

# A Platform for Data Sharing in an Open Health Environment

---

**Le Z., Ford J., Makedon F.**

*Dartmouth Experimental Visualization Laboratory (DEVLAB)*

*Dartmouth College, Hanover, NH, USA*

*{zhengyi.le, James.Ford, Fillia.Makedon}@dartmouth.edu*

## Introduction

In order to enhance data sharing among professional communities and individuals in health informatics, we are developing a platform and tools that allow secure data sharing in open environments like the Internet, where data sharing interactions often occur between entities that do not know each other beforehand. This flexible system supports not only registered members (such as members of a healthcare consortium), but also activities by any independent entity, provided there is mutual agreement between parties. In this context, building trust provides a method to authenticate a stranger, allowing a new entity to join an existing sharing agreement, provided he completes a trust building process. This approach can avoid traditional identity-based methods, which are often too rigid to suit the dynamic Internet environment. In essence, trust building is an automation of traditional human interactive registration processes. Two parties can build bilateral trust by exchanging credentials through automated negotiation [1, 2]. By verifying credentials, each party can determine the trustworthiness of the other prior to negotiating on what to share, and how to share. Each party, whether a user or agent, can tailor their negotiation strategies for the best individual results [3-5].

## Methods and Results

In our approach, data owners use a metadata database to describe their data. Finding matches in sensitive or complex data is difficult, and allowing a requester to access databases directly is often not appropriate. Through a standard of metadata representation, sensitive or complex objects can be securely and efficiently accessed, traded, and evaluated in a summary form. This not only protects the original data but also makes highly heterogeneous objects interoperable and amenable to comparison. Metadata information services can be used to dynamically extract and manage metadata; for more details, please see [7]. In general, each party may act as both a requester and a resource owner depending on circumstances.

When an entity finds a match in published metadata databases, it then sends a request to the data's owner. The owner checks its data access policies and sends a counter request for required credentials or agreements. The requester reads the counter request and checks its local credential release policies. Trust negotiation then continues. More details are in [1,2,6]. Traditional access control list technology is limited by the fact that the server only of-

```

DHMC.GeneData ← DHMC.prefered or Gov.authorized
DHMC.prefered ← DHMC.university.student
DHMC.prefered ← DHMC.institution.employee
DHMC.university ← UNH (University of New Hampshire)
DHMC.university ← Dartmouth
DHMC.institution ← New England Medical Reseach Center
DHMC.PatientsData ← DHMC.doctors or GOV.authorized
DHMC.doctors ← Bob
DHMC.doctors ← Carlo
.....

```

**Figure 1.** Access polices on DHMC database.

```

DHMC.GOVAuthorizedHospitalCert ← True

```

**Figure 2.** DHMC's credential release policy.

```

Alice.UNHstudentDiditalID ← True

```

**Figure 3.** Alice's credential release policy.

```

ContagiousDiseaseControlCenter.GOVauthorizedCert
← GovAuthorizedHospital

```

**Figure 4.** Contagious Disease Control Center's credential release policy.

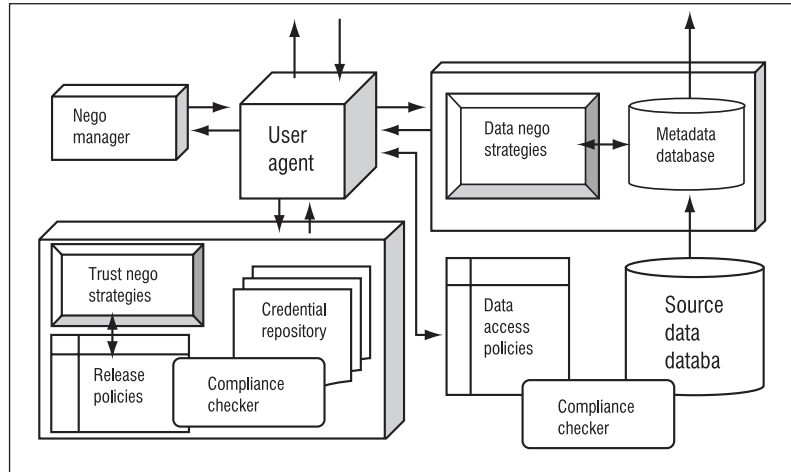
fers services to those clients it already knows. State-of-art policy languages provide a much more powerful tool for allowing resource owners to give access to stangers they don't know directly.

