Machine Learning

Progress
Fifteen persons have joined the Team Members: Italy (5), Poland (2), UK (2), Mexico (1), Netherlands (1), USA (1), Canada (1), Switzerland (1) and Belgium (1 from the Flemish Institute for Technological Research). Four members of the Blue-Sky Thinking ML team agreed to continue with the ML project.

Introductory documents were posted on the Machine Learning project Home. This included the position paper and the project proposal from the Blue-Sky Thinking Network team, and an introduction to the project and the project manager. The documents were circulated to members prior to a first virtual meeting held on March 19. The meeting was attended by Italy, UK, Netherlands, Canada, Poland, USA, Belgium and the UNECE. After the introductions, discussions took place on the high-level objectives of the project and potential pilot studies (work package 1).

It was decided to hold a virtual planning meeting at the beginning of April. After the meeting, a doodle poll was sent to set a 2.5 hour meeting between April 3 and April 5. The UK offered to host a face-to-face sprint meeting once the planning is completed and the work starts. We are looking at having this meeting early to mid-May. A questionnaire was sent to each member to collect information needed to prepare the April planning meeting and May face-to-face meeting.

Next Steps
Continue to welcome members from other organisations.

Hold a virtual planning meeting between April 1 and April 3 to finalize the scope of the project and its work packages, the composition of the WP sub-teams and, an initial list of pilot studies and their participating organisations.

Determine a soon as possible the feasibility and dates of a sprint meeting at the beginning to mid-May, and start planning accordingly.

Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The project has just started. The November workshop is not far away.</td>
<td>Set a reasonable scope and initial deliverables for 2019.</td>
</tr>
<tr>
<td>Number of participating organisations.</td>
<td>Organisations are welcomed to join before the virtual planning meeting at the beginning of April. Not a big issue yet, but if more organisations are needed, the project may raise it to the Executive Board before the next progress meeting.</td>
</tr>
</tbody>
</table>

Strategic Communication Framework phase 2

Progress
The first meeting of the project team was held on Monday, March 18, 2019. The business plan and related work packages were discussed. Participating countries expressed interest in the three work packages with the majority of participants expressing interest in work packages 1 and 2. The project team has fewer members than last year but those returning are very committed to achieving success with Phase 2.

With the project beginning in late March (rather than early February), we are approximately one-month behind schedule. However, with strong commitment by participating countries, we should be able to make up the time.
Next Steps

A sprint session has been proposed for the week of April 29th to be held in Geneva. At least five countries have agreed to participate. Participation in this session will be critical to the successful delivery of the three work packages.

A second face-to-face meeting has been recommended for June 2019. This meeting would be held in conjunction with the Workshop of Dissemination and Communication in Gdansk, Poland.

Depending on the outcome of the Conference of European Seminar on National Data Strategies in Summer 2019, additional members with different skill sets may be required for the project team.

Risks and Issues

A late start to the project has put it one-month behind. With full participation and commitment from project members, we should be able to deliver on all aspects of the project by end of year.

Any change in direction (based on the outcome of the CES Seminar on National Data Strategies) may require additional resources to the project team.

News from the Groups

| Blue-skies Thinking | Identifying Topics/Opportunities | The core team was established and now consists of 10 members. The first 13 potential topics were identified. The core team discussed the first 6 and decided on the follow-up. |
| Developing Organisational Capability | Skills and Capability Framework | No progress so far. |
| | Promotion Forum | We are aware that we need to make our outputs visible on the wiki, and promote and share our work. Moreover, we should prepare flyer or promotion materials for the CES conference in June. |
| | Setting vision in NSOs | No progress so far. |
| | Other | Currently, DOC Group is working on organisation of Culture Evolution Workshop held in September 2019, in Geneva. By 29th of March we plan to send first information notice together with the invitation letter to the countries. |
The task team, jointly chaired by Istat and Statistics Canada, had started its work. The discussion is still focused on agreeing on a template and the elements to be considered when mapping GSIM information objects to GSBPM sub-processes. A paper from 2013 (Unece Work Session on Statistical Metadata) describing metadata flows is a good starting point.

The task team is meeting every three weeks.

Main objective of the task team is to build common semantic model and vocabulary for official statistics. The team will first review the paper (COOS - A Core Ontology for Official Statistics. Draft Community Group Report -2018) and work on following activities:

- Activity 1: to provide precise definitions of the core ontology; to agree on scope
- Activity 2: to articulate this base ontology with more sectorial and in-depth works existing on GSIM, CSPA

The task team is meeting every four weeks.

The metadata glossary group has started working (even though without the secretariat support from Unece). The group intend to conclude the work initiated a few years ago and suspended in 2018.

The Group is helping in organising the ModernStats World Workshop 2019.

Unece modernisation program, the models and the new activities of Supporting Standards Group as well as two countries experiences (Norway and Statistics Canada) were presented at NTTS 2019 (12-14 March) in a special topic session.

The Features/Requirements concept will have to be included in the CSPA document. For the rest there will not be major changes to the content. The work will focus on making the information easily accessible and easy to navigate for the various user groups.

The group decided on various actions to find potential services (Investment Survey, surveys data collection workshop, internal plans etc.)

Not started yet.

A mapping was been done between the new Requirements/Features concept and information currently part of the documentation in the catalogue.
April 2019

Machine Learning

Progress
(The portions in blue below are links that should lead you to documents providing more details on the Machine Learning project Home page)

- There are now 29 Team Members from 14 organisations and 12 countries. A Responses to the questionnaire was produced and shared. The participants reported Machine Learning projects reported by team members on a wide range of data sources, processes, objectives and level of maturity. Another 5 persons were added to a distribution. They will also contribute to the project.
- Given the interest of the majority of members to work on pilot studies, the planning focused on WP1
- A planning meeting was held on April 4 2019 virtual planning meeting - documentation, followed by a progress meeting on 2019-04-16 - Progress meeting.
- Leads for the three work packages were identified.
- The Project context and objectives of the project were set. The ultimate objective is to advance the research, development and application of machine learning techniques to add value to the production of official statistics.
- It was agreed to conduct pilot studies (WP1) on four themes:
  - Classification and coding on more traditional data (survey, census, administrative register); led by UK - ONS
  - Edit and imputation on more traditional data; lead to be identified
  - Imagery; led by Mexico - INEGI
  - Web and sentiment data; led by Poland - Statistics Poland
- A lead organisation was identified for three of the four pilot study themes
- Each organisation was asked to Proposing Machine Learning Projects for pilot studies. A form was developed to capture Machine Learning Project - Baseline information.
- Inference with Big Data, namely sensor data, and related challenges was reported by several organisations; it will be addressed in WP2 on data quality
- The ONS have booked venues in London and Newport to hold a sprint meeting from May 13-15.

Next Steps

- Obtain baseline information on each machine learning project proposed by the organisations to finalize the specific objectives and tasks of each WP1 pilot study
- Set the detailed objectives and tasks of the WP2 on data quality
- Conduct a sprint hosted by the ONS May 13-15. So far, it would be attended by 11 participants from 8 organisations in 7 countries. Facilities for others to connect and participate will be arranged. The ONS has submitted a paper to provide an overview of the ML project and the outcome of the sprint to the ModernStats workshop.

Risks and Issues
### Strategic Communication Framework phase 2

**Progress**

**Next Steps**

**Risks and Issues**

#### News from the Groups

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blue-skies Thinking</strong></td>
<td>![IN PROGRESS] The BSTN Group will meet 26 April and continue discussing the first 14 topics identified so far. Planned to be ready for discussion in the May Executive Board meeting.</td>
</tr>
<tr>
<td><strong>Unknown Attachment</strong></td>
<td>![IN PROGRESS] Follow-up actions have been identified for three topics so far. Follow-up will start after discussion/selection in the Executive Board</td>
</tr>
<tr>
<td><strong>Developing Organisational Capability</strong></td>
<td>![IN PROGRESS] We are planning to start with identifying complementary skills. The top level of the GAMSO framework could be used as a starting point for the work.</td>
</tr>
<tr>
<td><strong>Promotion Forum</strong></td>
<td><strong>IN PROGRESS</strong></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>We need to update public wiki site of our Group regularly, and looking forward to getting proposals on fresh blog posts. We also need feedback on: what will draw you to this wiki site? and what you find the most useful? We can also consider to share national news on the blog and to add Discussion Forum, if useful. We will advertise public wiki page during the various UNECE Workshops. All Group members are also encouraged to share link to the wiki page with their colleagues.</td>
<td></td>
</tr>
<tr>
<td>We are preparing to prepare some handout about the work/outputs of our group for the CES session this June, maybe presentation at the CES with promotional video (it should be checked). In near future our promotional material should be interactive, short, sharp and action focused. We need to focus on the benefits we add to the NSO’s, and demonstrate the value of the outputs to the higher degree.</td>
<td></td>
</tr>
<tr>
<td><strong>Setting Vision in NSOs</strong></td>
<td><strong>IN PROGRESS</strong></td>
</tr>
<tr>
<td>We are planning to start work with the short questionnaire on how to define the best vision for NSO’s. We need more people interested to contribute to the work on Setting Vision in NSO’s.</td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td>We are under the process of forming Organising Committee of the workshop on Culture Evolution. Andrew Mann from ABS agreed to be a chair of the Workshop. All members of this Group are asked to volunteer to be co-organisers of one of the proposed sessions of the Workshop, and when OC is formed, we will schedule separate, more frequent, calls to prepare this Workshop. It is proposed to have 2 co-organisers for every session.</td>
<td></td>
</tr>
</tbody>
</table>

