Request for Proposal and Quote

Development of
Privacy-Preserving Record Linkage

The People-Centered Research Foundation

1399 New York Avenue NW

Suite 450

Washington, D.C. 20005

December 31, 2018

Table of Contents

Background

To improve our nation's capacity to conduct clinical research more efficiently and to answer important questions that patients and clinicians face, the Patient-Centered Outcomes Research Institute (PCORI) funded the creation of the National Patient-Centered Clinical Research Network (PCORnet). In March 2017, researchers from within PCORnet established the People-Centered Research Foundation (PCRF) as a private, nonprofit corporation to build on the successes of PCORnet and further its mission as a sustainable platform for efficient, innovative research. The Network is currently comprised 9 Clinical Research Networks (CRN's) that includes 69 network partner organizations, 2 Health Plan Research Networks (HPRN's), Coordinating Centers, and PCRF as headquarters.

PCORnet sites and networks have an established track record of collaboration and research success that offers a re-useable observational study and clinical trial infrastructure to launch studies faster, compress the study timelines, and produce higher-quality results, using:

- A Single IRB for multi-site clinical research;
- Resources to engage patient partners;
- A Master Clinical Research Agreement;
- Data sharing/usage agreements and privacy standards;
- Capacity to link claims and EHR data
- Study coordination for quality assurance and performance metrics; and
- Active and engaged clinicians.

The purpose of this Request for Proposal (RFP) is to solicit proposals and quotes from candidate organizations for a privacy-preserving record linkage solution for electronic patient records to be used by all networks participating in PCORnet and by other data holders who wish to link with PCORnet networks (or network partners) as part of specific research studies. The candidate organizations are asked to provide implementation services as well as maintenance and support for two years. The proposed solution should allow PCORnet network partners to use personal identifiers stored behind their institutional firewalls to generate cryptographic hashes that can then be shared with the PCORnet Coordinating Center, which will be responsible for determining whether a given patient has records present across multiple network partners. Candidate organizations are asked to provide a software solution that will support the management and distribution of keys/salts/shared secrets across network partners, allow those partners to generate the cryptographic hashes, and support the local implementation of algorithm(s) that can be used by the PCORnet Coordinating Center to determine whether a record is deemed a "match."

All application content will be confidential. Questions and corresponding answers will be shared without revealing the name of the potential applicant.

Awards will be made pending availability of funds.

Proposal Guidelines

This RFP represents the requirements for an open and competitive bidding process. Proposals will be accepted until **2/1/19 at 5pm PT.** Any proposals received after this date and time will not be considered. All proposals must be signed by an official agent or representative of the company submitting the proposal.

If the organization submitting a proposal must outsource or contract any work to meet the requirements contained herein, this must be clearly stated in the proposal. Any proposals which call for outsourcing or contracting work must include a name and description of the organizations being contracted. Additionally, proposals must be all-inclusive to cover any outsourced or contracted work.

Contract terms and conditions will be negotiated upon selection of the winning applicant for this RFP. All contractual terms and conditions will be subject to review by the PCRF and will include scope, budget, schedule, and other necessary items pertaining to the project.

Proposal Contents

Applicants should provide details of their qualifications for this activity based off the current version of functional specifications provided later in this document. The proposal should contain the following information and adhere to the stated page limits for each section:

- Organization background (2 pages; up to 15 pages for biosketches [maximum 5 pages each])
    - Describe your organization's background, history, track record for similar projects, and financial sustainability. Does your organization have the experience and dependability to initiate, complete, and support the project long-term? Also describe your organization's management of similar projects. Who will be the implementation team for this project? What is their experience with similar work? Provide up to three biosketches of project leadership and key personnel.
- Project scope (10 pages)
    - Provide answers to 23 functional specifications and questions listed below, spread across five domains – technical security, governance, implementation and use, hashing, matching, and ownership.
- Timeline (2 pages)
    - Provide a detailed timeline of initiation, implementation, and support. After the RFP winner is announced on February 28, we expect contracts to be signed by March 31, software installed across all PCORnet sites by May 31, and hashes created and populated into a Hash table by June 28 (The PCORnet Coordinating Center will work with PCORnet CRNs and HPRNs to ensure that these timelines are aligned between networks and the selected vendor.). We also expect support, maintenance, and upgrades for two years. How will your organization meet this project schedule? How have you worked under similar deadlines in the past? How will you manage the project to meet the tight timelines?
- Budget and pricing (3 pages)
    - Provide a detailed budget and time estimates using the template in Appendix A. Provide information on design and development, licenses, and support. Indirect costs are capped at 10% of direct costs.
- References (3 pages)

- Provide up to three references from current, and/or past customers (within the past 2 years) that attest to your organization's knowledge, dependability, and support for similar projects.

