# Open Call for **Essential Oncology Software for Research**

## Opportunity

Princess Margaret (PM) and the PM Data Science Program (PMDS) recognize that software is critical to modern scientific research, advancing medical discoveries while providing a framework for reproducibility and transparency. PM Research and Clinical labs have developed software packages, libraries, and applications that have become essential tools for research. Many of these software solutions do not have dedicated funding for maintenance, let alone for growth and maximizing their potential for transformational impact. In a targeted effort to bolster the use of PM-developed software for research, PMDS is seeking applications for -centered projects requiring support that are essential to biomedical research within PM. The software solution must be in active use, have already demonstrated impact and show need and/or potential for continued improvement and growth.

## Key Dates

* Jun 1, 2021 - Application Due
* Jul 1, 2021 - Earliest notification of decision (subject to change)
* Aug 1, 2021 - Earliest start date (subject to change)

## Budget

PMDS is looking to support three applications with PMDS funds to be used towards up to 30% FTE of a software developer. Funding support will be from **Aug 1, 2021 to Mar 31, 2022** with the possibility for renewal for another year dependent on deliverables, performance and budget allocation.

## Application Process

Please complete the application form and submit it via email to [info.pmdatascience@uhn.ca](mailto:info.pmdatascience@uhn.ca) by Jun 1, 2021 11:59pm EDT. Applications will be reviewed by the PMDS team based on the following criteria:

* Alignment with PMDS program
* Demonstrated active use and the potential for broad impact across research programs/labs
* Potential for continued improvement and growth, i.e., scalable architecture and development roadmap

# Essential Oncology Software for Research

Thank you for applying to the Princess Margaret Data Science (PMDS) Call for “Essential Oncology Software for Research.” Please adhere to the word limits and format outlined below.

|  |  |
| --- | --- |
| Software/Application Name | reportRx |
| PI/Lab | Wei Xu |
| First Release Date | Dec 12, 2013 |
| Software Developer(s) and other technical members and roles of the team | Jessica Weiss (developer)  Lisa Avery (developer)  Osvaldo Espin-Garcia (developer)  Tyler Pittman (technical support) |

Provide a description of the software (URL and links to documentation if available). *(limit 250 words)*

|  |
| --- |
| ReportRx is a comprehensive R package designed for biostatisticians, bioinformaticians, and research analysts to provide analytic and computational tools to generate reproducible statistical analysis reports of clinical and observational studies. The main objective of reportRx is to facilitate the reporting aspects of commonly performed statistical modeling and analyses while allowing for a reproducible pipeline. Thus, enabling systematic reporting, automatic tables and figures generation, and a faster turnaround, allowing biostatisticians and research analysts alike to focus on the statistical aspects of a project as opposed to spending time formatting and copying/pasting results, which is prone to error and highly non-reproducible. ReportRx also provides powerful computational tools to generate complex tables and figures for data visualization on comprehensive statistical results.  ReportRx automatically generates appropriate descriptive statistics (i.e. Table 1) based on variable type as well as between-group baseline comparisons. Univariate analyses are also automated based on the type of response variable and currently the package can conduct linear, logisitic, Cox proportional hazards, and competing risk models. A multivariable reporting function harmonises the output of multivariable models for seamless reporting.  Three plotting functions provide data visualizations: 1) a simple but flexible function for Kaplan Meier survival curves, 2) automated plotting of bivariate relationships to completement the univariate analyses, and 3) a forest plot to visualize and compare risk factors for multivariable logistic regression.  To date, the package has been downloaded 13,038 times according to the comprehensive R archive network (CRAN) log files. Link to current implementation:  https://github.com/biostatsPMH/reportRx  Intro document:  https://biostatspmh.github.io/IntroToReportRx/ |

Outline the Technology Stack for the Software

|  |  |
| --- | --- |
| Database/Data Storage Technology | Data integration, quality checking, imputation for missing data. Summary statistics presented in tables and figures. Survival data and longitudinal data summary. |
| Front end/User Interface Technology | Biostatisticians, bioinformaticians, and research analysts. Windows and Unix environment for PC and Mac machines and clusters. |
| Integrations/Interfaces | Different database, ACCESS, RedCap, Excel, R data files. |
| Standards | Different versions. |
| Other | R package. We may extent it to SAS macros and potential Python packages for high dimensional bioinformatics research studies |

Describe the current status of the software application and any planned improvements. *(limit 250 words)*

|  |
| --- |
| ReportRx has recently undergone a major update, now under version 1.03, which fixed various bugs and added functionalities including support for Rmarkdown, which is a modern interface for generating reproducible reporting pipelines.  Currently, the software can be used to generate reproducible analyses in R scripts and integrates with both the older R Sweave and the newer R Markdown literate programming packages for seamless and flexible output to HTML, Word and PDF documents. The package will detect the output format and generate tables in the appropriate syntax (latex. Markdown or html) to eliminate the time spent on table formatting. Similarly, plots produced by reportRx can be embedded into all document formats or saved to external files (.tiff, .png etc)  Need to list the major functions of the current package:  1.  2.  3.  A number of improvements are planned:   * support a wider variety of commonly used regression model, including but not limited to mixed effects models (packages nlme and lme4), joint models (packages JM, JMBayes, JoineR, joineRML), parametric survival models (packages survival and eha) * support for machine learning algorithms and methods such as penalized regression (e.g. glmnet package). * support for reporting streamlined predictive modeling, e.g. results from caret package * support for reporting and visualizing causal inference analyses such as propensity scores and mediation modeling * automated testing and reporting of model assumptions (e.g. proportional odds in ordinal regression) |

Provide description of demonstrated impact, include number of users (internal/external), data points, publications or other details that can help PMDS assess impact. *Letters of support can be attached to application (limit 250 words)*

|  |
| --- |
| ReportRx has a primary user base in the Department of Biostatistics at the Princess Margaret Cancer Centre where approximately 80% of the department members and most of the trainees at COMBIEL training program use the software on a regular basis (20 staff members and more than 30 trainees). In addition, a number of former trainees and visiting scholars have found the tools useful and have continued using the package for their biostatistical and bioinformatic modeling and analysis needs.  Need to give references for high impact publications that utilized the package (i.e. Lancet Oncology 2016 PMID: 26936027; Journal of Clinical Oncology 2015 PMID: 25667292; Cell Report 2018 PMID: 30380421; Bioinformatics 2020 PMID: 32449747. Can add more)  To date, the package has been downloaded 13,038 times according to the comprehensive R archive network (CRAN) log files. This number excludes more recent downloads from the GitHub repository.  From its inception, reportRx has facilitated a standard reporting system in the department and its utilization has directly impacted the number of publications supported by the Biostatistics department members. |

Please complete the application form and submit it to [info.pmdatascience@uhn.ca](mailto:info.pmdatascience@uhn.ca) by Jun 1, 2021 11:59pm EDT.