

# 2021 March-April

## Synthetic Data Guide

Progress

Next Steps

Output	Information Gathered	Tasks to Do	Due Date	Dependencies
Use Cases	Determine use case categories: <ol style="list-style-type: none"><li>1. Dissemination to the public</li><li>2. Testing analysis</li><li>3. Testing ML algorithms</li><li>4. Education</li><li>5. Testing systems</li></ol>	<ol style="list-style-type: none"><li>1. Generalization of the use case categories</li><li>2. Gradient on utility versus disclosure risk needs</li></ol>	June 1, 2021	
Methods	<ol style="list-style-type: none"><li>1. Sequential multimodel multivariate model method</li><li>2. Fully conditional specification</li><li>3. Special cases of a fully conditional specification</li><li>4. Microsimulation for generating synthetic data</li><li>5. Information preserving statistical obfuscation</li><li>6. Pseudo likelihood method</li><li>7. Method by Fleishman and Vale &amp; Maurelli to simulate multivariate non-normal random numbers from the original data</li><li>8. GANS</li></ol>	<ol style="list-style-type: none"><li>1. Identify any additional methods</li><li>2. Summarize the following methods:<ol style="list-style-type: none"><li>a. Sequential multimodel multivariate model method</li><li>b. Pseudo likelihood method</li><li>c. GANS</li><li>d. Special cases of a fully conditional specification</li></ol></li></ol>	June 30, 2021	
Utility and Risk Disclosure measures	Measures have been gathered and they are here on the project <a href="#">wiki page</a> .	Summarize/describe the methods	Ideally June 30  Drop dead date: August 30, 2021	
Recommended methods		A workshop will be held to recommend methods for each use case	July 2021	Use Cases
Recommended utility and disclose risk methods		A workshop will be held to recommend measures for each use case	August 2021	Use Cases
Tuning		Based on the methods gathered and use cases, highlight tuning methods and/or considerations	August 30, 2021	Use Cases, Methods and Measures
Synthetic data challenge	Virtual  Mixed NSO teams  Priority to synthetic project NSOs but can be open to others (i.e. ML group, HLG-MOS, etc.)	Test methods	Mid-end September	Use Cases, Methods and Measures Recommendations have been made
Review of data challenge outcomes		If any major flags on our recommendations were raised in the hackathon, they will be addressed in October	October 2021	
Bringing it all together		A final draft of the guide needs to be completed before the November workshop. Final touches and editing will be done in 2022.	October 31 <sup>st</sup> , 2021	

## Risks and Issues

Issue	Mitigation

## Input Privacy-preserving Techniques

### Progress

The work of workpackage 1 is finished. We have 5 use cases documented, discussed in the group and generalized the use cases so they can be used for other NSO's. Last step was to prioritize the use cases.

The work of work package 2 has been defined. In this work package, the work is divided into three tracks. The first two tracks are internal and are a continuation of the previous work. Work is continuing on two main scenarios, namely private set intersection and private machine learning. In the individual tracks best practices are exchanged and a mini use case is worked out and then actually implemented. In addition to experience, we also have some simple examples at the end that serve as inspiration to others. The third work package focuses on the outside world. What blind spots do we have. The description of the work package can be found on the wiki.

The subteams for the different tracks have been formed.



### Next Steps

For track 1 (private set intersection) and for track 2 (private machine learning), the following steps are the exchange of best practices and the definition of a mini pilot use case.


Track 3 is still on hold. For this we want to take time to think carefully about what we want. Starting too early carries the risk of plotting the wrong question.

## Risks and Issues

Issue	Mitigation

### News from the Groups

<b>Blue-skies Thinking</b>  <b>? Unbekannter Anhang</b>	Identifying Topics /Opportunities	<b>IN PROGRESS</b>	Work on identifying new topics for 2022 scheduled to start at the May meeting of the BSTN core team. Special attention will be paid to engaging with possible partners outside the official statistics community
	Network Data	<b>IN PROGRESS</b>	Bilateral and other contacts have been initiated following presentation by ABS
	Covid-19 Hotspot Joint Biosecurity Centre Platform	<b>IN PROGRESS</b>	Bilateral and other contacts initiated after ONS presentation, e.g. to replicate the approach on other cloud platforms in other environments. Possible connections with Serbian activity proposal (RSS) investigated
	User Research for Official Statistics	<b>IN PROGRESS</b>	Execution put on hold until after summer, OECD in preparatory contact with some interested NSIs