Evaluation

Proposals that provide the information specified above will be evaluated on the following criteria:

- Review criteria
  - Organization and project team – Does the application demonstrate that the organization and management team can perform and sustain the PCORnet common linkage approach?
  - Technical – Does the proposed solution sufficiently satisfy PCORnet's linkage requirements? Does the organization have the technical capacity and expertise to successfully implement the PCORnet common linkage approach (see Project Scope, below)?
  - Pricing/timeline – Is the proposed budget and timeline responsive to the RFP? Is there evidence that the organization has the capacity to implement the PCORnet common linkage approach within the stated timeline and budget?
- Review process
  - Review panel – Applications will be reviewed by PCRF staff and a panel of subject matter experts on the above criteria. The panel will be free of conflicts of interest and meet at least once to discuss the output of the quantitative review. The panel will present recommendations to PCRF.
  - Secondary review – Pending the outcome of the review panel's recommendations, top applicants might be requested to discuss their application to the review panel and/or PCRF via a webinar.
  - Selection recommendation – PCRF staff will present the successful application to the PCRF Board of Directors for approval before contract execution.
  - Applicant notification – All applicants will be notified of the final selection. Unsuccessful applicants may request a brief written summary of the review.

Submission Instructions

Applications will be submitted to PCRF.smartsimple.com. Applicants will need to register for an account in the SmartSimple system prior to submitting their application. To register for an account, contact info@pcrfoundation.org by **1/24/19 at 5pm PT**.

Please submit all other communications through Lindsey Petro at Lindsey_Petro@harvardpilgrim.org.

If you have any questions, please email them by **1/11/2019**. Questions will be addressed during an applicant Town Hall on **1/18/19**.

Project Purpose and Description

PCORnet includes nine Clinical Research Networks (CRNs), each with multiple health systems as sites, and two Health Plan Research Networks (HPRNs). Very few networks have complete capture of the variables and outcomes of interest within their datasets for their patient population(s), which is needed for many observational studies. For example CRNs may have deep clinical detail within a health system but lack longitudinal follow-up across health systems and HPRNs may have longitudinal follow-up across health systems but lack the deep clinical detail within a health system.

Linking across networks provides a way to bridge that gap. While individual networks – or in some cases pairs of networks –have developed solutions for linking data across their networks, PCORnet does not have a standardized linkage and de-duplication solution.

The goal of this RFP is to select a vendor-partner to work with and across PCORnet to implement a common linkage solution that will allow PCORnet to classify the network by (1) assessing patient overlap across clinical and health plan research networks, (2) identifying the number of unique patients across the network, (3) and producing a linked, de-duplicated PCORnet Table of Population Characteristics. Then, the solution will support PCRF's business and research goals by implementing common, network-wide linkage across multiple large observational studies and pragmatic trials.

Project Scope (Functional Specifications)

Terms and Definitions

1. Health Information Portability and Accountability Act. (HIPAA) – legislation in the United States of America that outlies privacy and security provisions for the protection of patient medical data. These safeguards are also outlined in the Health Information Technology for Economic and Clinical Health (HITECH) Act
2. Personal Health Information (PHI) – information that is considered a patient identifier under HIPAA.
3. Personally-Identifiable Information (PII) – any data that could potentially identify an individual.
4. Covered Entity – individuals, organizations and agencies that must comply with the HIPPA rules for privacy and security (https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html).
5. Business Associate Agreement (BAA) – a legal agreement between a covered entity an organization that creates, receives, maintains or transmit PHI on behalf of the covered entity to care out the covered entity's HIPAA-covered functions or to perform certain services on behalf of the covered entity (https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/index.html).
6. Institutional Review Board (IRB) – an administrative body that provides ethical and regulatory oversight of research involving human subjects.
7. Real-world Evidence (RWE) – clinical evidence regarding the usage and potential benefit of a medical product derived from an analysis of real-world data (https://www.fda.gov/scienceresearch/specialtopics/realworldevidence/default.htm).
8. Hash - A hash algorithm is a function that converts a data string into a numeric string output of fixed length. The output string is generally much smaller than the original data (http://www.digitizationguidelines.gov/term.php?term=hashalgorithm).
9. Salt - data appended to the input of a hash function, often to safeguard data in storage (https://en.wikipedia.org/wiki/Salt_(cryptography).

