. In-vitro research showed that inhibiting these BM-EPCs could impact tumor growth. Thus, drugs were invented to inhibit these. However, in clinical studies it was discovered that the impact on tumor growth was much smaller than expected and not sufficient to combat the growth of cancer. Thus, the hypothesis that BM-EPCs could be used to inhibit cancer growth needed to be revisited by the bench-side researcher.

B. *Bench to Bedside*

Due to the failure of the BM-EPCs to inhibit cancer growth to the expected degree, new bench-side research into EPCs was undertaken. A new local source of endothelial progenitor cells in the vascular zone (VZ) around blood vessels was discovered and studied by the co-authors. These new findings show that these vascular wall EPCs (VW-EPCs) are recruited for tumor vascularization to the degree that was missing in the BM-EPC results. This groundbreaking discovery has obvious applications in cancer treatment. However, there are also a number of other areas of medicine, such as atherosclerosis, wound healing, and recovering ischaemic tissue after heart attacks, that could benefit from this discovery. These basic science findings now need to be translated into medically applicable results. Therefore, drugs need to be researched to inhibit or stimulate the VW-EPCs in the different areas of application. For this, experts from the fields of oncology, dermatology and cardiology need to be involved in the respective drug research and clinical studies.

Currently, bench-side research results are usually disseminated manually through for instance publications, conference presentations or direct discussions between colleagues. There is very little in terms of IT infrastructure to facilitate collaborative information sharing, particularly across multiple medical domains.

III. PROBLEM STATEMENT

Both aforementioned scenarios would benefit from a digital ecosystem that allows researchers across several organizations and domains to collaborate. For this the digital ecosystem infrastructure must be able to handle data with and without identifying patient information safely. Thus, multiple organizational, legal and ethical as well as security and privacy related problems arise. While bed-side organizations are often centrally organized and have centrally defined security, privacy, and ethical policies (though often not centrally enforced), bench-side research is often a more self-organized and ad-hoc environment in which individual researchers can have far more control over security and privacy decisions than in a clinical environment. Thus, a digital ecosystem that incorporates both worlds needs to deal with the following challenges beyond what traditional translational research systems offer.

A. *Legal/Security*

One of the main issues hindering the adoption of IT solutions in the medical sector is security. Regan et al. studied the security issues in a clinical ecosystem, finding that it is an inherently difficult task to specify access control requirements: "Optimal access control specification requires an understanding of the nature of the task, the application systems that these tasks might affect, and the different access right configurations required to perform these tasks in the different configurable systems." [10]. This problem is magnified manifold in a translational research scenario. Not only are there highly conflicting legal and policy constraints in bench- and bed-side environments, the researchers also have conflicting

interests (e.g. gaining access to as much data as possible vs. protecting patient privacy).

Heterogeneous Access Control: The centralized approach of current translational research platforms does not fulfill the requirements of a translational research ecosystem. Current systems apply a single access control concept (usually RBAC, [11]). This works well for homogenous translational research environments, such as the translational cancer research project supported by caBIG [6], since here it was deemed acceptable that all users register in a central system and the homogenous nature of the data, roles and tasks allows RBAC to be applied effectively. However, in a more general digital ecosystem for translational research, much more diverse objects, users and rules need to be supported. While clinical organizations often have central identity management frameworks and organization-wide roles and policies (ideal for role based AC), bench-side researchers are often organized in a more ad-hoc manner and can own data and autonomously decide on its use (ideal for discretionary AC). Thus, a digital ecosystem needs to be aware of and facilitate the different AC requirements of the different worlds and not impose a single AC paradigm on all users homogeneously.