	Rapid survey systems	IN PROGRESS	Connection with ONS Joint Biosecurity Centre approach is being explored
	From experimentation to implementation in official statistics	IN PROGRESS	Online scoping workshop organised by StatsCan has identified "Culture" as first topic that most workshop participants considered worthwhile to explore further. Possible connections with other HLG-MOS groups will be explored. Invited session proposal for ISI/WSC on the 'Experimentation to Implementation' topic has been accepted
	Microdata for understanding declining response rates	IN PROGRESS	Slightly postponed until necessary resources at StatsNZ have been made available. First contacts with academic researchers interested in the topic have been established.
	Other	IN PROGRESS	
Capabilities and Communication			
	Future of work, future workplace and future skills	IN PROGRESS	<p>Selected Chair for the team - Jeremy Visschers, from Statistics Netherlands.</p> <p>On the base of different documents prepared by NSOs we selected first aspects of future work, like: social aspects, flexibility conditions, work profiles, analysis, mapping, and digitalization of processes and services, cost efficiency, management needs.</p> <p>The next step is first sprint on 4 May to decide on the scope of work.</p>
	Ethical leadership as part of culture evolution	IN PROGRESS	<p>New members joined the Task Team, will review proposed questionnaire and decide on new version, background materials and when it will be sent to the countries. UNECE will look for the information on ethics from the NSO's available on-line. Task Team still invites more colleagues to join this Task Team.</p> <p>The next call will be on Wednesday 26 May.</p>
	Role of market research, digital marketing & communication strategies and tools in managing a crisis communication situation and in promoting public engagement in surveys	IN PROGRESS	<p>New members joined this team, next call is on 5 May. Still deciding on the scope of work, collecting materials.</p>

	Strategic Communication Framework Publication	IN PROGRESS	Submitted to printing unit.
	HRMT Workshop 2022	NOT STARTED	No update
	Topic 6		placeholder
	Other		
<h3>Supporting Standards</h3>			
	Linking GSBPM and GSIM	IN PROGRESS	Following the harmonisation work done at the beginning of this year, the Task Team is now progressing with descriptions for the remaining sub-processes for GSBPM Phases 3, 7 & 8, using the refined template (objects harmonised with the work of the GSIM Task Team). Work is progressing as scheduled, no delays expected according to the timeline set in the activity proposal for this year.
	Core Ontology for Official Statistics	IN PROGRESS	Task Team is progressing as scheduled. The platform for the discussion is Github where information on the issues are accessible: <a href="https://github.com/linked-statistics/COOS/">https://github.com/linked-statistics/COOS/</a> . Work was also presented to the Supporting Standards Group at our last meeting (8th April 2021). The URI policy has been drafted and basically accepted by the Task Team (available on Github). Apart from the specific issues, the Task Team will focus more on the governance document next.
	Updating GSIM	IN PROGRESS	Task Team work is ongoing as planned. Proposals for removal of objects will be shared with the Supporting Standards Group as well at the next meeting (20th May 2021).
	Application of GSBPM for Geospatial Information	IN PROGRESS	The Task Team is currently reviewing the final document with the deadline of 30th April 2021 (next meeting for the Task Team). Following this round of internal review and finalisation of the document, the completed report will be shared with other teams and the Task Team will finish its work.
	GSBPM Task	NOT STARTED	Task Team is expected to start its work in the second half of the year, after other Task Teams have completed their work.
	CSPA	NOT STARTED	Planned to start in 2nd half of 2021. There was a first discussion with Márta, the previous champion of the group on what the activity should focus on, identification of opportunities and risks. A first version of an activity proposal is being drafted.
	ModernStats World Workshop 2022	NOT STARTED	
	Other		<p>The survey on the use of the ModernStats models concluded in mid-March and lessons learnt are being summarized. Valuable input is collected not only to have a better overview on the current situation of using our models but also on communication and some specific input for Task Teams (GSBPM Task).</p> <p>The Supporting Standards Group also identified potential actions for communication and how to have better visibility for our models. A short internal working paper is drafted, listing these ideas. The group discussed these ideas in written and we are to agree on specific actions to implement.</p>

