[redacted] reported: the next phase will see User Confidence Testing (Data Quality Testing) performed as part of the next Stepping Stone. This is due to start on 9 October. The Content PUT has now been formed and will carry out the User Confidence Testing.

[redacted] suggested at [redacted]'s LISTEN 08 briefing that linguists record occurrences of Pick Your Own Number (PYON) in B3M using the 'techref' method in the proc_cmnt_text field. So, if you have e.g. two speakers with UK-code tels but one speaker reveals he is based in another country, this may be a case of PYON and you could enter 'techref PYON' in B3M. [redacted] and [redacted] agreed to the suggestion.

4. B14 / MCS update (courtesy of [redacted])

a) Voice

Diariser: over the summer we have successfully deployed version 5.0 of the Diariser. The main difference between this and version 4.9 is the complete re-write of the modem arbiter, which now ensures more effective identification of various V series modems to assist OPC~CAP in de-modulating them.

Speaker ID: since the last VFUG in July we have deployed:

- Two systems for OPI~MENA: one against an Iraqi political target and one against a Saudi leadership target.
- One system for OPI~RCIT targeting a Gazprom official.
- A 5 way multi-speaker system for OPI~LANG/OPI~SC against Afghan counter narcotics targets

We have now cleared our order book for speaker id systems and so look forward to a well deserved, quiet, peaceful and reflective Autumn. Seriously though, if you believe we can help you with identifying your target of interest amongst the deluge of traffic that you have to wade through, feel free to approach us and we will happily discuss your requirements and hopefully offer a swift and accurate solution.

Language ID:
No new deployments. I have however, been concentrating on improving currently deployed models using non-sigint language data accrued from our Speech to Text (STT) corpora. Experimentation has shown that it is possible to improve accuracy to a certain extent by "brute-forcing" the language id algorithm with vast volumes of data, although this is by no means a substitute for actual "linguist-truthed" sigint language data accrued from the target set that the system will eventually run against.

It is also quite problematic to service a requirement where we are seeking to identify one target language from a number of others, some of which are unidentified. Inevitably this will lead to false positive identifications of the target language as the algorithm will assume that because a language is not included in the background

This information is exempt from disclosure under the Freedom of Information Act 2000 and may be subject to exemption under other UK information legislation. Refer disclosure requests to GCHQ on 01242 221491 x30306 (non-sec) or email infoleg@gchq