



Data Standards Body

Non-functional Requirements (NFR) Consultative Group

Minutes of the Meeting

Date: Thursday 22 May 2024

Location: Held remotely, via MS Teams

Time: 14:00 to 16:00

Meeting: Meeting # 4

Attendees

Participant Members

Mark Verstege, Chair
John Adshead, AEMO
Jon Denley, Basiq
Dhananjay Gourshettiwar, Westpac

Harish Krishnamurthy, ANZ
Michael Lin, NAB
Julian Luton, CBA
Mark Wallis, Skript

Observers

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

Ben Kolera, Biza
Stuart Low, Biza
Simon Pearce, Origin Energy

Apologies

Andrew Ferris, AGL



Chair Introduction

Mark Verstege, the Chair of the Non-functional Requirement Consultative Group (NFR CG) welcomed everyone to the meeting and acknowledged the traditional owners of the lands.

The Chair noted that following approval from the Chair of the Data Standards, Michael Lin from NAB has joined the group. He welcomed him to the meeting.

He also welcomed Stuart Low and Ben Kolera from Biza and Simon Pearce from Origin Energy who attending the meeting as Observers.

Minutes

The Chair thanked the group for their comments on the minutes from 24 April 2024 meeting. The Minutes were formally accepted and will be made publicly available on the Consumer Data Standards website.

Action Items

The Chair noted that a number of Action Items arose from the last meeting. An updated was provided below:

- Continuing discussions with Xero on how to share their traffic numbers and scenarios in a confidential way. They may attend a future meeting.
- Awaiting nominations from smaller energy DHs to join the group which will be forwarded to the Chair of the Data Standards for consideration
- Draft Standards Maintenance paper updated to include e-tag based and event-driven approach
- DHs to continue to investigate the feasibility on whether the last modified date exists at the transaction level in their systems and whether a change like that was feasible

The Chair asked DHs to provide a high-level architecture walk through of how event driven architectures and moving to different type feeds works at the next meeting.

ACTION: Member to provide an update at the next meeting on how they deal with bulk data

Problem Definition Statement

The Chair noted that the Problem Definition Statement has been revised following discussion as the previous meeting. The updated statement and child-statements was provided for consideration by members:

Problem Definition Statement #2: Are the traffic thresholds appropriate to accommodate known growth of CDR usage without imposing undue over-capacity requirements on Data Holders?

Problem Definition Statement #2a: What changes to NFRs are needed to ensure Data Holders can right-size their infrastructure without over-scaling based on current traffic?

Problem Definition Statement #2b: What guidance is needed to ensure Data Holders grow their infrastructure at the appropriate rate to accommodate traffic growth without constant tweaking of NFR threshold standards?



Problem Definition Statement #2c: At what point do traffic thresholds become onerous and unviable for Data Holders to support?

The group discussed constraints needed to ensure data holders can right-size infrastructure, guidance to ensure appropriate infrastructure growth rate, and determining when traffic thresholds become onerous. They further discussed revising the problem definition statement to focus specifically on assessing potential changes to traffic thresholds first before exploring other solutions.

The DSB noted that in regard to load distribution, could an event-based architecture help recipients spread load. Other potential solutions like a queuing system or enforcing thresholds through rejection were discussed. One member suggested the problem may require rethinking the dimensionality of thresholds by API rather than broad categories.

One member noted that peak transactions per second usage is often far higher than average usage, indicating idle infrastructure capacity most of the time and understanding traffic distribution across days and times and per API would help analyse infrastructure right-sizing and scaling.

The group discussed whether tiered traffic thresholds could be defined per API based on usage levels and the constraints on the number of requests per access token force load spreading. They noted the complexity of enforcing thresholds per API versus broader categories.

The Chair summarised the discussion noting that it's getting to the point as to whether it is a longer NFR or is it a different model to fetch the data. He queried what are the general traffic distributions across the day on an average, and also what API level. Also, do we need to look at NFRs from an actual data cluster perspective or an API perspective and as the load grows how do we manage a general umbrella kind of TPS limit.

