Data Institute
Univ. Grenoble Alpes

Health Data Challenge
Data Institute Univ. Grenoble Alpes

A. Arnaud, C. Décamps, F. Petitprez, Y. Blum & M. Richard

November 2019

Create a Codalab competition

- **Codalab subscription**
  - **Sign in to :** **https://competitions.codalab.org**

- **Go to "My competitions" / "Competitions I'm Running"**

  - **Click on the "Create competition" button**

- **Bundle upload**

  - **Zip file with all the files required by Codalab**

    https://github.com/codalab/codalab-competitions/wiki

    https://github.com/codalab/codalab-competitions/wiki/ User_Building-a-Competition-Bundle

Bundle file structure

■ **Zip file with :**

- **Configuration file :**
  competition.yaml

- **Html files for web pages :**
  overview.html, evaluation.html, terms_and_conditions.html, data.html …

- **Image :**
  logo.png

- **Zip files for data :**
  data_public.zip, data_input.zip, data_reference.zip …

- **Zip files for programs :**
  ingestion_program.zip, scoring_program.zip …

- **Zip files for starting kit :**
  starting_kit.zip …

Configuration file : competition.yaml

- **Section 1: Competition properties:**
  - title: < String between double quotes>
  - description: < String between double quotes>
  - has_registration: <True/False. If true, organizer must approve participants>
- **Section 2 : Phase properties**
  - phasenumber:
  - label: <String. Label used to identify the phase.>
  - start_date: <YYYY-MM-DD, Phase start>
- **Section 3: Leaderboard properties**
  - leaderboards:
      Results: &RESULTS
        label: RESULTS
        rank: 1
  - columns:
      set1_score:
        leaderboard: *RESULTS
        label: Set 1
        numeric_format: 4
    https://github.com/codalab/codalab-competitions/wiki/Organizer_Codalab-competition-YAML-definition-language

Html file and logo

Data Institute
Univ. Grenoble Alpes

Health Data Challenge 2019

Organized by Alexis_Arnaud - Current server time: Nov. 29, 2019, 9:26 a.m. UTC

Previous	▶ Current	End
CHALLENGE #2 // Phase 1 // Exploration	CHALLENGE #2 // Phase 2 // Validation	Competition Ends
Nov. 27, 2019, 8 a.m. UTC	Nov. 28, 2019, 10 a.m. UTC	Dec. 1, 2019, midnight UTC

Learn the Details   Phases   Participate   Results   Public Submissions   Forums

Overview
Evaluation
Terms and Conditions

Overview

[!] The subscription is only open to people physically attending the challenge in Aussois [!]

Learn the Details   Phases   Participate   Results   Public Submissions   Forums

Data
Files
Submit / View Results

Data

The goal of the challenge is to quantify tumor heterogeneity (how many cell types are present? In which proportion?) accounting for confounding factors.

Data files

- **Public data : *data_public.zip***
  - Available from the competition web page

- **Input data : *data_input.zip***
  - Hidden data that is used by the ingestion program or submitted program

- **Reference data : *data_reference.zip***
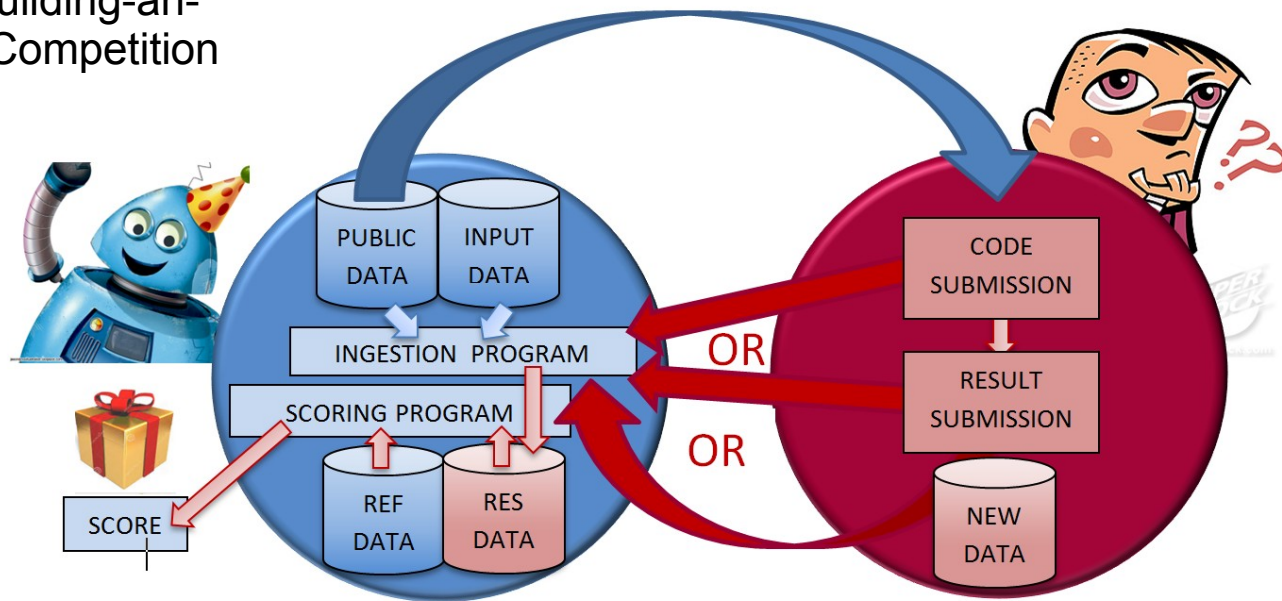 - Hidden data used only to compute the score

Programs

Ingestion program

- *ingestion.R*

- *metadata*   command: Rscript $ingestion_program/ingestion.R $input $output …

https://github.com/codalab/codalab-
competitions/wiki/User_Building-an-
Ingestion-Program-for-a-Competition

Scoring program

- *scoring.R*

- *metadata*

https://github.com/codalab/codalab-
competitions/wiki/User_Building-a-
Scoring-Program-for-a-Competition

Starting kit

■ **Zip file with a baseline method :**

- **Summary of the challenge :**
  starting_kit.html

- **Baseline method as a R markdown file with generation of the results (zip file) :**
  submission_script.Rmd

Download	Size (mb)	Phase
Starting Kit	0.270	#1 CHALLENGE #1
Public Data	173.181	#1 CHALLENGE #1
Starting Kit	0.270	#2 CHALLENGE #2 // Phase 1 // Exploration
Public Data	173.198	#2 CHALLENGE #2 // Phase 1 // Exploration

Bundle generator *(work in progress)*

- **R script to generate a full Codalab bundle**

  - based on input of the user from command line

- **GitLab repository of UGA**

  - **https://gricad-gitlab.univ-grenoble-alpes.fr/codalab/bundle_generation**