

StatsBots

Bringing reference facts to the online conversation

Report to the HLG-MOS Executive Board, October 2020

Authors

Eric ANVAR, Jonathan CHALLENGER, Organisation for Economic Cooperation and Development (OECD)

Reviewers

Graeme GILMOUR, Statistics Canada (STATCAN), Piet VAN DOSSELAAR Statistics Netherlands (CBS)

Contents

1. Background: Business case(s) for chatbots in official statistics	2
2. PoCs Results	4
2.1. PoC with INRIA/Polytechnique – A chatbot for interacting with SDMX databases	4
2.2. PoC with GOLEM.AI – StatsBot, a chatbot for official statistics.....	5
2.3. Assessment of the Golem.AI PoC by the core group (CBS, STATCAN, OECD).....	7
3. Lessons learnt and next steps	8
3.1. General lessons learnt in terms of approach to innovation in statistical organisations	8
3.2. Immediate next step: UNECE HLG-MOS workshop	9
3.3. Possible scopes of projects to consider for 2021 and beyond	9
4. Annex 1: Project Background information.....	11
5. Annex 2: Report from INRIA/Polytechnique	11
6. Annex 3: Report from Golem.AI.....	11
7. Annex 4: Joint user research to drive value and innovation (ARTIK).....	11

1. Background: Business case(s) for chatbots in official statistics

The idea of a collaboration among statistical organisations emerged in late 2018, after CBS presented their concept of a voice assistant responding to questions on socio-economic data, during the HLG-MOS meeting in Geneva¹. Bilateral discussions ensued, where several statistical offices shared results on their experiences in the field. Eventually, the project took the shape of an HLG-MOS activity (see project description in the Annex 1):

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 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) 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.

