



EUROPEAN COMMISSION
EUROSTAT

Deputy Director-General
Task Force Big Data

 Ref. Ares(2018)6073319 - 27/11/2018

Description of the European Big Data Hackathon 2019

Description of the European Big Data Hackathon

What is a hackathon?

Data hackathons, also known as data dives, are intense events where teams of data scientists, computer programmers, graphics and interface designers as well as project managers creatively tackle data problems and prototyped data analytics products. Hackathons typically last between one day and one week. Some data hackathons are intended simply for educational or social purposes, although in many cases the goal is to create usable data analytics products. Data hackathons tend to have a particular focus, which can include specific data sources, methodologies, technologies and applications; in other cases there is no restriction on the type of data analytics product being created.

What is the purpose of the European Big Data Hackathon?

The European Big Data Hackathon has three main objectives:

- to solve statistical problems by leveraging algorithms and available data, by engaging with developers and data scientists across Europe, giving them the possibility to work with relevant data sets in order to generate new ideas and potentially contrive novel algorithms;
- to produce innovative products, including visualisation tools, developing prototypes that official statistics will be able to integrate at European and national level;
- to promote partnerships with the research community and the private sector, by raising awareness about big data initiatives in Official Statistics in Europe.

Teams

National Statistical Institutes will be responsible to nominate, or select, teams to compete in this hackathon. Each team will be composed of three members, who may be part of the staff of the Statistical Office, academia, students or faculty of EMOS certified programs, or members of the national data science community (or partnership of several of these). We foresee a limit of 32 teams, one for each Member State of the European Statistical System. Additional teams, proposed by NSIs in the 2nd place, will be considered if the limit of 32 teams is not reached¹. The final list of competing teams and their members will then be announced.

Hackathon

Teams will compete for the best application of big data providing an answer to a pressing European policy or statistical. The policy domain and a catalogue of datasets will be provided in advance, while the specific policy question will be announced right before the beginning of the hackathon. The policy domain will be selected among the 10 priorities of the European Commission². After this selection, teams can prepare for the hackathon by

¹ For the 2017 European Big Data Hackathon, 22 teams were nominated.

² https://ec.europa.eu/priorities/index_en

collecting information about the policy domain announced and prepare their own datasets which they can bring to the hackathon under certain conditions (see below).

The hackathon will take place during three days in Brussels, back-to-back with the NTTS conference³.

During the hackathon, the teams will work on the creation of a data product's prototype, meant to be useful to the policy analysts working on the policy question announced.

At the beginning of the hackathon, an introduction to the policy question will be provided to the participants. The teams will then work for about two days developing their prototype. At the end of the hacking time, each team will have 5 minutes to present their data product. After the presentations, a panel of evaluators will evaluate the best proposed data products. The hackathon will finish with the announcement of the winner and five runners-up.

The winner and the first two runners-up will finally present their prototypes at a plenary session of the NTTS 2019 conference.

Data sources

The applications will have to use the provided big data source(s), and they should ideally also use European official statistics.

The main big data source will be the participants themselves. Eurostat will provide access to a framework for the collection of geo-coordinates and sensors data collected by smart devices. The framework will be composed of an APP (most probably for the Android operating system only) which will capture the data, a remote server (to where the data will be securely transmitted) and possibly some predicted variables (e.g. mode of transportation used). Eurostat may provide more than one alternative framework which would differ in a few aspects (e.g. predicted variables) and which the teams can use (and potentially combine).

The participants, and/or other volunteers they may recruit, will have to use their own devices and install the APP. The data will be transmitted securely from the device to the remote server, where it will be stored separately from the personal identification data gathered for the registration in a unique dataset. Finally, the anonymised dataset will be made available to all the teams. Every team should contribute with the data of at least three individuals for at least a period of one month.

Additionally, Eurostat will provide access to anonymised microdata of the Harmonised European Time Use Survey.

Eurostat will provide a catalogue of datasets consisting of a description of the datasets and instructions on how they can be accessed.

Teams are allowed to use additional data sources, which they may prepare in advance. These additional data sources should have an international applicability, i.e. they either cover all EU member-states or they are national data sources but may be available in

³ https://ec.europa.eu/eurostat/cros/NTTS2019_en

several member states. They should also allow the particular data product to be possibly released as a policy tool at European level.

Data product

The teams are free to propose any type of data products. The only conditions are that they need to help answer the policy question or address the statistical challenge and they need to include a front end (i.e. visualisation) which allows the policy analyst and /or statisticians explore the answers provided by the data product.

The prototype can be the subject of a scientific paper which will have the same publication possibilities as the ones available to the participants presenting at the NTTTS 2019.

IT infrastructure

The European Commission will assure internet connection and power plugs in the room where the teams will hack, as well as access to a cloud big data infrastructure. However, participants will need to bring their own laptops / portable computers.

Evaluation criteria

The panel of evaluators will evaluate the proposals of the teams based on the following criteria:

- *Relevance*: the extent to which the data product provides an effective answer to the policy question;
- *Methodological soundness*: the extent to which the data product provides an unbiased answer to the policy question;
- *Communication*: the ability of the product to provide the information in a clear and user-friendly way, including through data visualisation;
- *Innovative approach*: the extent to which the data product introduces or uses new ideas or methods;
- *Replicability*: the extent to which the data product can be implemented in several EU Member States.

Panel of evaluators

An independent panel of evaluators will be responsible for the evaluation of the data products proposed by the competing teams.

The panel will consist of two sub-panels, each composed of 5 to 10 persons having no links with the NSIs which nominated the teams. The statistical/scientific sub-panel will include individuals from both the statistical domain and the policy making domain. The industry sub-panel will include representatives from the sponsors.

Funding

The European Commission will reimburse the travel and accommodation cost for one team per country. The cost of participation of additional teams selected in the second phase will not be reimbursed by the European Commission and the NSI which nominated the team will need to secure its funding.

The cost reimbursed will include the return trip to Brussels and a maximum of 4 hotel nights in Brussels. Meals will be provided to the participants at the venue.

Sponsors

Organisations from the big data industry sponsor the Hackathon by fulfilling several roles:

- Provide the teams with access to their tools;
- Provide coaching before and during the event;
- Participate in the evaluation panel;
- Distribute small gadgets to the participants;
- Possibly support further development of prototypes;

Prize

Prizes are foreseen for the members of the winning team and five runner-ups. The prizes will be announced at a later point in time.

Venue

The hackathon will take place at the same time as the NTTS in Brussels during three days and a half, in a venue nearby the Charlemagne building where the NTTS will take place.

Follow-up

In the third day of the Hackathon, after the winners' announcement, all the participants of the Hackathon will gather for a session of networking and discussion on possible synergies between the prototypes and future collaboration.

Calendar

Registration of teams by NSIs	14 December 2018
Announcement of final list of competing teams	31 December 2018
Release of data catalogue	31 January 2019
Hackathon	8-12 March 2019