Trust Model for Efficient Honest Broker based Healthcare Data Access and Processing

Mauro Lemus Alarcon*, Minh Nguyen†, Saptarshi Debroy†
Naga Ramya Bhamidipati*, Prasad Calyam*, Abu Mosa*
* University of Missouri-Columbia, † City University of New York
Emails: {lemusm@umsystem.edu, mnguyen@gradcenter.cuny.edu, saptarshi.debroy@hunter.cuny.edu
nbny6@mail.missouri.edu, calyamp@missouri.edu, mosaa@health.missouri.edu}

***Abstract*—With the increased push to promote data-driven methods in modern healthcare, there is a tremendous need for fast access to clinical datasets in order to pursue medical breakthroughs in the areas of personalized medicine and big data knowledge discovery. However, the inherent lack of trust between the data custodians and data consumers/users has resulted in a fully manual honest broker approach to access and process protected healthcare data. Such a manual approach leads to slow data handling, and adds to overheads needed to address data auditability and assurance needed for compliance with healthcare data security standards. In this paper, we address these challenges by proposing a trust model to enable semi-automation of the honest broker process to increase its efficiency. The trust model is based on multi-dimensional risk management principles and considers risk associated with data identifiers, as well as requestor profile and reputation. We implement and evaluate a semi-automated honest broker that uses our trust model in a community cloud testbed using the SynPUF synthetic dataset. Our experiment results show that our multi-dimensional risk management approach consistently identifies the lower confidentiality risk configuration in the semi-automation in comparison with a one-dimensional strategy. Thus, our semi-automated honest brokering approach improves efficiency for data custodians and data consumers by facilitation of fast and secure data access, while also ensuring compliance in the processing of the protected datasets.**

***Index Terms*—Security, trust and privacy, Cloud data processing, Data-driven healthcare, Common data model, Resource brokering, Risk management**

I. Introduction

Cloud platforms facilitate a scalable and on-demand hosting of healthcare big data collections [1] [2]. The big data is obtained from multiple healthcare organizations and corresponds to protected datasets such as e.g., patients' data history, clinical diagnosis, laboratory results, medical imaging, data from wearable and IoT devices, clinical outcomes, and healthcare related financial data. Increasingly, researchers are demanding fast and secure access to such multi-source data to conduct knowledge discoveries that include finding rare patterns in heterogeneous datasets [3]. However, data custodians need to handle such data request brokering for protected data without violating healthcare data security standards [4]. Given the lack of inherent trust between the data custodians (health care organizations and hospitals) and data consumers (researchers, clinicians), a fully manual "honest broker" approach is used in the data governance practice that causes undesirable levels of bottlenecks in data handling.

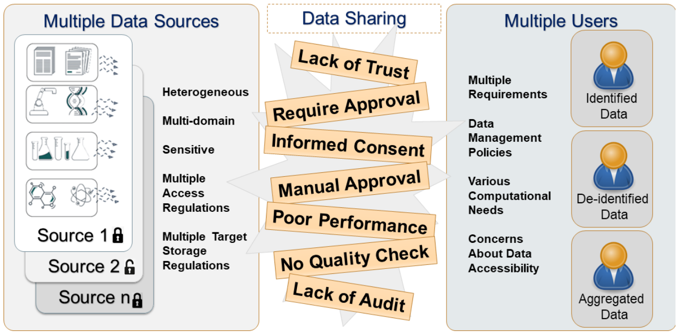


Fig. 1. Data sharing challenges faced by a fully manual brokering process

Fig. 1 shows the challenges and limitations faced by the current manual honest broker governance process. The limitations are amplified when there is a need to access multiple data sources with heterogeneous data types and varying access/processing policies provided by multiple organizations. The burden of a manual review and approval of protected data requests becomes more complex in these cases to ensure auditability (pertaining to authorization to access data) and assurance (tracking data security/privacy) needed for compliance with healthcare data security standards adds long delays (e.g., order of months). These long delays hinder innovation for researchers and clinicians who need timely information for their studies of disease management i.e., diagnosis, prevention, early prediction, personalized treatment [5] [6] for quality health care. Thus, there is a need to build trust through honest broker automation and allow for data handling of healthcare-community specific policies in a community cloud platform.