Solution Overview / Technical Security Requirements

1. Overview of solution
    - Describe each step of the methodology and process used to link patients' electronic health records: distribution of salts/keys/shared secrets, hashing process, aggregation, matching, disambiguation, deterministic/probabilistic/combination, etc. Has the proposed solution undergone an independent third-party security/risk assessment? If so, what type of assessment (e.g., penetration, line-by-line code

evaluation), and what were the results? What administrative, technical, and managerial safeguards are in place to protect PHI and/or the salts/keys/shared secrets? What are the inputs and outputs of each step in the process (e.g., public keys, crosswalk tables, hashes, global unique ID, etc.)?

2. Hosting / networking requirements
    - Describe how the solution is hosted/implemented by customers. How is it deployed on premises? Does the proposed solution require a service call to allow data transfer beyond the local network when hashes and/or identifiers are generated? What about when matching or disambiguating the hashes? Does the proposed solution require the use of a third-party cloud service vendor or another type of honest broker? How are configuration files stored?
3. Cryptographic method / encryption
    - Describe the cryptographic method that is used to generate the hashes. How are data protected at rest/in motion? What encryption standard, if any, is used to protect salts, hash files, and other data?

Governance Requirements

1. Regulatory requirements
    - Describe the typical steps needed to satisfy regulatory requirements for health plans and health systems when implementing your software. Are Business Associate Agreements signed with every participating site? Are there outputs from Information Technology security assessments or other audits that can be provided to customers as they perform their privacy and security reviews?
2. Human subjects protection
    - Describe any experience in obtaining approval from Institutional Review Boards to utilize your product to safely and securely link patient records across institutions to support research. Is there common language or an example Institutional Review Board protocol that can be provided to customers as they seek IRB approvals? Please provide examples of IRB approval that has been obtained at institutions that leverage your product?
3. Privacy/security
    - Describe whether the methodology used by the proposed system has been evaluated by an expert to determine that the outputs (e.g., hashes) satisfy the de-identification requirements described under the HIPAA Common Rule (164.514(b)(1)). Does the proposed solution generate crosswalk tables to facilitate re-identification? How is re-identification prevented?
4. Privacy risks
    - Describe the steps that are taken to minimize privacy risks to patients. Are the hashes generated using a salt as an additional input? What mechanisms are in place to mitigate the risk of revealing the hashes, salts, and other secrets to unintended third parties?

Implementation and Use Requirements

1. Technical Details
    - Describe the software and hardware requirements to implement, maintain and update the proposed solution. This includes database and server needs, schemas,

anticipated server storage, computing overhead, typical/anticipated support (% FTE by role), etc.. If applicable, delineate the requirements for implementing the solution versus day-to-day operation. Also, describe whether there are specific needs for certain steps in the process (e.g., encryption/hash generation). Examples might include SAS or R, an accessible environment for a custom executable, or that any of the above be placed in an environment with identifiable data in a particular format.

2. Implementation Timeline
    - Please describe the time needed to implement the proposed solution, both in setup and in configuration at the site level. Estimates should be based on real-world examples, if possible. Applicants should be mindful of the common linkage timeline – full PCORnet install by May 31, 2019 and CDM hash table populated by June 28, 2019.
3. Code stability / Software validation
    - Describe the overall maturity of your proposed solution as well as the process for validating that the software meets business requirements. How are defects identified and fixed? What defines success and how do the makers of the system and the customer determine the software is successful? How often is the software updated, and which components of the solution are typically included (e.g., user interface, hashing algorithm, etc.)? What is the process for testing and deploying updates across customers? Is the software source code available for inspection, and if not, has it been verified or validated by an independent third party?
4. Cost
    - Describe the cost required to implement the overall system across the network, including the Coordinating Center and all participating sites. What license fees (in USD) are required? What types of personnel (e.g., analyst, software engineer, security specialist) are required to implement the system, and how much effort is required? What is the cost for adding an additional data source/site?
5. Scalability
    - Describe how your product can scale to other research partners and data sources. Do you have experience linking clinical and/or claims data to other data sources like registries or across sectors (like housing or social services data, for example)? How do you see future research partners – registries, large studies, and/or other data networks (like Sentinel) – partnering in the future to link to PCORnet data, either as added infrastructure and capacity or through individual projects?
6. User interface
    - Describe the way in which users interact with the system. Is the system web-based, or does it require the installation of a thick client? If web-based, what components (e.g., HTML5) does it use? Is it possible to modify visualization of the user interface using stylesheets and/or templates? To what extent does the system's user interface meet federal Section 508 standards?

