

# 2021 January-February

## Synthetic Data Guide

Progress

## Membership

The membership of the synthetic data project is now at 38.

## Scope

The goal of the Synthetic Data Project is to develop a hands-on guide for creating and using synthetic data primarily geared towards data protection and disclosure control. The target audience of this guide includes both managers and methodologist at NSOs to provide a starting point for using and creating synthetic data.

After the HLG-MOS Executive Board recommendations, the guide will emphasize creating synthetic data from non-survey data sources. In addition, the use cases will highlight the circumstances in which to synthetic data is the appropriate solution to their output privacy problems.

## Work Packages

Work package 2 - methodology - has been under with presentations of new method from Statistics New Zealand and Australian Bureau of Statistics.

## Partnerships

The HLG-MOS Executive Board recommended partnerships with the ML 2021 group and the Input Privacy Preserving Techniques project. I met with both projects and synthetic data can be a suitable option for testing data.

## Next Steps

## Work Packages

**Work Package 1 - Use Cases:** the BSTN Synthetic Data Working Group in 2020 had created a foundation of use cases for synthetic data. After the EB's recommendations, the use case categories are being reassessed. The use case work package has two main outputs:

1. A description of different use cases and highlighting their analytical and disclosure risk needs
2. Creating a gradient of different use cases based on their analytical and disclosure risk needs

The use cases will form the basis for the methodology and quality measure recommendations.

The target deadline for work package 1 is end of May 2021.

**Work Package 2 - methodology:** The following methodology presentations have been scheduled:

- March 4, 2021: Joerg Drechsler presenting on *method used to synthesize US business data*
- April 1, 2021: Kenza Sallier presenting *Statistics Canada' experience creating public synthetic datasets using the FCS and the Synthpop package* and Gillian Raab and Beata Nowok will be presenting on new additions to the Synthpop package.
- May 6, 2021: Tentative Christine Task and Rolando Rodriguez

## Partnerships

The ML 2021 group will get back to me on their requirements for testing data and we will then determined if synthetic data is the right solution.

Risks and Issues

| Issue | Mitigation |
|-------|------------|
|       |            |
|       |            |

Input Privacy-preserving Techniques

Progress

The use cases submitted were analyzed by the project team and provided with advice on how they can be used for other NSOs (generalization). Subsequently, the project members have drawn up 5 generalized use cases. We are now in the midst of discussing and assessing the generalized use cases. The team has received an update from the UN global task team on the status of the new version of the PPT handbook, the work on legal matters and the option to submit the generalized use cases to the UN global task team has been discussed. Finally, we received a presentation from Fabio Ricciato (Eurostat) on: "Towards a reference architecture for Trusted Smart Surveys". This is for inspiration what it takes to make PPT a success.



Next Steps


In the next meeting, the assessment will be completed, and a joint decision will be made which use cases will be selected for the next phase (reference scenarios can be used for multiple NSOs. This means that WP1 has been concluded and the next WP is scheduled.

Risks and Issues

| Issue | Mitigation |
|-------|------------|
|       |            |
|       |            |

| News from the Groups   |                                    |             |  |
|--|------------------------------------|-------------|--|
| <div>Blue-skies Thinking</div> <div>? Unbekannter Anhang</div> | Identifying Topics / Opportunities | IN PROGRESS |  |
|  | Network Data                       | IN PROGRESS |  |

|  |   |             |  |
|--|---|-------------|--|
|  | Co<br>vid<br>-1<br>9<br>Ho<br>tsp<br>ot<br>Joi<br>nt<br>Bi<br>os<br>ec<br>uri<br>ty<br>Ce<br>ntr<br>e<br>Pl<br>atf<br>orm                               | IN PROGRESS |  |
|  | Us<br>er<br>Re<br>se<br>ar<br>ch<br>for<br>Off<br>ici<br>al<br>St<br>ati<br>sti<br>cs   | IN PROGRESS |  |
|  | Ra<br>pid<br>su<br>rv<br>ey<br>sy<br>ste<br>ms  | IN PROGRESS |  |
|  | Fr<br>o<br>m<br>ex<br>pe<br>ri<br>m<br>en<br>tati<br>on<br>to<br><br>im<br>ple<br>m<br>en<br>tati<br>on<br>in<br>offi<br>cia<br>l<br>sta<br>tist<br>ics | IN PROGRESS |  |