In this paper, we address the above challenges in the long-drawn data accessibility while ensuring auditability and compliance with healthcare data security standards by proposing a semi-automated "honest broker" that uses trust automation via a novel trust model in a healthcare community cloud setting. Our trust model approach uses multi-dimensional risk management principles and considers risk associated with

This material is based upon work supported by the National Science Foundation under Award Number: OAC-1827177. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

data identifiers in the request, as well as requestor profile and reputation based on prior successful handling of protected datasets. The reputation calculation is done using trust scores that are assigned to data users/requestors by adopting a conservative Dirichlet distribution [7] for handling highly sensitive datasets, alternately by adopting an optimistic Beta distribution [8] for comparatively less sensitive datasets. The conservative component of this trust score builds towards a long-term reputation for the data requestor that acts as a key system variable towards the overall risk assessment. Based on the overall risk assessment rating, the outcome of this compliance check can result in situations such as: (i) instantaneous 'automatic approval' (i.e., human-out-of-the-loop), or (ii) a 'semi-automated' process through data custodian assistance on filling the data form (i.e., human-on-the-loop), or (iii) a fully-manual 'custodian-in-the-loop' approach (i.e., human-in-the-loop). On the other hand, the optimistic trust score component in the honest brokering is used as part of an Entitlement Service [9] that seeks to reward trustworthy data requestors by e.g., giving more interactive data computation and analysis tools such as Jupyter Notebooks that can accelerate the data sharing, analysis/visualization process with satisfactory user experience [10], while ensuring requirements compliance with suitable auditability and assurance.

We implement and evaluate a semi-automated honest broker, and perform experiments to compare our multi-dimensional risk management approach against a one-dimensional strategy. The comparison is performed in terms of the consistency in the identification of lower confidentiality risk configurations in the honest broker process following the risk assessment method defined by the National Institute of Standards and Technology (NIST) [11]. Next, we perform experiments on comparing the efficiency of our trust model against different model variables. Specifically, we demonstrate how our proposed combination of conservative and optimistic trust models can guarantee a more efficient 'custodian-in-the-loop' approval process or sometimes even disapprovals in cases of lack of compliance for high-risk data requests.

The remainder paper organization is as follows: Section II details our risk-assessment-based honest broker semi-automation. Section III discusses our proposed underlying trust model. Section IV presents the semi-automated honest broker implementation and performance evaluation. Section V concludes the paper.

II. Honest Broker Semi-Automation

A. System Overview

Fig. 2 shows the "honest broker" system that we outline for semi-automation. Our proposed system architecture features a pipeline to speed up the compliance checking and data request handling to provide protected data access to researchers and clinicians. Multiple and disparate data sources are integrated via a common data model (CDM) for storing and cataloging healthcare data along with data consumer information using OMOP-CDM [12] guidelines. The community cloud governance handles the data sharing trustworthiness without compromising the access policy compliance of the data sources by using the IRB and data requests information, and also interfaces with the supplementary components. The supplementary components include the computation and analysis workspace for the data consumers, and the e-Management component for handling various forms in the governance process to assist with the data brokering. In the following, we provide more details on each of the above components that are important to provide the semi-automation in the honest brokering of the protected datasets.

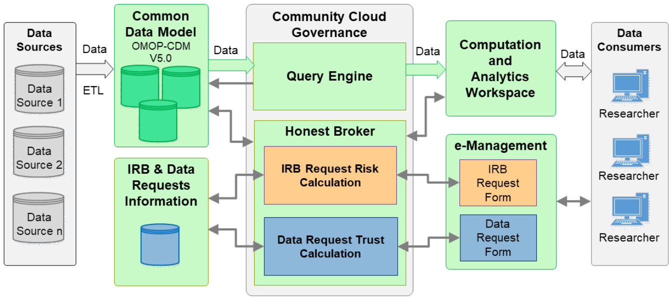
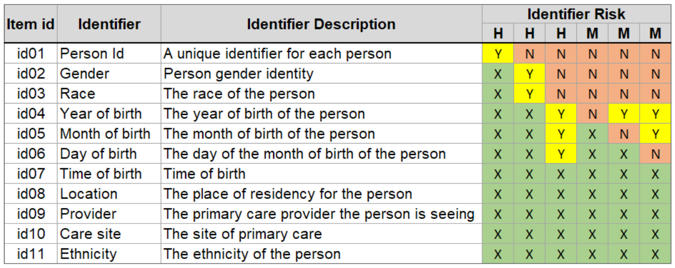


