

Business Case for StatsBots

Bringing reference facts to the online conversation

For HLG-MOS Meeting November 2020

Prepared and presented by Jonathan Challener, OECD

On behalf of the

Working group developing a Statistical Chatbot for Official Statistics

Data deluge continues to grow at an exponentially rate where no longer **Statistical organisations are the only source for data** that feeds the policy making machine.

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.**

This calls for a **need to innovate in the dissemination of official statistics**, so as to bring reference facts to the conversation.



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) need reliable facts that they can access easily, and rely on in their wider conversation with the public.

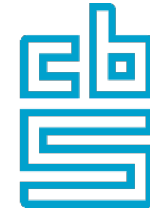
Statistical organisations can **enable this conversation around facts with new digital and emerging technology.**



“COULD STATSBOOTS HELP BRING REFERENCE FACTS TO THE CONVERSATION?”

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.

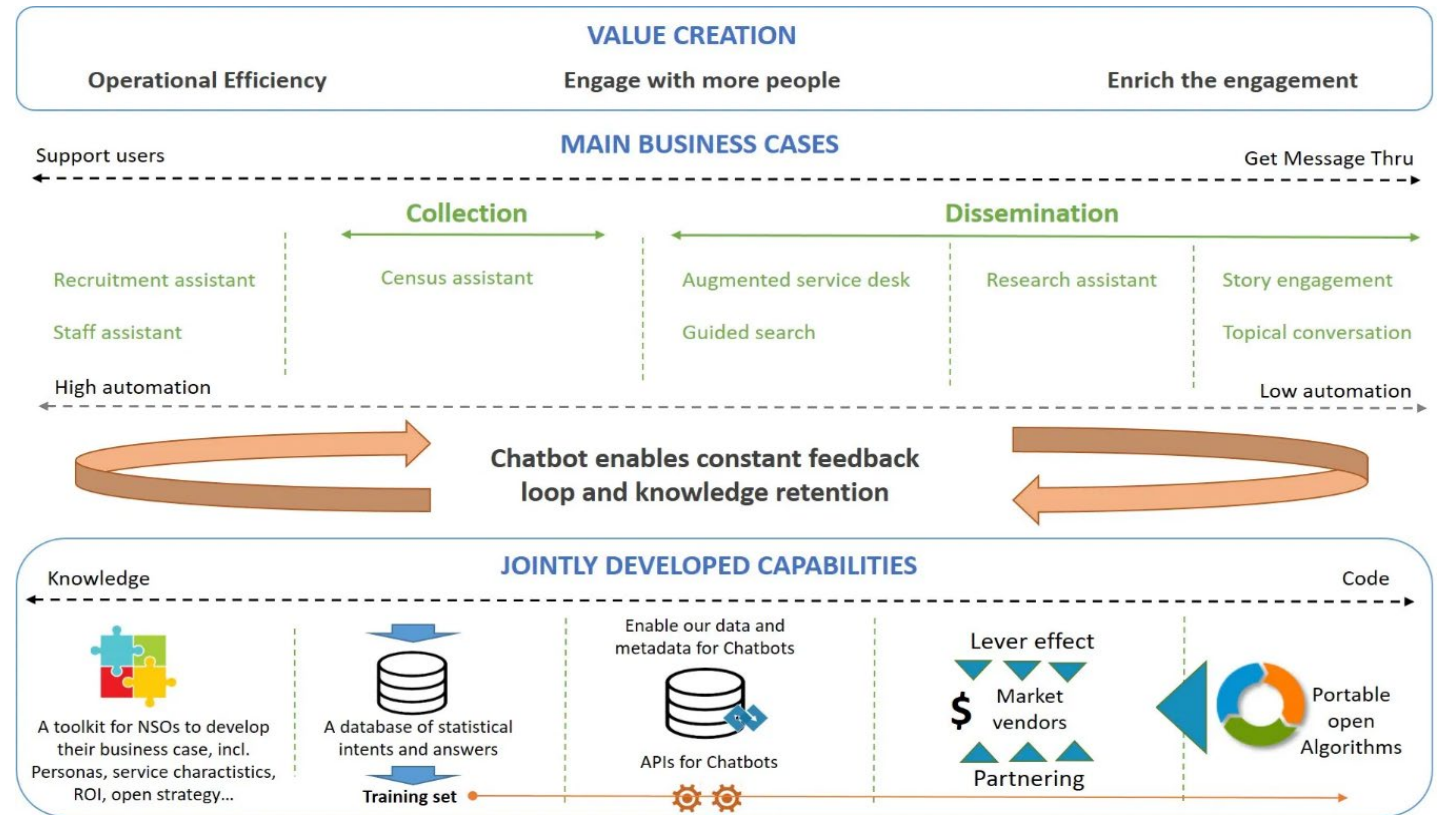
The **SIS-CC**, a Community of statistical organisations who collaborate to build open source tools and who share common needs, helped to push this agenda forward with the **coming together of 9 organisations at the OECD in Paris in March 2019**.



