Title: ArisLIMS: a simple portable open-source system to facilitate sample documentation

Authors (order to be decided): Michael Mironidis, Shiraz Bheda, Hila Milo Rasouly, Anna Mo, Sarath Babu Krishna Murthy, Ali Gharavi

Abstract

Summary: As more laboratories are gathering in-house biobanks, efforts spent on effective sample management and other administrative tasks scale poorly. To address the need of sample management, we developed ArisLIMS - a user-friendly, highly customizable, laboratory information management system (LIMS). This web-based management system allows sample-tracking and update of sample status by multiple users. Its flexible structure allows for seamless interaction with large sources of information, reducing time spent on manual query and data entry.

Availability and Implementation: (https://github.com/orgs/ColumbiaCPMG/repositories)

Contact:

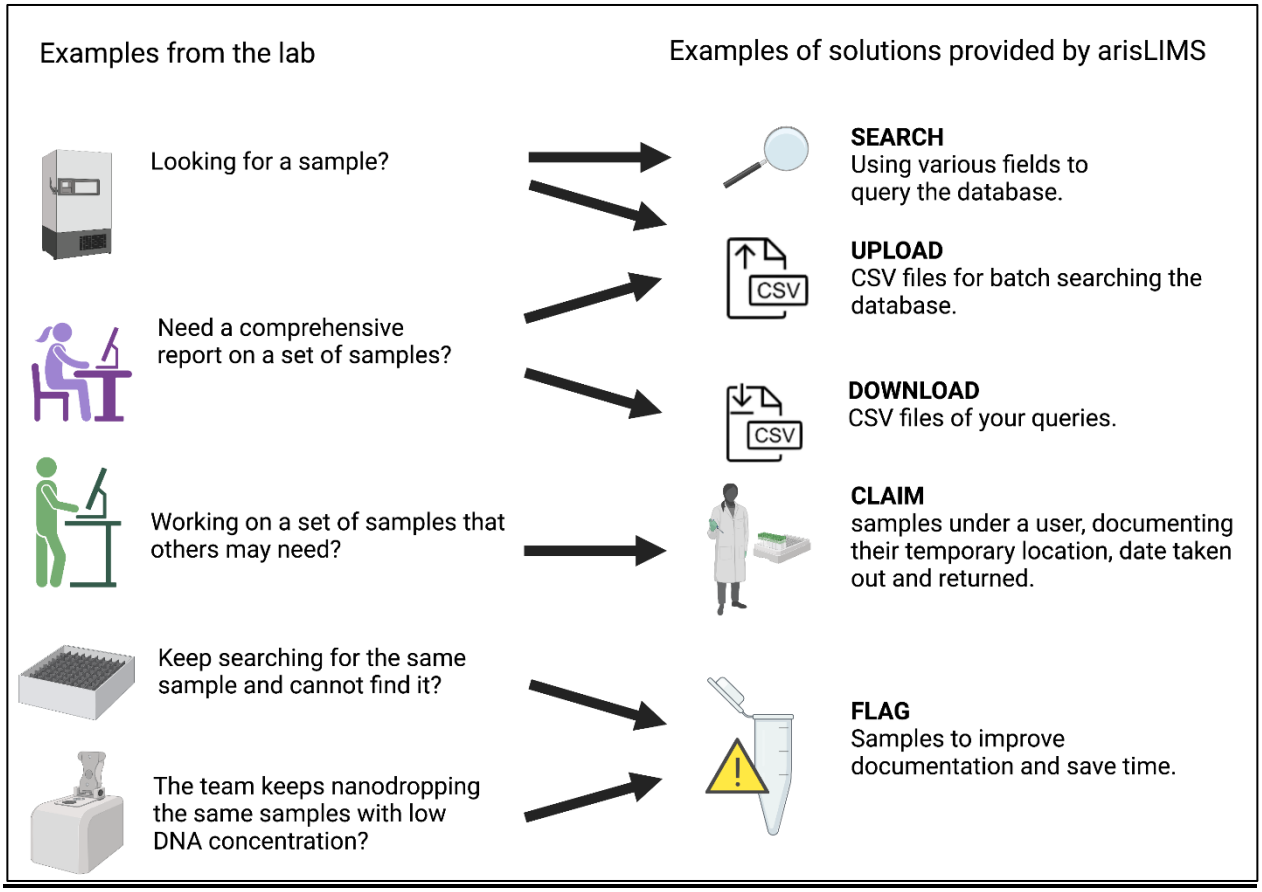
Supplementary information: (currently at the end of this document)