Access Control Negotiation: Another challenging aspect of a translational ecosystem, unique to transregional and transnational translational research projects, are the diverse legal and ethical policies governing the participating organizations. While traditional AC systems offer binary yes/no answers to access requests, this significantly falls short of what is ideally required in a translational research environment, since the diverse legal and ethical constraint create far more complex interaction patterns. If a bench-side researcher from organization A in the US tries to access pseudonymized clinical data from organization B in Australia, a traditional AC system would reject the request, since sharing of clinical data outside of organization B is forbidden by law. However, if ethical approval is granted, anonymized data may be shared. An AC system for a translational research ecosystem needs to be aware of such legal constraints and be able to negotiate the appropriate type of access, i.e. the request should be denied with the added information that anonymized access can be permitted and ask whether the access request should be modified accordingly. This concept can be taken a step further, to offer significantly greater benefit to the researcher. In the above scenario, an ethical committee must be involved in the AC request, since even anonymized clinical data may not be shared by Australian organizations without permission by the committee on a case-by-case basis (this can take months to acquire). However, clinical organizations in the federal state of Hamburg, Germany for instance are not as constrained. Here, a clinician has full personal control over data once it has been anonymized [12]. Thus, if similar data is present in this organization, gaining access to the data only requires the permission of a single person and thus can potentially be obtained in a much shorter timespan. This sort of knowledge about the legal and ethical constraints on the data needs to be incorporated into the AC system to enable optimal access to data in a legal and ethically compliant manner. This involves several aspects not currently present in AC systems, i.e. negotiation, reasoning and user interaction capabilities.

Furthermore, it is desirable that the underlying infrastructure facilitates the organizational aspects of these processes, such as automatic anonymization or pseudonymization of data, locating relevant alternative data sources, contacting persons and importantly the ability to offer an interactive AC process.

Sovereign Access Control: A further consequence of the diverse legal and ethical laws governing data access is that it is unrealistic to assume that in a translational digital ecosystem spanning multiple countries, all data can be managed by a central AC system. Therefore, it is necessary to offer a decentralized AC system in which each participating organization can retain sovereign AC authority over its data. This is a fundamental conceptual difference to what the traditional infrastructure approaches offer, where the aim is to centralize and homogenize data storage and access as far as possible.

B. Technical/Administrative

Decentralized Control: The major infrastructure requirement deduced from the legal and security requirements is that a digital ecosystem for translational research must allow for decentralized regions of sovereign control. Current ICT infrastructures for translational research, such as [6] [7] [8], offer central services for a subset of tasks in translational research, such as user and Virtual Organization (VO) management, access control, data storage, as well as search and retrieval. While these services can be of great benefit to the community, the central nature of these services require participating organizations to adapt to the rules and capabilities of these systems. In a translational ecosystem, a more flexible decentralized approach is desirable in which participating organizations can retain internal organizational structures and procedures. In particular, the diverse structures and operational procedures of bench-side research groups are ill suited to be unified into a single central management infrastructure.

Regan et al. found that technologically there is little reason to avoid integration of ICT methods in e-Health ecosystems. Most barriers arise from acceptance and organizational issues with the professionals [10]. The decentralized ecosystem approach allows the ICT solution to be tailored more closely to the needs of the participating organizations.

Human Centric Information Management: A further technical aspect of translational information management is data integration. Queries such as "Show me all pathology reports for Her-2/neu positive patients with a lobular carcinoma." or "Identify all lymphoma patients who have had good prognostic responses to rituximab/CHOP treatment regimes" are common in translational research and are currently often dealt with manually, i.e. the researcher queries the local database/software and/or requests the owner of a foreign data source to execute the query. Standardization approaches such as HL7 [13], DICOM [14] and caDSR [15] aim to create the basis for integrated information systems, which can execute queries on common data formats across multiple systems. Ontology based approaches are also widely used and allow queries across different data sources and domains by creating mappings between the different standards. While both these approaches are highly laudable and already

offer great benefits to early adopters, standardization is still far from complete and ontology based approaches are very work intensive. Thus, many incompatibilities remain. Also, the top down approach (use of standard X or ontology Y mandated by the IT system) does not sit well with the organizational constraints of many medical research organizations, where the integration/roll-out of a new standard can be extremely labor and cost intensive. In order to offer benefits to early adopters at a good benefit to effort ratio, a translational medical ecosystem needs a human centric, non-invasive infrastructure to assist researchers in locating relevant organizations, data and fellow researchers, without mandating all participating organizations to adopt a single format and make significant changes to their current infrastructure.