**Supporting Standards**

<table>
<thead>
<tr>
<th><strong>Linking GSBPM and GSIM</strong></th>
<th><strong>IN PROGRESS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>The task team, jointly chaired by Istat and Statistics Canada, had started its work. A template has been agreed to map GSIM information objects to GSBPM sub-processes as well as to provide concrete examples. Two GSBPM sub-processes have been mapped by Stats Canada and Istat. The Group has volunteered to produce the mapping for a few sub-processes to be discussed at next meeting. The task team is meeting every three weeks.</td>
<td></td>
</tr>
<tr>
<td><strong>Core Ontology</strong></td>
<td><strong>IN PROGRESS</strong></td>
</tr>
<tr>
<td>The task team chaired by France is meeting regularly. The issues found in terminology have been reported in a common wiki page. The group plans to finish the document for public review before November HLG meeting and to extend phase 2 into 2020 to incorporate the public review results.</td>
<td></td>
</tr>
<tr>
<td><strong>Alignment of GSBPM and GAMSO</strong></td>
<td><strong>IN PROGRESS</strong></td>
</tr>
<tr>
<td>The task team, chaired by Mexico, had started its work. Agreement was reached on the OPs of GSBPM to consider and that need further explanations. The different role in GSBPM and GAMSO of common OPs will be clarified and examples will be provided. Mexico prepared a paper summarising the work to be done. The task team is meeting every four weeks.</td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>Status</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Metadata at a Glossary</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Other</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Sharing Tools</td>
<td>Unknown Attachment</td>
</tr>
<tr>
<td>Digitalising the ModernSPA document</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Adding Services to Catalogue</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Communication restarted</td>
<td>NOT STARTED</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

**Events:**

- **HLG-MOS (3 March, NYC)**
- **Strategic Communication sprint (30 April-2 May, Geneva)**
- **Machine Learning Sprint (13-15 May, UK)**
- **Date Integration Workshop (21-23 May, Belgrade)**
- **Strategic Communication Sprint 2 (10 - 11 June, Gdansk)**
- **DissComm Workshop (12 - 14 June, Gdansk)**
- **ModernStats World Workshop (26-28 June, Geneva)**

**Page viewed times:**

- May 2019
Progress
Engagement in the group continues to be very high. Over 20 projects were proposed, either test ML techniques or share knowledge and experience.

A sprint was held on May 13 (see objectives and agenda). It was attended by 12 participants from 9 countries, and hosted by the ONS. Good progress on Work packages 1 and 2 was made:

- **WP1 - Pilot Study on Coding and Classification**: Building on US Bureau of Labor Statistics' successful implementation of machine learning to autocode injuries and illnesses, the code and practices has been shared and will be tested on different types of data in Serbia, Poland and Belgium. The types of data include data collected on the Web, notably to measure sentiment. Other applications may join this pilot study.

- **WP1 - Pilot Study on Editing and Imputation**: A sub-group will investigate the potential of ML in automating the editing process and determine how the statistical foundations of ML and traditional techniques differ, where ML techniques can add value and in what context this value added can be the most beneficial. Examples from the UK and Italy will be used to conduct these investigations. Knowledge and experiences in imputation will be shared. Communication with the [Statistical Data Editing](#) expert group will be assured by members sitting on each group, as well as a member of the SDE on the ML project's distribution list.

- **WP1 - Pilot Study on Imagery**: The scope of this pilot study has not been finalized because some key project members were not in attendance. At the sprint, the UK presented a successful application of ML to use street images to produce relevant statistics at a relatively local scale (two cities). It is relevant to the needs and interests of Belgium and Netherlands. We discussed an idea to produce a document to assist new users of imagery data by describing the processing pipeline (that calls on ML) in the use of such data, and its accompanying high-level ML-questions and aspects to consider, as well as proposed ML solutions/applications or places to find them. The pilot study may look at using satellite data to measure population (density, change). It will likely not look at satellite data for land use, as this topic, including its ML aspects, has been extensively covered a UN Task Team a couple of years ago.

- **WP2 - Quality**: It will identify quality indicators and performance indicators in two contexts: ML applied to carry out traditional processes on traditional data, and ML applied on non-traditional data sources. To do this it will consider quality features (definitions, dimensions, indicators) from the official statistics and ML communities. It will seek to apply some of these indications in one or two of the WP1 pilot studies. One of the challenges will be to remain focused on ML issues and not the broader issues that come with the various data sources. These are covered more extensively by the [ESSnet Big Data II WPK on Methodology and Quality](#). Communication with this group will be assured by two members sitting on each group and our respective wiki spaces.

- **WP 3 - Lessons learned**: This WP was not discussed directly. One of the intentions is to combine the experiences of organisations who have implemented or are close to implementing ML techniques with the experiences of organisations who will make advancements in implementing them through the WP1 pilot studies into lessons learned on topics such as: facilitators, obstacles, importance and costs of creating and maintaining learning datasets, role of manual operations, etc.

The project manager and participants at the sprint wish to thank the ONS and the UN Data Science Campus for their hospitality, facilities and overall accompaniment throughout the three days.

The documents produced at the sprint will be posted on the wiki and shared with all project members.

Next Steps

- Share documents presented at sprint will all project members
- Finalize and share the plans for each WP and pilot study: objectives, projects, deliverables, success indicators, timelines (most of this was set at the sprint; they now need to be written down, shared and agreed by all)
- Get the pilot studies in motion

Risks and Issues
### Strategic Communication Framework phase 2

#### Progress

UNECE hosted a three day sprint session April 30th - May 2nd to launch phase 2 of the project.  8 participants from 7 countries made good progress on Work packages 1 and 2 of the project plan.  The documents produced at the sprint have been posted on the wiki and shared with all project members.  Work package 3 was discussed but no real progress was made as all the appropriate project members were not in attendance.

#### Next Steps

The project team conducted its regular monthly webex meeting on Tuesday, May 21st.  Workplans were discussed and members agreed on their contributions.

A two day face-to-face meeting is scheduled for Gdansk, Poland on June 10 and 11.  The location and dates were chosen to maximize participation with those already attending the Workshop on Dissemination and Communication in Gdansk June 12-14.  7 members from 7 countries have confirmed their participation with two additional countries indicating interest but have not yet registered.

The Project members thank Statistics Poland for agreeing to host the face-to-face project meeting.  The objectives of this meeting will be to make progress on Work packages 1 & 2 and begin work on Work Package 3.

#### Risks and Issues

Successful delivery of the ambitious work plans require full participation from all members.  I will be in a better position following the Gdansk session to know if we have the right skill sets within the project to deliver on Work Package 3.

Discussions continue with Australia and New Zealand to encourage participation as their knowledge and experience will bring significant contributions to all three work packages but particularly work package 3.
DOC group members are working on paper showing connection between technical and complementary skills and increase awareness about the issue. It was prepared very initial document for further discussion among group members including to some extent alignments with GAMSO which is not easy task. The deadline is on 11 June, before next Webex call.

We prepared one-page draft flyer for the CES session in June. It is dedicated to senior and mid level management to attract attention to the outputs of our group.
### Setting vision in NSOs

After the Communications Sprint that took place in Geneva at the end of April, it was decided that the Strategic Communications Team will take over the work on the paper on setting vision in NSO’s. Our group will have possibility to give comments to draft paper.

### Other

The Organising Committee for the workshop on Culture Evolution will have first formal call at the beginning of June, just after the deadline to submit abstracts (end of May).