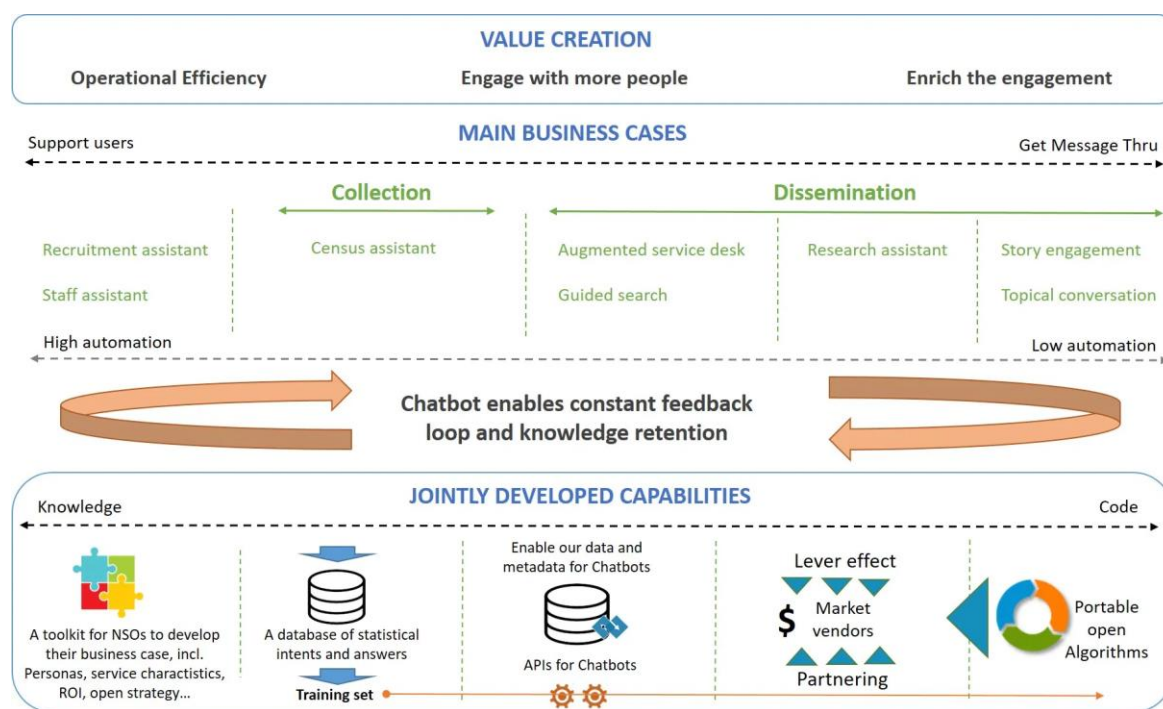


Figure 1: The business cases for chatbots in official statistics

¹ HLG-MOS operates under the aegis of UNECE, and drives the overall agenda of [modernisation of official statistics](#) (ModernStats) through a number of initiatives and projects, including through standardisation efforts. This project was managed under the supervision of the *Blue Sky Thinking Network* (BSTN), reporting to the HLG-MOS Executive Board.

The discussion crystallised during a workshop organised by the OECD, in Paris, in March 2019 (see [here](#) for workshop report), the result of which was summarised in the above sketch (Figure 1). The discussion – drawing on experimentations and production chatbots being rolled out in several countries (Netherlands, UK, Estonia...) – enabled to classify the main use cases for chatbot application in the area of official statistics (three directions: operational efficiency, engage with more people, and enrich the engagement). It also emphasised the notion that chatbots are not just another channel for dissemination or user support – they constitute potentially a new way of establishing a feedback loop with the ‘user’ through conversational mode, and henceforth, an opportunity to gather new knowledge on data users. The discussion also uncovered areas where organisations could potentially co-invest, that is capabilities that would benefit from positive network effects (vs organisation specific approach) and could scale: the concept of **scalability** (from one organisation / language / topical domain to the other) is a core dimension of ensuing discussions, the intuition being that a chatbot that is scalable enough, thanks to AI techniques, should be able to interpret at a relatively low cost data that is structured and semantically rich (e.g. in [SDMX](#) format).

In the experience of statistical offices who had already deployed or experimented with chatbots, the one use case that could be met with the mainstream technologies (typically, machine learning based approach with Google Dialogflow and similar offerings) was the one identified as ‘augmented service desk’: after feeding the chatbot with questions and answers between the users and the service desk, the chatbot could relatively quickly (after a few months) become capable of answering simple questions and switch to a human support when it could not answer². The one use case that would not work so well with such an approach however, is the ‘research assistant’ or ‘guided search’, whereby a chatbot would allow users to explore and find relevant data through an online conversation in natural language (vs existing usually complex search interfaces plugged onto databases). Experience in STATCAN and CBS, for example, resulted in the incapacity to scale to hundreds if not thousands of data sources, due to high cost of inception of the chatbots (they are not ‘scalable’ for that use case). The group then decided to focus on those particular use cases, which were named ‘**StatsBots**’, or chatbot specialised in exploring official statistics through online conversation – in other words, in **bringing reference facts to the online conversation**. The StatsBot could be a way, for users, to easily query and visualise data, coming from thousands of sources, through natural language, in order to fully exploit the potential of statistical data and its rich structure and semantics, as well as offering more exploration options to users (and dispelling the somewhat overinflated chatbot expectations).

An RFI (*Request for Information*, see RFI document in Annex 1) process was then organised, consisting mainly in an online interview where companies would expose their approach to develop a StatsBot, and to which CBS, STATCAN and OECD took part. This exercise enabled, with a minimal time investment, to get a feel of various options offered on the market. One offering came across as the most promising: that of [Golem.AI](#), based on linguistic AI approach³. CBS, STATCAN and OECD decided to co-invest in a Proof of Concept (PoC) with Golem.AI. In parallel, a collaboration started with the research group leader in INRIA⁴, [Ioana Manolescu](#), on the topics of improved accessibility of statistical

² Notably, Estonia, Netherlands, Canada have reported production-grade projects in this space. UK reported conversational engagement type of application.

³ See <https://golem.ai/en/technology>

⁴ *Institut National pour la Recherche en Informatique et en Automatique*, that is: National Institute for Research in Digital Science and Technology.

data through chatbots; this led to a small Proof of Concept (PoC) carried out by a student under Ioana's supervision for 6 months.