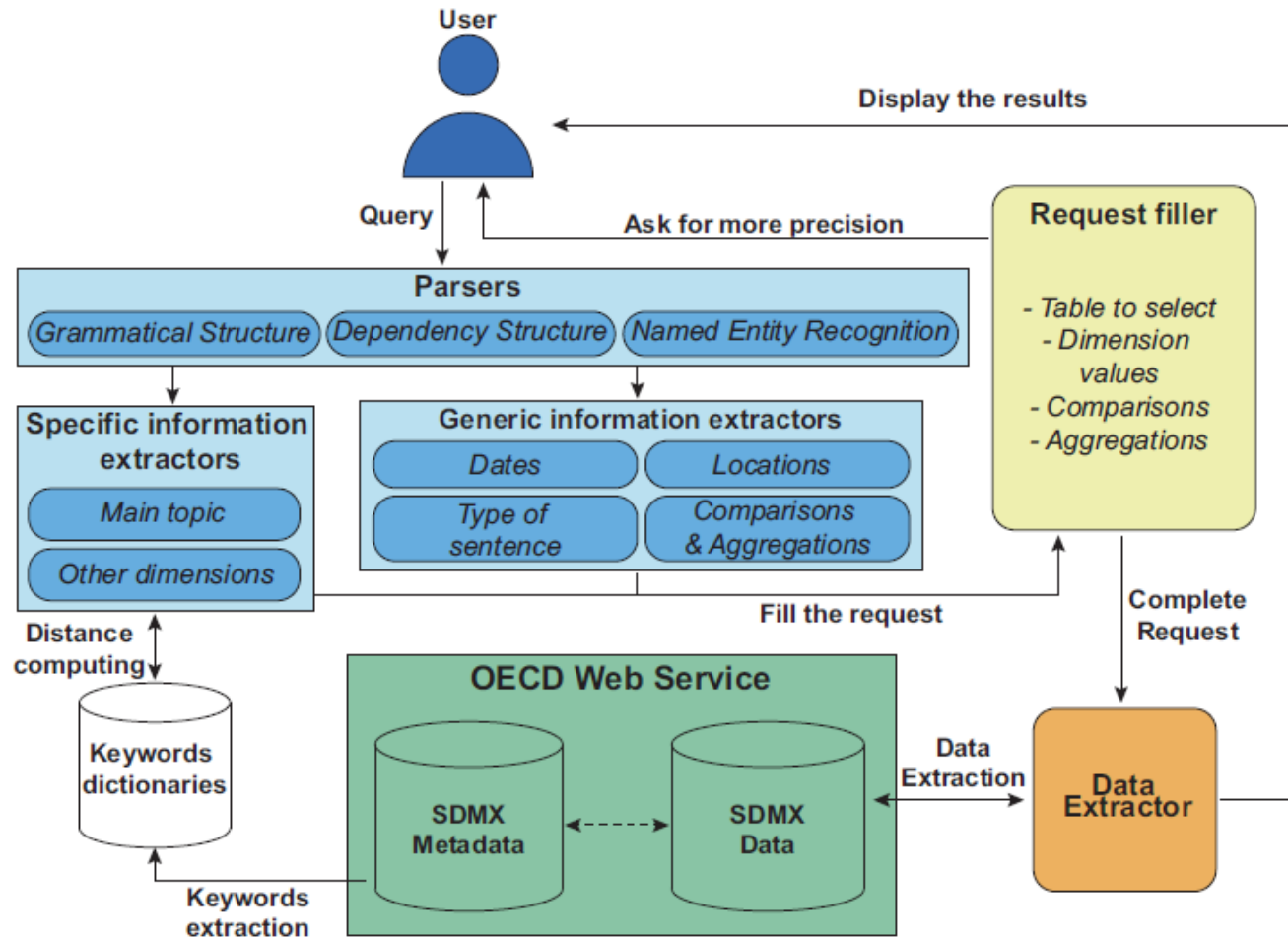
Statistics
Canada



Drawing on experimentations and production chatbots being rolled out in several countries (Canada, Netherlands, UK, Estonia...) – enabled to classify **the main use cases for chatbot application in the area of official statistics** with three directions: operational efficiency, engage with more people, and enrich the engagement.

