REQUEST FOR PROPOSAL

ISAC Background

The Mission of the International Society for Advancement of Cytometry (ISAC) is: "to serve a multidisciplinary community by leading technological innovation, scholarship, and the exchange of knowledge in the quantitative cell sciences" (isac-net.org). ISAC funded the initial development of FlowRepository V1 (though funding from the Wallace H Coulter Foundation) in collaboration with the Brinkman lab at the BC Cancer Research Center (BCCRC), and contributions from CytoBank Inc. (Santa Clara, CA).

Project Overview

This RFP describes the requirements for a public, online, Cytometry Data Repository (FlowRepository V2). FlowRepository V1 has reached the end of its life. It is built on now outdated technologies that cannot be reliably updated presenting security and stability issues and is hosted on a single server, making it susceptible to failure and presenting remote users with decreased performance.

FlowRepository hosts public flow and mass cytometry data sets used in publications that are linked to online manuscripts through many journals, including *Cytometry A* (Figure 1). It is critical that these data and journal links are preserved, both because they represent a substantial investment in data generation by many funding agencies, time and effort by PIs and researchers, but also because they contribute towards reproducible science. It currently hosts 147,495 FCS files (1136 GB) from 582 public datasets and 631 that are in the process of being released.

FlowRepository V2 should capture many of the features as well as all the data available in FlowRepository V1, but should leverage robust technologies, including distributed storage and hosting, allowing it to scale to many users and large data sets. Applications are due December 15, 2019. Consideration will be given to FlowRepository V2 proposals that either are ultimately managed by ISAC or by the vendor themselves. The major requirements are described below.
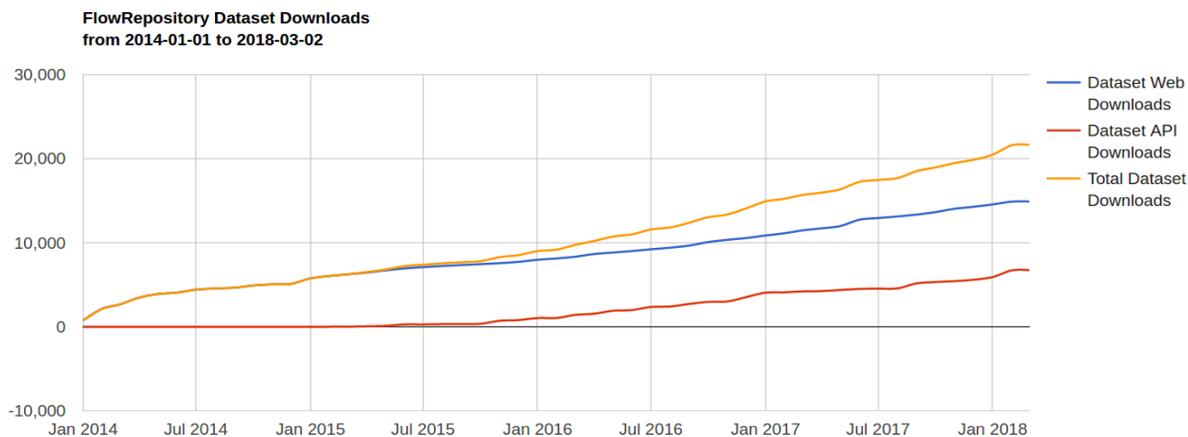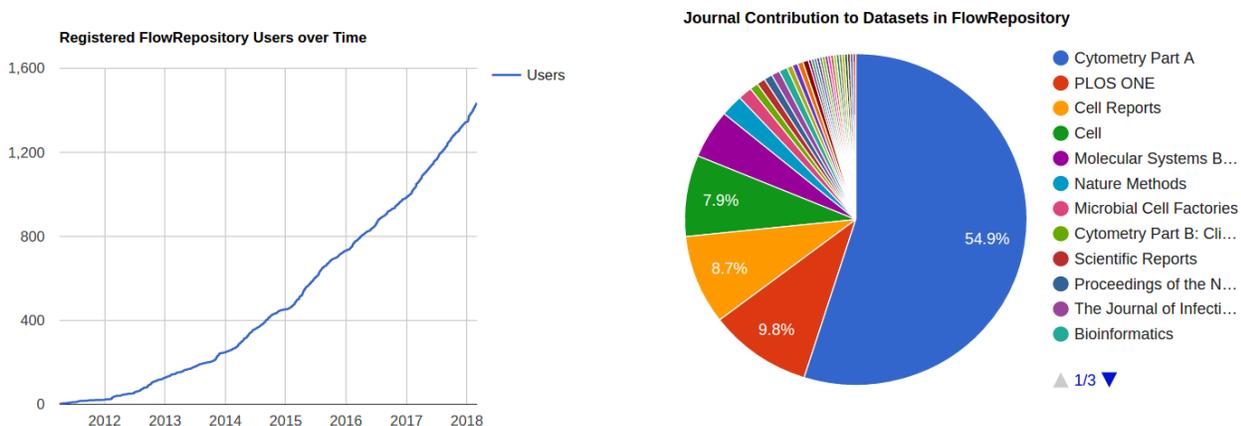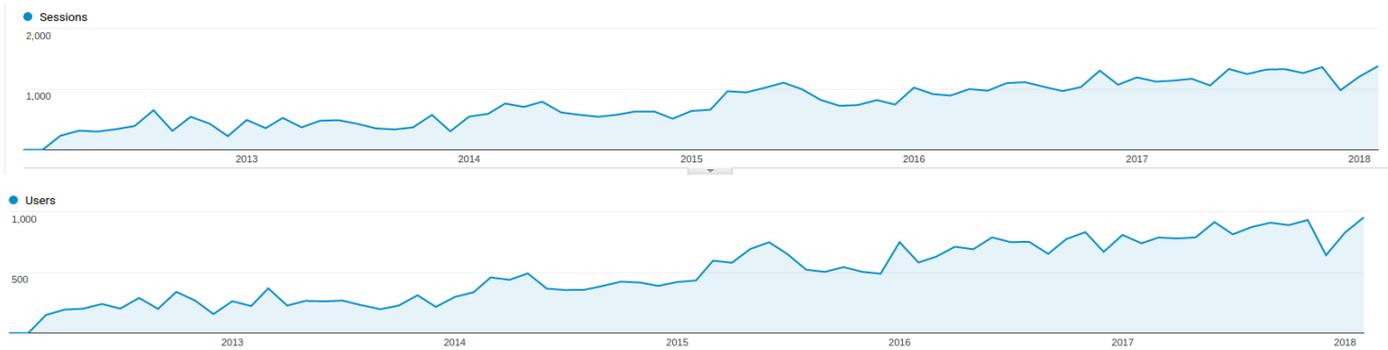
Figure 1. Flow Repository usage statistics.

Project Goals

There are three goals for FlowRepository V2:

1. Provide open, public access to raw and analyzed cytometry data sets linked to studies and manuscripts, either private, published, or under review. The purpose of the FlowRepository is analogous to other public data repositories such as the Gene Expression Omnibus (GEO) for gene expression microarray data, and the Sequence Read Archive (SRA) for sequencing data.

2. The repository should provide a resource to enable reproducible analysis and reanalysis of cytometry data sets and studies by tracking experimental and study metadata, study design information and other required information and data required to fully reproduce analyses cytometry data sets, either published or under review.
3. The repository should facilitate data sharing by providing suitable mechanisms for data submission, searching, and download via both a graphical interface (*i.e.*, through a web browser) and via a programmatic API that can interface with open-source data analysis tools and programming languages (*e.g.,* R / BioConductor, python, etc. via http or other open protocols).