The results of these two PoCs are attached in annex to this document. The objective of this document is to concisely review those results, and draw lessons from them and envisage possible next steps. The two PoCs are not comparable in scope and resources invested (on the one side, a modest research work with a 6 months traineeship, on the other an extensive experimentation with a start-up specialised in the field and significant time investment by CBS, STATCAN and OECD); with that caveat in mind, we sense reporting on both of them is interesting and illustrate how innovation can involve a various range of actors.

The audience of this report is the [HLG-MOS](#) Executive Board, as well as members of the [BSTN](#) network, [SIS-CC](#) community, and relevant managers in the respective interested organisations. The report is also to be shared with [SDMX](#) SWG and TWG.

2. PoCs Results

2.1. PoC with INRIA/Polytechnique – A chatbot for interacting with SDMX databases

Detailed outcome of this work stream is in Annex 2, article from which this section is sourced. This research work has been accepted as a contribution to the next [ISWC2020](#); see [here](#).

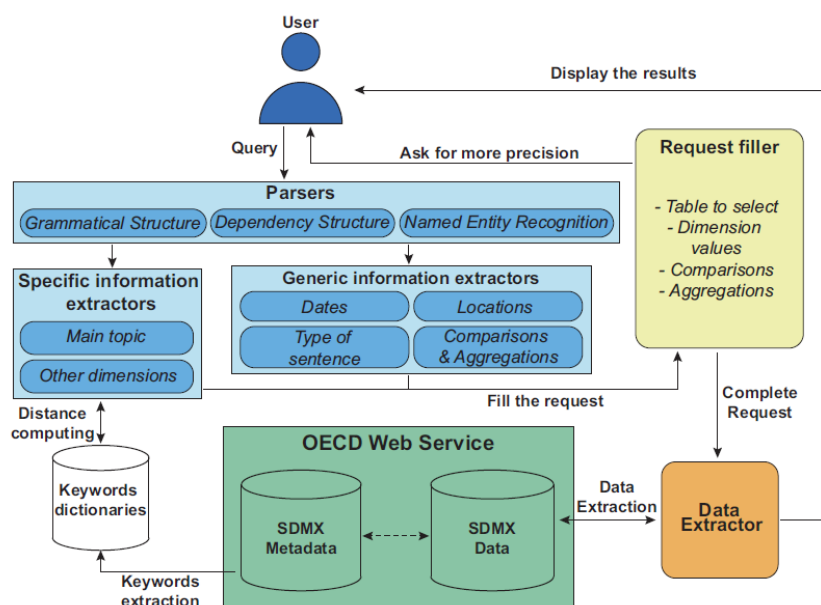


Figure 2: INRIA/Polytechnique PoC – Chatbot structure

Building on the StatsBots business case, the INRIA/Polytechnique team focused on how to develop a chatbot focused on a specific task: “understand information requests, when expressed in natural language”, which they categorised as the following: (type of question) direct questions, comparisons, aggregations; (type of result) topic and dimensions, date(s) and location(s) of the request. Each task is translated into a data query (against an SDMX source); or a precision request to the user. The project

leverages an open source stack⁵ to develop the experimental chatbot in Python. The structure and variances of the tasks, the limitations in the available libraries, and workarounds, are explained in the paper. A methodology to measure the chatbot performance is proposed, against test data sourced from OECD and Eurostat.

The INRIA/Polytechnique PoC defines well the user needs to be addressed and explains well the approach used to detect entities such as topics, dimensions, or the type of question. Given the modest resources invested, the comprehension level is good. Being based on open source and open knowledge, it could serve as a basis for future work, to improve the performance through application to a much wider range of sources and data, and develop conversational features currently absent (disambiguation, drill down, contextual request building on previous requests, suggestions, seeking user feedback...). Explainability could be another area of progress (e.g. explain to the user why a given result has been provided). Multilingualism has not been addressed at all at this stage and could also be subject for future research.

2.2. PoC with GOLEM.AI – StatsBot, a chatbot for official statistics

Detailed outcome of this work stream is in Annex 3, report by Golem.AI from which this section is sourced.