### Supporting Standards

#### Linking GSBPM and GSIM

The task team is meeting regularly every three weeks. A template for the mapping has been agreed. The mapping is being done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM. The task team is concentrating on phase 5 of the GSBPM both at a design and implementation level. The mapping exercise, including examples from different countries is quite demanding. For the time being, the task team will be able to do the mapping for phase 5 and 4 and maybe one additional GSBPM phase. However, the task team will probably not be able to complete the mapping by November.

#### Core Ontology

The development of the core ontology goes on at a steady pace, via virtual meetings (6 of them in the first semester) and offline exchanges. The construction of the model first focused on the integrated view between GSBPM and GAMSO, with discussions on the connection between the notions of activity and process. More recently, modelization of the statistical organizations and products was undertaken. A first version of the ontology will be available for presentation at the HLG meeting in November and afterwards submitted to public review.

#### Alignment GSBPM and GAMSO

The task team has produced a document specifying the activity to be done. Agreement has been reached. At next meeting in May, the task team will prepare and discuss first draft descriptions of overarching processes in GSBPM and their relationships to GAMSO corporate support activities.

#### Metadata Glossary

The work of the task team is proceeding regularly.

### Other

The Supporting Standards Group is highly involved in the preparation of the June ModernStats World Workshop.
<table>
<thead>
<tr>
<th>Sharing Tools</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Digitalizing/editing CS PA document</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Adding Services to Catalogue</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Communication Interest CSPA</td>
<td>NOT STARTED</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLG-MOS (3 March, NYC)</td>
</tr>
<tr>
<td>Strategic Communication sprint (30 April-2 May, Geneva)</td>
</tr>
<tr>
<td>Machine Learning Sprint (13-15 May, UK)</td>
</tr>
<tr>
<td>Date Integration Workshop (21-23 May, Belgrade)</td>
</tr>
<tr>
<td>Strategic Communication Sprint 2 (10-11 June, Gdansk)</td>
</tr>
<tr>
<td>DissComm Workshop (12-14 June, Gdansk)</td>
</tr>
<tr>
<td>ModernEO Workshops 28 Jur Genev</td>
</tr>
</tbody>
</table>
Machine Learning

Progress

- A follow-up was initiated on three of the four topics identified by the BSTN core group and directed to the ML project. One of the follow-ups led to Statistics Norway joining the project to collaborate on the use of ML to code NACE. We await a response from a member of the BSTN core group on another topic. The membership of the project now stands at 32 participants from 13 countries.
- The objectives, deliverables and timelines of each work package and pilot studies will be completed by the end of the month. Most countries are preparing the required data. The approach taken in the various studies is to share code and test them each organisations' data, rather that attempt to create common datasets that would be accessible and meaningful to all organisations.
- The pilot study on Coding and Classification now includes 7 projects from 6 countries. Statistics Poland has used code from the US-BLS on an ECOICOP classification and achieved an initial accuracy of 88%.
- The project manager wrote a short report describing the scope and progress for a meeting of chief statisticians at the CES plenary session in Paris. Two members will be introducing the project at the ModernStats workshop in Geneva.
- Discussions are taking place to set the date and location of a second face-to-face sprint. A decision will soon be taken.

Next Steps

- Set up an environment to share code (refer to risks below)
- Conduct the various studies, regularly monitor progress and gather results.
- Organize the next face-to-face meeting.

Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>We need to setup immediately an environment to share and manage machine learning code. GitHub is the preferred option. Given the number of participants, we cannot use a free account. One needs to be purchased. Discussions on this are taking place with the UNECE.</td>
<td>For now, we are testing a temporary approach using Confluence. This will also be necessary until a permanent solution is acquired and setup.</td>
</tr>
<tr>
<td>Some organisations are very interested in learning how to analyze imagery data. Mexico will be conducting a pilot project using Landsat open data to measure and monitor urban density. Other organisations could learn to use satellite data by collaborating with Mexico. Such a collaboration would be facilitated if they can acquire satellite Analysis Ready Dataset (ARD). These are not produced by NSOs.</td>
<td>This in not a risk, but rather a challenge. Resolving it would further augment the impact of the ML project. Exchanges are taking place between ABS, INEGI and third party data processors. We raise this as a &quot;heads-up&quot; in case we need the support of the EB to move this issue forward.</td>
</tr>
</tbody>
</table>

Strategic Communication Framework phase 2

Progress

We had an excellent two day sprint in Gdask Poland on June 10 and 11. A special thank you to Statistics Poland for being such gracious hosts to both the sprint and the Workshop on Dissemination and Communication.
Work is progressing well. We developed some content and have excellent work plans for the remaining work related to work packages 1 (Mission, vision and values / Staff Engagement) and 2 (Stakeholder Engagement). We also have sufficient and dedicated participants willing to contribute to the work required to complete these two activities. During the Workshop of Dissemination and Communication, we ceased the opportunity to promote the results of Phase 1 and present the work plans for Phase 2. This resulted in greater awareness amongst the broader international communications community. It also resulted in the addition of a new member to the project team from Oman.

As for work package 3 (National Data Strategies), we await the outcome of the CES Seminar later this month. In the meantime, we have begun to gather an inventory of countries engaged in national data strategies, examples of national data strategies and have commitment from a couple of countries to produce case studies on their experience to date. In addition, we have identified a strategy that we believe can best contribute to this exercise at this time. We have begun to develop content regarding the considerations NSIs need to explore to ensure public acceptance of this new business model.

Next Steps

Work will continue over the summer on work packages 1 and 2. We will await the results of the CES Seminar in a couple of weeks to determine whether adjustments need to be made to work package 3.

Eurostat has agreed to act as an editor on the final draft products and INEGI has agreed to take on the role of producing our final products. These commitments are very much appreciated by the project team members.

Risks and Issues

As we began work late this year, the timetable is very tight to deliver a solid draft product by the HLG Workshop in November. However, we believe we have established a solid workplan and have the necessary resources that will allow us to deliver a quality product on time.

<table>
<thead>
<tr>
<th>News from the Groups</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blue-skies Thinking</strong></td>
<td><strong>Identifying Topics / Opportunities</strong></td>
<td>14 Topics were identified and the next steps were decided on.</td>
</tr>
<tr>
<td>? Unknown Attachment</td>
<td>Follow-up selected topics</td>
<td>Project Proposals submitted in 2018 that were not selected, will now be considered by the core group.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The group is also considering to organize an IT Strategy meeting of CIOs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The following two topics will at first be discussed in more detail. A position paper will be prepared for each:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Synthetic data sets: Statistics Canada (lead), ABS, Istat, and Statistics Netherlands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Secure Multi-party Computation: ABS (lead), Stats Canada, ONS and Statistics Netherlands</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INEGI, CSO Ireland and Statistics Netherlands will do a further refinement on the Data Science Lab proposal.</td>
</tr>
</tbody>
</table>
**Developing Organisational Capability**

<table>
<thead>
<tr>
<th>Skills and Capabilities Framework</th>
<th>IN PROGRESS</th>
<th>We have got first initial draft of paper on importance of complementary skills. It was decided that day before Culture workshop in September we will have a session on it. All DOC group members are obliged to develop this paper by their experiences up to half of August.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Promotion Forum</th>
<th>IN PROGRESS</th>
<th>We prepared new leaflet for CES in June to promote our activities.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Setting vision in NSOs</th>
<th>IN PROGRESS</th>
<th>We are waiting for the results of Group on Communication.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th></th>
<th>The Organising Committee of Culture Workshop had a webex call. So far we have got around 10 papers so we need to fill gaps, especially in the area of vision and mission. On 28th of June will be next call. Our goal is to encourage all interested people to present their experiences in given areas. We would like to ask EB members to share this information with their colleagues.</th>
</tr>
</thead>
</table>

| Supporting Standards | IN PROGRESS | The task team is meeting regularly every three weeks. A template for the mapping has been agreed. The mapping is being done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM. The task team is concentrating on phase 5 of the GSBPM both at a design and implementation level. Last meetings were dedicated to reviewing the mapping and the examples provided by different countries. This implies an harmonisation process, on one side and a consolidation of the templates, on the other.

