Data Standards Body

Non-functional Requirements (NFR) Consultative Group

Minutes of the Meeting

Date: *Thursday 24 April 2024*

Location: *Held remotely, via MS Teams*

Time: *14:00 to 16:00*

Meeting: *Meeting # 3*

Attendees

Participant Members

Mark Verstege, DSB, Chair	John Adshead, AEMO
Jim Basey, Basiq	Joseph Lucas, ANZ
Andrew Ferris, AGL	Julian Luton, CBA
Dhananjay Gourshettiwar, Westpac	Mark Wallis, Skript

Observers

Elizabeth Arnold, DSB	Nils Berge, DSB
Hemang Rathod, DSB	Terri McLachlan, DSB

Apologies

Jon Denly, Basiq

Chair Introduction

Mark Verstege, the Chair of the NFR Consultative Group welcomed everyone to the NFR Consultative Group meeting and acknowledged the traditional owners of the lands.

Minutes

The Chair thanked the group for their comments on the minutes from 27 March 2024 meeting. As per the Terms of Reference, the minutes will be made publicly available. The Chair sought any last-minute feedback prior to approving.

The Minutes were formally accepted.

Action Items

The Chair noted that a number of Action Items arose from the last meeting. An update follows:

1. Invite Xero to a future meeting

    DSB met with Xero last week and they are keen to participate in an upcoming meeting. The group agreed.

2. Speak to AEC about a potential participant (larger energy DH) to join this group

    DSB to meet with them 2 May.

3. Draft a problem definition statement for discussion

    Tabled for discussion at April meeting

All other Action Items had been completed.

Meeting Schedule

The forward schedule of meeting dates was included in the papers for noting.

Update on Data Analysis

Nils Berge from the DSB noted the team is still working through analysing the data received but does not have any specifics to present at this time. He noted some high-level statistics around authentication flows, traffic volumes, and performance metrics that would be interesting to monitor and share in the future.

Some of the key points are:

*   Rejections from authenticated invocations is less than 0.5 % but much lower in most cases with an TSP around 3, average being 25 and the absolute maximum of 50 to 60.
*   Authenticated invocations are around 10-15 requests per session.
*   To work with ACCC to see what data, that is not published on the performance dashboard, they can make publicly available.

DSB advised the group they would be happy to review any particular metrics, ratios or percentages the ACCC receive through GetMetrics.

One member shared some metrics on conversion rates for open banking versus screen scraping, showing better performance for open banking. This was across all data holders and was measured from the point of redirection to a data holder. The trend is increasing, and consumers are becoming more aware that they should not be sharing their username and password for their online banking.

Problem Definition Statement

The Chair put forward problem definition # 1: frequently requesting a collection of resources that have a low velocity of change can create strain on Data Holders servicing other requests especially as request load grows.

The group discussed the need to define the core problem they are trying to solve related to reviewing non-functional requirements and appropriate thresholds. Several issues were raised including inefficient processes causing excessive calls for unchanged data, bulk historical data requests, multi-site requests, and authentication performance.

One member noted that historical access may not be a separate problem, but it could have a completely different solution. An overarching statement saying how we're trying to get the most out of the NFR limits because we're worried about increased load due to being forced to request data that is never changing. There is no efficient way of doing bulk data transfers when we need to do historical requests. Authentication is potentially another issue.

One member noted in the energy sector there is a problem with multiple requests for high volume low velocity data which is prevalent and increasing. There are also existing customers with 2 to 300 accounts wanting access to their datasets which is an issue.

There was agreement that a higher-level problem statement is needed that encompasses addressing growth and anticipated significant increases in load.

The Chair noted there are competing views around the threshold ceiling being too low for potential growth vs too high and the cost of the infrastructure too much for active authorisations. He asked the group whether we are solving for both or for the ceiling being a limit or potentially being more fine-grained lowering limits based on more evidence.

One member noted that volume traffic burden of repeat calls and potential failure are two very different problem statements.

One member noted that much of the banking load is still occurring via screen scraping and could shift quickly, posing risks of hitting traffic limits.

One member shared analysis showing small multi-site requests can be handled but performance degrades significantly beyond 5 sites, commonly hitting limits at 25+ sites.

The Chair suggested creating a document on GovTEAMS to capture the various problem definitions raised for group collaboration and further refinement.

ACTION: DSB to create a word document in GovTEAMS to capture various problem definitions

Any Other Business

One member posted a draft paper in GovTEAMS to start a discussion on Single requests with multiple service points and possible approaches to establishing NFRs.