The Chair asked if the DH could do some analysis on what TPS looks like, peak and average across days and weeks at an API level as well as an aggregate and bring back to the group for discussion.

One member agreed to bring it back to the group.

One member presented a typical day in energy to provide the group with a better understanding. They had peaks at 3am, 6am and 10pm which represented almost 50% of the load in about 30 minutes of the day. A ticket was raised and ACCC stepped in, and the loads lifted accordingly with more functional behaviour.

One observer volunteered to provide a breakdown per operation across all sectors.

Rate limiting approaches

The Chair would like to discuss rate limiting approaches to protect AEMO infrastructure whilst meeting obligations to share energy usage data for large non-individual consumers.

The Chair invited a member to provide an update on how to make CDR work for single requests with an infinite number of service points.

The member outlined several potential solutions including limiting number of service points per request, adjusting service level targets based on size of requests, throttling large requests and asking consumer to retry later, and rearchitecting systems for full scalability. The full scalability model would potentially require changes to standards around entitlement checks, response payload retrieval, and response payload page return.



The group discussed details around the issues including differences with banking APIs, problems caused by needing total page count upfront, whether non-retail commercial aggregators were the original intended users, and how energy data usage patterns exacerbate existing scalability issues that banks have already worked to solve. Potential solutions discussed included adding endpoints to optimize entitlement checks, using cursor-based pagination, throttling and retry logic, and trials to test approaches.

One member noted the rules written for energy are vastly different to banking, but all data holders face similar challenges around preparing entire datasets to meet short response time requirements and that the standards were designed with simplicity in mind rather than efficiency for the data holders.

The DSB noted following the discussion, does energy need a resynching of the way the endpoints are structured or the way they operate potentially to solve this, where it was maybe based on the banking structure, but it's not really applicable to energy and for example in banking you have transactions per account but for energy you transactions for all accounts which is a bit different and maybe it should be considered holistically.

The group further discussed details around the proposed solutions like cursor-based pagination, checking entitlements per page instead of full payload, use cases driving large requests, and how banking APIs face similar issues. They highlighted that the scalability issues faced in energy exist for banking as well, where significant infrastructure investment has been required to prepare dataset readiness.

The Chair asked with entitlement checks if there were issues when returning a 422 or 404 where one of the accounts out of a list that have been sent in a bulk request result, and is there a better way to deal with that?

One member noted most of their time is pulling all of the response, but it limits the scalability of the solution. They can't rely on the primary data holders as they all get informed of a change of retailer at different times.

The Chair asked from an ADR perspective not using total pages as part of how they look at their requests, does the group see any advantage or issues at the moment with banking APIs and total pages?

The group discussed the current observations of paging behaviour, including lack of usage of total pages count, parallel page requests but not jumping around pages, and rarely accessing last page directly. They note the standards explicitly require sorting in different ways across energy and banking, impacting ability to rely on ordering.

Meeting Schedule

The Chair noted that the next meeting is scheduled for Wednesday 19 June.

Any Other Business

No further business was raised.



Closing and Next Steps

The Chair summarised the meeting noting that the Problem Definition Statement was broadly agreed on. He also agreed to reach out to the banks who offered to look at usage and API calls including the observer, which will provide current patterns of usage and collection from a banking focus. The next steps are to collate a list of the issues and the solutions we're looking at potentially based on high value.

The DSB asked the Chair to consider conducting a trial of some proposed solutions like retry headers and throttling, while also considering trials of modifications like cursor-based approach and total page and reporting back to the group. A member cautioned against expanding trial scope too far without additional engineering bandwidth.

The Chair also noted that the DSB will reach out to the group regarding issues accessing GovTEAMS. He also encouraged the group to provide any further comments on the Problem Definition Statement.

ACTION: DSB to reach out to group regarding any issues access to GovTEAMS.

The Chair thanked the group for participating for attending the meeting.

The meeting closed at 15:58