Acknowledgements

Maddalena Marasa, Enrico Cocchi, Iman Ghavami, Gina Jin, Junying Zhang

Funding information

Figure 1:


Examples from the lab
Examples of solutions provided by arisLIMS
Looking for a sample?
Need a comprehensive report on a set of samples?
Working on a set of samples that others may need?
Keep searching for the same sample and cannot find it?
The team keeps nanodropping the same samples with low DNA concentration?
SEARCH
Using various fields to query the database.
UPLOAD
CSV files for batch searching the database.
CSV
DOWNLOAD
CSV files of your queries.
CSV
CLAIM
samples under a user, documenting their temporary location, date taken out and returned.
FLAG
Samples to improve documentation and save time.

INTRODUCTION: As datasets are growing, tracking samples can become increasingly difficult. Poor sample management slows and inhibits research outcomes. The importance of a centralized management system for laboratory organization cannot be understated, and many have been developed over the years to address different laboratory needs [1, 2]. While commercial solutions exist, they are accompanied by expensive licenses and fees. There are several open-source solutions with some features we were interested in [3, 4]; however, none implemented all the requirements needed for our laboratory, which includes a large number of team members managing diverse, global data [5]. We therefore designed ArisLIMS to provide an intuitive database structure for sample tracking, sample status updates, and documentation.

TOOL DESCRIPTION: The web-app was developed using Flask, a lightweight, Python-based web framework which simplifies and enables further development for years to come. The back-end is MySQL, a secure database management system best suited for the kind of structured data that is stored in a biobank. Laboratory personnel requested a system to keep track in real time of the location of stock samples, the location of samples while being used, the name of the person using the sample and for which project, and the date that they were taken out from their stock location (Figure 1). They also requested batch searches and export of search results. To enhance system compatibility, the system was dockerized and a manual was developed to guide its installation on any system (GitHub).

ArisLIMS allows simple, user-friendly sample management of the catalogue of samples using external inputs. Access to the application's features is protected by a login page. Accounts have tiered access to the LIMS. By default, these tiers are Admin and Standard. Admins have access to additional site functionality such as database updates and user access restriction. The admin

has the capability of updating the entire ArisLIMS database and track user activity, including the username, action type, sample ids, and date time of the action.

Once logged in, the login page becomes the Account page. ArisLIMS includes six pages: (1) Home (Figure S1a), (2) Account (Figure S1b), (3) Samples (which serves as the main catalogue, Figure S1c), (4) My Samples (unique for each user, Figure S1d), (5) Missing Samples Management (Figure S1e), and Database Upload (Figure S1f). **The Home page** includes a display of the total sample count by researcher and a message board to facilitate communication.

The Account page contains 3 key features: (1) a snapshot of the samples database, (2) a log of user activity (Figure S2), and (3) templates for batch sample queries. Users can query the entire database using either the Samples or the File Upload pages. In **the Samples page**, users can manually search for samples using a variety of fields and update the status of samples in real-time. In **the File Upload page**, users can upload a CSV input file that includes any relevant search fields for batch queries (Figure S3), and download a report providing a detailed report of samples found, samples flagged as in the database but missing from physical location, and sample queries that are not in the database or misspelled. Samples marked as taken out of stock and their temporary location appear on **the My Samples page**. The My Samples page is restricted to the list of samples taken out under the username of a specific lab personnel. Users can also mark a given aliquot as “missing” or as having “low DNA volumes". Missing aliquots will then appear on **the Missing Samples Management page**. In a survey of the lab team, respondents shared that ArisLIMS reduced the time spent searching for samples.

CONCLUSION: In summary, ArisLIMS is a cohesive, dynamic system that streamlines and simplifies sample management. It can serve as a solid base for other laboratories to implement a centralized information and management system and even synchronize with existing databases.

Supplemental

Figure S1a – Home tab

Figure S1b – Account tab

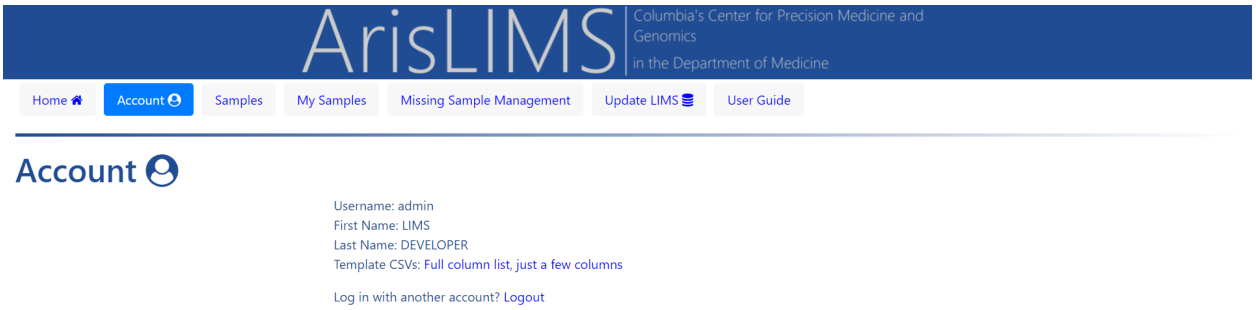


Figure S1c – Sample tab