IV. RELATED WORK

A. Translational Research Platforms

Winter et al. [9] present a study on integrated information systems for translational research. They identify several key issues that such a platform needs to deal with: Availability of relevant items, reliable definitions of items, appropriate level of data quality, unambiguous identifications of objects and legally compliant access. This study is used as a foundation the work presented in this paper.

The caTrip Project [7] is a translational research extension to the caBIG cancer research platform. The project's main focus is on aggregating clinical and molecular data in the caBIG cancer research network. The platform allows users to pose queries such as: "What are all the ER positive patients that have survived breast cancer after radiation treatment?" To achieve this, all data must be registered with the same access control policies in a standardized manner in the Cancer Data Standards Repository using the common data elements (CDE) format. An RBAC AC system is set up and managed centrally to prevent unauthorized access to the data. Due to the homogeneous and centrally organized nature of the caBIG cancer research program, this is a legitimate approach. However, it is not applicable to the broader and more ad-hoc world of translational ecosystems. The drawback is also visible in the security concept proposed by caTrip, which relies on the Globus Toolkit Authentication Service and the Common Security Module (CSM) authorization. Again, due to the focused nature of the project and the data, the binary AC decision support offered by this approach is sufficient for caTrip. It is however not flexible enough to support a translational ecosystem.

A competing translational research extension in the caBIG project is caIntegrator [6]. The caIntegrator project lets researchers and bioinformaticians access, analyze, and integrate clinical and experimental data across multiple clinical trials and studies. Like caTrip, the focus of caIntegrator is on data-integration and search. However, also like caTrip, the user must conform to the user management, data policies and a single AC system that enforces binary access control on the system.

Neither project offers significant support for the ad-hoc and self-managed nature of many benchside projects, enforcing the

same central administrative approach on all aspects of the research ecosystem. The same holds for the myGrid project [8]. Furthermore, the systems all focus on one specific translational track (e.g. cancer) and do not facilitate the collaboration of the different domains as is required by the scenario posed by the Lab for Translational Research at the Harvard Medical School and the Genomics Research Centre of the Griffith University.

B. Digital Ecosystems

Nankani et al. [16] describe universities and research facilities as prime examples of digital ecosystems. Research projects and their administration often thrive on dynamic, ad-hoc and diverse relationships among entities and organizations. Frigidis et al. [17] state that research ecosystems benefit from clear context and support for integrated macro- and microscopic views. They also acknowledge the temporary nature of collaborations and argue that research ecosystems need to be relieved from excessive complexity imposed by project management. Furthermore, they ask for the research ecosystem approach to investigate "alternative or improved solutions in the organization and implementation of research activities". Both these analyses show how the digital ecosystem paradigm offers benefits to the dynamic research environment and form the solid motivation for this work.

Katriou et al. [18] suggest support for EU research proposal creation via a digital ecosystem approach. The proposed EURPCCS system retrieves the research ecosystem's actors via a set of mapping web services. The EURPCCS enables the suggestion of appropriate partnerships based on actor profiles. They demonstrate how the digital ecosystem paradigm is well suited to assist in the ad-hoc creation of research teams. Their findings strengthen the case for the benefits of digital ecosystem based management of translational research projects, since the creation of a translational research project shares several attributes with the creation of an EU research project. Their work however does not deal with the execution of the project, project data or security issues. Furthermore, the portal based solution is centrally managed, requiring all participating organizations to accept and adhere to the central policies.

V. TRANSLATIONAL RESEARCH ECOSYSTEM

Digital ecosystems have emerged as a new paradigm for complex, interdependent, loosely coupled and demand-driven interactive environments [19]. Traditional approaches to translational research information management have focused on how to centralize and homogenize data storage as well as data access as far as possible. This runs contrary to the digital ecosystem paradigm and the natural usage pattern of translational research. In the following, we introduce Mind Mesh, a digital ecosystem infrastructure for translational research, which presents novel solutions to the challenges presented in Section III.