Fig. 2. Honest Broker system architecture for semi-automation

B. Common Data Model

The common data model (CDM), based on the OMOP-CDM framework [12], serves as a global catalogue that provides a standardized repository of the data which is shared by the honest broker system. The CDM exposes a common interface that allows the definition and execution of standard queries to fulfill an approved data request. As shown in Figs. 3 and 4, the data model mandates data to be stored as data identifier and data domain items that are defined in a patient-centric manner. The data identifier elements shown in Fig. 3, when included in a data requests (by the user/requestor), determine the identifiability of a particular individual's health records, and thereby determines the level of risk associated with such data release. Higher the level of identification, higher is the risk of sharing the data item. For example, if a data request includes "Person Id" (id01) item (represented with 'Y'), then the risk of such request is automatically 'High'. Fig. 3 only outlines 'High' (H) and 'Moderate' (M) risk combinations with all other combinations representing 'Low' (L) risk. Based on a given use case, additional combinations and risk values could be considered and updated in Fig. 3 accordingly. The data identifier items are defined by the CDM content, hence any identifier the user selects will necessarily fall into one of those categories.

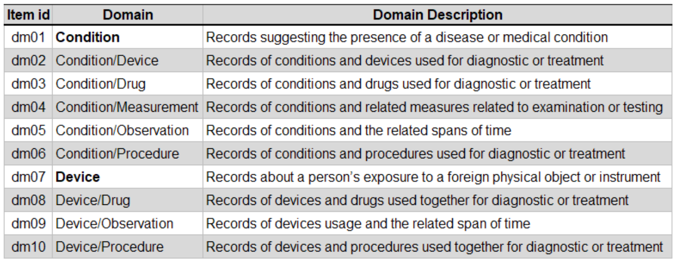
The data domain items as described in Fig. 4 define the set of allowable concepts for the standardized fields in the CDM repository. For example, the "Condition" domain describes the condition of a patient, whereas the "Device" domain contains information about the devices used for diagnostics and treatments of the patient. A data request could include one or more domain items in addition to identifier items. The domain item themselves have no risk factors associated with them as domain items independently cannot uniquely identify any



Item id	Identifier	Identifier Description	Identifier Risk					
			H	H	H	M	M	M
id01	Person Id	A unique identifier for each person	Y	N	N	N	N	N
id02	Gender	Person gender identity	X	Y	N	N	N	N
id03	Race	The race of the person	X	Y	N	N	N	N
id04	Year of birth	The year of birth of the person	X	X	Y	N	Y	Y
id05	Month of birth	The month of birth of the person	X	X	Y	X	N	Y
id06	Day of birth	The day of the month of birth of the person	X	X	Y	X	X	N
id07	Time of birth	Time of birth	X	X	X	X	X	X
id08	Location	The place of residency for the person	X	X	X	X	X	X
id09	Provider	The primary care provider the person is seeing	X	X	X	X	X	X
id10	Care site	The site of primary care	X	X	X	X	X	X
id11	Ethnicity	The ethnicity of the person	X	X	X	X	X	X

Fig. 3. OMOP-CDM data identifiers description and associated risk based on data sensitivity level

patient. Risk of any request as a combination of identifier and domain items is expressed by the requested identifier items.



Item id	Domain	Domain Description
dm01	**Condition**	Records suggesting the presence of a disease or medical condition
dm02	Condition/Device	Records of conditions and devices used for diagnostic or treatment
dm03	Condition/Drug	Records of conditions and drugs used for diagnostic or treatment
dm04	Condition/Measurement	Records of conditions and related measures related to examination or testing
dm05	Condition/Observation	Records of conditions and the related spans of time
dm06	Condition/Procedure	Records of conditions and procedures used for diagnostic or treatment
dm07	**Device**	Records about a person's exposure to a foreign physical object or instrument
dm08	Device/Drug	Records of devices and drugs used together for diagnostic or treatment
dm09	Device/Observation	Records of devices usage and the related span of time
dm10	Device/Procedure	Records of devices and procedures used together for diagnostic or treatment

Fig. 4. OMOP-CDM data domain items description

C. Community Cloud Governance

The risk management system provided by the community cloud governance component is the core of the honest brokering process. It provides a scientific procedure towards secure data sharing in terms of data identifiers and data domain request approval with minimal human intervention. As shown in Fig. 5, the data identifier approval process includes the risk calculation of the request comprising of three independent and equally important factors: (i) Risk of the request itself calculated based on what data elements have been requested, (ii) Risk of the requestor based on his/her roles and privileges, and (iii) Long term reputation of the requestor based on his/her prior history of handling sensitive data through different projects.

