

# July-August 2020

## Machine Learning

### Progress

- Two project members (Australia and Mexico) were panelists at a webinar on "ADDING VALUE TO STATISTICAL DATA PRODUCTION THROUGH MACHINE LEARNING: POTENTIAL USE OF MACHINE LEARNING FOR OFFICIAL STATISTICS" organized by the UN World Data Forum. Links to the recording of the webinar and the papers were shared to all members. Questions, answers, comments and references provided by participants in the ChatBox were also made available on the ML wiki.
- The pilot study reports have been updated and are being edited for public release.
- Draft theme reports on C&C, E&I and Imagery have been received and are being reviewed.
- Statistics Poland shared a very comprehensive library of ML methods to code product descriptions to ECOICOP (refer to: [https://github.com/statisticspoland/ecoicop\\_classification](https://github.com/statisticspoland/ecoicop_classification)). Their updated pilot study report describes the process of testing machine learning methods in products classification and all the steps of their study.
- Work continued on describing a use of the ML methods shared by Statistics Poland to code product descriptions also shared by Statistics Poland and translated in English and French.
- Replies from ABS, ONS and Statistics Flanders on questions related to the integration of ML (WP3) were shared with project members to seek their opinions and further insights on their respective experiences.



### Next Steps

- Project members (UK, Poland, USA) will participate in the following events:
  - UNSD Big Data conference <https://unstats.un.org/unsd/bigdata/conferences/2020/>. Three members will present "HLG-MOS Machine Learning Project: Sharing ML Techniques and Algorithms, how to tackle large data sets and build international capability" at a session on "Training in use of new data sources and new technologies"
  - BigSurv20 <https://www.bigsurv20.org/> (invited session)
- Release pilot study reports to the public
- Organize a virtual project meeting at the beginning of October
  - Presentations on pilot studies initiated in 2020
  - Updates to pilot studies initiated in 2019
  - Presentations on WP1 theme reports, WP2 Quality and WP3 Integration
  - Presentations on recent developments in ML, e.g. new studies, organisations, etc.
  - Discussion on ML group activities in 2021
- Organize an open event to communicate and share the outputs of the ML project, It is proposed to hold this event with the HLG-MOS November workshop

## Risks and Issues

Issue	Mitigation

## Input Privacy-preserving Techniques

### Progress

The project started 1 of July and now has 20 participants from Istat, Stats Canada, Statistics Netherlands, ONS and Eurostat. Members are presenting use cases and work taking place in their offices. The group is using this as a basis to identify interesting use cases.

Statistics Netherlands provided a (part-time) project manager and UNECE secured additional assistance to the project.



### Next Steps

New participants will be invited and use cases identified as well as further scoping of the work programme and expected output.

## Risks and Issues

Issue	Mitigation
Not enough active members will join the group	EB called to have their colleagues involved in the project.
Substantive expert to co-lead the project will not be provided by HLG-MOS member	EB called to provide (in-kind) a substantive expert to co-lead the group

## News from the Groups

<p><b>Blue-skies Thinking</b></p> <p><b>? Unbekannter Anhang</b></p>	Identifying Topics / Opportunities	IN PROGRESS	In September, the BSTN will have two dedicated meetings of pitched presentations of strategic innovation projects or specific innovative activities in Statistical Organizations to identify common grounds for potential collaboration and proposals for activities under the HLG-MOS.
	Synthetic Data Sets	IN PROGRESS	<p>The Synthetic Data Sets Group has seen great interest and it has grown and now has 25 members from 20 NSOs and three academic institutes. Two sub-groups were setup to further discuss various aspects of the topic:</p> <p>1) Methods and tools for producing synthetic data (chaired by Eric Deeben, ONS). This group sets out to provide an overview of methods available to produce synthetic data and establish consensus or recommendations on the most appropriate methods to use. This group will also explore the tools necessary to create synthetic data, what infrastructure or environment are necessary to support these tools and what are the limitations or considerations of the tools.</p> <p>2) Utility measures and communication (chaired by Kate Burnett-Isaacs, Stats Canada, and Alistair Ramsden, Stats NZ). This group will research, review and propose measures to evaluate the analytical and confidentiality quality of any synthetic data set. The goal of this group is to establish international consensus on how to evaluate the quality of a synthetic data set regardless of the method used. This group will also focus on how to communicate the risks and considerations required to appropriate use a synthetic data set.</p> <p>The group is further working on a project proposal for 2021.</p>
	Follow-up other selected topics	IN PROGRESS	<p><b>Chatbots:</b> the work is progressing under the leadership of OECD. Focus is at first on labour force statistics. It will be demonstrated in the BSTN in September. Stats Canada and Statistics Netherlands are now involved and Serbia is joining. Other countries are very much welcome to join.</p> <p><b>Data Science Lab:</b> Synthetic Data Sets and ML will be test cases as these groups are already a kind of communities and members are motivated to keep on collaborating. The EB was requested to provide staff to support the collaboration platform from the technical /information management side.</p> <p><b>DDDM support at local level:</b> no updates. Countries are requested to join the subgroup discussing this topic. Some offices might have new experiences coming from the Covid pandemic.</p>
	Other		The group has not received any project or activity proposal for evaluation but is ready to receive and evaluate them.
Capabilities and Communication	Culture Change and Internal Communications Strategy	IN PROGRESS	It hasn't started work yet.