Human Centric Information Management: Standards, such as HL7 [13], DICOM [14], and caDSR [15], aim to standardize medical data and systems to create a homogeneous environment in which systems can be linked and information can be stored and retrieved automatically. Even though much

progress has been made, clinical information systems are still far from plug-and-play compatible and collaborations across organizational boundaries are still a complex endeavor. To facilitate ad-hoc collaborations across multiple domains in which centralized systems and standardization are not advanced enough yet, we propose the concept of the Mind Mesh. The Mind Mesh is a conceptual extension of Mind Maps that allows the collaborative modeling of research environments. Mind Maps focus on intuitive visualization of ideas, reflecting the way in which humans tend to organize complex relationships between concepts. However, the strict tree-structure of Mind Maps does not lend itself to modeling research collaborations and the relationships between the participating entities. The Mind Mesh models information in a graph structure and allows arbitrary relationships between any two entities. Nankani et al. [16] already showed how automated graph based visualization of research ecosystems is beneficial for human interpretation of complex research ecosystems. We take this concept one step further and suggest that that users allow the user to actively create and maintain the information graph like a Mind Map and combine the visualization with active components to create a fully functional visual translational research ecosystem. The Mind Map concept helps users to find information in a human centric manner where automated search tools and central databases fail to bridge the gap between different translational domains. Figure 1 shows a partial screenshot of a Mind Mesh, depicting two organizations A and B. Organization A is a clinical research institution and contains three data sets organized into two study groups and one project group. Each node can be annotated with keywords in aid of search and navigation. Unlike in a Mind Map, the Mind Mesh allows objects to have multiple parents, thus allowing data set C to be assigned both to study B as well as project A. This is important both for navigation and search reasons as well as access control reasons. Organization B, a basic research organization, also contains three data sets organized into two projects. Again, due to the graph-based nature, projects B and C can share data set E. Cross-organizational data sharing is one of the major features of a translational research ecosystem. Four distinct data sharing methodologies are shown in the example:

- *Bidirectional Node to Node*: Two graph nodes share their resources. Users of either node can access resources of either node, e.g. Users connected to Study B or Project B can access data sets B, C, D and E.
- *Unidirectional Node to Node*: One graph node makes its resources available to another, e.g. Users of Project A can access data sets E and F from Project C.
- *Person to Node*: A node makes its resources available to a person, e.g. the user assigned to Project C gains access to data sets D and E.
- *Person to Data*: A person is granted access to a data set directly, e.g. the user from organization B is assigned to data set C from organization A.

To grant access, a user with administrative privileges for a node can simply drag a connection to the new node. This extremely simple operation is vital to a translational research

ecosystem, since most users do not have extensive ICT skills and granting correct access on the physical resources is too complex and requesting an administrator to execute the action each time is too inefficient and impractical. The technical execution of these graphical requests is dealt with by a plugin system described in the next subsection. All access is still constrained by organizational policies, discussed in Section VI.

Decentralized Control: Particularly the legal and security requirements make a decentralized approach for a translational research ecosystem vital for its adoption. Also, user management and legacy system integration make a decentralized approach far easier to adopt. Figure 1 shows the architecture of the Mind Mesh ecosystem for the example research collaborations. Both organizations operate a Mind Mesh server and retain full control over their user management and data. The clinical research organization A uses a central Identity Management System to manage its users. A plugin is required to export the user information into the Mind Mesh server. The benchside organization enrolls its users directly with the basic inbuilt user management of the Mind Mesh server. Data can be stored in multiple ways (e.g. SQL databases, HTTP, SMB or SVN servers, etc.) and each system requires a plugin to connect it to the Mind Mesh server. This involves an initial overhead, however the benefits of automated management at run-time are significant, since non-technical users can grant and revoke access to resources without administrative assistance. Vital components to each Mind Mesh server are the domain policies that are required to ensure legal and ethical compliance in the translational research environment. These will be discussed in the next section.

