Data Catalogue

Web scrapped data on job vacancies

This dataset consists of 7.4 Mill job vacancies scrapped from job portals in 5 countries (CZ, DE, IE, IT and UK) for the period from June 2015 to April 2016. This data was scrapped by CEDEFOP, the European Centre for the Development of Vocational Training.

More information can be found in the following presentation:
http://www.cedefop.europa.eu/files/webadsanalysis_02_approach_mezzanzanica_and_cesarini_0.pdf

EU-LFS, EU-SILC public use files

The EU Labour Force Survey (EU-LFS) and the EU Statistics on Income and Living Conditions (EU-SILC) are the two main European social surveys. The first one is an individuals' survey which collects data on the employment situation and related issues of the general population, including occupations. The second one is an individuals' survey which collects data on the income and other social characteristics of the population.

Early this year public use files of the EU-LFS and EU-SILC of 5 countries (DE, FI, HU, NL and SI) was released.

You can find more information on the EU-LFS here:
http://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey

and on the EU-SILC here:
http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions

PIAAC

The Programme for the International Assessment of Adult Competencies (PIAAC) is an international survey of adult skills launched as an initiative of the OECD. It collects data about skills in literacy, numeracy, and problem solving in technology-rich environments (PS-TRE) among adults between the ages of 16 and 65 in over 40 countries. Data from PIAAC provide a measure of cognitive and workplace skills, a rich evidence base for policy-relevant analysis, a better measure of a country's "stock" of skills, insights into whether education and training programs are focusing on the right competencies, and the capacity to compare skills across a broad sampling of countries around the world.

You can find more information on PIAAC here:
http://www.oecd.org/skills/piaac/

EURES

EURES (European Employment Services) is a cooperation network formed by public employment services. Trade unions and employers' organisations also participate as partners. The objective of the EURES network is to facilitate the free movement of workers within the European Economic Area (EEA) (the 28 members of the European Union, plus Norway, Liechtenstein and Iceland) and Switzerland.

The EURES data comprises two datasets, one on job vacancies (35 million) proposed by potential employers in European Public Employment Services, and another on curriculum vitae (4.7 million) proposed by jobseekers.

You can find more information on EURES here:
https://ec.europa.eu/eures/

ESCO

ESCO is the multilingual classification of European Skills, Competences, Qualifications and Occupations. It systematically shows the relationships between the different concepts. It provides one way to link the other datasets. In particular it allows linking occupations in the EU-LFS and EU-SILC with skills from web scrapped job vacancies, EURES and PIAAC.

You can find more information on EURES here:
https://ec.europa.eu/esco/portal/home

Additional datasets brought by the teams

The teams can use other datasets to build their prototypes. The only constraint is that they need to be real datasets (or synthetic datasets constructed from real ones for confidentially protection reasons) and not be so specific that it only applies to one country. In case of doubt you can contact the Hackathon organisation.

One potential candidate is Eurostat online database. It can be accessed in an automated way via API. You can find more information in the following pages.

SDMX format: http://ec.europa.eu/eurostat/web/sdmx-web-services/rest-sdmx-2.1

JSON format: http://ec.europa.eu/eurostat/web/json-and-unicode-web-services

How to obtain the datasets

The datasets are available via SFTP from the server *fts.ec.europa.eu* to which you can connect using an FTP client, for example FileZilla available on GNU/Linux, Windows and Mac OS.

In order to access the datasets via SFTP, you need to authenticate. You'll receive the login details after all members of the team have completed, signed and returned to the Hackathon organisation the agreement concerning the use of the database on scraped vacancies.