The Category of De-identified Data doesn't exist - or its severely limited

The Draft Report lists the Framework of recommended approaches which consists of 4 categories. Which are:

- Non-personal / non-confidential
- De-identified
- Identifiable
- Confidential / protected

A critical issue for "De-identified" data is users can be re-identified thru the combination of additional data.

As way of a specific example lets assume I personally live in Sydney and lets assume that my name is deleted but to have marketing meaning I leave the zipcode and shopping behavior. So while I am de-identified a data aggregator I does need to retain meaningful marketing data. By finding an obscure shopping behavior – I bought a garden sprinklers at Chatswood Bunnings last weekend I can re-identify a specific user. Motivated party can reidentify the users.

The notion of de-identified data is extremely limited use-case and in practice does not exist. If you take out all the meaningful data then it has no behavioural marketing value. There is an inherent tension between stored behavioural data and the ability to combine new data. With enough data one can always be re-identified.

Marketing data aggregators claim that their data has user de-identified. In practice this approach does not exist. All de-identified data can be used by motivated parties to re-identify the user, by combining with new data. So the original data aggregator will claim de-identified data but if I combine the Bunnings shopping dataset then I can re-identify myseld. De-identifed dataset + Bunnings enables my re-identification. In practice a de-identified dataset can always be used to re-identify a user.