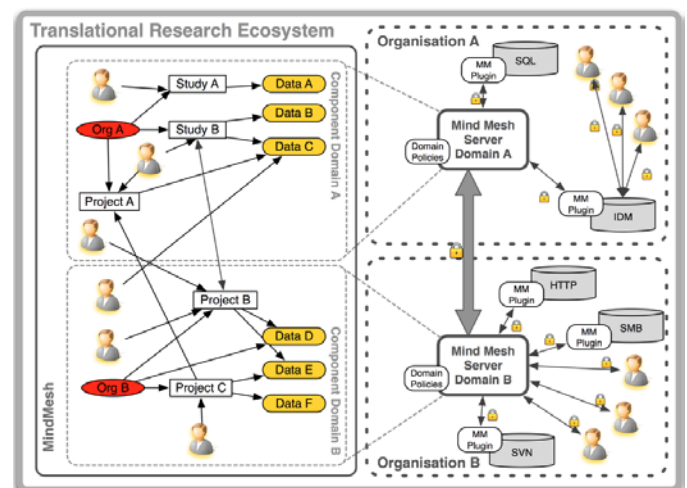


Figure 1: Mind Mesh Architecture

VI. SECURITY SYSTEM

Using the graphical Mind Mesh paradigm introduced in Section V, it is easy for a user in one organization to grant access to resources under his or her control to a user in another organization. However, as shown in Section III, there are complex legal and ethical constraints that must be adhered to. Since dealing with these issues manually is both time consuming and error-prone, our digital ecosystem infrastructure offers several novel approaches to automate

security and privacy relevant operations and ensure legal and ethical compliance (e.g. anonymization or pseudonymization of data). Two main issues need to be addressed: a policy framework for legal and ethical compliance and a usability concept. Each will be briefly sketched in the following.

The policy and access control framework is a vital part of the translational research ecosystem infrastructure. To summarize, the requirements identified in the previous sections are as follows:

- Enable owners of data and meta-data to specify who can access their data in which form a-priori;
- Enable organizations to specify who can access data and meta-data in which form within their domain a-priori;
- Enable users to request access to data and meta-data;
- Enable and automatically negotiate full, pseudonymized or anonymized access in cases where no policy or access control rule is violated;
- Enable owners and organizations to interactively respond to access requests that cannot be resolved automatically;
- Enable and support ethical committees to allow, reject or modify access requests that require ethical approval to comply with the legal and ethical framework.

The first three requirements are addressed using the Mind Mesh paradigm. Connections between nodes in the graph can be annotated with attributes such as *works_on* or *has_access_to*. Based on these annotations, access control rules can be specified globally or for sub-graphs, e.g. anybody who *works_on* a project can access all data connected to the project. Similarly, a data request is as simple as dragging a connection between the users node and the data node. Whenever a new connection is added or used, the active security policies are queried to check for compliance. If the breach of compliance can be resolved automatically, e.g. pseudonymized data was requested but only anonymized data may be shared, automatic policy based transformation of the request can be offered to the user. For a number of medical file formats (e.g. DICOM) automatic anonymizers and pseudonymizers exist, that allow an automated transformation of the data. The plugin architecture of the security model allows these transformations to be executed during the access negotiation phase.

However, not all requests can be dealt with automatically. This can either be due to particularly complex transformations (such as can occur for pseudonymization where re-identification is a problem) or due to non-reconcilable request/policy combinations (such as only anonymized data may be shared without ethical approval but pseudonymized data is required). In these cases, where the organizational policies or the user policies do not allow automated access, a user interaction event needs to be triggered in the AC framework and the blocking entity is presented with the request.

Addressing these needs requires a comprehensive but flexible policy engine. Especially user interaction, owner-based policy evaluation and data manipulation are not readily available in current approaches. Our work aims to extend the

Protune policy engine [21] to allow user interaction and policy negotiation, as is needed for the security requirements in the translational ecosystem. The following two examples show how policies that include interaction with a user (e.g. the ethical committee) are specified. MDAT stands for medical data.

```
allow(retrieve(MDAT)) :-
    isProtected(MDAT),
    userInteraction("Ethical Committee
        approval required. Approve?", MDAT),
    anonymize(MDAT).
```

The exemplary organizational policy above is taken from an Australian Mind Mesh site. The legal and ethical requirements for the Australian site require that all medical data sharing must be approved by an ethical committee and even then only anonymized data may be shared. Thus, user interaction and anonymization need to be hard coded into the policy, so that no accidental sharing of non-anonymized or unapproved sharing can take place. In the case of a German site, the policy options are more flexible:

```
allow(retrieve(MDAT)) :-
    isPublic(MDAT),
    anonymize(MDAT).
allow(retrieve(MDAT)) :-
    isPublic(MDAT),
    isProjectPartner(MDAT, $Requester),
allow(retrieve(MDAT)) :-
    isProtected(MDAT),
    userInteraction("User $Requester
        requests:", MDAT).
```