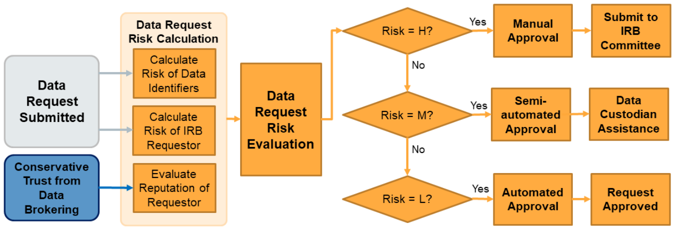
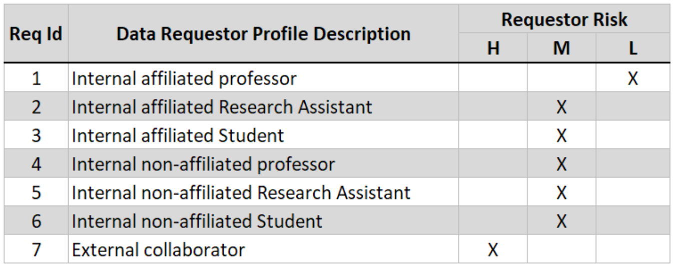

Data Request Risk Calculation
Data Request Submitted
Conservative Trust from Data Brokering
Calculate Risk of Data Identifiers
Calculate Risk of IRB Requestor
Evaluate Reputation of Requestor
Data Request Risk Evaluation
Risk = H?
Yes
No
Manual Approval
Submit to IRB Committee
Risk = M?
Yes
No
Semi-automated Approval
Data Custodian Assistance
Risk = L?
Yes
Automated Approval
Request Approved

Fig. 5. Honest broker data request approval process flowchart showing the three factors users to calculate the overall risk of the request: (i) risk of the data identifiers, (ii) risk related o the requestor profile, and (iii) risk related to the requestor reputation

The data identifier request risk computation is based on risk of requested data identifiers (represented as (I_r)) that follows the risk combinations of requested data identifiers (represented as 'H', 'M', and 'L') already explained in Fig. 3. The second component of data requestor risk (represented as (U_r)) is more subjective in nature and is based upon the requestor role and associated privileges defined within the community cloud ecosystem. One such example of different roles and how that translates to risk factors is illustrated in Fig. 6. Here we can see that a data identifier request submitted by an internal affiliated faculty member is lower ('L') than a request submitted by an affiliated collaborator from an external institution ('H'). The different roles illustrated in Fig. 6 serves only as an example which can be easily customized to fit the requirements of a particular community cloud platform. Consequently, based on a given use case, additional risk levels and roles can be added.



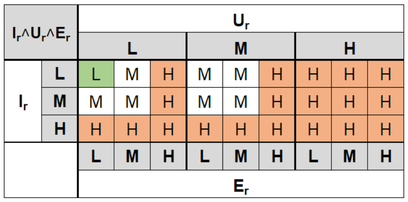
Req Id	Data Requestor Profile Description	Requestor Risk		
		H	M	L
1	Internal affiliated professor			X
2	Internal affiliated Research Assistant		X	
3	Internal affiliated Student		X	
4	Internal non-affiliated professor		X	
5	Internal non-affiliated Research Assistant		X	
6	Internal non-affiliated Student		X	
7	External collaborator	X		

Fig. 6. Data requestor profile descriptions within an ecosystem and associated risk factors

Additionally, the long term reputation of a data requestor (represented as (E_r)) is included as a key component in the approval process. Although this reputation is a numerical value, for the approval process, it is translated into 'H', 'M', and 'L' ratings. The details about reputation computation and the corresponding translation is explained in Section III. In order to calculate the overall risk of the data identifiers request, we employ the multi-valued Kleenean logic algebra [13] that is suitable to compute results involving multivariate ternary systems composed of three truth values, i.e., Low (L), Moderate (M) and High (H). In general, Kleene's algebra defines the regular logic operator AND, OR, and NOT. In our case we use the AND operator on all three input parameters $(R_t = I_r \wedge U_r \wedge E_r)$ to ensure that all the input parameters have the same weight and to yield conservative High (H) risk as the most relevant value. Kleene's AND logic operator for three valued parameters is defined as:

$$R_t = \begin{cases} H: \text{if ANY parameter is H} \\ L: \text{if ALL parameters are L} \\ M: \text{in all other cases} \end{cases} \tag{1}$$