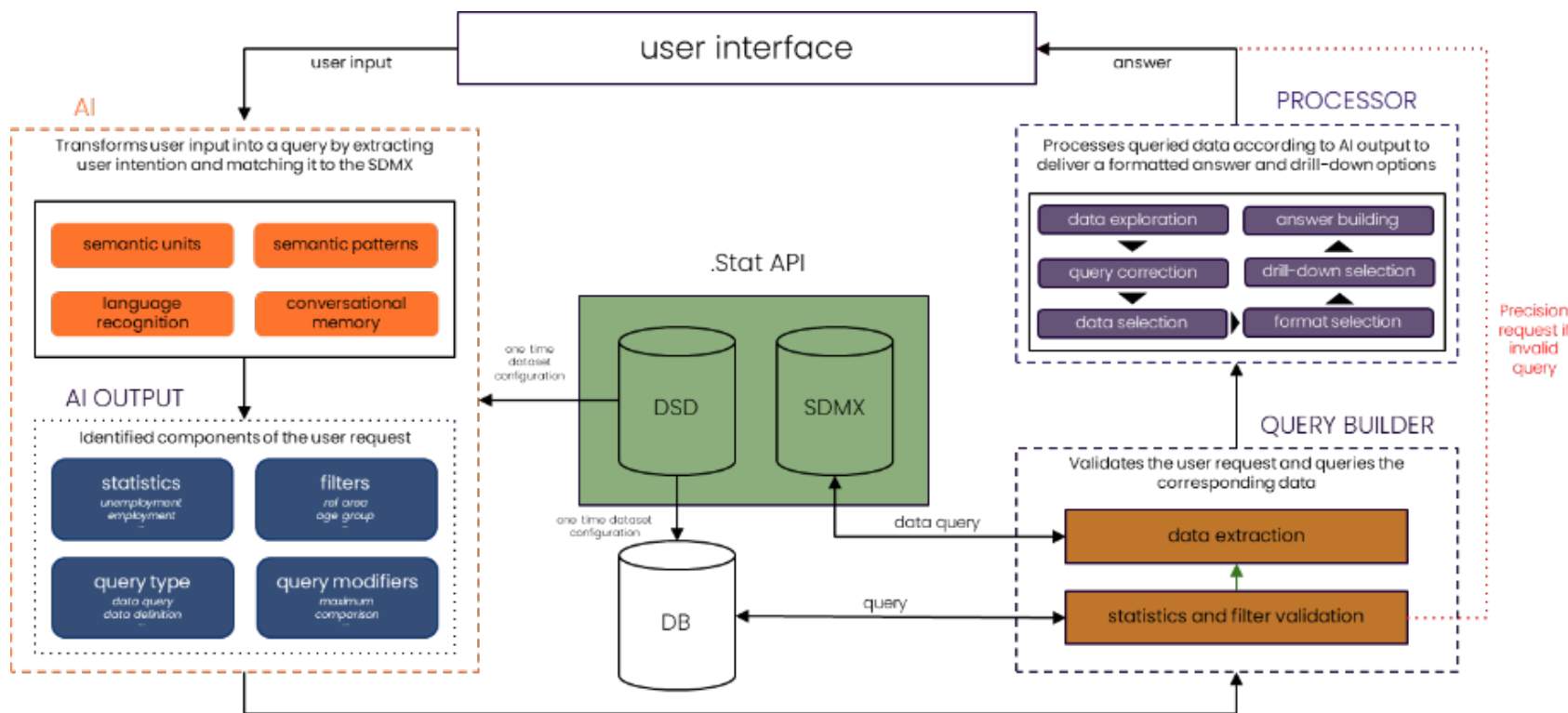
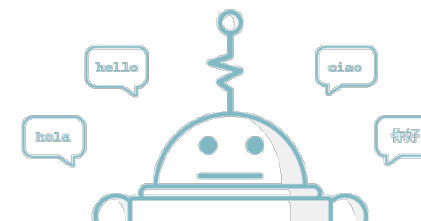