As already mentioned to the Executive Board meeting in May, the task team will be able to do the mapping for phase 5 and 4 and maybe one additional GSBPM phase. It is likely that the task team will not be able to complete the mapping by November. |
|---------------------|-------------|--------------------------------------------------------------------------------|

| Supporting Standards and GSIM | |}

---

**Supporting Standards**

- CSPA
- GBSIM
- GAMSO
- GSBPM
- Modern Stats

**Linking GSBPM and GSIM**
<table>
<thead>
<tr>
<th><strong>Core Ontology</strong></th>
<th><strong>IN PROGRESS</strong></th>
<th>The development of the core ontology goes on at a steady pace, via virtual meetings and offline exchanges. The construction of the model first focused on the integrated view between GSBPM and GAMSO, with discussions on the connection between the notions of activity and process. More recently, modelization of the statistical organizations and products was undertaken. A first version of the ontology will be available for presentation at the HLG meeting in November and afterwards submitted to public review.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alignment GSBPM and GAMSO</strong></td>
<td><strong>IN PROGRESS</strong></td>
<td>The task team has produced a document specifying the activity to be done. The document is regularly updated after each meeting. Agreement has been reached on the content and the structure of the document. The premises of the document were consolidated. The discussion on the overarching processes in GSBPM and their relationships to GAMSO corporate support activities, started with the OP “Quality management”.</td>
</tr>
<tr>
<td><strong>Metadata Glossary</strong></td>
<td><strong>IN PROGRESS</strong></td>
<td>The work is proceeding regularly</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td>The Supporting Standards Group was highly involved in the preparation of the June ModernStats World Workshop. The Workshop will be the occasion to present the ongoing work of the Group and receive feedbacks. Some small group discussions are being organised about topics on “Terminology”, “Statistical Activities in GAMSO” and items on “GSIM”.</td>
</tr>
<tr>
<td><strong>Sharing Tools</strong></td>
<td></td>
<td>First the CSPA Document will be updated and edited. The Application Architecture section has nearly been completely rewritten. The ‘Common Statistical Production Architecture’ section had some significant changes. Smaller changes were made on: Business Architecture, Information Architecture, Enablers.</td>
</tr>
<tr>
<td><strong>Digitalizing CSPA document</strong></td>
<td><strong>IN PROGRESS</strong></td>
<td>CSPA V2.0 Overview (New) was added.</td>
</tr>
</tbody>
</table>

Unknown Attachment
Several new services were added to the Catalogue. During the ModernStats World Workshop we will have interactive activities in which potential services will be identified and some selected will be tried to add to the catalogue.

Several members were active with preparations for the ModernStats World Workshop. We’ll have several presentations and interactive sessions to promote and advance CSPA.

*HLG-MOS (3 March, NYC)*
*Strategic Communication sprint (30 April - 2 May, Geneva)*
*Machine Learning Sprint (13-15 May, UK)*
*Date Integration Workshop (21-23 May, Belgrade)*
*Strategic Communication Sprint 2 (10 - 11 June, Gdansk)*
*DissComm Workshop (12 - 14 June, Gdansk)*
*ModernStats World Workshop 28 Jul, Genev*

**Machine Learning**

**Progress**

- The project membership and distribution lists kept growing. The membership now stands at 35 participants from 16 organisations and 13 countries. The followers and additional collaborators list now stands at 20 persons from 8 countries, with Finland recently joining.
- Work is progressing on all work packages and pilot studies. Refer to the documentation on the July meeting held last week. Wiki pages have been setup to document pilot studies as they are executed. See: Summary documentation - C&C, Summary
and E&I. The pages were created last week and contain limited information for now. Some preliminary results and deliverables are available at Working documents.

- Mexico has managed to obtain satellite data for the Netherlands with NASA / USGS. This request is very advanced, hopefully, hard drives will be coming to Holland soon.
- An introduction to the ML project presented at the ModernStats workshop in Geneva was attended by over 50 people and was very well received. Plenty of further discussion took place during the coffee breaks and over dinner. This led to Statistics Finland joining the ML distribution list. The presentation and other final documents are available at Project documents.
- In response to suggestions provided by the Exec Board at last month’s meeting and the ones before, a Related international activities wiki page was setup to list international activities and their potential links to the ML project. Most of these activities have at least on ML team member or person on the ML distribution list. A meeting took place with Jon Blake to learn more about the Global Platform for Data.
- A solution was found to enable members to share code.
- Initial information on the Machine Learning project - Second sprint was communicated to team members. The sprint will be held in Belgrade, Serbia, from September 18 to 20. Some e-mail invitations have been sent.
- The project continues to accumulate papers related to the use of ML for statistics. They are available at the ML Reference documents wiki page. They are listed in a manner to direct users to papers that refer the most to certain key themes of the ML pilot studies.

Next Steps

- Work on the work packages and pilot studies, and document them along the way
- Finalize the agenda and logistics of the second sprint
- Look up the Global Platform for Data’s Marketplace for potential collaborations or synergies.

Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each work package and pilot study is progressing, but their objectives remain ambitious and their deadlines are tight.</td>
<td>Progress is regularly reported. Studies will be documented as they are executed. The planning of the September sprint will focus on identifying and delivering key priorities.</td>
</tr>
</tbody>
</table>

Strategic Communication Framework phase 2

Progress

Work continues to progress albeit at a much slower pace due to summer vacation for project team members. The monthly webex meeting was held on July 16 and the leaders of the work packages have reached out to their respective colleagues to confirm commitment to the work that has to be completed by the end of July. Response has been positive with team members confirming their participation and input.

Draft documents are expected by the end of July/beginning of August for wider circulation and consultation within the project team.

A template has been created by the Netherlands to gather case studies and best practices for inclusion in the final products.

Feedback from the CES Seminar on national data strategies has not altered the direction of work package 3. This will continue, as described last month along with the other two work packages.

Next Steps

- Continue to work on the project plan as described last month.
- Draft versions of the three work packages are scheduled for wider circulation and input by mid-August.

Risks and Issues
Work is progressing slowly. Without the dedicated contribution of team members, deadlines will be hard to reach.

Consultation with other HLG-MOS groups/activities (ie Developing Organization Capability) is underway to incorporate any information or case studies that have already been collected.

### News from the Groups

<table>
<thead>
<tr>
<th>Blue-skies Thinking</th>
<th>Identifying Topics/Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>? Unknown Attachment</td>
<td>IN PROGRESS</td>
</tr>
</tbody>
</table>

### Developing Organisational Capability

![Tree Image]

Skills and Capability Framework  

<table>
<thead>
<tr>
<th>One day sprint session to advance this work was set up on 10 September in Geneva, just before the Workshop on Culture Evolution</th>
<th>IN PROGRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN PROGRESS</td>
<td></td>
</tr>
</tbody>
</table>
### Promotional Forum

**IN PROGRESS**

Prepared new flyer of the Group that was distributed during the CES at the end of June in Paris.

### Setting Division in NSOs

**IN PROGRESS**

Work is being done by the Communication team on this activity.

### Other

Workshop on Culture Evolution will take place on 11-13 September in Geneva. We got a confirmation recently, that we will have a session by Consulting company "Like Minds Advisory" about the ABS experience of culture change.

## Supporting Standards

![Diagram of Supporting Standards]

### Linking GSIM and GSBPM

**IN PROGRESS**

The task team is meeting regularly every three weeks. A template for the mapping has been agreed. The mapping is being done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM. The task team is concentrating on phase 5 of the GSBPM both at a design and implementation level.

A Diagram depicting the GSIM objects used in the template at the conceptual level together with notes on their definitions and usage was prepared and discussed by the task team. A similar Diagram will be developed at the implementation level. Task team members are reviewing the mapping done for sub-processes of phase 5 according to the definitions and notes in the Diagram.

As already mentioned to the Executive Board meeting in May, the task team will be able to do the mapping for phase 5 and also phase 4 and/or 6 of the GSBPM. It is likely that the task team will complete the mapping in 2020.

### Core Ontology

**IN PROGRESS**

The development of the core ontology goes on at a steady pace, via virtual meetings and offline exchanges. The construction of the model first focused on the integrated view between GSBPM and GAMSO, with discussions on the connection between the notions of activity and process. Currently, the work is focusing on the mapping of GSIM objects. A first version of the ontology will be available for presentation at the HLG meeting in November and afterwards submitted to public review.

### Alignment GSBPM and GAMSO

**IN PROGRESS**

The task team has produced a document specifying the activity to be done. The document is regularly updated after each meeting. Agreement has been reached on the content and the structure of the document. The premises of the document were consolidated. The discussion on the overarching processes in GSBPM and their relationships to GAMSO corporate support activities, started with the OP "Quality management".
### Metadata
The work is on-going at a regular pace both through webmeetings and offline work. The task team is currently reviewing the definitions for GSIM v1.2 information objects. The review checks whether the definitions provided are compliant with rules (for definitions), precise and exhaustive. In case a definition shows some inconsistency or need for revision, the task team provides suggestions for changes but do not actually change them.