Applying the Kleene's logic on the three input parameters I_r, U_r and E_r generates the resulting truth map for (R_t) as indicated in Fig. 7. The truth map includes the result for all combinations of the three input parameters. Applying this logic we have a conservative approach to generate a single risk state, where the result is Low if and only if all the inputs are Low, and the result is High if one of the inputs is High irrespective of other input values. A data identifier request with Low risk is automatically approved, whereas a Moderate risk triggers a semi-automated approval process that involves a data custodian assistance helping the requestor to refine the request and details. Finally, a High risk request triggers a 'custodian-in-the-loop' approval process requiring manual intervention for review, which takes considerably more time than automatic approval. However, due to the detailed audit trail left by the process, even the 'custodian-in-the-loop' approval takes considerably less end-to-end processing time than the legacy fully-manual compliance checking process.



$I_r \wedge U_r \wedge E_r$		U_r								
		L			M			H		
I_r	L	L	M	H	M	M	H	H	H	H
	M	M	M	H	M	M	H	H	H	H
	H	H	H	H	H	H	H	H	H	H
		L	M	H	L	M	H	L	M	H
		E_r								

Fig. 7. Overall risk computation through Kleene's logic

III. Trust Model Design

Our trust model is used to compute two different trust values (i.e., conservative data identifier trust and optimistic data domain trust) during the data brokering process. First, we compare the approved data identifier and domain items with the requested data identifier and domain items during the lifetime of a project. Next, we compute a conservative trust value using Dirichlet model [7] that builds towards a long term reputation of the requestor that is later used for a new data approval process shown in Fig. 5. At the same time, a more optimistic trust score is also computed using Beta model [8] that is used to encourage data usage 'best practices' through allocating 'Computation and Analytics Workspace' resources and tools to trustworthy users.

A. Data identifier trust

The purpose of data identifier trust is to compute a requestor's long-term reputation that would eventually be used as a deciding factor towards the requestor's future data approval requests. In our proposed honest broker, trust and reputation for a new requestor always starts with 0.5 i.e., moderate trust. In some cases, it might be suitable to start a new requestor with 1 i.e., high trust. In order to estimate long term reputation based on a requestor's data usage and consumption history, we monitor how data requests submitted for an approved data request adhere to its original approval. Thus, we can gauge if a requestor did the due diligence in requesting only those data identifiers that are absolutely necessary for their research purposes.

For this, we compare the approved data identifiers for the project (during the data identifiers request approval process) against the data identifiers actually requested (through data requests) during the project lifetime. Our honest broker saves approval information of all 11 data identifiers (from Fig. 3) after a project has been approved as explained in Section II. The outcomes of such comparison can be a *Match* (i.e., item was approved and requested or item was denied and never requested), or a *Mismatch* (i.e., item was approved but never requested), or an *Undecided* (i.e., if the data request specified that item to be uncertain but it was originally approved). Apart from these, a fourth outcome of *Violation* is possible if such an identifier item is requested or specified as uncertain that was originally never approved.

Herein, we detail how we apply the Dirichlet model to trust evidence. In order to design a conservative trust model for data identifiers, the honest broker risk management system applies Dirichlet distribution [7] on the trust evidences. This way the Dirichlet data parameters are defined as $d_1 = \eta_\varphi + Ca(x_1)$, $d_2 = \eta_\beta + Ca(x_2)$, and $d_3 = \eta_\mu + Ca(x_3)$ where C represents an a-priori constant [7]. Since there is no particular reason to believe a requestor's prior intentions during data request, we assume a uniformly distributed non-informative prior, which leads to $d_1 = \eta_\varphi + 1$, $d_2 = \eta_\beta + 1$, and $d_3 = \eta_\mu + 1$. With this, we can express the expected degrees of belief associated with the events of match, mismatch and undecided in terms of the observed trust evidence as:

$$E_\varphi = \frac{\eta_\varphi + 1}{\eta_\varphi + 1 + \eta_\beta + 1 + \eta_\mu + 1} \qquad (2)$$

Similarly, $E_\beta = \frac{\eta_\beta + 1}{\eta_\varphi + \eta_\beta + \eta_\mu + 3}$ and $E_\mu = \frac{\eta_\mu + 1}{\eta_\varphi + \eta_\beta + \eta_\mu + 3}$. Thus, for each data request j, we have $E_\varphi = E_j^b$ representing degree of belief, $E_\beta = E_j^d$ representing degree of disbelief, and $E_\mu = E_j^u$ reflecting degree of uncertainty where $E_j^b + E_j^d + E_j^u = 1$. This belief or the lack of it is from the honest broker's point of view in the requestor's ability to be compliant with the healthcare data confidentiality policies computed from the trust evidences and reflected through the data request.

