|  |  |  |
| --- | --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** | | |
| **Reference Number**  *(to be assigned by CC)* | NT317 | |
| **Submission Date** | 12/17/2018 | |
| **Project Title** | ARBOR: An Identity and Security Solution for Clinical Reporting | |
| **Tentative Lead Investigator** *(first author)* | Eric Venner (co-lead) Mullai Murugan (co-lead) | |
| **Tentative Senior Author**  *(last author)* | Richard Gibbs | |
| **All Other Authors** | Interested eMERGE sites | |
| **Sites Participating** | Baylor College of Medicine – Human Genome Sequencing Center | |
| **Background / Significance** | **Motivation**: Clinical genome sequencing laboratories return reports containing clinical testing results, signed by a Board Certified clinical geneticist, to the ordering physician. This report is often a pdf, but can also be a physical paper copy or a structured data format. The reports are frequently modified and re-issued, due to changes in variant interpretation or clinical attributes. To precisely track report authenticity, we developed ARBOR, an application for tracking the lineage of versioned clinical reports even when they are distributed as pdf or paper copies. ARBOR employs a modified blockchain approach and instead of relying on a computationally intensive consensus mechanism for determining authenticity, we allow supervised writes to an encrypted ledger, which is then exactly replicated to many clients.  **Results**: ARBOR was implemented for clinical reporting in the HGSC-CL Clinical Laboratory, initially as part of the NIH’s Electronic Medical Record and Genomics (eMERGE) project. This system has provided us with a simple and tamper-proof mechanism for tracking clinical reports with a complicated update history. | |
| **Outline of Project** | 1. Background on clinical reporting 2. Describe current identity management and security tracking solutions 3. Architecture and design of our system 4. Common use cases for this system 5. Benefits of this approach and extension to de-centralized clinical reporting networks. | |
| **Desired Data - Common Variables\***  *(Available from the CC)* | ☐Demographics  ☐ICD9/10 codes  ☐CPT codes  ☐Phecodes  ☐BMI | ☐Common Variable Labs  ☐Common Variable Meds  ☐Other: Case/Control status on Phase I and Phase II phenotypes |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*  None | |
| **Desired Genetic Data** | ☐eMERGE I-III Merged set (HRC imputed, GWAS)  ☐eMERGE PGx/PGRNseq data set  ☐eMERGEseq data set (Phase III)  ☐eMERGE Whole Genome sequencing data set  ☐eMERGE Exome chip data set  ☐eMERGE Whole Exome sequencing data set  ☐Other (not listed above):  N/A | |
| **Does project pertain to an existing eMERGE Phenotype?** | ☐Yes, if so please list  ☒No | |
| **Planned Statistical Analyses** | None, paper will focus on a tool for security and identity management we developed while carrying out clinical reporting for eMERGE. | |
| **Ethical Considerations** | None | |
| **Target Journal** | JAMIA | |
| **Milestones**  *(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | Initial Draft: Jan 2019  Revisions by Co-Authors: Feb 2019  Final Draft reviewed by Co-Authors: March 2019  Final Draft Submitted: March 2019 | |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, decade of birth, race, ethnicity
* Codes: (repeated values & age at event): ICD, CPT, Phecodes
* BMI: (repeated value & age at event) height, weight, BMI
* Labs: (lab name, repeated lab value & age at event) Serum total cholesterol, LDL, HDL, Triglycerides, Glucose fasting/non-fasting/unknown, & White Blood Cell count
* Medications: (medication name, repeated, & age at event) Cerivastatin sodium, Rosuvastatin, Simvastatin, Fluvastatin, Pravastatin, Lovastatin, Atorvastatin, & Pitavastatin
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants