



**HAL**  
open science

# The backstage of the third Health Data Challenge: Introducing an Organizational Template and Custom, Temporary and Local Codabench Instances.

Nicolas Homberg, Magali Richard, Florent Chuffart

## ► To cite this version:

Nicolas Homberg, Magali Richard, Florent Chuffart. The backstage of the third Health Data Challenge: Introducing an Organizational Template and Custom, Temporary and Local Codabench Instances.. Jobim 2024, Jun 2024, Toulouse, France. 10.1016/j.patter.2022.100543 . hal-04664469

**HAL Id: hal-04664469**

**<https://hal.science/hal-04664469v1>**

Submitted on 30 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Backstage of the Third Health Data Challenge:



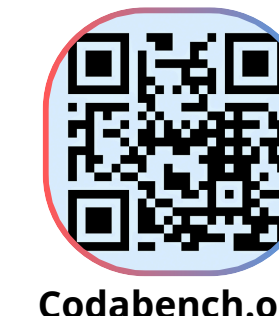
Introducing an Organizational Template and Custom, Temporary and Local Codabench Instances.



Nicolas Homberg<sup>a</sup>, Magali Richard<sup>a</sup> and Florent Chuffart<sup>b</sup>

<sup>a</sup>. Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France  
<sup>b</sup>. Univ. Grenoble Alpes, INSERM U1209, CNRS UMR5309, IAB, 38000 Grenoble, France

## Introduction to Codabench



Codabench.org

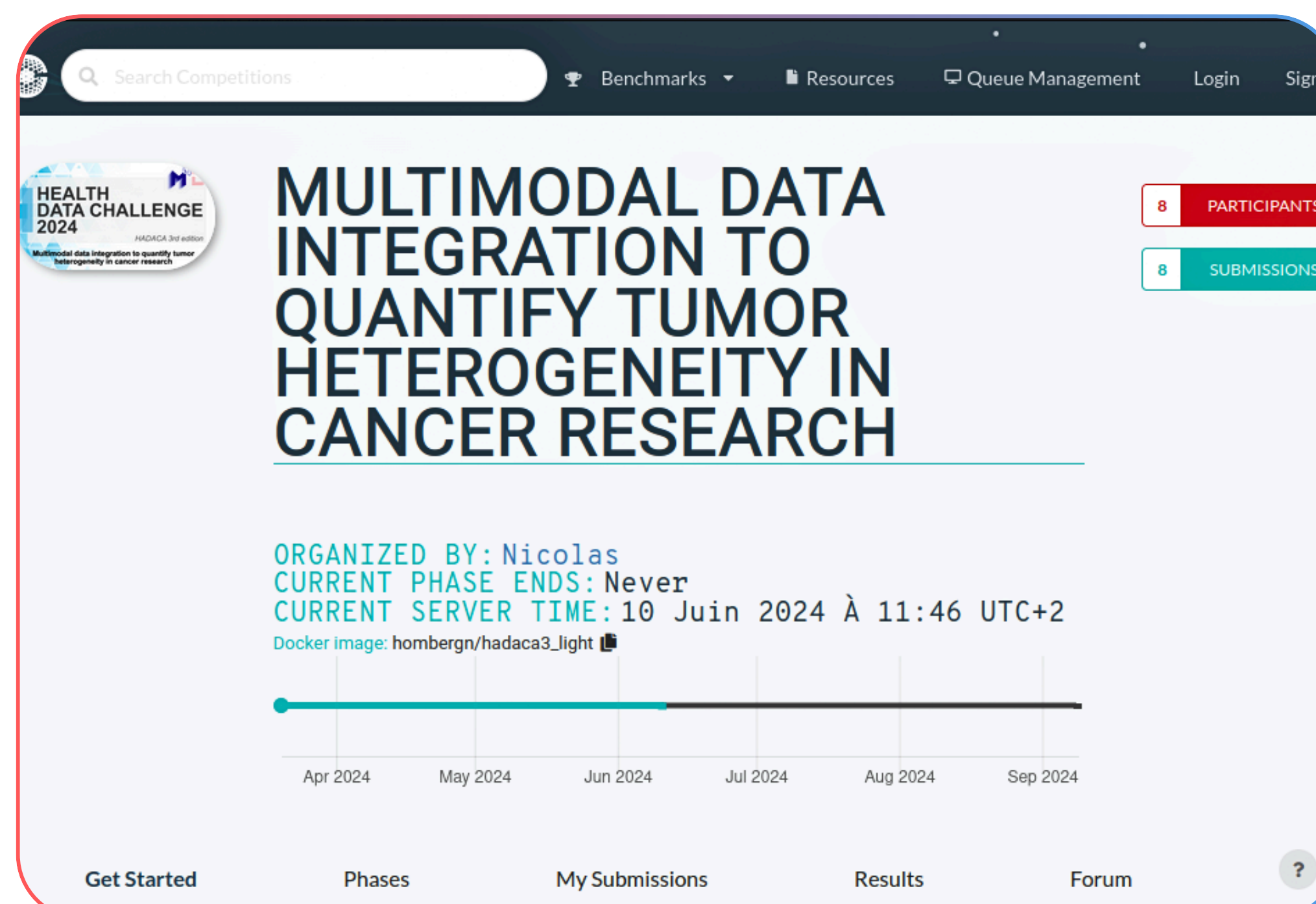
Codabench [1] is an open-source web platform that aims toward reproducible research and benchmarking. It represents the upgraded iteration of the competition platform **CodaLab**. Alongside EvalAI, Alcrowd, and Kaggle, CodaLab stands out as one of the leading platforms for academic competitions today [2].

### Codabench @LISN & Paris-saclay

- 4 computing servers, 4 storage servers and backup servers (main instance)
- 3 dozens of virtual machines (compute workers and pre-production)

### Codabench @UGA & Gricad

- **Custom-sized, temporary**, and local Codabench instances (compute workers and pre-production)

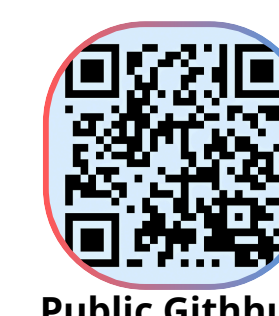


The beautiful and user-friendly **Codabench** interface.

### Features:

- Create a challenge from
  - a graphical User Interface
  - a zip bundle and a yaml recipe
- Submit a result or a function
- Rank submission in a leader board
- Set opening and ending dates
- Set different phases
  - with several tasks per phases
- Create and manage queues for
  - GPU workers
  - CPU workers

## Challenge Organizational Template

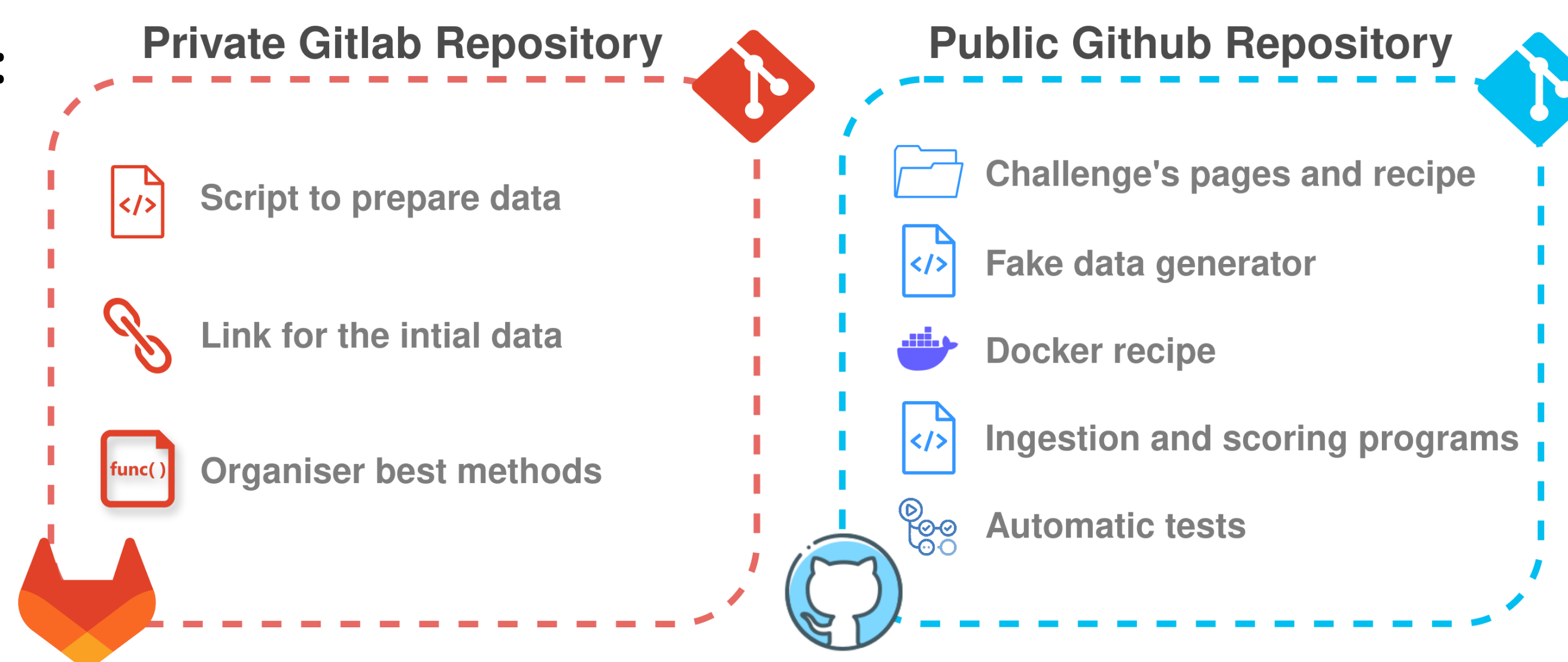


Public Github

Our organizational template relies on a bipartite git repositories devoid of binary files: zip, data files... This organization follows the Agile software development practices and guarantees ease of reproducibility and maintenance.

### The private Gitlab repository:

- Ensure fairness and confidentiality
- The recipe to prepare, clean and create target datasets
- Organiser best methods including easier methods to use as baseline methods.



This template enhances teamwork by Facilitating task distribution and using GitHub Issues to track problems and desired features.

### The public GitHub repository:

- Ensure reproducibility and fairness
- Create the challenge zip bundle
  - Challenge materials
  - Starting kit with **multiple baselines**
  - Ingestion and scoring programs
- Manage challenge's datasets
  - Simulate **a fake dataset**
  - Fetch **target dataset**
- Automatic tests of submissions to guarantee the challenge's health
  - Every push via GitHub Actions
  - Locally on the docker container

## Third Health Data Challenge-hadaca3: 2-6 Dec 24

Health Data Challenge (hadaca) aim to explore and potentially discover new methods to resolve cell-type deconvolution problems: An explanation of deconvolution problem is provided in the next poster ;)

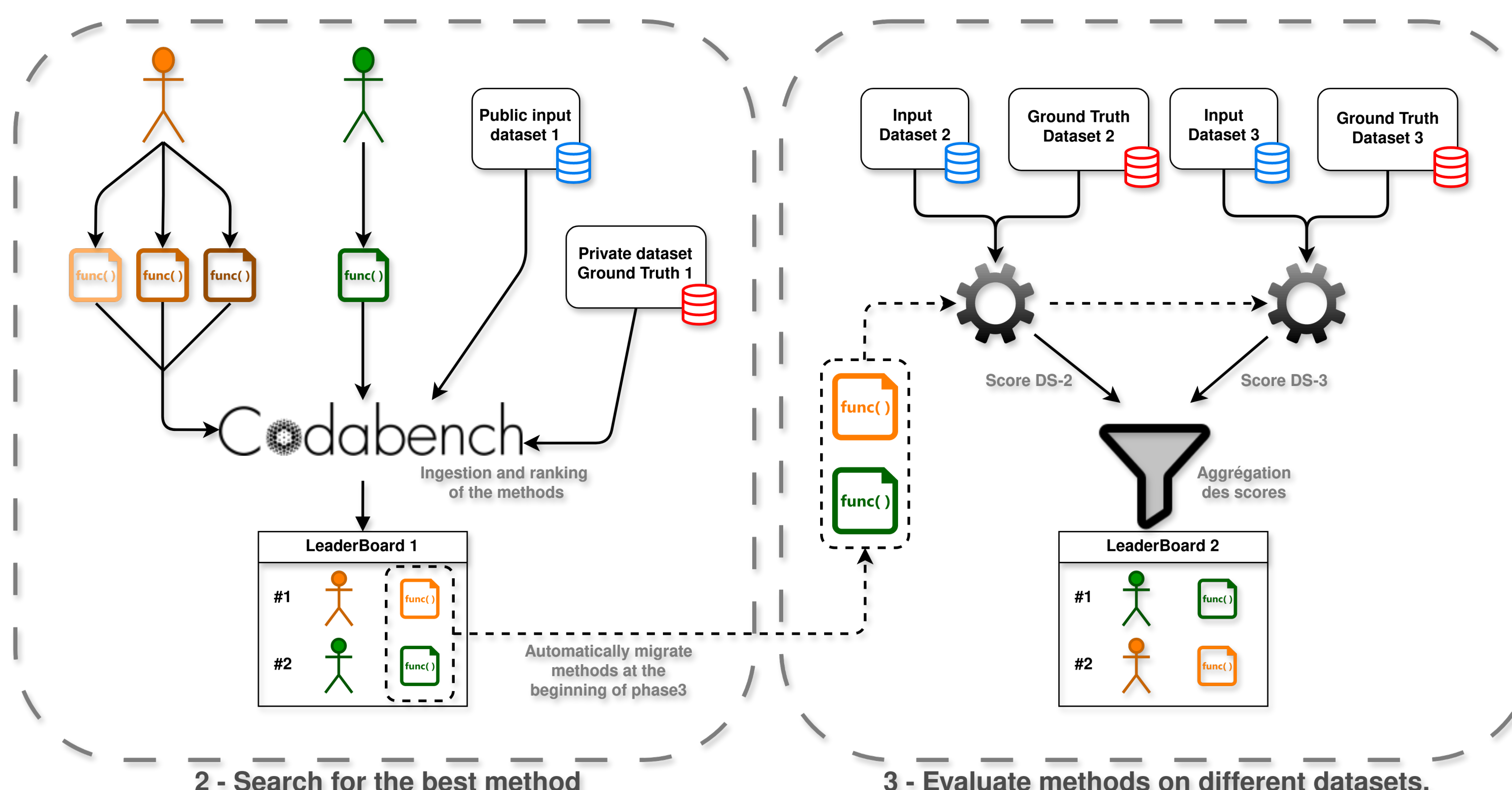
### Hadaca 3 Program:

#### Phase 1:

Toy deconvolution challenge to test the Codabench framework and familiarize with the platform and dataset.

#### Phase 2:

Estimation of cell type heterogeneity from pancreatic adenocarcinoma matching bulk methylomes and transcriptomes using multimodal references profiles, for five different cell types



#### Phase 3:

Auto-migration from phase 2 best methods and evaluation on previously unseen validation dataset.

#### Post challenge:

Aggregate the data, further trial the methods, and publish the results in collaboration with the challenge participants.



Hadaca3 website