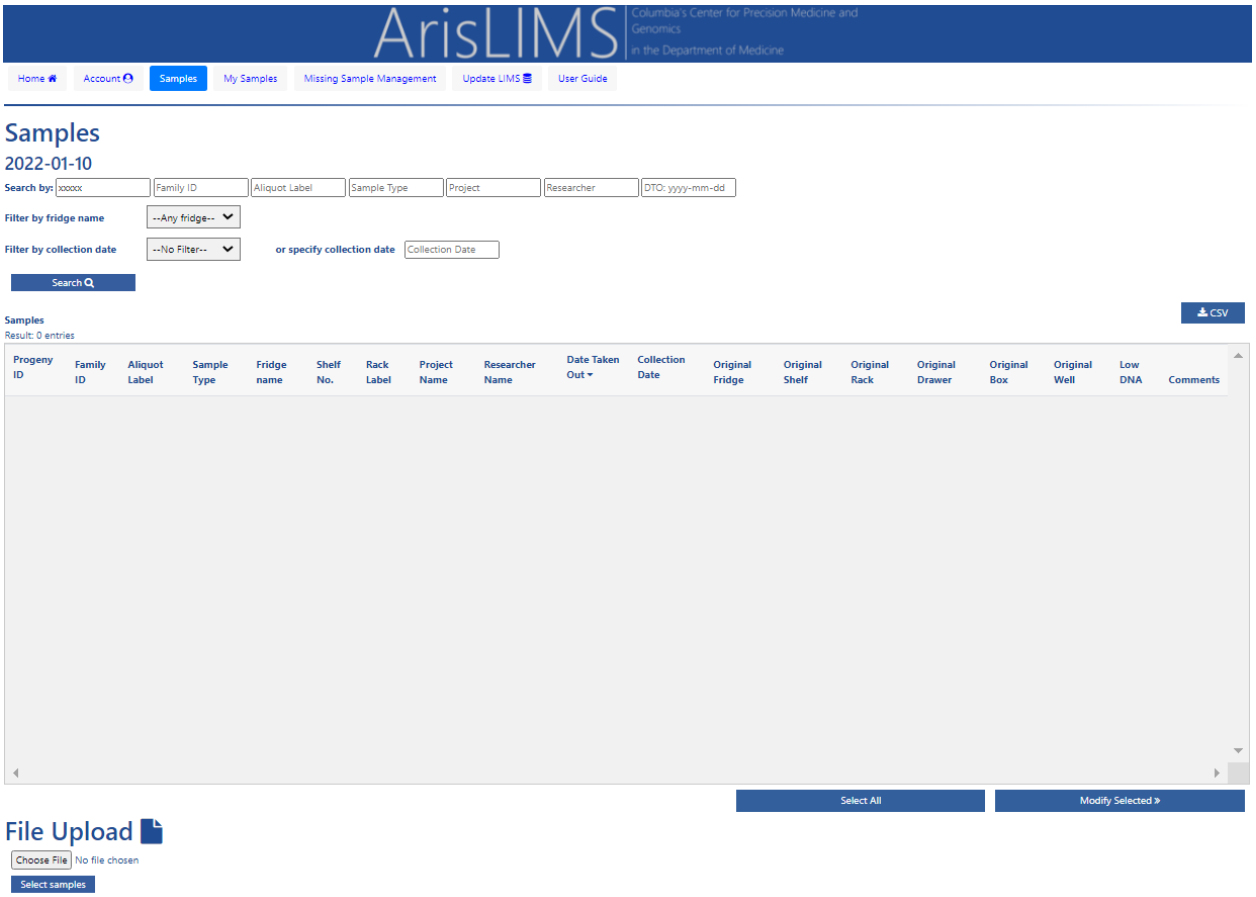


Figure S1d – My Samples tab

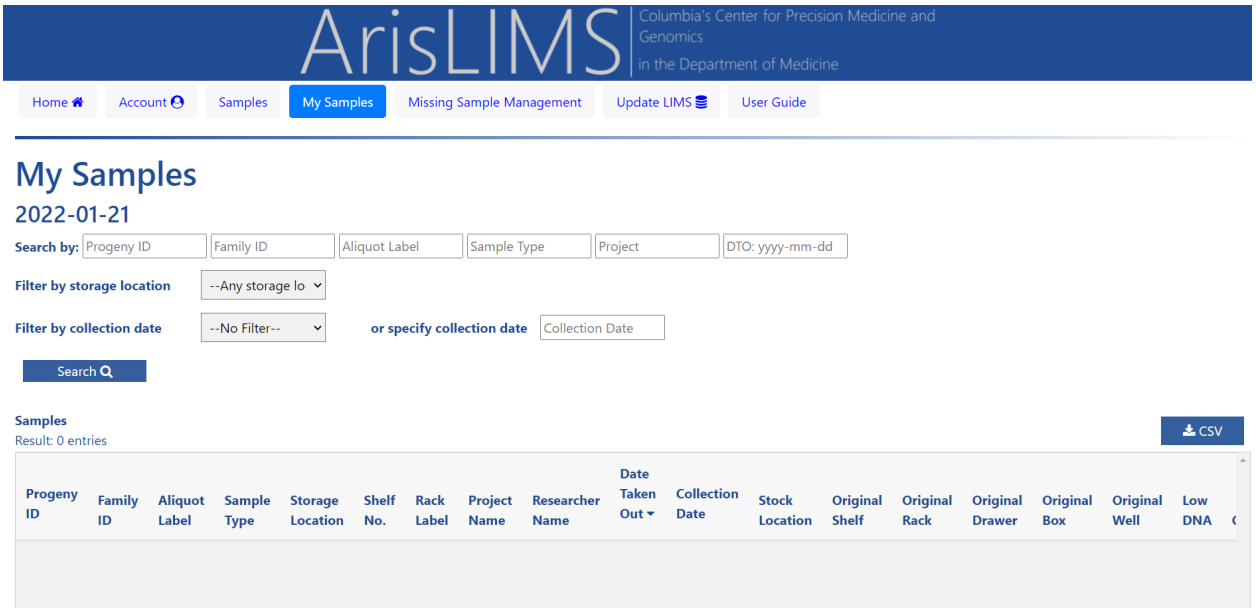


Figure S1e – Missing Sample Management tab