B. Data domain trust

The purpose of computing the data domain trust is to encourage data request responsibly. More specifically, the honest broker encourages requestors who ensure secure usage of data domain items that was originally requested during the data identifier approval process and reward their current and future projects with 'Computation and Analytics Workspace'. For data domain request, the requestor chooses from a list of data domain items (Fig. 4). In order to enable such logic, we employ an 'approval-data request' comparison process similar to that of data identifiers. Here the outcomes of *Match* and *Mismatch* are the same as it is for data identifier items. However, due to the less sensitivity of data domain items than data identifiers, any *Uncertain* choice in data request will lead to an *Undecided* outcome with no concept of *Violations*. The total numbers of *Matches*, *Mismatches*, and *Undecideds* for each requestor j are demoted as η_{φ^j}, η_{β^j}, and η_{μ^j} respectively.

IV. Implementation and Performance Evaluation

We implement the proposed semi-automated honest broker on a community cloud testbed as shown in Figure 8. In this testbed, the e-Management platform is implemented on an Open Cloud [14] instance that provides user interface functionalities to data consumers as well as system administrator to review and update system and user information. The Open Cloud instance interacts with an Amazon EC2 instance that implements the honest broker along with its risk management and trust model components. The CDM and relational repository (containing user and projects information)

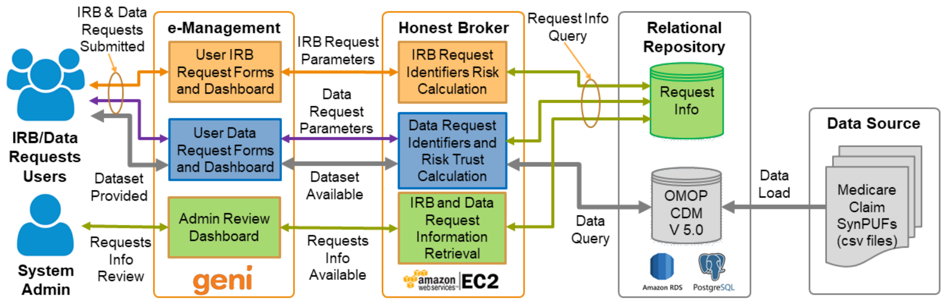


Fig. 8. Cloud-based testbed for honest broker implementation

are implemented using Amazon Relational Database Service (RDS). Using this tesbed setup, we evaluate the end-to-end processing time, determine the advantage of our proposed multivariate risk management system, and evaluate the trust model's efficiency.

A. *Risk management system evaluation*

We evaluate the performance of the honest broker's multi-dimensional risk management system in ensuring confidentiality of the requested data and compare the results against more one-dimensional methods typically adopted by manual brokering approaches. In order to achieve that, we compare the Loss of Confidentiality (LoC) risks for three types of healthcare data (identified, de-identified, and aggregated) using NIST [11] based risk assessment method. Using this method, we identify five possible most relevant (to healthcare community clouds) NIST Personally Identifiable Information (PII) factors [15] to evaluate LoC risk against. The NIST definitions of these factors and the relative impact are shown in Table I. We use a pre-defined semi-quantitative scale of 0-10 as guided by NIST for the impact assessments, with $10-9$ indicating *very High*, $8-7$ indicating a *High*, $6-5$ indicating a *Moderate*, $4-2$ indicating a *Low*, and $1-0$ indicating a *very Low* levels of impact. The overall LoC risk value is calculated for different competing strategies using the NIST-guided method [11] where we use $\lceil avg() \rceil$ function on the PII impact and the likelihood of each event. The comparison pits our proposed Honest Broker's multi-dimensional risk management system against: a) an approval strategy that only considers data request risk, b) a strategy that only considers IRB user/requestor risk, c) a strategy that only considers requestor long term reputation.

The results for identified, de-identified, and aggregated data are shown in Fig. 9(a), Fig. 9(b), and Fig. 9(c). The results show that although the LoC risk of identified data is higher than other data types, the honest broker risk management system consistently ensures lower risk against its one-dimensional counterparts. The results illustrate that although relying just on the risk of requested data identifiers effectively protects the identifiability, sensitivity, and quantity of the data, it does not prove effective in protecting data usage and accessibility. Similarly, the user/requestor risk and reputation are reliable parameters to make an effective decision about usability and accessibility of the data, but do not help much on the other three PII factors. Overall, the results justify the proposed honest broker's multi-dimensional risk management system in successfully mitigating the LoC risk of unauthorized data access.