### Other
The Modernisation World Workshop 2019 was quite successful. Besides presenting the work of UNECE on modernisation (from Supporting Standards and Sharing Tools Groups), countries presented their experiences and current challenges. The interactive sessions proved to be highly appreciated by participants. They provided a lot of suggestions for further work and improvements of the models. A particularly worth mentioning issue raised, was the need to set up a task team collecting countries needs and challenges in using GSIM to avoid countries adopting different solutions to the same problem. It was also suggested to have a sprint on GSIM next summer aimed at a light revision of the model for handling most urgent issues without waiting for the 5-year revision.

### Sharing Tools
#### Digitizing CSPA document
First the CSPA Document will be updated and edited. The Application Architecture section has nearly been completely rewritten. The 'Common Statistical Production Architecture' section had some significant changes. Smaller changes were made on: Business Architecture, Information Architecture, Enablers.

CSPA V2.0 Overview (New) was added.

#### Adding Services to Catalogue
Several new services were added to the Catalogue. During the ModernStats World Workshop we will have interactive activities in which potential services will be identified and some selected will be tried to add to the catalogue.

#### Communication re stated CSPA
Several members were active with preparations for the ModernStats World Workshop. We had several presentations and interactive sessions to promote and advance CSPA. Results will be used.
August 2019

Machine Learning

Progress
(Texts in bold blue colour provide links to more detailed information)

People continue to join our project. The project now has 36 members and 23 "followers and additional collaborators" (see: Team Members).

Members continue to send Reference documents. We now have 73 documents.

Good 2019-08-20: Meeting has taken place in spite of the summer vacation period.

Statistics Poland completed a first study on coding ECOICOP and E&I using ML. Statistics Canada have provided a paper on the ML application that they added to their G-CODE generalized system and are in the process of sharing some code (FastText) with Statistics Norway. A discussion was initiated on the possibility of sharing their Statistics Poland's ECOICOP data. It appears that with a bit of editing and the required anonymization, the data could be shared. Furthermore, a test was conducted to translate the data to English using an application available on the Internet. There are still many things to work out, but things look very promising.

The ONS has contacted Statistics New Zealand on their research in using ML for editing.

INEGI drafted a Imagery-Pipeline.pptx that has generated a lot of constructive exchanges to generalize and improve it. As a test, it is being applied to Statistics Canada’s crop yield estimation (as an example of using satellite data for regression with continuous variable; this case study is found in UN GWG Report). Statistics Canada is providing any additional information required.

With the assistance of INEGI and USGS, Statistics Netherlands have received pre-processed satellite data (30 m, 11 channels, EPSG:4326 (WGS 84 -- WGS84 - World Geodetic System 1984, used in GPS). They now have several sources of imagery data to combine with geotagged income data per household and aggregated data on various statistics per square (100 x 100 m and 500 x 500 m), including the level of urbanization (1-5). Their challenge now is to find the tools to cut the images into rectangles, and to scale up this process. The next steps will be to link the statistical information and to build a neural network.

An updated version of a quality framework was circulated and discussed. A literature review on ML quality was initiated by Statistics Poland (for continuous data) and Australia Bureau of Statistics (for classification).

Follow-up discussions took place on the UN Global Platform’s Marketplace. It lists a few projects that are related to ML. However, “The information used in the marketplace was drawn from UNSD project information, the users have not given their explicit consent to have their contact details revealed so this information will be added in time as each record updated and claimed by the Collaborative teams”.

Next Steps
A sprint will be held in Belgrade from September 18 to 20. Work will be conducted towards the objectives of the ML project and decide on the final deliverables and timelines leading to the November 19-21 HLG-MOS workshop.

The project made very good progress during the summer, but much more will be carried in the next three months.
Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each work package and pilot study is progressing, but their objectives remain ambitious and their deadlines are tight.</td>
<td>Progress is regularly reported. Studies will be documented as they are executed. The planning of the September sprint will focus on identifying and delivering key priorities.</td>
</tr>
</tbody>
</table>

**Strategic Communication Framework phase 2**

**Progress**

Progress continues on the three work packages albeit at a very slow pace. The summer months have proved quite challenging with very little additional new input contributed in July and August. New deadlines were established at the monthly Webex meeting on Monday, August 26th. Individual contributions will be confirmed over the next week to ensure the original work plans are completed and products developed for the HLG-MOS Workshop in November.

Members of the Developing Organizational Capabilities group joined the monthly webex to discuss common work objectives.

**Next Steps**

Special meetings will be scheduled with the two work groups to ensure deliverables are met within the required timeframe. Follow up with countries who have promised contributions will be conducted within the next two weeks.

**Risks and Issues**

Several project members have agreed to work on specific tasks but have not yet contributed. They are citing summer holidays and heavier than expected workloads within their home organizations as reasons for not contributing as planned.

There have also been concerns raised related to competing priorities with requests from the HLG and the CES.

**News from the Groups**

<table>
<thead>
<tr>
<th>News from the Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blue-skies Thinking</strong></td>
</tr>
<tr>
<td>? Unknown Attachment</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Synthetic Data Sets and Secure Multi-party Computation are the first topics to follow-up on. There was limited progress made during the summer months but members are assigned to work on the topics that were selected. In separate webexes the topics will be further discussed and a standard template will be used to report on them.

The topics of ‘Data Science Lab’ and ‘Integral Platform for IT Support’ are being refined.

Expected progress will be after sprint session on 10 September.

Since preparation of new flyer of the Group that was distributed during the CES at the end of June in Paris, there is no progress.

We are finalising preparation of Workshop on Culture Evolution which will take place on 11-13 September in Geneva.

The task team is meeting regularly every three weeks. A template for the mapping has been agreed. The mapping is being done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM. The task team is concentrating on phase 5 of the GSBPM both at a design and implementation level.

A Diagram depicting the GSIM objects used in the template at the conceptual level together with notes on their definitions and usage was prepared and discussed by the task team. A similar Diagram was developed also for the implementation level.

Task team members are reviewing the mapping done for sub-processes of phase 5 according to the definitions and notes in the Diagram.

As already mentioned to the Executive Board meeting in May, the task team will be able to do the mapping for phase 5 and maybe for one additional phase 4 or 6 of the GSBPM. Given the great amount of work needed to conceptualise the framework for the mapping and harmonise the country examples, it is likely that the task team will complete the mapping in 2020.
**Core Ontology**

The development of the core ontology goes on at a steady pace, via virtual meetings and offline exchanges. The construction of the model first focused on the integrated view between GSBPM and GAMSO, with discussions on the connection between the notions of activity and process. Currently, the work is focusing on the mapping of GSIM objects. A first version of the ontology will be available for presentation at the HLG meeting in November and afterwards submitted to public review.

**Alignment GSBPM and GAMSO**

The task team has produced a document specifying the activity to be done. The document is regularly updated after each meeting. Agreement has been reached on the content and the structure of the document. The premises of the document were consolidated. The discussion on the overarching processes in GSBPM and their relationships to GAMSO corporate support activities, started with the OP “Quality management”.

The task team is now completing the description of the other OPs in GSBPM and the corresponding activities in GAMSO. The task team expects to have a consolidated version of the document to be discussed in the meeting in September.

**Metadata Glossary**

The work is on-going at a regular pace both through web meetings and off line work. The task team is currently reviewing the definitions for GSIM v1.2 information objects. The review checks whether the definitions provided are compliant with rules (for definitions), precise and exhaustive. In case a definition shows some inconsistency or need for revision, the task team provides suggestions for changes but do not actually change them.

**Sharing Tools**

The draft CSPA 2.0 document is ready. All countries represented in the group will now revise the document within their offices. Finland was also prepared to go through the document. Comments are welcome until 27 September.

**Other Activities**

Services that were added recently need to be checked for completeness. The group will continue to add new services. Services developed within ESS Shared Tools Expert Group will also be added.

