DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the National Coordinator for Health Information Technology

Announcement of Requirements and Registration for "Patient Matching Algorithm Challenge"

Authority: 15 U.S.C. 3719

AGENCY: Office of the National Coordinator for Health Information Technology, HHS

AWARD APPROVING OFFICIAL: Jon White, Deputy National Coordinator for Health Information Technology

ACTION: Notice

SUMMARY:

The goal of the Patient Matching Algorithm Challenge is to bring about greater transparency and data on the performance of existing patient matching algorithms, spur the adoption of performance metrics for patient data matching algorithm vendors, and positively impact other aspects of patient matching such as deduplication and linking to clinical data. Participants will be provided a data set and will have their answers evaluated and scored against a master key. Up to 6 cash prizes will be awarded with a total purse of up to $75,000.00. The statutory authority for this Challenge is Section 105 of the America COMPETES Reauthorization Act of 2010 (Public L. No 111-358).

DATES: (Submission period and Announcement dates will be posted to the Challenge Website.)

- Announcement of Challenge: April 28, 2017
- Registration Period Begins May 10, 2017
- Submission period Begins: Upon availability of test data.
- Submission period Ends: 3 months (90 days) from availability of test data.

- Winners notified: 1 Week from the end of the submission period.
- Winners Announced: 1 week from winner notification date.

FOR FURTHER INFORMATION CONTACT: Debbie Bucci, Debbie.Bucci@hhs.gov or Adam Culbertson, Adam.culbertson@hhs.gov

SUPPLEMENTARY INFORMATION:

Subject of Challenge:

In late 2015, the Office of the National Coordinator for Health Information Technology published "Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap" (the Roadmap). The final Roadmap reflects the combined contributions of dozens of experts and hundreds of public comments received during its drafting phase. The Roadmap includes "Section L," which was specifically framed to reflect the challenges health care faces with respect to accurate individual data matching. This section highlights data matching's overall importance to interoperability, the nation's health IT infrastructure, and future research capabilities. Indeed, health care providers must be able to share patient health information and accurately match a patient to his or her data from a different provider in order for many anticipated interoperability benefits to be realized. Conversely, matching mistakes can contribute toward adverse events, compromised safety and privacy, and increased health care costs due to repeat tests and other factors. The cost to manually correct mismatched patient records is estimated to be $60 per record[1] not including the potential harm that could be caused due to a patient receiving the wrong treatment and potential legal fees.

Given the substantive impacts poor patient matching can have on care delivery, population health analyses, and research, it is important for organizations to be able to quantify

[1] https://intermountainhealthcare.org/~/media/Files/Research/medical-informatics-reduce-duplicate-pt-creation.pdf accessed Wednesday, December 21, 2016.

their patient matching algorithm's performance and compare the results to industry standard benchmarks and performance metrics. To date, the absence of such benchmarks as well as a baseline for different use cases has made it difficult to make advances in patient matching. Reports such as ONC's Patient Identification and Matching Final Report list patient match rates in the range of 50%-Mid 90%.[2]

Every patient matching algorithm has blind spots and there are methods to calculate the performance of a patient matching algorithm and help identify these blind spots. This is accomplished by giving an algorithm a known data set in order to see how many of the known linkages the algorithm can correctly identify. Matching algorithms can make two types of errors. The first error is the failure to find a matching pair (often referred to as a "false negative"), which is measured by "inverse recall" in the field of information retrieval. The second type of error is a record that is matched when it should not be (often referred to as a "false positive"), which measured by a metric known as "precision." The weighted average of precision and recall generates the final metric pertinent to this Challenge is known as "F-Score."

Challenge Summary

This Challenge uses a large test data set, provided by ONC, against which participants must run their algorithms and provide their results for evaluation. A small set of true match pairs (that have been created and verified through manual review) exist within the large data set and will serve as the "answer key" against which participants' submissions will be scored.

Participants will unlock and download the test data set at the time of registration. Participants will then run their algorithms and submit their results to the scoring server on the Challenge Website. These submissions will receive performance scores and may appear on a

[2] https://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf accessed Wednesday, December 21, 2016.

Challenge leaderboard. Upon submitting results, participants will receive objective evaluation metrics (F-scores) that can be used to guide system improvements; a total of 100 re-runs will be allowed. Up to six participants will be selected as winners for this Challenge and awarded cash prizes. Top prizes will be awarded to participants' algorithms that generate the highest F-Score(s). Additionally, algorithms with the best recall, best precision, and best first F-Score run performance will also receive a cash prize.

Eligibility Rules for Participating in the Challenge:

To be eligible to win a prize under this Challenge, an individual or entity:

1. Shall have registered to participate in the Challenge under the rules promulgated by ONC. Any publicly-displayed team data, including team names, may be changed if deemed inappropriate by HHS in its sole discretion. Each Participant must consent to be bound by these Rules. Individuals may be members of multiple teams. All entry information and submissions shall be deemed collected and judged in the United States and responses should be in English
2. Shall have complied with all the stated requirements of this challenge.
3. In the case of a private entity, shall be incorporated in and maintained a primary place of business in the United States, and in the case of an individual, whether participating singly or in a group, shall be a citizen or permanent resident of the United States.
4. Shall not be an HHS employee.
5. May not be a federal entity or federal employee acting within the scope of their employment. We recommend that all non-HHS federal employees consult with their agency Ethics Official to determine whether the federal ethics rules will limit or prohibit the acceptance of a COMPETES Act prize.

