

SoBigData research infrastructure

The workshop will present an overview of the important aspects a community must take into consideration in order to design and implement a distributed, pan-European, multi-disciplinary research infrastructure for big social data analytics such as SoBigData RI (www.sobigdata.eu).

This RI is the result of a first project called SoBigData ended in 2019 and the base of the newest SoBigData++ started in January 2020 with the objective to consolidate and enrich the platform for the design and execution of large-scale social mining experiments accessible seamlessly on computational resources from the European Open Science Cloud (EOSC) and on supercomputing facilities. SoBigData++ integrates a community of 31 key excellence centres at a European in Big Data analytics and social mining. We will use this experience to present in the workshop the following key aspects:

Data Science, Multidisciplinary & AI.

We believe that the necessary starting point to tackle the challenges is to observe how our society works, and the big data originating from the digital breadcrumbs of human activities offer a huge opportunity to scrutinize the ground truth of individual and collective behaviour at an unprecedented detail and at a global scale. Ethics & Privacy.

There is an urgency to develop strategies that allow the coexistence between the protection of personal information and fundamental human rights together with the safe usage of information for scientific purposes by different stakeholders with diverse levels of knowledge and needs. There is a need to democratise the benefits of data science and Big Data within an ethical responsibility framework that harmonizes individual rights and collective interest.

Training the next generation of Data Scientists for Social Goods

There is an urgency to thoroughly exploit this opportunity for scientific advancement and social good as currently the predominant exploitation of Big Data revolves around either commercial purposes (such as profiling and behavioural advertising) or –worse– social control and surveillance. The main obstacle towards the exploitation of Big Data for scientific advancement and social good –besides the scarcity of data scientists– is the absence of a large-scale, open ecosystem where Big Data and social mining research can be carried out.

TagMe: A success story of how the integration boosted a research result

SoBigData has succeeded in boosting the usage of some services it offers, with highly relevant peaks of daily accesses, for example in the TagMe system for automated semantic annotation of short texts. The future direction is to develop more tools that are similarly simple and effective to use.

We will include in the workshop a round table moment with the participants going in details and open discussions about the presented aspects.

Primary authors: Mr PAPAPLARDO, Luca (CNR); Mrs PRATESI, Francesca (CNR); Mr COTÈ, Mark (CNR); FERRAGINA, Paolo (UNIFI)

Presenters: Mr PAPAPLARDO, Luca (CNR); Mrs PRATESI, Francesca (CNR); Mr COTÈ, Mark (CNR); FERRAGINA, Paolo (UNIFI); Mr TRASARTI, Roberto (CNR)

Session Classification: The SoBigData Research Infrastructure