Request for Proposal: Data Cleaning, Curation, and Harmonization

Sponsor	Accelerating Medicines Partnership for Parkinson's Disease (AMP PD)
Main Contact	Rosa Canet-Aviles Scientific Program Manager, Neuroscience Research Partnerships Foundation for the National Institutes of Health (FNIH) 301-402-5326 Rcanet-aviles@fnih.org
RFP Issue Date	13 June 2018
Submitted by	Lead Name: Lead Title: Organization: Lead Phone #: Lead Email:
Submit Questions	EOD on June 29, 2018
Proposal Return Date	EOD on July 9, 2018

1. Introduction

Accelerating Medicines Partnership for Parkinson's Disease (AMP PD) is a $24-Milllion public-private partnership to advance understanding, measurement, and treatment of PD through broad molecular profiling of biological samples from well-characterized PD cohorts. The PD cohorts included in AMP PD are the Parkinson's Progression Markers Initiative (PPMI), BioFIND, the Parkinson's Disease Biomarkers Program (PDBP), and the Harvard Biomarkers Study (HBS). All data generated through AMP PD will be housed in a Knowledge Portal built by AMP PD partner company Verily. Verily will include a subset of existing clinical and biologic data from the included cohorts in the Knowledge Portal. AMP PD established the Clinical Data Harmonization (CDH) Sub-Group to determine which data elements to include in the Knowledge Portal.

2. Description of the Initiative

The clinical and biologic data from the cohorts included in AMP PD are currently located in the individual study databases. Especially for ongoing studies, the studies often make the clinical trial data available to the research community in as close to real time as possible, which may mean only after minimal cleaning. In addition, the existing data from the individual studies are not harmonized to an external standard or to each other. Therefore, some pre-processing of the data is required before Verily can load the data into the AMP PD Knowledge Portal, including aggregating the data from the disparate study databases, streamlining the full study datasets to the subset of variables that will be included in the Knowledge Portal, cleaning the data (e.g. logic checks, correcting for missing values), and harmonizing the four datasets to the same standard. After pre-processing, Verily will upload the final harmonized datasets into Big Query for release into the AMP PD Knowledge Portal to enable researchers to integrate these data with the molecular profiling data being generated through AMP PD (e.g. whole genome sequencing, RNA sequencing).

After release of the Knowledge Portal to the AMP PD partners and the subsequent public launch of the Knowledge Portal, the CDH Sub-Group anticipates biannual updates to the clinical data that will include updated data from ongoing studies and may include the addition of new variables and/or datasets. Therefore, it is critical to establish a flexible curation/harmonization pipeline that is well documented for application to future datasets and where applicable uses data programming solutions or scripts that enable automated conversion to a standardized format. The CDH Sub-Group hopes to establish a relationship with

a curation/harmonization vendor that can establish the up-front pipeline and continue to engage with AMP PD for biannual data updates.

3. RFP Procedure

We invite you to provide a strategy and competitive cost proposal to provide data cleaning, curation, and harmonization services for AMP PD, with corresponding documentation that will enable the strategy to be applied to future datasets. The CDH Sub-Group anticipates that the scope of this project will include development of 1) a data retrieval and aggregation solution, 2) a curation/harmonization pipeline, 3) a harmonized data dictionary, incorporating CDISC standards when possible, and 4) documentation of the standardization/harmonization strategy for use with future datasets. This project will require close communication with AMP PD stakeholders – particularly, Verily and the CDH Sub-Group - and incorporation of feedback from subject matter experts throughout the development and testing phases.

The databases for PPMI, BioFIND, and PDBP are open access. As you may find it helpful to review the structure of these data as you're preparing your proposal, please find instructions for registering for access to the study databases below:

