

# EUROPEAN RESEARCH COUNCIL (ERC): ERC DMP

## SUMMARY

Project Acronym: OpenClick  
Project Number

## DATASET SUMMARY

OpenClick project aims to discover physical status, skills and abilities of people based on human – computer interaction tests. There are currently 9 different sets of data within the OpenClick project. Each set of data corresponds to a single test that exists within the project.

Testing tools measure specific user performance in human – computer interaction. Each test is designed to measure one, exactly defined, type of interaction. During the test, several samples are taken, the average value is calculated, and then this average value is recorded in the database as a result of the test for specific person.

Recorded data sets are stored in text files, in the CSV (comma separated value) format.

Common data for all sets is user data. For each user, gender, the year of birth, dominant hand and identification number are given. This data is the same in all datasets, which makes it possible to link the results.

In addition to the common data, each dataset contains the test results. These mostly refer to the time needed to perform an action related to the use of a graphical user interface (for example, the time required to select the mouse button, the time required to select an auxiliary key from a keyboard, etc.).

The data sets will be updated on a monthly basis.

The size of data files will depend on the number of users who perform the tests.

## FAIR DATA AND RESOURCES

### 1. MAKING DATA FINDABLE

The data will be described using the following metadata:

General metadata

Title of the project – Test results

Description – Data generated as a result of performing tests on the OpenClick portal.

Research topic – Analysis of the use of graphical user interface

Publisher – Faculty of Mechanical Engineering, Nis

Creation date

Location information – Data is collected in Serbia (the city of Niš), but the tests will be performed by people from all over the world.

Software used – Web application, available online

Access or rights policies – Data is available on the basis of Apache License, version 2.0

Each of the 9 datasets contains the following metadata:

The address from which the data set is to be downloaded

Title

Description

Originator

Contact person

Keyword

Distribution

Data files downloaded are automatically created using the software. Dragan Mišić will be responsible for putting them on the portal.

File names will be chosen so that they reflect the test from which these results come from, but also the date of origin. Based on the date of creation, it will also be possible to determine the versions of the data files.

File names will be in the format:

OpenClick\_TestName\_YYYY.dd.MM.csv

## **2. MAKING DATA OPENLY ACCESSIBLE**

All data are openly available. The data will be located on the OpenClick Web portal ([www.openclick.rs](http://www.openclick.rs))  
The files with the test results will be successively uploaded on the portal. Setting up new files will be done once a month.

No special tools are required to access the data, because the data are in text files (csv).

In the later phase of the project, consideration will be given to setting up data in one of the public repositories, such as Zenodo.

All data on users in the database are anonymous and will be published as such. The user is identified by a randomly assigned identifier. This identifier only serves to connect different results (data) with a particular person, which would facilitate data analysis.

## **3. MAKING DATA INTEROPERABLE**

In order to enable interoperability between applications, each data set will be described in accordance with the DCAT-AP standard. JSON-LD script will be placed on each page. This enables interoperability, but also improves the ability to find information by the search engine.

As mentioned, the data will be available through the OpenClick portal.

The idea behind the creation of such data is to enable data analysis (primarily by different methods of machine learning). The format most commonly used in the application of machine learning is the CSV file and is therefore chosen as the format for storing data from the project.

In addition to unlabeled data in the project, it will also be possible to create tagged data, for example, if a diagnosis is made for the user. In this way, the data may be able to be used to assist in the diagnosis or monitoring of the patient's recovery.

## **4. INCREASE DATA REUSE**

The data will be available throughout the project, and later. The data will be available immediately after their collection and processing.

Data can be downloaded in accordance with APACHE LICENSE, VERSION 2.0.

In order to ensure the quality of data, prior to their publication, an analysis of the obtained test results will be carried out. Tests with detected irregularities will be discarded.

## **5. ALLOCATION OF RESOURCES AND DATA SECURITY**

The portal on which the data is stored will remain available after the official completion of the project. The team from the Faculty of Mechanical Engineering in Niš will take care of the maintenance of the portal.

The data is stored in the database, so the daily backup of the database is planned. For security reasons, backup data will be stored in multiple locations.