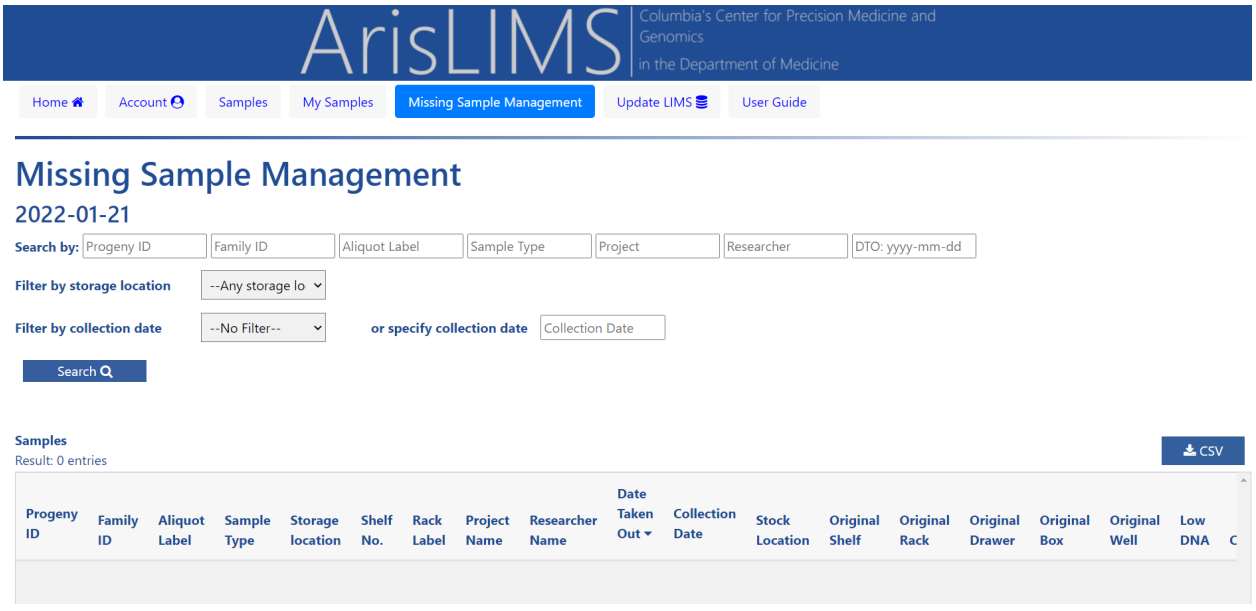


Figure S1f – Update LIMS tab

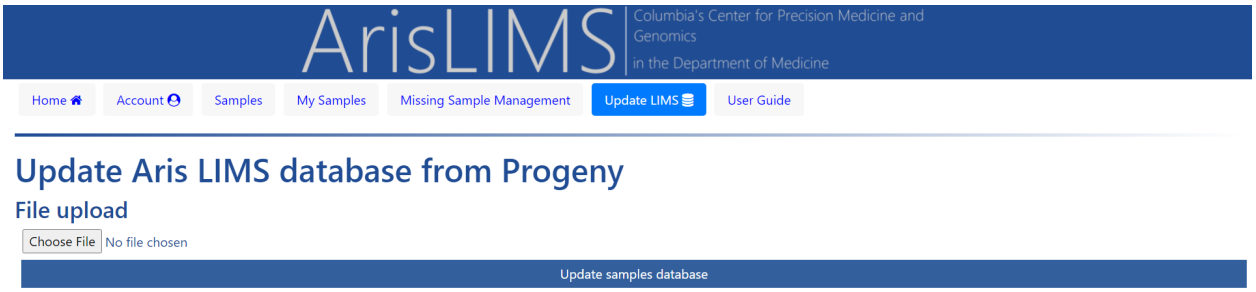
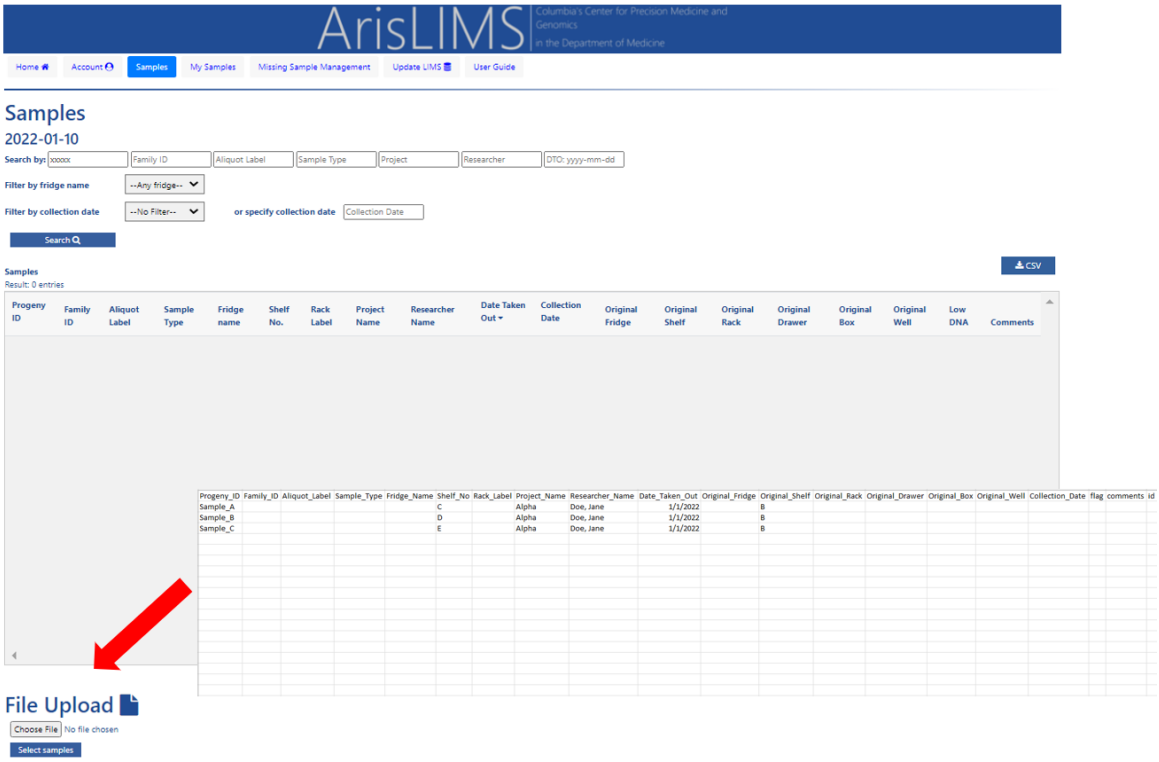