|  |   |             |   |
|--|---|-------------|---|
|  | Mi<br>cr<br>od<br>at<br>a<br>for<br>un<br>de<br>rst<br>an<br>din<br>g<br>de<br>cli<br>nin<br>g<br>re<br>sp<br>on<br>se<br>rat<br>es | IN PROGRESS |   |
|  | Ot<br>her   | IN PROGRESS |   |
| <p>Capabilities and Communication</p>  | Fu<br>tur<br>e<br>of<br>wo<br>rk,<br>fut<br>ur<br>e<br>wo<br>rk<br>pla<br>ce<br><br>a<br>nd<br>fut<br>ur<br>e<br>ski<br>lls         | IN PROGRESS | <p>Task Team met on 9<sup>th</sup> February 2021. We had a brain storm on the scope of work. It was proposed to divide work into parts: 1) organisational matters: switch to work from home instead of office, technical requirements for work from home (Webex, Zoom, Lync, etc) and 2) what kind of skills are necessary for employees for remote working? It was also suggested, as a starting point, to review presentations from the September 2020 virtual workshop to get some input for the work of the team, and decide which presenters we would like to invite to join this team. Another proposal is to conduct the survey among NSOs to address two main points: 1) what have we learned from the pandemic? 2) what we would like to have in the future? As a following steps it was also suggested that we can have two sprint sessions soon 1) exploratory sprint, to discuss what we want to do as a Team and 2) in-depth sprint to produce some concrete output.</p> <p>Next call will be on Monday 8 March at 12.00 Geneva time.</p>  |
|  | Et<br>hic<br>al<br>lea<br>de<br>rs<br>hip<br><br>as<br>pa<br>rt<br>of<br>cul<br>tur<br>e<br>ev<br>olu<br>tio<br>n                   | IN PROGRESS | <p>Task Team met on 10<sup>th</sup> February 2021. The <a href="#">first results of the preliminary poll</a> were presented. We received 36 responses, and 29 organisations work on Ethics Policies, and 24 of them work in this area for more than 3 years. Total of 12 organisation were interested to be involved in the drafting of the survey results.</p> <p>It was suggested to invite countries that expressed interest to participate in the analysis of the survey results to participate in the next call of this Task Team, so they can contribute to the design of the survey and not just analysis. For the next meeting first proposal for the survey should be prepared and the plan is to complete survey design by April/May and to collect responses by July. First results of the survey should be ready in September.</p> <p>It was also discussed proposal to have a virtual meeting focusing on the topics we are working on. We can organise Webinar on Ethical Leadership (Ethics Camp), that will present results of the survey.</p> <p>Next call of this group will be on 10 March at 14.00 Geneva time.</p> |

|  |   |                        |   |
|--|---|------------------------|---|
|  | <div>Role of market research, digital marketing &amp; communication strategies and tools in managing a crisis communication situation and in promoting public engagement in surveys</div> | <div>IN PROGRESS</div> | <div>It was agreed to review all presentations of the 2020 different virtual meeting and select information that will be relevant to the work of this team, and to contact authors of presentations to ask them to join this Task Team. It will be also reviewed past activities and documents of UNECE.</div> <div>It was agreed that next call of this group will be on 3 March at 14.00 (CET).</div> |
|--|---|------------------------|---|

|  |   |             |   |
|--|---|-------------|---|
|  | Strategic Communication Framework Publication | IN PROGRESS | The various documents from the 2018 and 2019 SCF project have been combined (and common parts have been rewritten/edited). UNECE is now working on producing high-quality versions of all figures in the document.  |
|  | HRM Workshop 2022                             | NOT STARTED |   |
|  | Topic 6                                       |             | placeholder   |
|  | Other   |             |   |
| <div> <div>Supporting Standards</div> </div> | Linking GSBPM and G SIM                       | IN PROGRESS | The Task Team had a sprint session in January with the aim to harmonise the work done so far and move forward with the description of the remaining sub-processes in a more harmonised manner. The current work is focusing on the harmonisation of sub-processes at the specification level (Phases 1, 2, 4, 5, 6) after which, authors of each sub-process example can reflect the changes at the implementation level. Division of work for the remaining sub-processes is also currently ongoing.   |
|  | Core Ontology for Official Statistics         | IN PROGRESS | <p>The Task Team started its work in January. Currently a roadmap is being developed for our activities in 2021. Current plans for the Task Team:</p> <ul style="list-style-type: none"> <li>January: Roadmap for 2021 activities</li> <li>February-April: Articulation with work on GSIM and Metadata Glossary</li> <li>May-June: Work on governance (can start in Feb in parallel with other works)</li> <li>July: Expert review</li> <li>August: Final version 1 of the ontology</li> <li>September-October: Document on governance</li> <li>November: Submission to HLG-MOS for adoption</li> </ul> |
|  | Updating G SIM                                | IN PROGRESS | Task Team work is ongoing as planned. Currently issues around the Business Group of GSIM are being discussed.   |