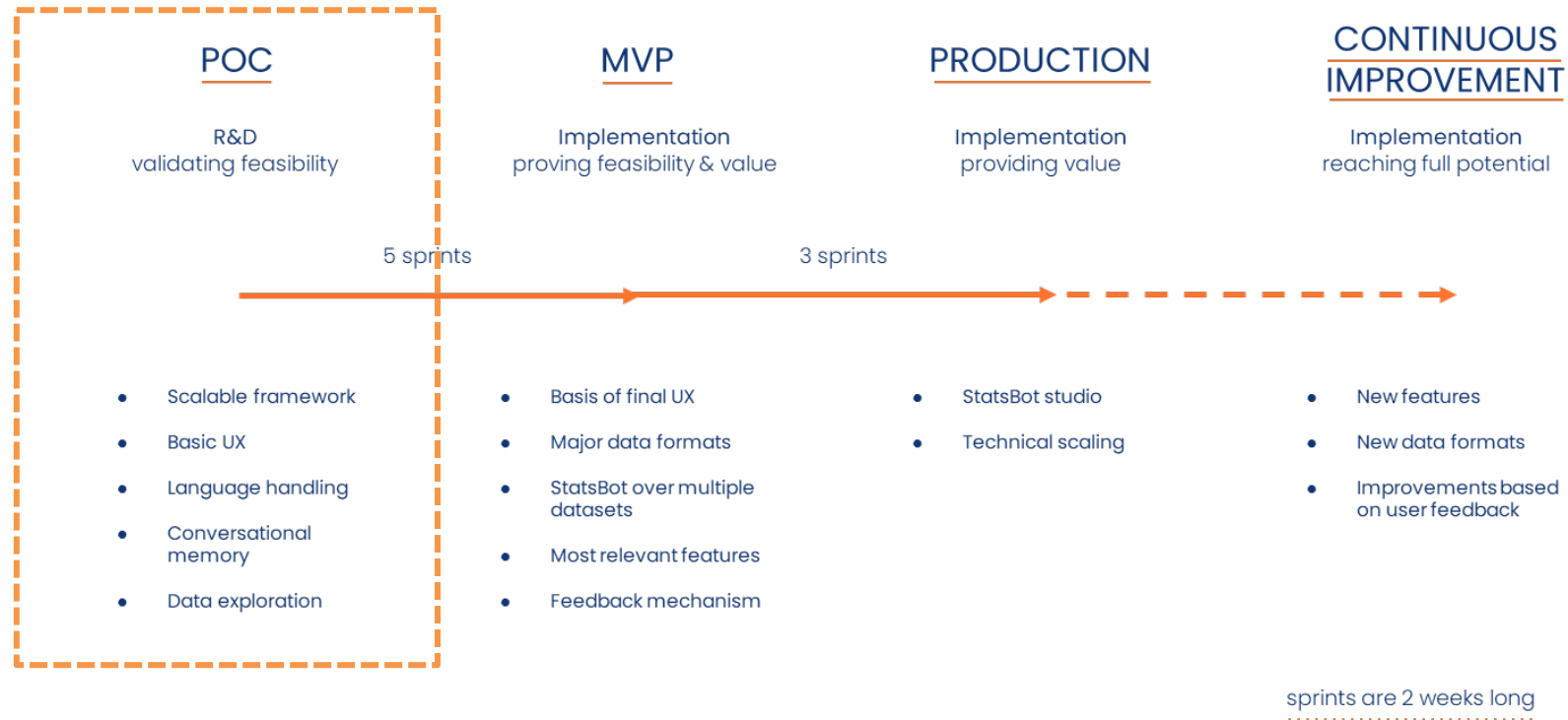
Implemented in Python, based on the nltk toolkit and parsers from the Stanford NLP Group. It consists of 12 files and about 1800 lines of code.





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





Drawing on the experiences and results of a proof of concept to develop a Chatbot for Official Statistics undertaken by the OECD, CBS NL, and StatCan, it is proposed to **launch a large scale project for the joint development of a generic StatsBot that draws on statistical structured sources and rich semantics.**



Carry out a larger scale market consultation that would lead to a) selection of a target technology (either the one selected for the PoC, or another), b) joint development of the generic StatsBot in a second step, and, c) maintenance over time of the StatsBot.



Estimated budget for the project is in the range of 300k€ (recurring cost in the range of 20% of the initial investment), but the resource burden could be split and maximal knowledge sharing and cross-fertilisation could happen if 10 organisations would unite to develop the generic StatsBot.

Accepting this proposal would entail to position the project no longer as an exploration activity, but as **a project to deliver a production grade StatsBots**, and involving fund raising and coordinated execution.

An international organisation could typically host the project (e.g. support project coordination, fund raising, and procurement and delivery activities) under the **HLG-MOS governance**.

The **OECD could take on that role**, leveraging know-how and infrastructure developed in the context of the .Stat Suite project and under the **SIS-CC**.





modernstats
BY HLG - MOS

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.

StatsBots - Bringing reference facts to the online conversation - Report to the HLG-MOS Executive Board, October 2020:

<https://statswiki.unecce.org/download/attachments/290358402/StatsBots%20project%20-%20report%20to%20the%20HLG-MOS%20Executive%20Board%20Oct%202020.docx?version=2&modificationDate=1604418907316&api=v2>

Project repository for PoC - chatbot for interacting with SDMX databases: <https://github.com/Guillaume-thiry/OECD-Chatbot>

Report from workshop organised by the OECD, in Paris, in March 2019: <https://siscc.org/workshop-to-conceive-a-statistical-chatbot/>

Statsbot Webinar - Side event of the 2020 Workshop on the Modernisation of Official Statistics:

<https://statswiki.unecce.org/display/hlqbas/Statsbot+Webinar+20+November+2020>