



EUROPEAN COMMISSION
EUROSTAT

Deputy Director-General
Task Force Big Data

Description of the European Big Data Hackathon

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, graphic and interface designers and project managers try to creatively tackle data problems and prototype data analytics products. Hackathons typically last between a day and a 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 specific focus, which can include specific data sources, methodologies, technologies and applications but 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 five objectives.

- Solve statistical problems by leveraging algorithms and available data, by engaging with developers and data scientist across Europe to build relevant algorithms and by exposing relevant data sets to participants to play around and come up with ideas.
- Identify best of class European data scientists, by targetting local developers' and data scientists' communities to attend the events, by connecting with the best participants during and after the events for possible future collaboration, and by creating networks for capacity building between statisticians and data scientists.
- Promote and accelerate big data for statistics initiatives in Europe, by developing prototypes that European countries will be able to integrate, and by generating buzz about big data within local communities of scientists entrepreneurs.
- Promote partnerships with research community and private sector, by raising awareness about big data initiatives in Official Statistics in Europe, by starting partnerships with the private sector and Universities.
- To produce innovative products and tools, including visualisation, to stimulate the use of open data and public use files and to engage with new audiences, thereby contributing to the objectives of the DIGICOM project.

Teams

National Statistical Institutes will be the ones responsible to nominate, or select, teams to compete in this hackathon. Each team will be composed of three members who may be staff of the statistical office, members of the national data science community or a partnership of both. In an initial phase, running until 7th of November, NSIs can nominate one (or two) team(s). We foresee a limit of 32 teams, one for each country 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 reached. The final list of competing teams will be announced by 9 January 2017.

Hackathon

Teams will compete for the best application of big data providing an answer to a pressing policy challenge facing Europe. The policy domain and a catalogue of datasets will be provided in advance and the specific policy question will be announced at the beginning of the hackathon. The policy domain will be selected from among the 10 priorities of the European Commission¹. After this selection, foreseen to take place in January 2017, teams can prepare for the hackathon by 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, at the same time as the NTTS 2017 conference².

At the hackathon, the teams will work on the creation of the prototype of a data product directed to 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 around two days on their prototype. After the hacking time, each team will have 5 minutes to present their data product. After the presentations, and a panel of evaluators will decide on the best data products. The hackathon will finish with the announcement of the winner and two runners-up.

The winner and the two runners-up will finally present their prototypes at a plenary session of the the NTTS 2017 conference.

Data sources

The applications have necessarily to use a big data source and should ideally also use European official statistics.

Eurostat will provide a catalogue of datasets including Eurostat online statistical database, the open data portal, some open big data sources and possibly private big data sources made available specifically for the hackathon. The catalogue will consist of a description of the datasets and instructions on how they can be accessed.

Teams can 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 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 and they need to include a front end (i.e. visualisation) which allows the policy analyst explore the answers provided by the data product.

¹ https://ec.europa.eu/priorities/index_en

² <http://www.NTTS2017.eu>

The prototype can be the subject of a scientific paper which, like in the case of NTTS papers, can be submitted for consideration for a special NTTS issue of the Journal of Official Statistics³, as well as a special NTTS section of the Eurostat Review on National Accounts and Macroeconomic Indicators⁴.

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 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.

It will be composed of between 5 and 10 persons with no links to the NSIs which nominated the teams and coming from both the statistical domain and the policy making domain.

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.

³ <http://ec.europa.eu/eurostat/cros/content/special-ntts-2017-issue-jos>

⁴ <http://ec.europa.eu/eurostat/cros/content/special-ntts-2017-section-eurona>

Prize

Prizes are foreseen for the members of the winning team and two 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 2017 in Brussels during three days, in a venue nearby the Charlemagne building where the NTTS will take place.

Calendar

| | |
|---|------------------|
| Nomination of teams by NSIs | 7 November 2016 |
| Announcement of final list of competing teams | 9 January 2017 |
| Announcement of policy domain and release of data catalogue | 27 January 2017 |
| Hackthon | 13-15 March 2017 |