International Rett Syndrome Foundation

Request for Proposals

Overview

International Rett Syndrome Foundation (IRSF) is requesting proposals for a database management platform that can store data from multiple and disparate sources and combine these data to support a registry focused on clinical outcomes and quality improvement, as well as research by clinical principal investigators and industry (pharma) partners.

The desired database platform should be multi-faceted and broadly support IRSF data needs with functionality for data governance, data acquisition, data aggregation, reporting, and data distribution.

Data Governance

The Medical Advisory Board (MAB) of IRSF advises on definitions of required data that IRSF Centers of Excellence (CoE) and other clinical partners could submit to the IRSF data platform. These definitions include the following:

1. Extracts from EHRs - Clinical data that may be extracted from the provider's electronic health record. The metadata include the list of data elements and data definitions for each data element. Criteria for the data capture period, patient cohort population description, and inclusion/exclusion criteria for the data extract. The criteria may reference code value sets for various clinical domains such as diagnoses, labs, medications, and procedures etc. The criteria may also describe desired syntactical formats for data files. Data elements should also include the preferred coding system (e.g., RxNorm, SNOMED CT, ICD-10-CM, LOINC, etc.). If the provider sends the requested data without codes or in a different coding system than desired these data will need to be translated to the preferred coding system as part of the data aggregation process (see Data Aggregation below).
2. Surveys and Questionaries - Surveys and questionaries will be designed for patient caretakers to complete via a web-based form before the visit (in some cases), as well as for clinicians to collect data during patient encounters in a standardized way. Survey

questions and allowable responses should be stored in the data platform as master data. These tools will include questions in narrative text for each data element to capture, as well as standard options for responses to each question. These responses could be defined as a numerical value (such as temperature in Celsius), or a list of allowable values code and description) such as a pain scale (select one choice) or a list of symptoms (select none, one, or multiple). The surveys and questionaries will be versioned and will likely evolve over time to meet the changing needs of the IRSF community and should have new versions for the survey as well as each of the survey questions so that responses to the same question can be aggregated across survey versions. While the MAB will advise on the core or required questions for a survey, clinicians may want to extend the survey to collect additional data. These data definitions should be submitted to IRSF, assigned a data element version, and be available as optional data elements for clinical sites that want to capture these additional data elements. The purpose for creating master data for survey questions and allowable question responses is to standardize data collection for mandatory surveys and questions while offering clinicians flexibility.

3. Data schema and definitions – Data definitions for common data models and schema used for data aggregation, reporting, and data distribution. Data models should include a data dictionary for each table or data view object that describe the purpose of the data set and each data element it includes.
4. Report Definitions - Report definitions should include meta data that describe the Report Title, Purpose, Intended Audience, Measures, Measure Calculations, Inclusion/Exclusion Criteria, Report Version, and any user configurable filters or parameters. Additional meta data may include Measure Goals, upper and lower control limits, as well as related effective start/stop dates.
5. Data Distribution - Data distribution meta data should include definitions which are similar to Report definitions to describe data sets that may be distributed to IRSF partners.

Data Acquisition

Data will be submitted to the IRSF data management platform from numerous disparate sources for a variety of clinical and financial data domains. The data management platform should allow data providers to submit data in the method that is easiest for them. Data acquisition and submission may be via upload of data files or direct data entry via a secure web data collection form.

1. Data file meta data – All data files received should be linked or tagged with meta data that describes the organization/site that submitted the file, date the file was received, if the file contains protected health information (PHI), data domains(s), links to EHR extract specifications or survey versions, etc.

2. IRSF constituents should be able to submit files to the data platform via upload to SFTP, or file upload to a secure website.
3. Web-based data entry - Clinical sites and/or patient families may submit data by completing data collection forms (surveys or questionnaires) for one patient at a time using a secure website form. Survey data collection tools developed for web-forms would implement the data collection specifications established by the MAB. The vendor solution must include the ability to create and host web-based data collection forms that align with Surveys and Questionnaires defined by the MAB (see Data Governance section, item #2).
4. Data quality assessment and remediation planning– all data files should be evaluated for conformance to the data collection specifications established by the MAB for the applicable data set. This assessment will identify data quality issues and data remediation plans that determine if data quality issues must be addressed by the data submitter or if they can be resolved on the data management platform via rule-based data transformation. If rule-based transformations are utilized, the original data must be preserved, and the data transformation documented and create an audit trail of data lineage and transformations.

Data Aggregation

1. Master Patient Index – A unique identifier for each patient should be created by the IRSF data platform. The data management platform should provide the ability to create an index of the related identifiers from source systems that allow patient data to be linked across disparate data source systems.
2. Semantic Normalization – The target data schema for aggregation should specify coding systems (RxNorm, SNOMED CT, ICD-10-CM, LOINC, etc.) for each coded data element. The data platform should provide the ability to create semantic maps from source terminologies to the preferred target terminologies. In most cases source terms should be mapped to their semantic equivalents in the target terminology.
3. Common Data Schema – Common data schemas should be created to support the Reporting, Research, and Data Distribution needs of the IRSF community. Refer to the "Reporting" section below for general requirements. Research and Data Distributions requirements will be defined in the at a future date by IRSF.