|   |  |  |   |
|---|--|--|---|
|   | Ap<br>pli<br>cat<br>ion<br>of<br>G<br>SB<br>P<br>M<br>for<br>Ge<br>os<br>pa<br>tial<br>Inf<br>or<br>m<br>ati<br>on | IN PROGRESS  | With the last meeting, the Task Team has covered all phases and overarching/corporate-level activities aimed. Currently the integration of comments into the report is ongoing and the draft report is expected to be shared around mid-March for review. After that, there will be a dedicated meeting to finalise the report. |
|   | G<br>SB<br>P<br>M<br>Ta<br>sk  | 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.   |
|   | C<br>SPA   | NOT STARTED  | Planned to start in 2nd half of 2021  |
|   | M<br>od<br>er<br>nS<br>tat<br>s<br>W<br>orl<br>d<br>W<br>or<br>ks<br>ho<br>p<br>20<br>22                           | NOT STARTED  |   |
|   | Ot<br>her  | The SSG updated the previous questionnaire (Survey on the use of ModernStats models (GSBPM, GAMS0, GSIM and CSPA)) on the use of ModernStats models and the collection of information from countries is currently ongoing. Deadline for submission: 10th March.<br><br>Ideas, proposals from the SSG members on communicating the ModernStats models better and to increase visibility are currently being collected until mid-March. The exact actions on these topics will be agreed upon after. |   |
| Machine Learning 2021<br><br><div>Mcl r<br/>Mch l r m<br/>Mchn l r n n i<br/>Mache leann<br/>Machne learing<br/>Machine learnig<br/>MACHINE LEARNING!!!<br/></div> | W<br>S1<br>-<br>Pil<br>ot<br>stu<br>die<br>s:<br>fro<br>m<br>Id<br>ea<br>to<br>Va<br>lid<br>sol<br>uti<br>ons      | IN PROGRESS  |   |

|   |             |  |
|---|-------------|--|
| W<br>S2<br>–<br>Fr<br>o<br>m<br>Va<br>lid<br>So<br>luti<br>on<br>to<br>Pr<br>od<br>uct<br>ion   | IN PROGRESS |  |
| W<br>S3<br>–<br>Da<br>ta<br>Et<br>hic<br>s<br>an<br>d<br>Go<br>ve<br>rn<br>an<br>ce   | IN PROGRESS |  |
| W<br>S4<br>–<br>On<br>Th<br>e<br>Qu<br>alit<br>y<br>of<br>Tr<br>ain<br>ing<br>Dat<br>a  | IN PROGRESS |  |
| W<br>S5<br>–<br>On<br>Th<br>e<br>Qu<br>alit<br>y<br>Fr<br>a<br>m<br>ew<br>or<br>k<br>for<br>St<br>ati<br>sti<br>cal<br>Al<br>go<br>rit<br>h<br>ms | IN PROGRESS |  |



|       |   |
|-------|---|
| Other | <p><b>ONS-UNECE ML 2021 Group</b></p> <p>Led by the <a href="#">Data Science Campus</a> in partnership with the <a href="#">UNECE HLG-MOS</a>, the ML 2021 Group aims to demonstrate the added value of ML in the modernisation of Official Statistics. We will achieve this by facilitating the development and implementation of ML research projects and capacity building activities that meet the global statistical community needs. With nearly 200 members from 36 countries, the Group is delivering 17 research activities grouped into 5 workstreams: pilot studies, integration into production, ethics, quality of training data and quality of statistical algorithms. The ML 2021 outputs for 2021 include:</p> <ol style="list-style-type: none"> <li>1. Monthly meetings, quarterly workstreams reports and open webinar in December</li> <li>2. Engagement of the <a href="#">ML group</a> under the UN Global Network of Data Officers and Statisticians</li> <li>3. Co-ordination and delivery of skills building activities, such as short courses, masterclasses, hackathons, coffee and coding, virtual/ physical sprint, etc</li> <li>4. Publicity of ML 2021 via blogs and articles on key channels and platforms</li> </ol> <p>The inaugural ML 2021 meeting was hosted on 29th January 2021. The second meeting will be hosted on 23rd February and will include update from all 5 workstreams leads and three presentations. More information below.</p> <p>23 February 2021 Meeting Documents</p> <ul style="list-style-type: none"> <li>• <a href="#">Final Agenda</a></li> <li>• <a href="#">WS2 Monthly Update Slides</a></li> <li>• <a href="#">Presentation: Best practices in maintaining the quality of data in ML developments</a> <ul style="list-style-type: none"> <li>- Riitta Piela, Statistics Finland</li> </ul> </li> <li>• <a href="#">Presentation: Multilingual Classification of Economic Activities</a> <ul style="list-style-type: none"> <li>- Casper Eriksen, Danish Business Authority</li> </ul> </li> <li>• <a href="#">Presentation: WS1 Theme 5: Transferring Knowledge and Experience</a> <ul style="list-style-type: none"> <li>- Michael Reusens, Statistics Flanders</li> </ul> </li> </ul> |
|-------|---|

●

**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**