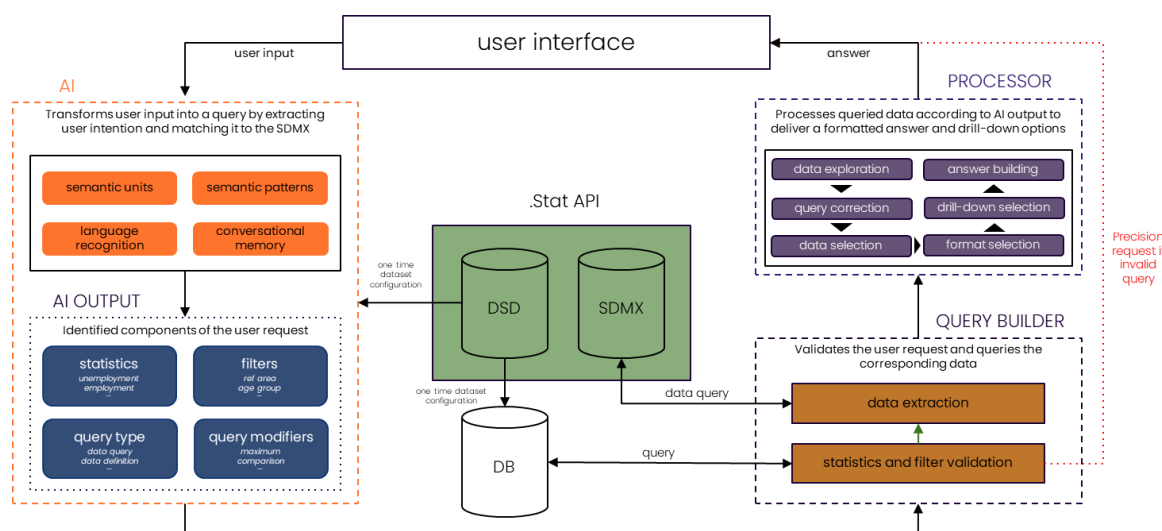


Figure 3: StatsBot workflow in the Golem.AI PoC

The focus of the PoC was not to develop a nice and good looking chatbot, but rather to focus on the strategic question of scalability, as stated in the introduction – that is, how could linguistic / symbolic AI capabilities, combined with data richness (semantic structure according to SDMX especially), lead to an approach that can scale, from one topic / language / organisation to another, with limited cost. The PoC focused at an early stage on one domain (Labour statistics, made available in SDMX format through a [.Stat Suite API](#)⁶, sourced from the 3 organisations) in order to identify roadblocks to scalability, and derive from the analysis an approach to build a generic, scalable StatsBot that could expand across topic, languages and source organisations with limited additional cost.

⁵ The chatbot has been implemented in Python, based on the nltk toolkit and parsers from the Stanford NLP Group. Code is available online at <https://github.com/guillaume-thiry/OECD-Chatbot>.

⁶ .Stat Suite is an open source SDMX native platform built as a collaborative development by SIS-CC.

Based on the assumption that the user is focused on a statistical based interaction (find relevant data, in the relevant format) , the bot leverages existing linguistic AI capabilities (identifying semantic pattern through language recognition, in a semantic context e.g. conversation memory) in order to identify components of the expected output (identification of the query type, statistical domain, filters, calculations). The output processor would, in a second step, either ask questions if the request is deemed incomplete, or format an answer based on extracted data, offering drill down options to the user.

After an initial assessment and configuration, the PoC was organised around 4 sprints, whereby CBS, STATCAN and OECD staff would test the StatsBot and bring feedback where there were issues – approximately 300 issues were reported and corrected during those 4 sprints. At the end, the remaining functional gaps were assessed, along with recommendations in terms of optimal data modelling (in SDMX) to enable efficient data sourcing by the StatsBot.

The PoC demonstrated good results in following areas:

- Comprehension: 3 languages managed; interpretation of various comparison patterns (min, max, comparison against various dimensions), taking into account the context (conversation memory, building on the previous request, conversation reset when a new query is detected);
- Data exploration: adapting queries to closest match, proposing available drill down filters or alternative filters or filter values;
- Restitution: display of relevant filters and units; text, chart or table formatting; guiding the user in case of non-comprehension.