- PPMI
  - Visit http://www.ppmi-info.org/access-data-specimens/download-data/
  - Click APPLY FOR DATA ACCESS radio button
- BioFIND
  - Visit http://biofind.loni.usc.edu/download-data.php
  - Click Apply for Data Access radio button
- PDBP
  - Visit https://pdbp.ninds.nih.gov/cas/login?service=https%3A%2F%2Fpdbp.ninds.nih.gov%2Fportal%2Fj_spring_cas_security_check%3Bjsessionid%3DA6C19B92C867A207C8762B6C3CF94568
  - Click "Request A New Account"
  - Fill in the form with your personal information
  - Click the box to agree to the Data Use Certificate (this will serve as an electronic signature)
  - On the next page, please request access to be a "Data contributor and retriever with ProForms"
  - Request access to the permission groups "PDBP Consortium and PDBP Clinical Coordinators"
- HBS
  - Visit https://pdbp.ninds.nih.gov/cas/login?service=https%3A%2F%2Fpdbp.ninds.nih.gov%2Fportal%2Fj_spring_cas_security_check%3Bjsessionid%3DA6C19B92C867A207C8762B6C3CF94568
  - Click "Request A New Account"
  - Fill in the form with your personal information
  - Click the box to agree to the Data Use Certificate (this will serve as an electronic signature)
  - On the next page, please request access to be a "Data contributor and retriever with ProForms"
  - Request access to the permission groups "PDBP Consortium and PDBP Clinical Coordinators"
  - Would you like access to Genomics Data? – select "Yes" if you are requiring access to genomics data.
    - Requester will need to complete the PDBP Genomic DUC.
    - This document will need to be signed by an administrator from your institution, then scanned and attached to your account.

- Once the document is uploaded, you can submit your request.

As mentioned previously, the AMP PD Knowledge Portal will only include a subset of data from the studies. Please consult **Appendix A** for a summary of clinical data elements prioritized for the initial release of the Knowledge Portal. In addition, please consult **Appendix B** for a detailed list of clinical data elements prioritized for the initial release of the Knowledge Portal and example instructions for how to locate the selected variables in one of the open access databases. Please note that the CDH Sub-Group anticipates integrating additional clinical data for subsequent Knowledge Portal releases.

Please include the following details in your proposal:

- Project execution plan, including:
  - Data retrieval and aggregation solution that addresses the spread of data across multiple study databases. This solution must be scalable to allow for the incorporation of updated data from ongoing studies and the addition of new variables and/or datasets over time.
  - Curation/harmonization pipeline. If proposing a scripting solution for curation/harmonization, please confirm that you would be comfortable sharing these scripts with the research community so that AMP PD Knowledge Portal users could apply these scripts to curate additional data for upload into his/her workspace. If proposing the use of a licensable curation/harmonization tool, please fully describe any costs associated with licensing the tool.
- Project team structure, including the profiles (CVs) of key proposed team members expected to be assigned to the project. Include a description of each member's role with respect to budget allocation (i.e. his/her expected scope of contribution to the overall project). Please include a description of the proposed team's experience and capabilities in this area to demonstrate prior success.
- Project management strategy
  - Risk assessment and mitigation plan
  - Tools used for internal task management and milestone tracking
  - Tools used for internal and external communication
- Description of relevant expertise
  - *Required:* Clinical trial data curation
  - *Preferred:* PD/neurodegenerative diseases, Google Big Query, CDISC
- Budget (see Section 5)

Please submit any **questions** to Rosa Canet-Aviles (rcanet-aviles@fnih.org) no later than **5pm ET on June 29, 2018**. Please submit your **final proposal** to Rosa Canet-Aviles (rcanet-aviles@fnih.org) no later than **5pm ET on July 9, 2018**.

4. Timeline Table

Project Milestones	Target Date
Project kick-off call and vendor contract finalization	8/1/18
Curation/harmonization solution finalized and curation/harmonization started	8/30/18
First Big Query test upload	9/13/18
Second Big Query test upload	9/27/18

Curation/harmonization complete	10/11/18
Subject matter expert formal review	10/15/18
Final Big Query upload and curation/harmonization documentation delivered	11/1/18
Internal AMP PD Knowledge Portal launch	1/30/19
Set project scope for next clinical data release	3/31/19
Public AMP PD Knowledge Portal launch	7/31/19

5. Budget

As part of your proposal, please submit a budget in U.S. dollars utilizing the template attached as **Appendix C**. We expect this budget to enable the CDH Sub-Group to compare costs across received proposals. Please provide your costs and any unique assumptions, including hourly rates by position type, that you applied during budget preparation.

6. Confidentiality

All information included in this RFP is confidential and only for the recipient's knowledge. No information included in this document or in discussions connected to it may be disclosed to any other party. Through the acceptance of this invitation you further agree to keep your involvement in this RFP process confidential at all times unless the AMP PD Steering Committee co-Chairs agree in writing to allow such a disclosure. We agree to keep all of the information provided by you confidential, unless we are required by law to disclose it or it becomes part of any legal process.