In this case, anyone can access the data if the owner of the data ensures that the data is anonymized and marks it as public. As an example, if the data is marked as protected, only project partners may access it. Any other access attempts will trigger a user interaction informing the owner of the data about the request. A rule definition including constructs to address these concerns directly instead of using external mechanisms enables concise and flexible specifications. To increase the usability of the security mechanisms especially for non IT personnel, the semantic information contained in the Mind Mesh graph can be used to specify, query and also explain policies visually, which is of great benefit to the non-technical users of our translational research ecosystem.

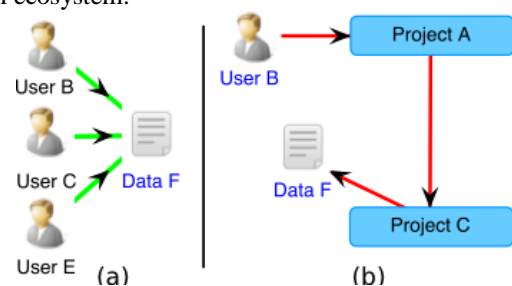


Figure 2: Access Control Visualization

Figure 5 shows two concept drawings of two such policy queries, based on the project graph in Figure 1. The first query was executed by the owner of F to find out who has access to the resource. Since User B belongs to another organization and is unknown to the owner, the second query shows over which

policy path User B has access to the resource. This visualization of access rights is a significant aid, especially for non-technical users in complex digital ecosystems.

VII. TECHNOLOGY

The early infrastructure prototype underlying our translational research ecosystem is based on several components. The Mind Mesh servers are built on top of the Openfire XMPP implementation. XMPP provides secure real-time server-to-server and server-to-client communication and offers a robust and highly extensible basis for the ecosystem. We currently support X.509, LDAP and/or self managed username/password authentication. Additional IDM and authentication systems can easily be integrated via a standard security layer. The Mind Mesh graph itself is stored as Resource Description Framework (RDF) triples, since graph relationships between nodes directly correspond to the *subject – predicate – object* paradigm of RDF. This also has the benefit that we can use SPARQL to query the graph for search operations. We have currently implemented two security plugins, one allows simple rules to be defined in the Mind Mesh and then translated into SQL database access control rules and the other allows the automated anonymization of DICOM records. The Mind Mesh UI is browser based, since this significantly reduces the deployment and management effort and facilitates adoption of the digital ecosystem paradigm in a medical environment.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we introduced a novel digital ecosystem design and infrastructure for transnational, translational medical research. Several key contributions were made – we identified why traditional IT infrastructures approaches fall short of fulfilling the needs of translational researchers and deduced the key issues for our digital ecosystem based approach. We presented a novel XMPP based decentralized support infrastructure for the management of translational research ecosystems. A new security system encompassing the different security requirements of bench and bedside research environments was designed to reduce the manual effort of legal and ethically compliant collaboration. The security design allows each participant in the digital ecosystem to retain control over their data and specify custom policies to ensure legal and ethical compliance in this complex environment.

There are several areas for future work: Currently, the entire research ecosystem is visualized in its entirety with all details. As the networks grow, it will become necessary to reduce the amount of information shown at any given time. For this, novel approaches to select relevant information are needed as well as the capability to define and select views within the ecosystem. Currently, the prototype is not capable of graphically defining policies. We aim to create a smart selection mechanism to create new access control policies, which will further increase the intuitivity and usability of our system.