**Communication related CSPA**

Work will start once the CSPA 2.0 document and digital version are ready.
Other September 2019

**Machine Learning**

Progress

- The sprint was held in Belgrade, Serbia from September 18 to 20.
- It was attended by 16 team members from 11 countries or organisations, one person from Russia, 3 persons from the Czech Republic and several employees from the Statistical Office of the Republic of Serbia (SORS).
- We wish to thank immensely the SORS for setting up the sprint and, in particular, Mira Nikic, Branko Josipovic and Sinisa Cimbaljevic for their warm hospitality, as well as the employees of SORS who provided us with drinks and sweets and worked in the background to make this sprint pleasant and successful. We also warmly thank Dr Miladin Kovaevi, Director of the Statistical Office of the Republic of Serbia, for his very thoughtful opening remarks.
- The meeting was also attended by at least 8 other persons at different times by Webex.
- Fourteen studies and other topics were presented and discussed; another presentation as well as four working papers were also provided in advance. These documents are all available on the project's Working documents page. A draft pipeline on the use of satellite imagery and a dataset of product descriptions in Polish coded using Machine Learning were shared and discussed. The dataset was also automatically translated in English (please note that the translations were not reviewed).
- The presentations highlighted the significant progress achieved on many fronts in a relatively short amount of time (four months since the first sprint focused on planning). Organisations have gained knowledge and ability to classify data using ML. Early studies on the use of ML for editing and imputation are more positive than expected. The imagery project in Mexico using satellite data to measure urban/rural growth provided an opportunity to test, improve and generalize the pipeline. Imagery data was acquired by the CBS with the assistance of Mexico, and a coded dataset has been shared within the group to practice ML solutions.
- The sprint also enabled further progress, e.g. NACE classification by Serbia with the collaboration of USA-BLS. It also initiated new collaborations, e.g. Poland sharing classification developments with UK, Mexico is now involved in sentiment analysis with Belgium and Poland. Canada will consider Support Vector Machine techniques in their coding applications based on the experiences of others. Results from studies on the use of ML for imputation generated considerable discussions on the quality of the predictions (better than expected) and of the distributions (at times, problematic). More work on this will be conducted in collaboration with the work package on quality. Furthermore, the project not only generated collaboration within the team, but numerous new collaborations within organisations were reported.

**Next Steps**

- While the work by the team continues, the project manager and the work package leaders will meet at the beginning of October in Washington and Ottawa to prepare the material for the workshop.
- Numerous potential deliverables from the project were discussed, including:
  - Reports on the results of studies (demonstration of value-added from ML), knowledge acquired and advancement of ML within the organisation
  - Collective conclusions and recommended best practices
  - A pipeline describing the main steps in setting up and maintaining solutions based on ML
Documented code for others to used
- Instructional products combining best practices and code to enable other organisations to consider ML solutions. These products could take the form of documents, a workshop, etc.
- A draft quality framework and reference documents supporting all the above.
- The subject of how a statistical office is organized to bring a ML solution from an idea to full implementation was mentioned several times before and during the sprint. The situation where an organisation has several pockets of employees looking into ML solutions with more or less efficient communication is not uncommon. As a possible extension of lessons learned, the project will propose looking at this aspect and compare different models present in organisations.

## Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The project has enabled participating organisations to advance in the use of ML solutions. It is also providing more demonstrations what ML can and cannot do. By the end of this year, the project will be in a position share the results of many of its studies and work. However, more time will be needed to produce the list of deliverables listed above to fully shared the work of the team with other organisations.</td>
<td>Propose the extension of the project to a second year at the November workshop.</td>
</tr>
</tbody>
</table>

## Strategic Communication Framework phase 2

### Progress

Works is progressing steadily on the development of content for the work packages. A final draft is expected for mid to late October. Initially we had hoped to have the content finalized for the end of September but delays have occurred due to heavy workloads for project team members within their respective statistical institutions. Competing priorities with other international activities also continue to be a challenge.

### Next Steps

Work will continue on the content of the work packages. We expect to have a final draft of the products ready for presentation to the HLG-MOS workshop in November. The case studies will not be ready for November but will be included in the final products schedule to be completed by the end of the calendar year or early January 2020.

### Risks and Issues

Project team members are experiencing heavy workloads within their respective national statistical institutions. This pressure combined with competing priorities at the international level have significantly impacted the speed at which we are able to develop content for the intended products. We have adjusted timelines and shifted deliverables to allow project team members more time to develop their contributions.

## News from the Groups

**Blue-skies Thinking**

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>Unknown Attachment</td>
<td>IN PROGRESS</td>
</tr>
</tbody>
</table>

No new topics have been identified. Neither project proposals have been received.
### Teams have been created to write a 3-5 pager on Synthetic Data Sets and Secure Multi-party Computation that will answer the core questions. Information has been shared on Chat bots. This topic will be presented at the Modernisation Workshop to discuss follow up. The topics of 'Data Science Lab' and 'Integral Platform for IT Support' are being refined.

### We had a sprint session in Geneva, on 10 September. We decided to prepare document on November workshop. It will be short one presenting framework of core skills, leadership skills and expert skills. For next year we will propose next stage of our work including possible aligned with GAMSO model.

### No progress. We are in touch with communication group. For next year we will propose joint meeting between workshops on HRMT and communication.

### The task team is meeting regularly every three weeks. A template for the mapping has been agreed. The mapping is being done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM. The task team is concentrating on phase 5 of the GSBPM both at a design and implementation level.

A Diagram depicting the GSIM objects used in the template at the conceptual level together with notes on their definitions and usage was prepared and discussed by the task team. A similar Diagram was developed also for the implementation level.

Task team members are reviewing the mapping done for sub-processes of phase 5 according to the definitions and notes in the Diagram. The task team is also discussing the degree of harmonisation of countries' examples.

Given the great amount of work needed to conceptualise the framework for the mapping and harmonise the country examples, it is likely that the task team will complete the mapping in 2020.
**Core Ontology**

The development of the core ontology goes on at a steady pace, via virtual meetings and offline exchanges. The construction of the model first focused on the integrated view between GSBPM and GAMSO, with discussions on the connection between the notions of activity and process. Currently, the work is focusing on the mapping of GSIM objects. A first version of the ontology will be available for presentation at the HLG meeting in November and afterwards submitted to public review.

Despite the great amount of work already done, it is likely that the task team will ask the EB to complete its work during 2020.

**Alignment of GSBPM and GAMSO**

The task team has produced a document specifying the activity to be done. The document is regularly updated after each meeting. Agreement has been reached on the content and the structure of the document. The premises of the document were consolidated.

The task team has completed the description of the OPs in GSBPM and the corresponding activities in GAMSO. The task team has produced a consolidated version of the document to be discussed at next meeting in September.

It is expected that the task team will conclude its work in 2019 and present it to the HLG Workshop in November.

**Method at a Glossary**

The work is on-going at a regular pace both through web meetings and off line work. The task team is currently reviewing the definitions for GSIM v1.2 information objects.

The review checks whether the definitions provided are compliant with rules (for definitions), precise and exhaustive. In case a definition shows some inconsistency or need for revision, the task team provides suggestions for changes but do not actually change them.

**Other**

The Supporting Standards Group started discussing the future activities for 2020 to be presented at the HLG workshop. The proposals will be further discussed during next meeting of the group at beginning of October.

The Group will also take into account the needs and requests emerged last June during the ModernStat World Workshop in Geneva.

**Sharing Tools**

CSPA V2.0 is currently being reviewed by country members of the Sharing Tools Group. This will continue until the end of September (all have agreed to this date). All comments will be reviewed and responded to and the document will be finalized through October. Digitalisation will occur after this is completed.

Requests for changes to the catalogue are being gathered and reviewed and will be prioritized in Jira. Eurostat have advised that they may be able to make the highest priority changes.
In October/November we will be discussing what will be handed over to the Standards Group.

Other

October 2019

**Machine Learning**

**Progress**

Eric Deeben traveled to Washington to meet Alex Measure at the Bureau of Labor Statistics on October 7 and 8. Both Eric and Alex then traveled to Ottawa to meet Wesley Yung and Claude Julien at Statistics Canada to work on the material needed for the November Modernisation workshop.

