Chatbots in the context of Official Statistics
On behalf of the working group developing a Statistical Chatbot for Official Statistics
HLG-MOS – Geneva, November 2019
Could StatsBots help bring *reference facts* to the conversation?
Background

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 was made.

The SIS-CC, a Community of statistical organisations who collaborate to build open source tools and who share common needs, helped to push this agenda forward with the coming together of 9 organisations at the OECD in Paris in March 2019.
Working group was formed

<table>
<thead>
<tr>
<th>Name</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike Ackermans</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>Piet van Dosselaar</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>D.E.A. (David) Kloet</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>Stéphane Dufour</td>
<td>Statistics Canada</td>
</tr>
<tr>
<td>Andrew Sinkinson</td>
<td>Statistics Canada</td>
</tr>
<tr>
<td>Maarja Jämsa</td>
<td>Statistics Estonia</td>
</tr>
<tr>
<td>Daniel O'Donnell</td>
<td>Federal Statistical Office of Germany</td>
</tr>
<tr>
<td>Lorenzo Bruni</td>
<td>Central Statistical Office of Ireland</td>
</tr>
<tr>
<td>Eric Anvar</td>
<td>OECD</td>
</tr>
<tr>
<td>Eileen Capponi</td>
<td>OECD</td>
</tr>
<tr>
<td>Jonathan Challener</td>
<td>OECD</td>
</tr>
<tr>
<td>Lisa Thiel</td>
<td>Statistics Sweden</td>
</tr>
<tr>
<td>Hakim Sjöström</td>
<td>Statistics Sweden</td>
</tr>
<tr>
<td>Soph Warnes</td>
<td>ONS</td>
</tr>
<tr>
<td>Rob Fry</td>
<td>ONS</td>
</tr>
<tr>
<td>Jure Stabuc</td>
<td>ONS</td>
</tr>
<tr>
<td>Vic Chenery</td>
<td>ONS</td>
</tr>
<tr>
<td>David Rawnsley</td>
<td>UKData Service</td>
</tr>
</tbody>
</table>

To define the scope, mission and vision for the project, prepare the business case(s), identify and showcase technologies needed for this project, list the skills and competences needed, with a goal to develop a ‘sharable’ (or common) solution through different business models, and look at ways in which to develop non-NSO partnerships in the wider data ‘ecosystem’.
VALUE CREATION

Operational Efficiency

Engage with more people

Enrich the engagement

MAIN BUSINESS CASES

Support users

Recruitment assistant

Census assistant

Staff assistant

High automation

Collection

Dissemination

Get Message Thru

Augmented service desk

Research assistant

Guided search

Story engagement

Topical conversation

Low automation

Chatbot enables constant feedback loop and knowledge retention

JOINTLY DEVELOPED CAPABILITIES

Knowledge

A toolkit for NSOs to develop their business case, incl. Personas, service characteristics, ROI, open strategy...

A database of statistical intents and answers

Enable our data and metadata for Chatbots

Lever effect

$ Market vendors

Partnership

Code

Portable open Algorithms

Training set

APIs for Chatbots
Exploring **together** the market

An RFI was sent to potential technology partners as a result of the discussions in Paris with 5 shortlisted for a webinar. Each webinar consisted of:

1. a brief overview of their company and project(s) relevant to the RFI,
2. a response to one or more of the business cases and how they would go about addressing these,
3. an opportunity for the WG members to ask questions.

The webinars were recorded and made available for playback to the Working Group members.

**More webinars should be planned in the future, as part of ongoing joint effort to map market players with a relevant value proposition for statistical organisations.**
More of the same, or something new?

The technology providers could be categorised either as inside the box thinkers (industry standard offering), or out of the box thinking (with the exception of Artik, a consultancy with a very good view of the ecosystem as whole).

Innovative approach ‘Symbolic AI based on Linguistics’ able to reproduce and automate existing processes.

No need to duplicate data, or create vast numbers of potential questions (intents) that data consumers may ask.

Speaks our language with ability to add any number of languages with minimal effort.

...
Next steps (tbc with the group in Dec)

1) The Working Group proposes to move forward with a **StatsBots Proof of Concept in 2020**. The PoC would be centered on the data exploration uses cases (research assistant / guided search) which has not yet been tackled with by NSOs. The PoC would be co-funded by interested organisations.

2) A tentative statement of work has been put forth by OECD – for a collaboration with Ecole Polytechnique / INRIA research group on fact checking. Research topic would be:

**How can SDMX data semantics enable (or not) the implementation of StatsBots by statistical organisations?**

If interested contact jonathan.challener@oecd.org, SIS-CC community manager.