## Machine Learning 2021

Mc I r  
Mch I r n  
Mchn Irnni  
Mache leann  
Machne learing  
Machine learnig  
**MACHINE LEARNING!!!**



WS1  
– Pilot  
studies:  
from  
Idea  
to  
Valid  
solutions

IN PROGRESS

Please access the link. [Click here for the slides.](#)

WS2  
–  
From  
Valid  
Solution  
to  
Production

IN PROGRESS

Please access the link. [Click here for the slides.](#)

WS3  
–  
Data  
Ethics  
and  
Governance

IN PROGRESS

Please access the link. [Click here for the slides.](#)

WS4  
– On  
The Quality  
of  
Training  
Data

IN PROGRESS

Please access the link.

WS5  
– On  
The  
Quality  
Framework  
for  
Statistical  
Algorithms

IN PROGRESS

Please access the link.

Other	<p>.</p> <p><b>ONS-UNECE ML 2021 Group</b></p> <p>March-April update:</p> <p>After the launch of the ML 2021 in January, 18 research projects distributed across 5 workstreams are in progress. On 22<sup>nd</sup> March, workstream and theme leads uploaded a quarterly project report on Wiki describing initial steps taken towards progressing their projects. A summary is below, and detailed information for each workstream can be found in the links above.</p> <p><b>WS1 Pilot Studies</b> – Activity leads are reporting in regular meetings about the progress made in their activities to get feedback from peers</p> <ul style="list-style-type: none"> <li>• <b>WS2 Production</b> – There are 3 activities under WS2 which are looking at different components important for production, namely: service development, process modelling, data architecture. A sub-team was created to work on more detailed description of journey from PoC to production. This will provide more concrete guideline on steps stats organisations follow to successfully launch ML application into production. The final report of this sub-team will be fed into the UNECE publication on ML planned for December 2021.</li> <li>• <b>WS3 Ethics</b> – After performing a landscape review, the team have drafted ethical principles for ML. They intend to consult more widely with ML2021, starting on the April meeting on 26<sup>th</sup> April, aiming to determine which principles are worth delving deeper into. They are also planning an Ethics-ML workshop in the summer which will apply the principles to existing ML applications.</li> <li>• <b>WS4 Quality of Training Data</b> – The team did a literature review to identify types of model drift (circumstances where model performance deteriorates), and are working on metrics to be used to monitor drift. Given that data could not be shared, the team is testing their metrics using open data provided by Statistics Poland during ML project (ECOICOP).</li> <li>• <b>WS5 QF4SA</b> - Initial meeting took place for the introduction of QF4SA from ML project and use-case where the framework will be tested. The activity lead is working on the “explainability” dimension using INEGI C&amp;C activity as use case.</li> </ul> <p>Outreach: UNECE <a href="#">article</a> is now live promoting the ML 2021 initiatives and a ONS/DSC blog will be live next week promoting ML 2021 progress made in the first quarter.</p> <p>Collaboration and events – The Group receives a summary write-out monthly. The latest information and previous presentations can be found on <a href="#">Wiki</a></p> <p>The Group’s next monthly meeting is on 26<sup>th</sup> April and will consist of updates from the coordination team, workstream leads and a presentation from Kate Isaac-Burnett, Statistics Canada, about the HLG-MOS Synthetic Data project and potential collaboration with ML 2021.</p> <p>It is important to note that ML 2021 is adopting a “community” approach, rather than a “project” approach, which brings new challenges to monitor and follow all activities and maintain community engagement. The team is working hard to ensure all projects receive input and progress smoothly.</p>
-------	--

●

HLG-MOS, 29 January 2021, online

●

Linking GSBPM and GSIM Sprint, 2-3 February 2021

●

Expert Meeting on Dissemination and Communication of Statistics, 13-15 Oct, Lisbon (tbc)

●

Expert Meeting on Statistical Data Collection, 27-30 September, online

●

Expert Meeting on Statistical Data Confidentiality, Oct-Dec, tbc

●

HLG-MOS workshop on the Modernisation of Official Statistics, 15-16 November 2021, Geneva

