O.N.E Data - Going Smart Data Project

Purpose

The project aims at setting the ground for OECD to progressively mainstream smart data practices into the OECD regular data lifecycle. It leverages dozens of experimental projects carried out in all policy domains over the 2016-17 period, that illustrate the great potential of smart data, in order to develop new, more granular, more timely and relevant evidence to inform policy decisions (see the reference document reviewed by DSWG in Jan 2017: OECD Smart Data Framework).

Objectives and outputs

For precise project description see the O.N.E Data Going Smart Data project scoping paper.

Main Developments for 2018

General aspects:

Develop smart data practices leveraging the Smart Data Sandbox, as the reference data science platform for statisticians and analysts to easily experiment with data (exploring voluminous micro-data or satellite data, running complex regressions or machine learning algorithms…) in a flexible, scalable and multidisciplinary way: Delivered through co-funding and collaboration between EXD/DKI and substantive directorates. In 2018: 1) Set-up of the OECD Algorithm bank leveraging promising open source technologies; 2) Data science assistance to directorates: Build capacity across OECD data teams for using the Sandbox efficiently through assistance to algorithm optimisation (e.g. use of distributed computing, code refactoring) and migration to new coding open source technologies (R and Python) leveraging the Algorithm bank. This assistance, if organised globally and in a synchronised way, will enable an accelerated pace of adoption and benefits. 3) Sandbox consolidation: Acquire hardware required for archiving and optimise current infrastructure. The opening of the platform to external expert is postponed to 2019+, pending more resources. 4) Data science platform pilot in: following the outcome of the Data Services & Solutions CfT, acquire licenses required for and implement the data science platform pilot, and carry out a first set of experimentations. 5) Animation of smart data communities of practitioners: Organisation of events, sharing of good practices and experiences, through engagement with 5 communities of practice by end of 2018 involving OECD staff and external experts. 6) In parallel to these actions, based on the outcomes of the CfT, EXD/DKI will engage in 2018 with directorates, mobilising the newly built smart data ecosystem, in order to propose scopes of work leveraging smart data to potential funders, and raise VCs for the next biennium – and possibly fill the wide gap in investment in the area of smart data.