B. *Performance evaluation of the trust model*

For evaluating the trust model, we synthetically generate data requests with skewed number of matches and mismatches between approved data requests and data retrievals. We do this step in order to observe how the conservative data identifier trust and optimistic data domain trust models react in terms of computed trust values. Fig. 10 illustrates the change in data identifier trust with increasing matches and mismatches with different ratios of undecideds to matches and mismatches. We see that as expected, the trust value increases with increasing number of matches and decreases with increasing number if mismatches. However, the slope of increase is conservative in nature due to the inherent characteristics of Dirichlet's distribution. We also see that the overall trust value is lesser if there are more undecided cases in the data requests irrespective of the relative ratio between the matches and mismatches. This conservative nature is exactly what we desire from data identifier trust, i.e., due to the potentially serious consequence of any misuse/mishandling of data identifiers, a user/requestor should only be able to slowly gain back the trust (when lost) and subsequently get quicker data request approvals.

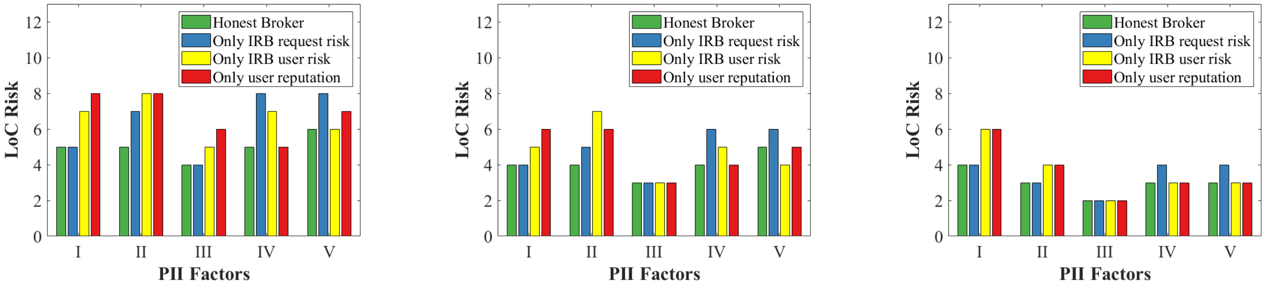
Finally, Fig. 11 shows the optimistic behavior of data domain trust that is based on Beta distribution against increasing matches and mismatches. We can observe that - unlike conservative identifier trust, the rate of data domain trust increase with matches is much higher. In other words, for someone with the same number of data domain and data identifier matches will have higher domain trust than identifier trust. This nature of domain trust is also desirable as the domain trust is used for encouraging data usage best practice by offering 'computation and analysis workspace' tools for requestors with higher domain trust.

V. Conclusion

In this paper, we motivated the need for a faster, secure and more efficient brokering by data custodians to handle data consumer requests for healthcare data sharing within community clouds. We demonstrated how our proposed semi-automated "honest broker" can accelerate the end-to-end brokering process that involves data discovery, security compliance checking, and data request processing. We developed

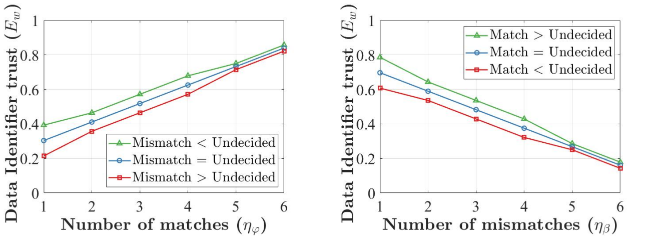
TABLE I
NIST PERSONALLY IDENTIFIABLE INFORMATION (PII)

No.	PII Factor	Identified Impact	De-identified Impact	Agg-regated Impact
I	**Identifiablity**: PII can be used to identify specific individuals.	High (8)	Moderate (6)	Moderate (6)
II	**Data Field Sensitivity**: Sensitivity of each individual PII data field.	High (8)	High (7)	Low (4)
III	**Quantity of PII**: Number of individuals identified.	Moderate (6)	Low (3)	Low (2)
IV	**Context of Use**: The purpose for which PII is collected, stored, used, processed, disclosed, or disseminated.	High (8)	Moderate (6)	Low (4)
V	**Access to and Location of PII**: Nature/location of authorized access.	High (8)	Moderate (6)	Low (4)