Problem Statement: NFRs for the first page of an API response are currently set at 4.5s for 95% of requests for secondary data holders for the more extensive energy data requests, although other data holders across other industries have slightly different NFRs the issue is similar – the NFRs do not limit the number of service points (read also NMIs or accounts) that may be bundled within one request. Therefore, a single service point request is measured against the same criteria as a hundred service point request (or many thousands of service points in one request).

They assume these NFRs are in place to support the rapid response to a legitimate API request for an end customer. This in turn establishes the architectural parameters needed to achieve the rapid response.

This is a comparatively new problem with substantial volumes from Feb-Mar 24 that increased by 300%. Typical month is a 95th percentile – the time taken to process 95% of requests for the Usage API. Their systems are currently optimised at just under 4 service points per request and would need to be constantly reoptimised to meet the changing demands ad infinitum. Service requests with 5 or more would normally fall outside the NFRs.

Further analysis showed individual requests with more than 4 service points may be well over the NFRs, when measured together with other single service point requests, the overall durations are marginally worse yet still well within the 4.5s NFR.

Some positions based on initial thoughts are:

1.   Calculate the NFRs as normal;

2.   Place a stated limit on the number of service points expected within the NFR;

3.   Set NFRs dependent on number of service points; and/or

4.   Throttle requests with large numbers of service points providing a return time and possibly a continuation token.

The group discussed the points related to inefficient API calls to refresh data and requests with multiple service points. They suggested improvements like adding filters to request only updated data and limiting the number of service points per request.

One member shared a paper on "Standards Maintenance – ability to filter transactions based on 'modified' date" for discussion.

There is a desire from both ADRs and DHs to reduce the number of API calls that need to be made from the ADR to the DH as part of business-as-usual product offerings. Unfortunately, there are parts of the standards which currently force ADRs to make more API calls than they would like. One of those areas is the 'oldest time' filter on the Banking Get Transactions for Account endpoint.

If the ADR wishes to obtain all new transactions, or transactions that have changed since they last collected transactions, there are only 2 options available:

1.   Obtain a full list of all transactions every time; or

2.   Obtain a sliding window of transactions.

Both solutions require more API calls to the DH than optimal, increasing system load and affecting TPS requirements.

They suggested adding a "modified since" parameter to transaction data requests so data holders only return updated data rather than full refreshes.

There was discussion around feasibility and merits of the 'modified since' approach vs event-driven webhooks. For data holders adding 'modified since' it would be potentially better but unknown how hard it would be to implement.

The data recipients shared concerns about hitting traffic limits as adoption increases, both gradually over time and suddenly if screen scraping is banned. Possible solutions discussed included reducing API calls, increasing limits, using bulk APIs, adding filters like modified date, and event-driven approaches.

The member agreed that he would update the paper following this conversation and include e-tag based approach and an event-driven, webhook- based approach as options. He would also like to get some wider data holder opinions offline.

The Chair suggested getting feedback and insight from data holders within this group initially.

ACTION: Member to update discussion paper

ACTION: Data Holders in group to provide feedback on discussion paper

The Chair asked Data Holders to investigate whether any bulk commercial API data feeds they supply have any patterns which can be leveraged through these APIs and are fit for purpose?

ACTION: Data holders to review lessons learnt for commercial batch API calls that they could bring to the CDR thinking

One member discussed an [issue](#) raised in GitHub about traffic limits being too low to support a data recipient serving the Australian consumer.

They shared projections of potential future load and concern that current traffic thresholds may be too low to handle a mass migration from screen scraping to open banking. There was discussion around current traffic patterns and spikes, and whether the core issue is the throughput threshold itself or the number of API calls required.

The group acknowledged they are focusing on the right topics - reducing the number of API calls and/or increasing traffic limits.

Closing and Next Steps

The Chair noted that the action items from this meeting include:

1. DSB to create a word document to collaborate on the problem definition statement via GovTEAMS

2. Work with banks to look into the lessons learnt for commercial batch API calls that could contribute to CDR solutions

3. Data holders to look at the 'modified-since' date and see what implications there are, if any

4. Data holders to analyse the traffic spikes hit during the day/week and whether there any symptoms for concern

5.   Member to update draft to include both e-tag and event driven approach

The Chair thanked the group for active participation and attending the meeting.

The meeting closed at 15:58