Figure S2: Log tracking sample location over time

```
2021-12-29 220320,520 INFO mod_history  DATA MODIFIED
User John Doe TOOK OUT SAMPLES
---------------
id,Progeny_ID
90,SAMPLE1
91,SAMPLE2
92,SAMPLE3
93,SAMPLE4
94,SAMPLE5
95,SAMPLE6
96,SAMPLE7
---------------
To temporary location
        Fridge A -80C   Shelf A Rack 1
Projects
        project1
############################################################
2021-12-29 220424,387 INFO mod_history  DATA MODIFIED
User Jane Doe RETURNED SAMPLES
---------------
id,Progeny_ID
114,SAMPLE50
115,SAMPLE51
116,SAMPLE52
---------------
Returned from temporary location
        Fridge B 4C     Shelf A Rack 1
Projects
        project2
############################################################
```

Figure S3: Batch query function



References

1. Anatskiy E, Ryan DP, Grüning BA, Arrigoni L, Manke T, Bönisch U. Parkour LIMS: high-quality sample preparation in next generation sequencing. Bioinformatics. 2019 Apr 15;35(8):1422-1424. doi: 10.1093/bioinformatics/bty820. PMID: 30239601.

2. Calabria, Andrea et al. “adLIMS: a customized open source software that allows bridging clinical and basic molecular research studies.” *BMC bioinformatics* vol. 16 Suppl 9,Suppl 9 (2015): S5. doi:10.1186/1471-2105-16-S9-S5

3. Craig T, Holland R, D’Amore R, Johnson JR, McCue HV, West A, Zulkower V, Tekotte H, Cai Y, Swan D, et al. (2017) Leaf LIMS: A flexible laboratory information management system with a synthetic biology focus. ACS Synth Biol 6: 2273–2280

4. Scholtalbers J, Rößler J, Sorn P, et al. Galaxy LIMS for next-generation sequencing. *Bioinformatics.* 2013;29:1233–4.

5. Marceddu, Giuseppe et al. “appMAGI: A complete laboratory information management system for clinical diagnostics.” *Acta bio-medica : Atenei Parmensis* vol. 91,13-S e2020015. 9 Nov. 2020, doi:10.23750/abm.v91i13-S.10521.