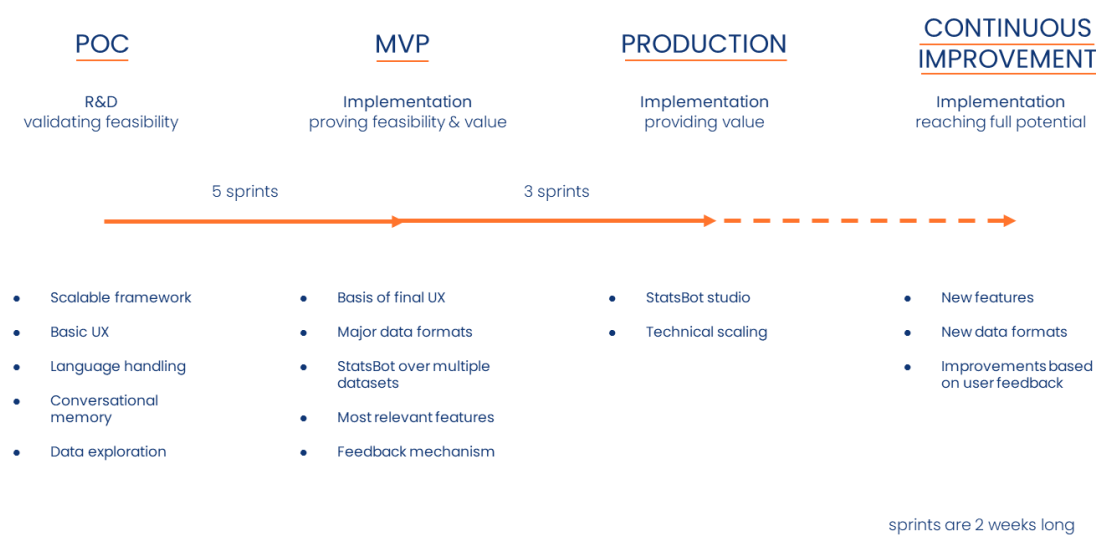


Figure 4: Journey to production grade StatsBot, according to Golem.AI

Golem.AI estimates the need for approximately quadruple the time spent for the PoC (so, 8 sprints of 2 weeks vs 4 sprints of 1 week) in order to reach a production grade StatsBot, including much improved user interface that could adapt to different sources, languages and topics with limited incremental cost. It argues that the technology under the hood (linguistic AI, as opposed to traditional ML in most chatbot technologies on the market) allows for capitalisation (e.g. when one organisation has proofed the chatbot for tourism statistics in a given language, other organisation should be able to leverage the knowledge with limited additional cost) – leveraging

the StatsBot studio, e.g. back office for non-technical experts to assess the chatbot performance and improve the configuration of it over time.

2.3. Assessment of the Golem.AI PoC by the core group (CBS, STATCAN, OECD)

The core group invested significant expert time over the June-July 2020 period, in order to test the StatsBot, imagine typical requests by users and expected outputs (this knowledge, formulated as a backlog of approximately 300 issues, constitute in itself and interesting by-product of the PoC which could serve as an input to a more systematic specification of a StatsBot).

The interactions with Golem.AI have been overall smooth. Golem.AI proved to be responsive, fixing issues quickly, responding to questions – and accepting to invest in the PoC beyond the initially agreed scope. The group did not have time to assess how the StatsBot should be maintained and improved over time using back office tools, as this was not in scope.

Overall, the PoC result is positive in that the claim by Golem.AI to be able to reach production grade StatsBot in 8 sprints is deemed plausible, as well as its claim to provide for a scalable solution (in the sense conveyed in this document); however, it is assumed that applying StatsBot to a completely new topic will still require some additional work, which can be capitalised on when the StatsBot is applied to a data source on topics comparable to those already covered. The StatsBot in its current version does provide many relevant answers –getting stuck still in relatively simple conversation patterns – but that is to be expected, just as is to be expected the simple user experience (as UX was not the point of focus).