REFERENCES

- [1] B.C. Drolet and N.M. Lorenzi, "Translational research: understanding the continuum from bench to bedside.," *Translational research : the journal of laboratory and clinical medicine*, vol. 157, Jan. 2010, pp. 1-5.
- [2] D.G. Contopoulos-Ioannidis, E. Ntzani, and J.P.A. Ioannidis, "Translation of highly promising basic science research into clinical applications.," *The American journal of medicine*, vol. 114, Apr. 2003, pp. 477-84.
- [3] E.A. Zerhouni, "Translational and clinical science--time for a new vision.," *The New England journal of medicine*, vol. 353, Oct. 2005, pp. 1621-3.
- [4] E.A. Zerhouni, "US biomedical research: basic, translational, and clinical sciences.," *JAMA : the journal of the American Medical Association*, vol. 294, Sep. 2005, pp. 1352-8.
- [5] L.A. Green and C.M. Seifert, "Translation of research into practice: why we can't "just do it".," *The Journal of the American Board of Family Practice / American Board of Family Practice*, vol. 18, pp. 541-5.
- [6] J. Saltz, T. Kurc, S. Hastings, S. Langella, S. Oster, D. Ervin, A. Sharma, T. Pan, M. Gurcan, J. Permar, R. Ferreira, P. Payne, U. Catalyurek, E. Caserta, G. Leone, M.C. Ostrowski, R. Madduri, I. Foster, S. Madhavan, K.H. Buetow, K. Shanbhag, and E. Siegel, "e-Science, caGrid, and Translational Biomedical Research," *Computer*, vol. 41, Nov. 2008, pp. 58-66.
- [7] P. McConnell, R.C. Dash, R. Chilukuri, R. Pietrobon, K. Johnson, R. Annechiarico, and A.J. Cuticchia, "The cancer translational research informatics platform.," *BMC medical informatics and decision making*, vol. 8, Jan. 2008, p. 60.
- [8] R.D. Stevens, A.J. Robinson, and C.A. Goble, "myGrid: personalised bioinformatics on the information grid.," *Bioinformatics (Oxford, England)*, vol. 19 Suppl 1, Jan. 2003, pp. i302-4.
- [9] A. Winter, G. Funkat, A. Haeber, C. Mauz-Koerholz, K. Pommerening, S. Smers, and J. Stausberg, "Integrated information systems for translational medicine.," *Methods of information in medicine*, vol. 46, Jan. 2007, pp. 601-7.
- [10] D.B. Regan, O.T. Pusatli, E. Lutton, and D.R. Athauda, *Securing an EHR in a health sector digital ecosystem*, IEEE, 2009.
- [11] R.S. Sandhu, E.J. Coyne, H.L. Feinstein, and C.E. Youman, "Role-based access control models," *Computer*, vol. 29, 1996, pp. 38-47.
- [12] "Hamburgisches Krankenhausgesetz (HmbKHG)," <http://www.landesrecht.hamburg.de/jportal/portal/page/bshaprod.p.sml?showdoccase=1&doc.id=jlr-KHGHArahmen&st=lr>, 1991.
- [13] "Health Level Seven International." <http://www.hl7.org>
- [14] "DICOM." <http://medical.nema.org/>
- [15] P.A. Covitz, "caCORE: A common infrastructure for cancer informatics," *Bioinformatics*, vol. 19, Dec. 2003, pp. 2404-2412.
- [16] E. Nankani, S. Simoff, S. Denize, and L. Young, *Enterprise university as a digital ecosystem: Visual analysis of academic collaboration*, IEEE, 2009.
- [17] G. Frigidis, A. Mavridis, A. Vontas, A. Koumpis, and K. Tarabanis, *A proposed conceptual framework for the study of research ecosystems*, IEEE, 2008.
- [18] S.-A. Katriou, E. Toliaas, and A. Mavridis, *A european union research partner collaboration creation system*, IEEE, 2009.
- [19] H. Boley and E. Chang, *Digital Ecosystems: Principles and Semantics*, IEEE, 2007.
- [20] C.-M. Reng, P. Debold, C. Specker, and K. Pommerening, *Generische Lösung zum Datenschutz für die Forschungsnetze in der Medizin*, MWV Medizinisch Wissenschaftliche Verlagsgesellschaft OHG, 2006.
- [21] P. Bonatti, J.L. De Coi, D. Olmedilla, and L. Sauro, "A Rule-Based Trust Negotiation System," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, Nov. 2010, pp. 1507-1520.