Required Features

The new FlowRepository should have the following features and capabilities:
1. The ability for users to log-in, submit, upload, and annotate cytometry data in the form of raw FCS files as well as experiment annotations in the form of assay, sample and study metadata files (e.g. sample annotations in simple formats such as csv text files, experiment annotations describing marker staining panels, study annotations describing study designs, hypotheses and so forth).
2. Support for upload of data analysis information that describes how data are gated (e.g. "workspace" files associated with tools like FCSExpress, FlowJo, DIVA, Kaluza, or Cytobank), or clustered (e.g. in the form of cluster annotation files), and that enable the public consumers of the data to reproduce these analyses (e.g. via workspace files, or via programmatic code to reproduce computational analysis).
3. Users should be able to group uploaded data into "studies" together with associated raw data, annotations including study designs, hypotheses, and descriptions necessary to provide context to the data and the analyses performed.
4. The ability to have data sets and studies be private, and only accessible to the data "owner" and designated individuals (e.g. manuscript reviewers), or public and accessible to the community.
5. The ability to lock data sets in a "public" state.
6. The ability to generate "permanent links" to data sets for inclusion in manuscripts, in a format compatible with previous versions of FlowRepository.
7. The ability to generate Digital Object Identifiers (DOI) linking to data sets for inclusion in manuscripts.
8. The ability to link into journal publication workflows including easy to manage public release upon publication (e.g. event management review process of Cytometry A).
9. It should migrate and preserve data in the existing FlowRepository together with the permanent links to those datasets that are already linked from published manuscripts.
10. The ability for users to download public data sets to their local machine, together with associated annotations, metadata, and analysis information necessary to reproduce the analysis of the public data set in a local environment.
    a. The ability to download datasets via a graphical web-based interface.
    b. The ability to download, search, and submit datasets via an open, free, documented programmatic API in order to interface with open source cytometry analysis frameworks (e.g. R/BioConductor), web-based repositories (e.g. Cytobank), and other tools (e.g. FlowJo), freely and in an open manner.

11. The ability to store experimental design information for reproducible analysis, for example:

        a. Free text, statements of study goals, hypothesis, etc.

        b. Sample information including treatment groups, randomization, outcomes, etc.

        c. Sample type information including tissue type, source organism, etc.

12. The repository should be scalable to store, query, and transfer large data sets. It should leverage cloud technology for scalability.
13. The repository should be easy to maintain, leveraging proven technologies.
14. The repository should implement a controlled vocabulary as needed to describe subsets of the MIFlowCyt standards, including protein markers, channels, dyes, and so forth needed to describe flow cytometry experiments and enable cross-experiment data analysis. Depending on the project costs, preference will be given to proposals that fully implement the MIFlowCyt standards.
15. The repository should enable search of data sets to identify samples matching various search criteria, based on the staining panel, sample, experiment, and study annotations provided for each data set.
16. An index of available data sets should be crawlable by search engines so that they are directly accessible via Google and other engines.
17. Login should not be required to search and download data, only for upload.
18. Where possible, the repository should enable extraction and visualization of reproducible data analyses by importing, processing, and producing visualizations for datasets with available workspaces (that describe the gating procedures).
19. The repository should support open standards.
20. The repository should support cross-linking to CytoBank, ImmuneSpace, ImmPort and other repositories where cytometry data are publicly available.
21. Software should be available as open source. Alternative options where the software is placed under escrow allowing the code to be released to ISAC in the event that the licensor is not able, or willing to support the software will be considered.

Deliverables

1. Description of the architecture and implementation.
2. Timeline for the beta version, import of all existing data, interface with *Cytometry A*, and anticipated live date.
3. Breakdown of costs for development, ongoing maintenance, hosting, and other deliverables as appropriate.

Evaluation:

The proposal will be evaluated by the ISAC Flow Repository Task Force and ISAC Leadership. Emphasis will be placed on the experience of the respondent and the completeness of the application. The respondent should identify the individual person(s) who will manage the project including their qualifications/experience. The proposed budget will be a major consideration.

Questions about and the Response to this Request for Proposal:
Responses should be mailed/emailed to:
Ryan Brinkman, Ph.D.
Email: rbrinkman@bccrc.ca
Subject line: FlowRepository RFP