**Decisions I am making.**

* I think we want the final data to be entirely raw values merged together
  + Ie. we don’t create % private in the raw data, just present the data as it comes
* I think we want it to be panel data with long form years but and long form variables
* Where APIs are not available from Gov – we will publish the raw ODS, XL, CSV files on our github and clean it all directly from there.
  + We could build machines that scrape the data quite easily – but I actually think the reproducibility of this is very minimal gains
    - And I don’t know if dockerised functions of our code will handle downloading data into memory – which I think will be necessary without APIs.
* The final sheet will work by pulling different functions so that the reproducible file looks like eg.
  + Spend\_Data <- clean\_expenditure\_data()
  + Children\_Data <- clean\_SSDA\_data()
  + Housing\_Data <- clean\_homeless\_data()
  + Full\_Data <- merge\_all\_data(Spend\_Data, Children\_Data, Housing\_Data)
    - Magic
* We should include all data available at LA level.
  + Eg. education outcomes by LA