The PoC was also an opportunity for the group to reflect on its own objectives. As it was driven by an opportunistic techno drive ('what could be done with a chatbot?') and, it led us to question ourselves more fundamentally:

- a What exactly are we trying to achieve? Which are the users' problems we are trying to solve? What are they trying to achieve with our data?*
- b How could conversational mode contribute to improve data accessibility in real life – and not end up just being a nice techno feature? How would it fit into a broader 'data experience'?*
- c What is our level of understanding of users and users' needs? How do we assess them? Are we at edge of user research techniques available and used on the market?*
- d Do we have a structured, systematic way of capturing their degree of satisfaction with our services, our data products? Do we have a systematic feedback loop established with the data users – for example, to measure if they found the data they were looking for?*

Those are the types of 'back to basics' fundamental questions the PoC raised – which essentially inspired the second part of the HLG-MOS workshop agenda, as proposed below. The lessons learnt and next steps we could draw from these conversations are captured in the next, and last, section of this document.

3. Lessons learnt and next steps

3.1. General lessons learnt in terms of approach to innovation in statistical organisations

One meta-lesson that we draw from the PoC, that is transposable to many other frontier topics statistical organisations are willing to explore, is the relatively successful approach to innovation adopted in the context of these PoCs, which could be summarised as follows:

- I. Leave room for serendipity: the whole approach was initiated through informal discussions at HLG-MOS meeting in Nov 2018, leading to an informal workshop taking place in March 2019 with a dozen organisations. In the same way, informal contacts through a trainee led to connect to the INRIA/Polytechnique research group and triggered the PoC at close to zero cost or time investment.
- II. Don't be shy on consulting the market: the RFI process, carried out in Q4 2019, took very limited resources (drafting the business case document; identifying a dozen chatbot players at AI Paris 2019; holding a half-dozen webinars). In a record time we could build a conviction on what the market has to offer, and on the idea that investment in a limited PoC is worthwhile.
- III. Mobilise seed money and federate goodwill: none of the 3 organisations (CBS, STATCAN, and OECD) who invested seed money in the PoC could have done so alone. Doing it together was a way of sharing the burden, the 'risk of failure', and above all cross-fertilising through interactions around the PoC.
- IV. Engage with players from outside the statistical domain: academic groups, start-ups, GAFAs, etc. specialised in frontier technologies should be engaged much more often, much more early in projects, and at the cost of seed resources to enable experimentations (that can fail or succeed, but will always contribute to advance our strategic thinking).
- V. To invest in user research in order to drive innovation: the PoCs led participants to question more fundamental assumptions (summarised as questions a, b, c and d, see section 3.3.) regarding their approach to user research, their knowledge of the user and their practices in terms of establishing a systematic 'feedback loop' to inform service development. These interrogations led to the second part of the workshop (see agenda below), to take stock of the best practices and techniques available and used on the market⁷, in the area of user research driven innovation, and assess gaps in statistical organisations, and possibly envisage joint scopes of work or investments in order to fill that gap.

International organisations have a role to play in offering the forum for this type of innovation to emerge, and provide for the coordination, possibly seed resources. They have the operational model in place to raise funds from countries and transform that into joint deliverables and projects, under the supervision of the appropriate governance⁸.

⁷ See, for example, LEAN Start-Up methodology – <http://theleanstartup.com/principles>.

⁸ On this topic, see OECD Paul Schreyer's foreword in the [SIS-CC 2020-25 strategy](#).

3.2. Immediate next step: UNECE HLG-MOS workshop

As a way to conclude the work, the project group has proposed to organise a workshop at HLG-MOS 2020 (virtual) gathering in Nov 2020⁹. The draft workshop agenda is the following (for the second part of the workshop, the justification for it are provided in the previous sections):

1. (5') Introduction: the case(s) for chatbots in official statistics and workshop agenda

Background: StatsBot project – report to HLG-MOS Executive Board

2. (60') StatsBot, a chatbot for official statistics - Two proofs of concept and their lessons learnt

- (10') Chatbot for interacting with SDMX Databases (INRIA). *Background: Annex 2*
- (20') The business case for StatsBots (GOLEM.AI). *Background: Annex 3*
- (30') Q&A with the participants

(5') Transition (OECD): lessons learnt in terms of approach to innovation in statistical organisation, driven by systematic user research.

