UNECE High-level Group for the Modernisation of Official Statistics

Project Proposal: Data Science Lab

Prepared by the Blue Skies Thinking Network / Andrew O'Sullivan, Barteld Braaksma, Juan Muñoz, Taeke Anton Gjaltema

1 Purpose

To create an expert group and a virtual collaboration community to facilitate the analysis of data, the support for emergent sources of information, and the introduction of new technologies to modernise the statistical processes and the products and services derived.

2 Project description

The main goal of the project is to create an expert working group supported by a virtual collaborative Internet platform to develop, integrate, and share knowledge about emergent sources of information and new technologies to improve the production and outputs (products and services) of official statistics.

The project outputs will be:

1. An experts group to share knowledge, practices and efforts to promote the development of methodologies, techniques and technologies to improve the statistical processes, products, and services based on the out-comings from the data science fields and the integration of traditional and non- traditional statistical and geographical data sources.
2. A systematized library of research publications, documented experiences, data resources, video tutorials and algorithms developed in collaboration by the statistical community.
3. A virtual space to collaborate on data science projects where the parties can share efforts to develop solutions, training sets, provide feedback, etc.
4. A virtual community. Its initial configuration can be implemented as a special section of the WikiStats using the collaboration tools already in use by the HLG-MOS. In a more evolved model, this community might grow to become a federated network sharing hardware and software resources from participating institutions and from other efforts made by other international organisations like UNSD and Eurostat.

3 Alternatives considered

1. UNSD Global Platform. The project of the Global Platform coordinated by UNSD has similar goals and when developed might have connections for the exchange and to establish collaborative projects on a broader way.
2. Doing nothing. If the project does not go on there will still be opportunities to collaborate but these efforts will be still being something disperse and information will still being difficult to find.

4 Expected Benefits

☒	Reduced costs

☒	Increased efficiency
☐	Reduced risks
☒	New capabilities to meet user needs

5 Which key priorities in the HLG-MOS Strategic Framework does the proposed project relate to?

☒	Take cost out of our organizations to reinvest in more value-added areas
☒	Explore new areas collectively and leverage each other's' research investments in specific areas
☐	Provide whole of government data ecosystems based on international standards, for better estimates in key policy areas
☐	Renew our governance and operating processes

Justification:

Statistical Offices have already made some investments to develop knowledge, methodologies and technologies in the field of data science to improve its process and the products and services they offer.

Under the HLG-MOS umbrella, the statistical community have been working in some related projects to explore the Use of Big Data for the production of statistics, the integration of different traditional and non-traditional sources of statistics, the big data Sandbox, and more recently, the machine learning development project.

Although those projects have already delivered some value, the real potential of them has not been reached. The need still exists and there are many more applications to apply that may drive to modernise the official statistics.

6 How does the proposed project relate to other activities under the HLG-MOS?

This project will consolidate and provide continuity to efforts that the HLG-MOS have been developing in the following areas:

- Use of Big Data for the production of statistics
- Integration of different sources of statistics and data
- The Sandbox
- The machine learning development project (HLG-MOS ML Project)

As some of the products of the project will have some software as output, it might be seen as related to the efforts made by the Sharing Tools Group. However, as the scope is broader, the project is considered as the foundation of an evolution of this group with the participation not only of ICT people but also of statisticians and Data Scientist on formation in the different statistical offices.

The project intention is not to replace other efforts like the one made by UNSD neither Eurostat's one but to develop the field to improve the modernisation of official statistics while creating links to conform a federated experts network and share results potentializing the value of the outputs from different projects.

7 Proposed timetable

The first stage of the project will be to constitute the group and to establish the structure of the community; this task may take the first quarter of 2020. The participation of experts on information and wikki management, together with methodologists, data scientists, statisticians and ICT people will be fundamental to develop a useful working and sharing platform.

The instrumentation of this virtual space is a good opportunity to apply the CSPA and CSDA principles and identifying capabilities needed to develop systems focused on sharing information and software solutions.

While the virtual space is being organised and instrumented, other members of the group will be working on collecting information about prior or existing efforts developed in the data science field that may be related to the production of official statistics. In this way, the Wiki will start providing services very early.

During the rest of the year, the working group will be in charge of developing a more in-depth working program, compiling more information to constitute a rich data-science library, improving the collaborative working space, and working on agreements with statistical offices and international organisations to constitute a federated working network.

Participation of this group in the HLG-MOS workshops will bring activities that stimulate thinking as a community that shares things from the start.

8 Expected resources and costs

The following are the expected resources required for 2020:

- Experts on data science, statistical methodologies, statistics, geography, information management and information technologies
- Space in the UNECE's Wiki to facilitate the community
- Virtual conference facilities to support monthly meetings
- Resources to organise a physical sprint
- Resources to organise a physical workshop
- Support from the UNECE's HLG-MOS Secretariat to help in the coordination of the group