Problem statements

Teams are encouraged to pick one of the proposed data tracks. You might choose to focus on some of the prompts rather than trying to address them all. You should also consider effective ways to communicate the results (e.g. a report, an interactive visualisation, a map, and so on). Please allow time to submit all of your work including the video submission.

Useful acronyms

CHA	Community Hubs Australia
LGA	Local Government Area
AEDC	Australian Early Development Census

INTERNAL DATA TRACK

What is a ‘typical’ hub and what is an ‘outlier’?

Input data: Program, activity dates and attendance for each hub (file provided, hub is anonymous – not linked to LGA)

Opportunities: In this track there is potential to do some exploratory data analysis and look for outliers across a number of dimensions such as the programs being run by the hub and attendance rates

Challenges: It may be a challenge to present the results in an engaging way given hubs will be anonymous.

Potential outputs: A report for each (anonymised) hub, storytelling (e.g. some common ‘personas’ like Fledgling Hubs, Focused Hubs, Growing Hubs), some kind of visualisation/dashboard for benchmarking...

Prompts:

- What is a ‘typical’ hub? Based on the data is it possible to describe what the activity in a typical hub is, each term?
    - What program types
    - How often are activities run and
    - How many attendances
- What is the deviation from this ‘typical’ hub?
    - What are the outliers?
    - How do they differ from the norm?

- Can deviation be ‘good’?
- Can you identify any other data that might be useful in answering some of these questions? (data that CHA could consider capturing)?

EXTERNAL DATA TRACK

Where should CHA consider opening new hubs?

Input data:

- Information about hubs (schools and locations) as available from the CHA website (file provided)
- PHIDU Social Health Atlas by LGA
- School Profiles
- AEDC time series

On request:
We have data on Migration Settlements the last few years. It is not clean but if you need it talk to Paola Oliva-Altamirano (mentor/organiser).

Opportunities: In this track there is potential to link a number of public datasets to provide a clearer picture to CHA about the populations they hope to serve (through existing and new hubs).

Challenges: CHA already have a set of criteria for selecting locations (LGAs) for hubs so to be useful, participants need to go beyond the work already done. For this challenge some background (or interest) in geospatial analysis might be advantageous.

Potential outputs: A report for each LGA, interactive dashboard/map, a set of criteria/scoring system/thresholds to determine in which LGAs and/or schools CHA could consider opening a new hub.

Prompts:

1. Given the locations where hubs already exist, what is the profile of the area (LGA) over time?
    - Consider:
        - What is the social profile of the LGA in which the hub is based?
            - Female age distribution
            - English proficiency
            - Humanitarian and family migration
            - Early childhood education enrolments

        - AEDC results, e.g. children at risk or vulnerable
        - etc
2. Given the schools that already have hubs, what is the profile of the ‘typical’ school?
    - Type of school
    - ICSEA, SEA distribution (from school profiles)
    - Staff numbers
    - Student numbers
    - Language other than English
    - Proximity to other schools
3. Based on your analysis, where would you suggest CHA consider opening new hubs?
    - Which LGAs?
        - What is your rationale, which factors did you take into account?
    - Which schools?
        - What is your rationale, which factors did you take into account?
4. Can you suggest a set of criteria that CHA should apply when selecting new hub locations, e.g. thresholds/triggers CHA could set based on the data?)
5. Would the same criteria apply for rural and regional locations, or should there be different thresholds?
6. Can you identify any other data that might be useful in answering some of these questions? (either data that already exists or data that CHA could consider capturing)?