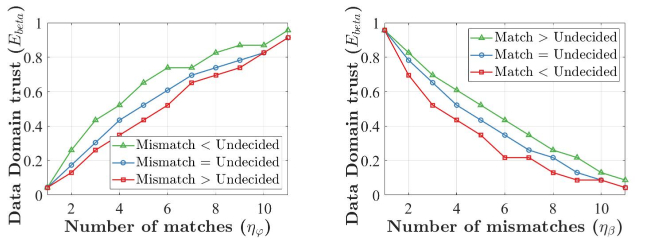
(a) Identified (b) De-identified (c) Aggregated

Fig. 9. Risk assessment comparison between the honest broker approach against other on-dimensional approaches



(a) With increasing # of matches (b) With increasing # of mismatches

Fig. 10. Data identifier trust characteristics



(a) With increasing # of matches (b) With increasing # of mismatches

Fig. 11. Data domain trust characteristics

a novel multi-dimensional risk management system taking a scientific approach towards the security compliance checking and data request processing steps in comparison to legacy one-dimensional processes. We used underlying conservative and optimistic trust models to ensure secure data access. Using experiments on an implementation of our semi-automated honest broker in a community cloud testbed, we demonstrated that our proposed approach significantly reduces in overall risk of data access and sharing.

Our future work is to implement the 'Computation and Analytics Workspace' and related functions through an *Entitlement service* in order to help data consumers manage large data and perform high-speed data analytics tools in the community cloud platform. Additionally, options to enforce privacy rules can be included on cloud-based healthcare data brokering functions based on data custodian requirements.

REFERENCES

[1] J. Andreu-Perez, Y. Poon, R. Merrifield, S. Wong, G. Yang, "Big Data for Health," *IEEE Journal of Biomedical and Health Informatics*, 2015.

[2] M. Vassell, O. Apperson, P. Calyam, J. Gillis, and S. Ahmad, "Intelligent Dashboard for augmented reality based incident command response co-ordination", IEEE Annual Consumer Communications Networking Conference (CCNC), 2016

[3] U. Fayyad et al., "Knowledge Discovery and Data Mining: Towards a Unifying Framework," *Association for the Advancement of Artificial Intelligence*, 1996.

[4] "An Introductory Resource Guide for Implementing the Health Insurance Portability and Accountability Act (HIPAA) Security Rule," NIST Special Publication, 2013.

[5] S. Aronson, L. Heidi, "Building the foundation for genomics in precision medicine," *Nature 526.7573*, 2015.

[6] K. Suh et al., "Tissue Banking, Bioinformatics, and Electronic Medical Records: The Front-End Requirements for Personalized Medicine," *Journal of Oncology*, 2016.

[7] A. Josang, J. Haller, "Dirichlet Reputation Systems", *Intl. Conf. on Availability, Reliability and Security*, 2007.

[8] A. Josang, J. Haller, "Beta Reputation Systems", *Intl. Conf. on Availability, Reliability and Security*, 2002.

[9] R. Akella, S. Debroy, P. Calyam , A. Berryman, K. Zhu, M. Sridharan, "Security Middleground for Resource Protection in Measurement Infrastructure-as-a-Service", *IEEE Trans. on Services Computing*, 2019.

[10] A. Sukhov, P. Calyam, W. Daly, and A. Ilin, "Towards an analytical model for characterizing behavior of high-speed VVoIP applications", Computational Methods in Science and Technology, 2005.

[11] Joint Task Force Transformation Initiative, "Guide for Conducting Risk Assessments", *NIST Special Publication 800-30*, 2012.

[12] OMOP Common Data Model (CDM) V5.0. by Observational Health Data Sciences and Informatics (OHDSI) at: https://www.ohdsi.org/data-standardization

[13] Y. Hata, K. Nakashima, K. Yamato, "Some fundamental properties of multiple-valued Kleenean functions and determination of their logic formulas," *IEEE Trans. on Computers*, 1993.

[14] M. Berman, J. Chase, L. Landweber, A. Nakao, M. Ott, D. Raychaudhuri, R. Ricci, I. Seskar, "GENI: A federated testbed for innovative network experiments", Computer Networks, 2014.

[15] E. McCallister, T. Grance, K. Scarfone, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)", *NIST Special Publication 800-122 - Technical Report*, 2010.