- At the BLS, Eric presented an overview of international activities at the ONS which included an overview of the ML project and Claus Sthamer presented his work on editing and Alistair Edwards presented the Data Science Campus and Earth Observation examples via Webex.
- At Statistics Canada, Claude presented a summary of the project to senior managers at Statistics Canada. Claude and the work package leaders met with Mr. Anil Arora, Chief Statistician of Canada and chair of the HLG-MOS, and Stéphane Dufour, Assistant Chief Statistician and chair of the HLG-MOS Executive Board. Mr. Arora and Mr. Dufour shared their thoughts and expectations on machine learning and the project.
- Eric also visited the International Monetary Fund who expressed a strong interest in joining the ML project. Eric and Claude will be discussing this with the IMF.
- The work package leaders agreed on the following project deliverables based on the work currently underway:
  - A report on each pilot study by team members conducting the studies
  - A summary report and documented code on each pilot study theme (coding, E&I and Imagery) by pilot study theme leaders
  - A machine learning summary report by the project manager and work package leaders:
    - Best practices on the development and implementation of ML
    - Quality framework with applications to two pilot studies
    - Organisational aspects
  - A workshop on machine learning and an application
- Statistics Poland coded their web-scraped product descriptions that had been translated to English and French. They achieved an accuracy rate that was just a few %-points lower than the original data in Polish. Statistics Canada also coded the descriptions in each language using their FastText application and achieved similar results.
Statistics Poland coded a file that combined some of their web-scraped product descriptions with other descriptions provided directly by supermarkets. The descriptions covered more ECOICOP classes and finer classes (6-digit rather than 5-digit). Again, they attained an accuracy rate that was just 1.5% lower than on their web-scraped data only.

Following the latest sprint, the Statistical Office of the Republic of Serbia collaborating with the USA and Poland on industry coding. They achieved 70% accuracy with logistic regression without using some special lemmatization and stemming tools.

Wesley Yung organized an invited session on machine learning for the Fourth International Conference on Establishment Statistics (ICES-IV). The session was accepted.

Statistics Finland joined the project team, and four persons were added to the followers and additional collaborators list.

Next Steps

- Continue the pilot studies
- Prepare the material to be presented at the Modernisation workshop
- Hold a project meeting prior to the Modernisation workshop

Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Same as the September update) The project has enabled participating organisations to advance in the use of ML solutions. It is also providing more demonstrations what ML can and cannot do. By the end of this year, the project will be in a position share the results of many of its studies and work. However, more time will be needed to produce the list of deliverables listed above to fully shared the work of the team with other organisations.</td>
<td>Propose the extension of the project to a second year at the November workshop.</td>
</tr>
<tr>
<td>Beyond the demonstrations, codes, best practices and quality framework that will be delivered by the project, discussions taking place within the team and with some stakeholders highlight the need to address organisational aspects that would facilitate the full integration of ML into the production of official statistics. To address this adequately, senior level managers will be required.</td>
<td>Make a call for people with the required knowledge and level to address organisational aspects that wouold facilitate the integration of ML.</td>
</tr>
</tbody>
</table>

Strategic Communication Framework phase 2

Progress

Significant progress has been made on the project since September. Of the two planned output documents, the content has been fully developed for one (stakeholder engagement and strategic communications considerations for national data strategies). The content related to mission, vision, values and internal communications is expected to be finished by the end of this week. The documents will be circulated to the full project team for input and comments by the end of October, first week in November.

Next Steps

Finalize the content of the documents and circulate for feedback.

The October meeting of the project team is delayed until October 29th due to scheduling conflicts. The focus of this meeting will be reviewing the content of the documents and a discussion about what will be presented in November at the HLG-MOS Workshop.

Development of business cases and best practices is still underway.

Risks and Issues
As long as the remaining content is delivered on schedule, the project be able to deliver as presented in September. However, any further delay in content development of the remaining work package will put the project outputs at risk.

### News from the Groups

#### Blue-skies Thinking

**Identifying Topics/Opportunities**

The focus is first on the two identified priority topics.

**Follow-up selected topics**

A first paper has been drafted on Secure multi-party computation. It will be expanded further and a presentation and potentially a project proposal will be prepared.

For 'Synthetic data sets', a plan has been outlined on how to move forward.

**Other**

The five project proposals that were submitted so far will be discussed on Thursday 25 October.

#### Developing Organisational Capability

**Skills and Capability Framework**

We are finalising document for November Workshop. Next week, we will have additional webex call.

**Promotion Forum**

No progress

**Setting vision in NSCs**

No progress.

**Culture Evolution Workshop**

FINISHED

**Other**

Unknown Attachment
### Supporting Standards

| Linking GSBP M and GSIM | IN PROGRESS | The task team is meeting regularly every three weeks. A template for the mapping has been agreed. The mapping is being done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM. The task team is concentrating on phase 5 of the GSBPM both at a design and implementation level.

A Diagram depicting the GSIM objects used in the template at the conceptual level together with notes on their definitions and usage was prepared and discussed by the task team. A similar Diagram was developed also for the implementation level.

Task team members are reviewing the mapping done for sub-processes of phase 5 according to the definitions and notes in the Diagram. The task team is also discussing the degree of harmonisation of countries’ examples.

The task team is preparing a document describing the activities done in 2019.

Given the great amount of work needed to conceptualise the framework for the mapping and harmonise the country examples, it is likely that the task team will complete the mapping in 2020.

### Core Ontology

| IN PROGRESS | The development of the core ontology goes on at a steady pace, via virtual meetings and offline exchanges. The construction of the model first focused on the integrated view between GSBPM and GAMSO, with discussions on the connection between the notions of activity and process. Currently, the work is focusing on the mapping of GSIM objects. A first version of the ontology will be available for presentation at the HLG meeting in November and afterwards submitted to public review.

Despite the great amount of work already done, it is likely that the task team will ask the EB to complete its work during 2020.

### Alignment GSBPM and GAMSO

| IN PROGRESS | The task team has finalised the document A model to support the synergistic implementation of GAMSO Activities and GSBPM Overarching Processes describing the activity done.

The task team will present the results of the activity at the HLG Workshop in November.

### Metadat a Glos sary

| IN PROGRESS | The work is on-going at a regular pace both through web meetings and off line work. The task team is currently reviewing the definitions for GSIM v1.2 information objects.

The review checks whether the definitions provided are compliant with rules (for definitions), precise and exhaustive. In case a definition shows some inconsistency or need for revision, the task team provides suggestions for changes but do not actually change them.

### Other

| | | | | |
Sharing Tools

Digitizing/Editing CSP A document IN PROGRESS
Several comments have been received and they are now being compiled/processed to improve the CSP A 2.0 document and make it ready for consultation/endorsement. Digitization has not started yet.

Adding Services to Catalogue IN PROGRESS
Several services were added to the catalogue. Team members will check for missing information. Service developed by the ESSnet on Sharing Tools will also be added once ready. See: https://www.statistical-services.org/

Communication postponed

Other

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLG-MOS (3 March, NYC)</td>
<td>3 March, NYC</td>
</tr>
<tr>
<td>Strategic Communication sprint (30 April-2 May, Geneva)</td>
<td>30 April-2 May, Geneva</td>
</tr>
<tr>
<td>Machine Learning Sprint (13-15 May, UK)</td>
<td>13-15 May, UK</td>
</tr>
<tr>
<td>Date Integration Workshop (21-23 May, Belgrade)</td>
<td>21-23 May, Belgrade</td>
</tr>
<tr>
<td>Strategic Communication Sprint 2 (10 - 11 June, Gdansk)</td>
<td>10 - 11 June, Gdansk</td>
</tr>
<tr>
<td>DissComm Workshop (12 - 14 June, Gdansk)</td>
<td>12 - 14 June, Gdansk</td>
</tr>
<tr>
<td>Modern$^{!!}$ Worksho (28 Jun, Geneva)</td>
<td>28 Jun, Geneva</td>
</tr>
</tbody>
</table>

**November-December 2019**

**Machine Learning**

Progress

- A draft background document on the progress of the project and a proposal to extend it was written for the Workshop on Modernisation. Team members will have time to provide comments, changes and additions to the document that will then become a end of year report that will be delivered early in January 2020.
- A half-day meeting attended by 10 team members, some via Webex, took place on November 18.
Proposed project outputs and timelines were discussed and approved by work package and pilot study theme leaders. The scope and definition of Work Package 3 on Integration of Machine Learning in Organisations was discussed, clarified and agreed.

A study on the use of ML to interpret aerial photos and combine them with satellite images was presented by the Switzerland Federal Statistical Office. Developments and results on pilot studies continue to be presented and discussed at the monthly update meeting. In November (25 attendees), Imputation of energy statistics at VITO was the topic. Following the discussion, the potential for this project to produce a shared dataset to practice and learn ML in the context of official statistics. In December (29 attendees), the Aerial Imagery - Australia.pptx at the ABS was the topic. It may lead to collaboration with the Switzerland who are experimenting with ML for a similar purpose.