Reporting

The data management platform should support web-based reports in the form of graphical visual displays and tabular data as implementations of reporting criteria defined in the report meta data (see Data Governance, Report Definitions). Initial reports will include report specifications and criteria for the following:

1. REPORT 1: Population statistics for age of onset for various symptoms or functional outcomes of Rett Syndrome by genetic variant.
2. REPORT 2: Communication/Language functionality by age group and genetic variant.
3. REPORT 3: Distribution of Rett population by genetic variant.

REPORT 4: Lists of patients expressing a particular symptom. **Data Distribution**

Data sets may be created from common data models for distribution to IRSF partners to support research, clinical trials, and other IRSF approved data use cases. Data sets created for distribution should be inventoried and linked to meta data (see Data Governance, Data Distribution above).

Scope

The scope of the project is defined by the following items:

Item #	Item Description
1	Select a cloud hosted, HIPAA certified database platform that will be the future repository for the IRSF data management platform. The selected database, reporting, website, and data visualization technologies must allow for authorized users to have role-based access privileges. The data management platform must be 21-CFR HIPAA compliant.
2	Implement an initial common data model to support the data acquisition, data aggregation, and reporting requirements described in the Reporting section of this document. Detailed specifications are described in Appendix A.
3	Create a standardized process for documentation of data governance, multiple data acquisition pathways including web-based data entry and file upload to SFTP or a web page.
4	Demonstrate the core requirements of the IRSF data management platform by creating a web-based reporting system to implement the Reporting features described in this document, which build upon the data governance, data acquisition, data aggregation, and reporting sections of this document. Reports will be delivered via an interactive web page where users can select from applicable report parameters to update the visual display.
5	Create a web-based data entry form to implement a clinical data survey or questionnaire linked to the data governance requirements of this document.
6	Create a web-based utility to allow IRSF partners to upload data files.
7	Create a SFTP site to allow IRSF partners to upload/download data files. The data upload functionality should complete a basic data quality check and provide a feedback mechanism to the data sender to confirm that upload meets data quality standards, or alert them is the file does not conform to expected standards.
8	Database hosting and management costs. Provide an annual estimate of fees for hosting the database and related database administration.

Response requirements

Please respond with your overall approach to the project, and how you would deliver each of the required items listed in the Data Governance, Data Acquisition, Data Aggregation, Reporting, Data Distribution, and Scope sections of this document. Include a not-to-exceed proposal for Scope items 1-7, and an estimate of costs for Scope item 8 (database hosting and administration).

Response Submissions and Due Date

Responses should be sent by email to mkennedy@rettsyndrome.org by May 7st, 2021 at 5pm EST.

Evaluation and Selection of Response Vendors

Rettsyndrome.org will evaluate vendor responses and will award the project by June 1st, 2021 and notify the vendor by email and telephone.

Timeframe for Starting and Completing the Project

This project should start by June 15th, 2021 and be completed by September 17th, 2021.

Success Criteria

The functionality of each of the Scope items is met, all related reports and artifacts are provided by the project due date and are acceptable to RettSyndrome.org.

Appendix A

Report Definitions:

REPORT 1: Population statistics for age of onset for various symptoms or functional outcomes of Rett Syndrome by genetic variant from a predefined list. The list of genetic variants would be populated from the variants identified in the Registry population. As new variants may be added from time to time, the list would automatically update to include them.

Purpose: Answer the question “What symptoms are common in children with the same variant as my child, and at what age do they typically start?”.

Description: Bar chart histogram with symptoms along the x axis and age category on the y axis. The height of the bar indicates the frequency of the symptom onset for the related age category. Only symptoms with a frequency > 0 for the selected genetic variant should be included in the chart. Age Categories are as follows: 0-5, 6-10, 11-18, 19-30, 31-40, 40-50, 50-60.

Parameters: Users may select the genetic variant.

REPORT 2: Communication/Language functionality by age group and genetic variant.

Purpose: Answer the question “How is communication ability typically impacted by her variant, and how might this change over time?”.

Description: Bar chart histogram with communication/language categories along the x axis and age category on the y axis. The height of the bar indicates the highest level of communication ability for the related age category. Age Categories are as follows: 0-5, 6-10, 11-18, 19-30, 31-40, 40-50, 50-60. Communication categories are as follows: Eye gaze, Sentences, Phrases, Words, Babble, No Words, Uses computer/tablet/iPad.

Parameters: Users may select the genetic variant.

REPORT 3: Distribution of RETT population by genetic variant.

Purpose: Answer the question “How common is my child’s genetic variant?”.

Description: Pareto chart (or sorted/ranked bar chart) showing the frequency of each genetic variation among the population in the registry. The x axis includes a bar for each variant and the y axis indicates the number of people in the registry with the genetic variant.

Parameters: None

REPORT 4: Lists of people in the registry with a particular symptom.

Purpose: Identify a potential cohort of people with a given symptom for a clinical trial.

Description: Tabular report listing the IRSF coded patient identifier, age, genetic variant, gender, and age of onset of a given symptom.

Parameters: Age category (single category or multi-select), genetic variant, and symptom from a predefined list. Symptoms may be added to the registry and documented in registry participants from time to time. The predefined list of symptoms should automatically update to include any new symptom as it is added to the registry.