Competencies Training and Development	IN PROGRESS	It hasn't started work yet.
Future of work in the context of Modernisation of the workplace	IN PROGRESS	It hasn't started work yet.
Ethical leadership	IN PROGRESS	
Follow up Strategic Communication Framework	IN PROGRESS	Task Team on social media strategy will need to invite more people in the team (old members of the Strategic Communication Framework) as we will need to review and prepare Strategic Communications Framework Publication for print (early 2021).
Training of staff in communication	IN PROGRESS	It hasn't started work yet.
HRM T Workshop	IN PROGRESS	OC is working very hard. So far we have received 19 presentations. Now we have to allocate all of them properly within 3 days. And we need to organise small group discussions having in mind technical possibilities.
Other		
Linking GSBPM and GSIM	IN PROGRESS	<p>The task team is continuing the work on describing how inputs and output flow through GSBPM sub-process using GSBPM constructs. In this work, the task team faced some issues concerning GSIM objects (both in concepts and definitions). The team discussed the issues and prepared concrete proposals that have been submitted to the attention of the GSIM task team.</p> <p>The team has covered about 30-40% of GSBPM sub-process (most commonly used sub-processes) and has developed use cases for the sub-processes that have been mapped.</p> <p>Given that the work has covered a variety of different types of sub-processes, the team assumes that most issues have already been encountered and tackled. The team will try to finalise the work next year.</p>
Core Ontology for Official Statistics	ON HOLD	

## Supporting Standards



	GSB PM for Geospatial data	IN PROGRESS	<p>The task team has covered until phase 4(included). The team agreed on the structure of final report and prepared a preliminary report based on the works done so far. In the report, geospatially related activities are embedded in original texts of GSBPM (rather than having a separate document).</p> <p>The final report will be a separate document from GSBPM and it is not going to replace it. The GSBPM revision team that will be set up for the next GSBPM revision, will decide whether to include additional texts into GSBPM or not.</p>	
	Meta data Glossary	FINISHED		
	GSIM	IN PROGRESS	The GSIM team commenced working (after the Metadata Glossary team finished).	
	Other	<p>The ModernStat World Workshop 2020 will take place virtually from 27-30 October 2020 (from 14:00 to 17:00).</p> <p>The Organising Committee was set up with members from the Supporting Standards and Sharing Tools Groups.</p> <p>The Organising Committee is going to finalise the Programme during next meeting that will take place at the end of September.</p>		
<div>Sharing Tools</div> <div>? Unbekannter Anhang</div>		Finalizing CSP A 2.0 /Digitizing/ & Communication restated CSPA	IN PROGRESS	Further comments have been collected on the CSPA 2.0 document. It will be finished before the November workshop. Further activities are planned as a Task Team under the Supporting Standards group.
		Adding Services to Catalogue	ON HOLD	On hold but services are being developed under the Eurostat sponsored I3S project that will be added to the catalogue
		Integrating Innovations in Architecture and technology	ON HOLD	On hold but work is continued under the Eurostat sponsored I3S project
		Other	The group is working on providing presentations for the ModernStats World Workshop.	

HLG-MOS, 1 March, New York

Machine Learning Sprint 1-3 April, Virtual

Workshop on Statistical Data Editing, 31 August- 2 September 2020, Virtual

Covid-19 response in Dissemination and Communication of Statistics 7-11 September, Virtual

Covid-19 Response in Human Resources Management and Training, 7-11 September, Virtual

Statistical Data Collection in times of Covid-19, 5-9 October, Virtual

Machine Learning Sessions, 13-1 October, Geneva