Request for Proposal (RFP):
Automated Data/ETL Pipeline for Network Mapping Project

The Opportunity

The Myotonic Dystrophy Foundation (MDF) is seeking proposals from qualified vendors to create an automated or semi-automated data pipeline to extract, transform, and load publicly available datasets into a single database for our Myotonic Dystrophy Research Map. The goal of this project is to streamline and automate the process of collecting, cleaning, and transforming data from multiple sources – which has thus far been a manual process—and preparing it to be merged into our existing network map database and schema. The pipeline should be able to handle large, complex, and unstructured data, and should be able to run on a regular schedule.

About the Myotonic Dystrophy Foundation

Founded in 2007, MDF is the leading global advocacy organization helping families and professionals understand myotonic dystrophy (DM), a rare, genetic, multisystem, highly variable neuromuscular disease. MDF helps constituents identify resources and support, improve quality care, and advance research for management and cures. MDF has a global reach, assisting families and championing other advocacy organizations in more than 121 countries around the world. As many as 150,000 individuals may be at risk for the most common form of DM in the United States alone. Tens of thousands of affected families as well as providers, industry partners, researchers, and donors, together, form the MDF.

Vision:

We envision a world with treatments and a cure for myotonic dystrophy (DM).

Mission:

The MDF mission is Community, Care, and a Cure.

- We support and connect the myotonic dystrophy community.
- We provide resources and advocate for care.
- We accelerate research toward treatments and a cure.

663 Thirteenth Street, Suite 100, Oakland, CA 94612
www.myotonic.org | 415.800.7777 | info@myotonic.org

Myotonic Dystrophy FOUNDATION

About the Myotonic Dystrophy Research Map

To fulfill our mission of accelerating research toward treatments and a cure, one of MDF's goals is to lower barriers to DM research and drug development. One such barrier is a lack of a centralized database of research tools, studies/trials, publications, and experts in the DM space. This gap increases the burden of accessing public research and knowledge, which creates a steep and tedious learning curve for newcomers. A true understanding of the big picture of DM research is limited to a few Key Opinion Leaders (KOL), who have dedicated their careers to the study of DM, and whose time is consistently in high demand.

While there is currently no treatment or cure for DM, four drug companies have active clinical trials that have started within the last few years. With the "race to a cure" for DM in full swing, pharmaceutical and biotechnology companies new to the space want to be able to learn as much as possible as quickly as possible about the field. Additionally, drug developers who have already invested in research need to run landscape scans and automate the surveillance of the latest findings and progress in DM.

To address these needs and increase collaboration & general knowledge across the DM ecosystem, MDF created the Myotonic Dystrophy Research Map.

The map is an interactive, visual database of the current DM research ecosystem in the form of a network map. Though the first iteration started with just a few hundred datapoints, the project has grown in scope and now includes almost 8,000 nodes, 16,000 edges, and over 68,000 data points by which to search, sort, group, and filter. There is a limited-scope sample of an older model of the map, available to the public at www.myotonic.org/myotonic-dystrophy-research-map. With the dataset projected to continue growing and expanding at an accelerated rate, and because of the time-consuming nature of manual data management, we are seeking a partner to work closely with us to ensure the integrity, sustainability, and longevity of the project by developing and automating an optimized ETL pipeline.

Project Goals

Our current project goals include:

1. To refine and replace our existing manual data collection, validation, and preparation process with an automated or semi-automated system that can be run easily, efficiently, and frequently to ensure all data is up-to-date and accurate.

663 Thirteenth Street, Suite 100, Oakland, CA 94612
www.myotonic.org | 415.800.7777 | info@myotonic.org

2. To ensure the integrity of the project and the accuracy and consistency of the dataset by refining and improving our data processing system.
3. To ensure the sustainability and longevity of the project by building a transparent data pipeline that can easily be handed off to new Foundation staff or future contractors.

Scope of Work

The scope of work for this project includes the following tasks:

- Develop an automated or semi-automated data/ETL pipeline to collect and integrate data from various sources, including PubMed and ClinicalTrials.gov, using myotonic dystrophy related search terms.
- Clean and preprocess data to handle missing or inconsistent data scrapes and other data harmonization issues.
- Transform data into a schema and format that can be loaded into our existing network map.
- Develop a system for data deduplication and fuzzy matching to ensure accurate data association and long-term database integrity.
- Test and debug the pipeline to ensure it is working correctly and efficiently.
- Provide project progress updates to Foundation staff via regular emails and a weekly zoom meeting held during business hours of 9am-5:30pm Pacific Mon-Fri.
- Document the pipeline and provide user training for our team to run and maintain it indefinitely.
- To be able to provide further support, debugging, feature additions, or consultation for an agreed-upon amount of hours for up to 6 months after the completion of the initial project.

Qualifications

The vendor should have the following qualifications:

- Strong understanding of social network analysis and the representation of connections in a dataset.
- Experience working with graph databases and visualization tools such as Cytoscape or Tableau.
- Fluency in at least one programming language such as Python, R, and JavaScript.
- Experience with data cleaning and preprocessing.
- Experience with data deduplicating and fuzzy matching.
- Experience with cloud solutions such as AWS and setting up CloudWatch monitors.

663 Thirteenth Street, Suite 100, Oakland, CA 94612
www.myotonic.org | 415.800.7777 | info@myotonic.org

- Knowledge of statistical analysis, machine learning, and natural language processing.
- Good communication and problem-solving skills.

Timeline

The Myotonic Dystrophy Research Map is currently in beta-phase and is targeting a late Spring 2023 release pending the completion of this project. The vendor should be able to complete the pipeline-building process and testing within *one month* and will be available for further debugging, updating, or consultation at an hourly rate for the following 6 months.

Proposal Guidelines and Submission Process

Proposals will be accepted through February 17th with interviews held through March 1st. We hope to select a consultant and begin the project by March 3rd.

Proposals or questions regarding the project should be sent directly to Kate Beck, MDF's Director of Development, at kate.beck@myotonic.org.

Proposals should include the following information:

- Detailed project plan and timeline including:
  - Budget
  - Projected time spent by you on the project
  - Projected time spent by MDF staff in assisting you in the project
  - A summary of your approach for tackling data deduplication and fuzzy matching on network map datasets
- Your resume or CV, including:
  - Your work and/or educational history
  - The programming languages, visualization platforms, or cloud solutions you are proficient in
  - Any relevant libraries or applications you currently use and have a license to access/use
- A portfolio with samples of past work, with a specific focus on social network analysis or network mapping and data pipelines for unstructured data.

663 Thirteenth Street, Suite 100, Oakland, CA 94612
www.myotonic.org | 415.800.7777 | info@myotonic.org