Hashing Requirements

1. Identifiers (PII used to create hash)
    - List the identifiers needed to create the hashes. Which identifiers are required, and which are optional? If one of the required inputs is missing, what behavior does the

system exhibit? Does your method generate multiple hashes from different subsets of identifiers?

2. Data preparation
    - Describe the expected format of the inputs to the hashing program (e.g., tabular format, JSON, XML, etc.)? Have pre-existing extract routines been created for any EHR or clinical information systems? Is any pre-processing required (e.g., convert to uppercase, remove punctuation, etc.), and if so, are there programs that can be provided to complete these tasks?
3. Performance
    - Describe hash generation performance for the proposed solution. Detail the system specifications (e.g., processor, memory, disk configuration) and the performance on that system (time per million patient hashes if linear, appropriate estimates otherwise). Please describe the storage requirements per million patient hashes.
4. Hashing validation (RWE and/or peer reviewed, or by standard test)
    - Describe whether the proposed hashing solution has undergone any independent, third-party validation – including any privacy expert certification. Has the method been peer-reviewed or been assessed through standard methods? Can you highlight any peer-reviewed publications, white papers, or any other examples of real-world evidence? Describe the data sources used in these real world and/or validation projects.

Linkage/Matching Requirements

1. Methodology (deterministic or probabilistic)
    - Describe the approach used to determine patient matches across data partners. Is the method deterministic or probabilistic? Are multiple methods supported? Can the parameters be configured to identify potential matches at varying levels of confidence?
2. Performance
    - Describe matching/linkage performance for the proposed solution. Detail the system specifications (e.g., processor, memory, disk configuration) and the performance on that system (time to match between one million patient hashes if linear, appropriate estimates otherwise). Please describe any storage requirements induced by matching per million patient hashes.
3. Linkage validation
    - Describe any linkage/matching validation studies undertaken or underway. What is the documented sensitivity and specificity of your product? Can you highlight any peer-reviewed publications, white papers, or any other examples of real-world evidence? Describe the data sources used in these real world and/or validation projects. Do you have evaluations of projects linking clinical data to other clinical, claims, registries, and/or social services data, etc.?

Ownership / Licensing

1. Source code licensing / Intellectual property
    - Describe your approach to intellectual property and source code licensing. Is your product's source code released under an open-source license? What license do you intend to use, and is there any flexibility in your choice? Besides source code

licensing, do you claim any proprietary processes or other intellectual property that would need licensing (e.g. Software-as-a-Service)?

2. Ownership/reuse of code
    - Describe the licensing model of your software solution. Is an ongoing subscription required, or does PCORnet maintain a perpetual license to use the solution? If selected for use in PCORnet, would individual networks or network partners (e.g., sites) be free to use it for their own internal purposes? Would there be any additional costs involved in this reuse?
3. Transparency
    - Describe whether the source code can be shared with PCRF, network partners, etc., either under a Non-Disclosure/Confidentiality Agreement or some other approach. Describe any reference implementation that can prove your software's functionality and cryptography. Will the interfaces to your software be well described and documented so that other implementations can inter-operate with your proposed approach?

Request for Proposal Timeline

Request for Proposal Timeline:

Register for an account in the SmartSimple system no later than **1/24/2019 at 5pm PT**.

All proposals in response to this RFP are due no later than **2/1/19 at 5pm PT**.

Evaluation of proposals will be conducted from **2/2/19 to 2/28/19**. If additional information or discussions are needed with any applicants during this window, the applicant(s) will be notified.

The selection decision for the winning proposal will be made no later than **3/1/19 at 5pm PT**.

Notifications to applicants who were not selected will be completed by **3/1/19 at 5pm PT.**

Upon notification, contract negotiation with the winning applicant will begin immediately.

Contact Information

Your primary point of contact is Lindsey Petro [Lindsey_Petro@harvardpilgrim.org].

Appendix A: Budget Templates

Please prepare a budget justification (<2 pages) to accompany the following templates:

YEAR 1	
	COST
1. Software License	
2. Implementation Support	
a. Sites (Hashing Function)	
b. Coordinating Center (Matching Function)	
TOTAL	

YEAR 2	
	COST
1. Software License	
2. Ongoing Technical Support	
TOTAL	

PCRF has a 10% cap on indirect costs. For purposes of calculating the indirect costs, use the total direct costs, consisting of all direct costs and up to the first $25,000 of each subaward for the entire period of the award.