For a simple example (Figures 1-4), suppose DHMC (Dartmouth-Hitchcock Medical Center) has two kinds of data: Gene Data and Patient Data (sometimes patients' personal data are very useful for disease analysis). Gene data is considered safe for general research after de-identification, and DHMC gives gene data access to preferred universities and institutions (Figure 1), with the restriction that they must limit access to their students or employees. If Alice, who is a student of University of New Hampshire, queries the Gene Data, the DHMC server can send back a request for her student digital ID. Alice checks her credential release policy, which says she does not mind showing her student ID to anybody (Figure 3). Thus, this data sharing negotiation succeeds. In another

case, suppose a contagious disease control center sends a request for patient data to the DHMC server; in this case, the server requires a proof that the recipient is a DHMC doctor or a government authorized center. A center may be sensitive about the distribution of this kind of credential (perhaps exposure will make the center a target of attacks), and therefore, the center sends a counter request that says it only shows this credential to an authorized hospital (Figure 4). DHMC is always willing to release its authorization certificate (Figure 2), so once each party verifies the credentials it received, this trust negotiation ends.

The two parties may continue negotiating on how to trade data after trust negotiation succeeds. Our framework supports dynamically assigning an appropriate negotiation strategy to an agent and creating new negotiation rules by learning from past negotiation (details are in [4]). Although credentials could be viewed as another form of protected data, we consider trust negotiation and data sharing negotiations separately because (1) an entity may act only as a simple requester, and may wish to use his own authorization database and (2)

**Figure 5.** Our data sharing system.



strategies for the two kinds of negotiation are not interchangeable. However, we do use a unified negotiation manager module to manage and trace the progress of these two kinds of negotiations.

## Discussion

We use this combination of methods (metadata, trust management, and trust negotiation) to allow a data owner to share directly with a stranger he doesn't know directly, which is the primary lack of traditional identity-based approaches.

There are many requirements for trust management and negotiation languages, such as well-defined semantics, monotonicity, credential combination, etc. However, no current policy language, such as TPL, X-sec, PSPL, and KeyNote, meets all these requirements. In addition, there are some practical concerns, such as credential validity, ownership, and chain discovery.

## Acknowledgement

This work was supported in part by the National Science Foundation under grants ITR-0312629 and IDM-0083423.

## References (excerpt only)

- [1] S. Ye, F. Makedon, and J. Ford, Collaborative Automated Trust Negotiation in Peer-to-Peer Systems, *Peer-to-Peer Computing* 2004:108-115.
- [2] E. Bertino, E. Ferrari, and A. Squicciarini, Trust-X: A Peer-to-Peer Framework for Trust Establishment, *IEEE Trans. Knowl. Data Eng.* 16(7): 827-842 (2004).
- [3] S. Ye, F. Makedon, T. Steinberg, L. Shen, J. Ford, Y. Wang, Y. Zhao, and S. Kapidakis, SCENS: A System for the Mediated Sharing of Sensitive Data, *JCDL* 2003: 263-268.
- [4] S. Zhang, S. Ye, F. Makedon, and J. Ford, A hybrid negotiation strategy mechanism in an automated negotiation system, *ACM Conference on Electronic Commerce* 2004: 256-257.

- [5] F. Makedon, S. Ye, S. Zhang, J. Ford, L. Shen, and S. Kapidakis, Data Brokers: Building Collections through Automated Negotiation, SETN 2004: 13-22.
- [6] N. Li and J. Mitchell, RT: A Role-based Trust-management Framework, In Proceedings of The Third DARPA Information Survivability Conference and Exposition (DISCEX III), Washington, D.C., April 2003. IEEE Computer Society Press, Los Alamitos, California, pp. 201--212.
- [7] F. Makedon, J. Ford, L. Shen, T. Steinberg, A. Saykin, H. Wishart, and S. Kapidakis: MetaDL: A Digital Library of Metadata for Sensitive or Complex Research Data. ECDL 2002: 374-389.