

Business Case for StatsBots

This business case was prepared by **Eric Anvar**, and is submitted to the HLG-MOS for their approval.

Type of Activity			
<input type="checkbox"/>	New project	<input checked="" type="checkbox"/>	New activity
<input type="checkbox"/>	Extension of existing project	<input type="checkbox"/>	Extension of existing activity
<i>Projects are undertaken by separate project teams. Projects are expected to produce a significant contribution to achieving the HLG-MOS vision</i>		<i>Activities are undertaken by Modernisation Groups. These activities produce smaller, more detailed outputs to help achieve the HLG-MOS vision</i>	
<i>See here for more details: https://statswiki.unece.org/display/hlgbas/HLG-MOS+Strategy</i>			
Purpose			
<p>Data deluge continues to grow at an exponentially rate where no longer Statistical organisations, at national and international levels, being the main source for data that feeds the policy making machine. This growth, driven by user demand for data, where users have the ability to search and consume information of any kind through online channels and digital devices, calls for a need to innovate in the dissemination of official statistics, so as to <i>bring reference facts to the conversation</i>.</p> <p>Statistical organisations, as producers of official statistics, must continue to ensure that the right data is available to the right person, at the right time, in the right way. Policy-makers and policy shapers (notably the media) have a key role to play in this regard, as they are the ones who structure the conversation on policy and expose the facts underlying a particular political issue. They need reliable facts that they can access easily, and rely on in their wider conversation with the public. Statistical organisations have a particular role in enabling this conversation around facts with new digital technology emerging that can greatly support this.</p>			
Description of the activity			
<p>Following on from the work to date of the 'working group developing a Statistical Chatbot for Official Statistics' it is proposed to undertake a proof of concept to deliver a StatsBot to address the 'data assistant use case' with a goal to respond to data consumers requests through an automated guided assistant 'StatsBot'. The StatsBot will allow a data consumer to query and retrieve data from a large statistical database (SDMX) without the need to be an expert and perform searching, browsing and customising on statistical tables.</p> <p>The WG engaged with a number of technology providers through a series of webinars. This exercise proved of great value leading to identify a partner (Golem.ai) that promises to potentially take a giant leap forward in this space. Their proposition limits the need to duplicate data, create vast numbers of potential questions (intents) that data consumers may ask, and support any number of languages with minimal effort.</p> <p>The scope of the PoC is to enable a data consumer to enter natural language textual strings into a data assistant interface that connects to a statistical database through a standard (SDMX) interface (API), and retrieve the correct information based on the query entered. The PoC will also test the ease of multi-linguality of the solution. For example, for a dataset in a major language (like French or English) and a solution in a minor language (like Dutch). Multi-linguality has been identified as one of the big challenges for shared Working Group benefits.</p>			

In parallel to the PoC, joint webinars with relevant technology providers shall continue, in order to continue the market exploration. Also, a research work with Ecole Polytechnique / INRIA fact checking research group has been scoped, which could provide further insights on how data semantics (in SDMX) can facilitate (or not) the implementation of StatsBots revolving around the research assistant / guided search use cases.	
Alternatives considered	
The proposed PoC is small scale and cannot be scaled back. The funding of it should be shared among participant organisations. If we do nothing then organisations will work on this topic alone, and hence at a slower pace and with lower capacity NSOs not being able to catch up with the potential opportunity that StatsBots represent. Ultimately, not working on this topic together will mean slower and less relevant progress in how statistical organisation can <i>bring reference facts to the conversation</i> in a more efficient manner.	
How does it relate to the HLG-MOS vision and other activities under the HLG-MOS?	
<p>The ‘working group developing a Statistical Chatbot for Official Statistics’ was created following the joint presentation by CBS NL and Statistics Canada at the HLG-MOS in November 2018 as part of the BSTN. A call to action soon followed and the SIS-CC, a Community of statistical organisations who collaborate to build open source tools and who share common needs, help to push this agenda forward with the formation of the WG during their first meeting at the OECD in Paris in March 2019.</p> <p>With a main goal to support the modernisation of official statistics through new and innovative ways in which our statistical organisations operate as well as disseminate data, this aligns well to the HLG-MOS vision.</p>	
Proposed start and end dates	
Start: February 2020	End: July 2020

Guidance notes for completing the template

Business cases should be **strictly no longer than 2 pages**. Keep your text short and to the point. You can delete these guidance notes when you have completed the business case.

Business cases should not go into technical details. Remember, the audience for your business case will be chief statisticians. They will not accept a proposal just because you think it is a good idea. They will need to be convinced!