Project and activity proposals for 2022

UNECE HLG-MOS Blue Skies Thinking Core Team, October 2021

Introduction
During the summer and autumn of 2021, the BSTN core team has organized sessions where several ideas for projects and/or activities were pitched that might be undertaken during 2022, given sufficient interest. These pitches have led to a number of proposals for projects and/or activities that are briefly enumerated below. Being proposals, they are open for discussion and amendment, both in the HLG-MOS Executive Board, the upcoming HLG-MOS workshop and the respective modernization committees and other bodies they have a relation with. Depending on appetite for participation and endorsement by the HLG-MOS we decide how to proceed and organize follow-up.

Some activity proposals have logical connections with other HLG-MOS related work. In principle, the BSTN core team is willing to take care of launching and overseeing the activities proposed; but if another group is interested in adopting an activity that could be an option as well.

In some cases activities have been suggested that have not been further elaborated. In those cases the BSTN will further articulate these activities early 2022 and try to launch them (low profile) under its own responsibility.

An open question is whether the new community proposed should report to the Exec Board or elsewhere.

Project proposals
- Data Interoperability and Data Governance. Proposal upcoming
- Meta Academy. Proposal available
- Input privacy preserving techniques. Continuation proposal from current project (not through BSTN)

Withdrawn proposal
- StatsBot. Despite the fact that the HLG-MOS workshop in November 2020 expressed quite some support for continuing this project, it has been decided not to submit a new proposal at this stage. See the Annex for some further background

Activity proposals
- Digital Twins. Proposal available (BSTN)
- Future of Work. Proposal upcoming (CapComm)
- Interacting with Generation Y. Proposal available (CapComm)
- Mobile Survey Data Collection for Climate Change. Not yet articulated
- Nowcasting. Not yet articulated

Community proposal
- Data Virtualisation. New community idea modelled loosely after ML community.

Annex: the StatsBot project

The StatsBot project has been carried out in 2019-20 by a core group of institutions (OECD, StatsCan and CBS) with the objective of testing the possibility of a bot leveraging SDMX semantics, in order to be able to assist in data search through a conversational experience. The Proof-of-Concept was part of a broader discussion on use cases involving bots for official statistics. The project involved mainly a collaboration with a start-up specialised in bots (golem.ai), and focusing on the implementation of conversational access to labour statistics. All background on the project and its outcomes can be found in the report to the executive board.

One of the main conclusion of the project is the following one (see report): "*Drawing on the PoC results, and the outcome of the HLG-MOS workshop, potential project proposals could be scoped, that would benefit from a collective approach (e.g. co-investment by several organisations, vs siloed, organisation-specific investments), in order to achieve our common goal: make our data more accessible and bring reference facts to the online conversation: I) Launch a large scale project for the joint development of a generic StatsBot drawing on statistical sources and semantics. Based on the PoC results, a larger scale market consultation could be carried out, that would lead to the joint development of such a StatsBot (with golem.ai or another provider). The workshop is an opportunity to identify interested organisations. With an investment in the range of 150-300k€, shared amongst interested organisations, a production grade, scalable StatsBot could be envisaged in the short term; …*"

In order for the project to continue as an HLG-MOS project, a co-funding scheme should be agreed between a core group of stakeholders, aiming to mobilise a budget in the range of 300k€ (tbd), which was deemed as the capacity needed for achieving the development of a production grade, generic StatsBot able to tap into any SDMX source, and able to efficiently address multilingual adaptations. Such a development would require a market consultation and an organisation playing the role of coordination (eg receiving the funds, executing the market consultation and managing the project delivery under the supervision of appropriate governance). OECD is able to play that role, especially in the context of the [SIS-CC](SIS-CC) (Statistical Information Collaboration Community) which has precisely as a raison d'être of supporting such projects – if that is wished by interested parties, of course. In fact we could reflect more generally on how OECD could second BSTN in undertakings requiring co-funding to develop IT tools; SIS-CC could potentially host such undertakings. UNECE is comfortable with the idea, especially as rules for managing such co-funded projects are better adapted in OECD.

An interesting development over the past months is that of Switzerland, where at national and local levels, and in collaboration with Swiss academia, it is planned to develop a StatsBot that is open source. An open source approach in this field would be highly desirable to enable an open collaboration, possibility of reuse in a wide variety of contexts, and sustainability over time through contributions of multiple organisations (as we can see for example, for the open source .Stat Suite project). But we don't yet have much information on the Swiss project and should learn more about it in the coming time. Depending on its outcome, perhaps an open source approach to the project could be envisaged.

Conclusion – way forward: drawing on the previous and the general support expressed for a StatsBot project at the HLG-MOS last year, the project could take shape in 2022 but under the precondition of sufficient budget available (eg 300k€) – a precondition not yet met, and which could take time and discussions before it is met. That is why we suggest not to propose the StatsBot proposal this year and rather continue to investigate informally appetite of particular stakeholders (such as Switzerland). Should we get closer to identify interested parties able to finance the project, the Executive Board would be informed in order to seek advice and guidance.