3. (60') Joint user research to drive value and innovation – Assessing where we are, what we could do together

Background: Annex 4

- Setting the scene: overview of user research best practices to drive value and innovation (ARTIK)
→ Self-assessment by participants of their organisation's maturity in user research
- Examples of envisaged joint user research projects (ARTIK).
→ Assessment by participants of their appetite for proposed joint user research projects
- Concluding remarks: self-assessment results and early identification of priority areas of work.

4. (10') Conclusion– identifying calls for action and possible next steps

Background: StatsBot project – report to HLG-MOS Executive Board, section 3.3.

General note: The main purpose of the workshop is to identify areas of joint work and agree on the approach to scope them better that could lead ultimately to relevant project proposal. Participants are invited, ahead of the meeting and based on the documentation provided, to reflect and consult within their organisations and join the workshop with ideas and proposals. Background reports and information are shared ahead of the meeting. Participants are expected to have read them, so that presentation will focus on key findings, so as to save time for discussion.

3.3. Possible scopes of projects to consider for 2021 and beyond

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

⁹ As of Oct 1st, tentative date has been set at Friday 20th Nov, afternoon session.

an investment in the range of 150-300k€, shared amongst interested organisations, a production grade, scalable StatsBot could be envisaged in the short term;

- II. **Continue the open research work stream on StatsBot.** Based on results of the initial research work by INRIA/Polytechnique, more research could be envisaged in the future with this and other groups focusing on fact checking or web semantics issues; this would contribute to make official statistics more visible in this area of research. With limited investment over a multi-year period (consisting mainly in staff time and making data available), this could potentially lead to production grade open source StatsBots in the mid-long term;
- III. **Launch one or several joint user research initiative(s), meant to help statistical organisations accelerate in developing the ‘next generation data experience’** and establish best of breed practices in the area of user research to drive innovation and value.– the scale and scope of these initiative(s) to emerge from the interactions during the workshop. A list of initiative will be proposed as input; of course participants could bring in their proposals as well. Examples of such initiatives:
 - a) Create and maintain the reference ‘stats personas’ knowledge base, eg evidence and analysis regarding the typical profiles and use cases in the area of official statistics (policy maker, researcher / student, data journalist, NGOs, citizens, businesses...) and ongoing capacity building activities for the reference network of UX/UI experts in statistical organisation;
 - b) Carry out a cross national research project on ‘data journalists’ (or another typical group user segment), in order to get much deeper, evidence-based knowledge of their needs, their expectations from data providers in terms of data products and services – the constitution of a reference, worldwide sample of data journalists could in itself enable different opportunities for engagement and joint consultation with that community;
 - c) Develop and maintain a toolkit for ‘Search Engine Optimisation (SEO) in official statistics’, that is, based on the analysis of opportunities and roadblocks to visibility of official statistics in search engine, identify tools, techniques, standards to measure and improve that visibility, including by engaging – together and at a strategic level;

For the three potential lines of action (I, II and III), a typical action could be for the group to identify volunteer organisations who would be interested in either or the two areas of work, and mandate a small task force to develop detailed project proposals inspired by the workshop discussions.

Depending on the outcome of the workshop, project proposals could be put forth for review by interested parties and/or at the level of the HLG-MOS Exec Board in due time.

4. Annex 1: Project Background information

HLG-MOS activity description:



Microsoft Word
Document

Request for Information: Business case for StatsBots



Business Case for
Stats Bot - RFI.pdf

5. Annex 2: Report from INRIA/Polytechnique



Chatbot-SDMX-Thiry
-et-al.pdf

This paper has been accepted as a contribution to the next [ISWC2020](#); see [here](#).

6. Annex 3: Report from Golem.AI



Project report -
StatsBot.pdf

7. Annex 4: Joint user research to drive value and innovation (ARTIK)



2020-10-28T1 -
OECD - Joint user res

The background information for the workshop is still work-in-progress. However, it is shared as background information ahead of the workshop. A questionnaire for self-assessing maturity of participant organisations during the workshop is being developed (see [draft questionnaire](#), work in progress).