6. Federal grantees may not use federal funds to develop COMPETES Act Challenge applications unless consistent with the purpose of their grant award.
7. Federal contractors may not use federal funds from a contract to develop COMPETES Act Challenge applications or to fund efforts in support of a COMPETES Act Challenge submission.
8. All individual members of a team must meet the eligibility requirements.
9. An individual or entity shall not be deemed ineligible because the individual or entity used federal facilities or consulted with federal employees during a Challenge if the facilities and employees are made available to all individuals and entities participating in the Challenge on an equitable basis.
10. Participants must agree to assume any and all risks and waive claims against the Federal Government and its related entities, except in the case of willful misconduct, for any injury, death, damage, or loss of property, revenue, or profits, whether direct, indirect, or consequential, arising from my participation in this prize contest, whether the injury, death, damage, or loss arises through negligence or otherwise
11. Participants must also agree to indemnify the Federal Government against third party claims for damages arising from or related to Challenge activities.

General Submission Requirements

The patient matching Challenge website will manage the submissions and provide the scoring results back to the Participant. The answer key can be submitted in CSV, XML, or JSON files. In order for a submission to be eligible to win this Challenge, it must meet the following requirements:

1. No HHS or ONC logo – The product must not use HHS' or ONC's logos or official seals and must not claim endorsement.
2. Functionality/Accuracy – A product may be disqualified if it fails to function as expressed in the description provided by the Submitter, or if it provides inaccurate or incomplete information.
3. Security – Submissions must be free of malware. Submitter agrees that ONC may conduct testing on the product to determine whether malware or other security threats may be present. ONC may disqualify the submission if, in ONC's judgment, it may damage government or others' equipment or operating environment.

How to Enter:

To participate in the Challenge, Participants must accurately and truthfully complete the Participant registration information request online at www.patientmatchingchallenge.gov. To register, Participants must provide a team name, identify a team leader, and provide a list of all team Participants, including each Participant's full name, e-mail address, affiliation, if any, and country.

Registration Process for Participants:

Challenge participants will register their team on the Challenge website. Participants can also access the Challenge on http://www.Challenge.gov and search for "Patient Matching Algorithm Challenge."

Prize:

Highest F-Score

- First Place: $25,000
- Second Place: $20,000

- Third Place $15,000

Best in Category Supplemental Prizes (1 prize for each category at $5,000):

- Best precision
- Best recall
- Best first F-Score run

Total Prize Purse: Up to $75,000

The cash-prize winning team(s) and up to ten runners-up may be invited to present at a Technical Exchange Meeting with interested government and industry partners (attendance would be at the participants own expense).

Payment of the Prize:

Prize will be paid by contractor.

Basis upon Which Winner Will Be Selected:

Judging will be based upon the empirical evaluation of the performance of the algorithms. For the purposes of scoring, all scores will be assessed to three places after the decimal (i.e., 0.xxx).

- F-score judging and award conditions:
    - In general, the top three algorithms with the highest F-scores will be selected as the winners.
    - In the event that a team achieves F-scores that would consecutively occupy 1st, 2nd, and/or 3rd place (or any consecutive combination or otherwise (e.g., (1st and 3rd) or (2nd & 3rd)), the team will be awarded only one prize, which will be the highest available prize. After that award, that teams remaining score(s) will be skipped until a different competitor can be awarded the next prize level.

- In the event of a tie in F-scores, the winners will be awarded their prizes based on least amount of "runs"/tries that it took to get such score. For example, if two teams tied for first place and one got its high F-score on its 31st try and the other team on its 57th try, the first team would be awarded 1st place and the other team 2nd place. The same tie-breaking approach will be applied for all award positions. If the least amount of tries method subsequently results in a tie, the teams will be awarded that place and the prize will be split evenly to all who won that place.

- Best in category prizes
  - In general, best in category prizes will be made for precision, recall, and best first run.
  - Unlike the F-score award rules, a team is permitted to win one or all of these "best of" category prizes.
  - In the event that a tie occurs for the precision and recall categories, the same least amount of tries method (as noted in F-score) will be used to break the tie and same even award split method will be used if the least amount of tries method does not break the tie.
  - If the best first F-score run category results in a tie, the award will be split evenly to all who won the category.

Additional Information:

General Conditions: ONC reserves the right to cancel, suspend, and/or modify the Challenge, or any part of it, for any reason, at ONC's sole discretion.

Intellectual Property: Each participant retains title and full ownership in and to their Submission. Participants expressly reserve all intellectual property rights not expressly granted under the Challenge agreement. By participating in the Challenge, each entrant hereby irrevocably grants

to the Government a limited, non-exclusive, royalty-free, perpetual, worldwide license and right to reproduce, publically perform, publically display, and use the Submission to the extent necessary to administer the Challenge, and to publically perform and publically display the Submission, including, without limitation, for advertising and promotional purposes relating to the Challenge. This may also include displaying the results of the Challenge on a public website or during a public presentation.

Representation, Warranties and Indemnification:

By entering the Challenge, each applicant represents, warrants and covenants as follows:

(a) Participant is the sole author, creator, and owner of the Submission;

(b) The Submission is not the subject of any actual or threatened litigation or claim;

(c) The Submission does not and will not violate or infringe upon the intellectual property rights, privacy rights, publicity rights, or other legal rights of any third party;

The Federal sponsors reserve the right to disqualify any Submission that, in their discretion, deems to violate these Official Rules, Terms & Conditions.

Dated:
April 18, 2017 Paul Jonathan White

Paul Jonathan White, MD

Deputy National Coordinator for Health Information Technology.