Following suggestions expressed in discussions at the September sprint in Belgrade, some pilot studies have improved the results obtained with the use of ML. Some pilot study results are being presented to senior management in participating organisations. Some pilots studies and their results have been presented at Geo Week in Canberra and the Istat Advisory committee. Most of the pilot studies on E&I have been submitted to the UNECE Workshop on Statistical Data Editing.

Following the November workshop and its support on the extension of the ML project, a review of participants was initiated. The vast majority of the team members have confirmed their engagement in the project. Several organisations have joined the project (Statistics Sweden, INSEE, Statistics New Zealand) and a few more people from participating organisations have joined the project as active members and collaborators/ followers. After the review, some participants will be disconnected from the project. Other collaborations are being established.

Istat have proposed a new pilot study on using ML to edit administrative data. This is a good addition and complement to the other studies on ML for E&I.

The project’s planned outputs and delivery dates have been established. A Plans for 2020 page has been initiated. More information will be provided in the next weeks.

Next Steps

- Complete the review of project participants, including the addition of new participants
- Identify the space where the project outputs will be shared. Discussions on the possibility of using the Global Working Group - Global Platform have resumed.
- Finalize the scope and planned activities for the work package on organisational aspects in the integration of ML
- Finalize and communicate the expected content and structure of the pilot study reports due at the end of February
- Organize a sprint to be held likely in early April (between two key delivery dates)

Risks and Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>Mitigation</th>
</tr>
</thead>
</table>
| Satisfying the needs and expectations of new participants and organisations, while ensuring the timely delivery the outputs based on all the work currently in progress. | • Strong continued engagement and focus from participants  
• Clear communication of expectations from new participants  
• Allow for review of outputs as any new work progresses and highlight new results, best practices, etc. |

Strategic Communication Framework phase 2

Progress

The content for the three work packages has been completed and produced in the two remaining components of the Strategic Communications Framework for Statistical Institutions. The two documents, Stakeholder Engagement and Internal Communications, were presented to the HLG-MOS Workshop in November. The presentation was well received and excellent comments during the discussion will be incorporated in the final update.

The documents are currently in the final editing stage as the project team members await comments from the HLG-MOS Executive Committee. Once this process is completed the documents will be updated and prepared for consultation with CES member countries.
Next Steps

Project members continue to seek additional case studies to include in the final output.

Documents will be updated with input received from the HLG-MOS Executive Committee. Once the consultation with CES member countries is completed, the final documents will be prepared for presentation to the CES in Summer 2020.

The communications work will continue under the direction of the Developing Organizational Resilience Working Group. Strategic Communications Project members have been encouraged to join this working group to ensure continuity and knowledge transfer. To date, the Project manager has received positive feedback from members regarding continued commitment to the HLG-MOS work and several will likely join the working group.

The project members have also been encouraged to participate in the virtual sprint (should it take place) to identify priority work for 2020 for the Developing Organizational Resilience Working Group. The project team is developing a list of activities to support the continued development of the Framework to be considered during this sprint session.

Risks and Issues

There is a risk that the excellent work of the communications community loses momentum now that the project has ended. There is also a risk that the knowledge gained during the two-year project will disappear and the Framework will not evolve as a living document.

To mitigate these risks, all members of the project team have been encouraged to continue their participation by joining the Developing Organizational Resilience Working Group. Also, with endorsement by the CES of all components for the Framework, it will receive broader support and exposure internationally. The working group should continue to gather case studies of countries using the Strategic Framework to show how the framework is applied in various statistical environments.

<table>
<thead>
<tr>
<th>News from the Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue-skies Thinking</td>
</tr>
<tr>
<td>? Unknown Attachment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identifying Topics/Opportunities</th>
<th>IN PROGRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Follow-up selected topics</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>
As far as you know we will be merged with communication group. So we need to discuss and prioritize our activities. The former DOC group presented several activities proposals during November workshop. It was decided that in January or at latest at early February we will have sprint session together with communication colleagues. After this event it will be possible to indicate our activities for 2020. For sure we will organize two workshops (on communication and on HRMT) in September in Lisbon with one common day.

The majority of DOC group participants confirmed their participation for next year. As for communication experts we will have still wait for their confirmation as communication project phase 2 was concluded.

<table>
<thead>
<tr>
<th>Skill and Capability Framework</th>
<th>IN PROGRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promotion Forum</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Setting Division in NSOs</td>
<td>IN PROGRESS</td>
</tr>
<tr>
<td>Culture Evolution Workshop</td>
<td>FINISHED</td>
</tr>
<tr>
<td>Other</td>
<td>IN PROGRESS</td>
</tr>
</tbody>
</table>

**Supporting Standards**

As far as you know we will be merged with communication group. So we need to discuss and prioritize our activities. The former DOC group presented several activities proposals during November workshop. It was decided that in January or at latest at early February we will have sprint session together with communication colleagues. After this event it will be possible to indicate our activities for 2020. For sure we will organize two workshops (on communication and on HRMT) in September in Lisbon with one common day.

The majority of DOC group participants confirmed their participation for next year. As for communication experts we will have still wait for their confirmation as communication project phase 2 was concluded.

| Linking GSBPM and GSIM        | IN PROGRESS |
| Core Ontology                | IN PROGRESS |
| Supporting Standards          | IN PROGRESS |

The task team is finalising a document describing the work done in 2019.

The mapping has been done for GSBPM sub-processes of phase 5 “Process” according to a conceptual schema and definitions developed by the task team. The mapping is done at two different levels of GSIM: a) a more conceptual level, corresponding to the specification level in GSIM; b) a less conceptual level, corresponding to the execution level in GSIM.

The work done was presented at the HLG Workshop in November.

The task team will continue its work in 2020.

A first version of the ontology was presented at the HLG meeting in November and will then submitted to public review.

The task team will complete its work during 2020.
### Alignment GSBPM and GAMSO

- **IN PROGRESS**
- The task team has concluded its activities.
- The achieved results are described in the document *A model to support the synergistic implementation of GAMSO Activities and GSBPM Overarching Processes*.
- The activities done and the achieved results were presented at the HLG Workshop in November.

### Metadata at a Glossary

- **IN PROGRESS**
- The work is on-going at a regular pace both through web meetings and off line work. The task team is currently reviewing the definitions for GSIM v1.2 information objects. In case a definition shows some inconsistency or need for revision, the task team provides suggestions for changes but do not actually change them.
- The activity was presented at the HLG Workshop in November.
- The task team will continue its work in 2020 and plans to conclude it by April 2020.

### Other

#### Sharing Tools

- **IN PROGRESS**
- The online version of CSPA 2.0 will be ready by the end of this year. It will be saved as a word document as well. The next step will be to submit it to the Bureau of the CES to decide how it can be endorsed by the CES.
- The CSPA 2.0 will not be a static document. The foundational part will be relatively stable but for example the application architecture part needs continuous updates to incorporate new developments in IT/Architecture. We will therefore include a roadmap for future updates to keep it up to date with recent developments.
- The Wiki version will be turned into a Word document as well. The aim is to create an interactive version that can be navigated easily depending on your level and role. It will be aligned with the work done under the ESS Sharing Tools project. The ‘building’ representation from Jakobs (Prezi) presentation might be a good structure to be used for this.

#### Digitizing/editing CSPA document

- **IN PROGRESS**
- The CSPA catalogue now has over 30 services listed. The Sharing Tools group will keep on adding services and assist others to do so.

#### Adding Services to CSPA catalogue

- **POSTPONED**
- Postponed to 2020. Depends on the availability of communication experts as well.
<table>
<thead>
<tr>
<th>Event Name</th>
<th>Date/Dates</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLG-MOS</td>
<td>3 March, NYC</td>
<td></td>
</tr>
<tr>
<td>Strategic Communication Sprint</td>
<td>30 April-2 May, Geneva</td>
<td></td>
</tr>
<tr>
<td>Machine Learning Sprint</td>
<td>13-15 May, UK</td>
<td></td>
</tr>
<tr>
<td>Date Integration Workshop</td>
<td>21-23 May, Belgrade</td>
<td></td>
</tr>
<tr>
<td>Strategic Communication Sprint 2</td>
<td>10 - 11 June, Gdansk</td>
<td></td>
</tr>
<tr>
<td>DissComm Workshop</td>
<td>12 - 14 June, Gdansk</td>
<td></td>
</tr>
<tr>
<td>ModernStats Workshop</td>
<td>28 Jun, Geneva</td>
<td></td>
</tr>
</tbody>
</table>

Rosemary McGrath will step down as a chair by the end of 2019. The group is still looking for a new chair (or co-chairs).