

# Data Integration Knowledge @ Stats NZ

16<sup>th</sup> Mgt Seminar ESCAP-SIAP: Session B  
December 2021

Gary Dunnet, Deputy Chief Methodologist

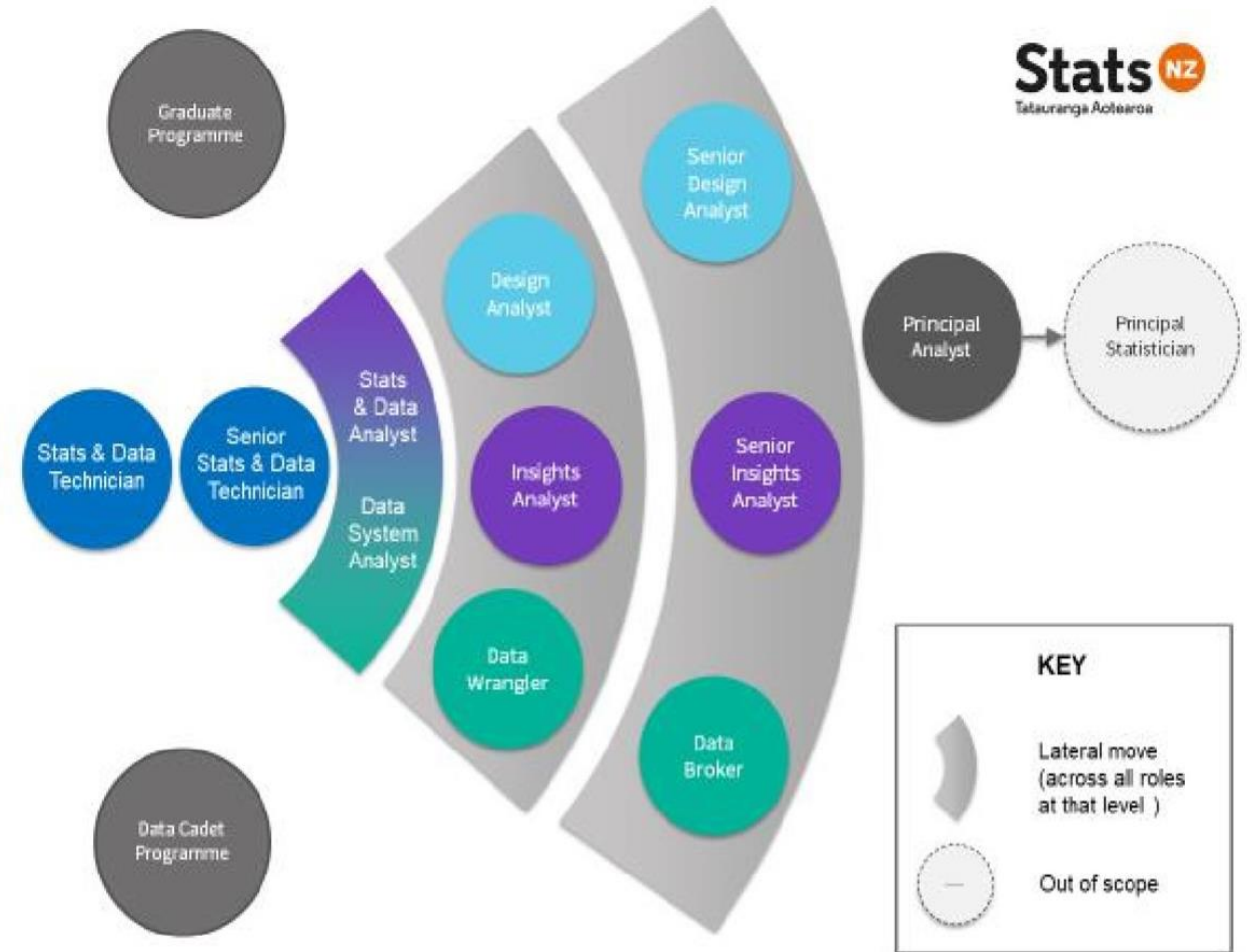


# A few areas of specific focus

- ⦿ Data preparation:
  - ⦿ Often overlooked, but requires the bulk of the work.
  - ⦿ Developing knowledge of the data can be tricky and relies on relationships.
  - ⦿ Need to have clear & initiative documentation.
  - ⦿ Relationship between Subject Matter, Methods, DBS, Integrated Data curators.
- ⦿ Use of integrated data:
  - ⦿ Skill in itself – methods & data knowledge
  - ⦿ Data Integration Researchers Community of Practice & GitHub
  - ⦿ An important area of expertise that needs development
- ⦿ Two ‘new’ roles (Data Broker & Data Wrangler) and extensions to existing roles of Insights Analyst & Design Analyst.

# Statistical Job Family Review

- ⦿ Redesign of workforce capability
- ⦿ ‘Lift & Shift’ - July 2019
- ⦿ Data Wrangler & Data Broker
- ⦿ Coaching & mentoring
- ⦿ Investment in technical leadership



Statistical, Data, and Analytic Job Family Career Pathway (aka ‘wif’ diagram)

# Ngā Tikanga Paihere



Mā ngā tikanga e arahina - Be aided by good principles

# Data Integration roles

- ⦿ Data integration enables the use of admin data to supplement surveys / censuses / registers.
- ⦿ All those involved need to consider conceptual issues, population definition, units, data relationships, quality assessment, methods for producing estimates, data extraction requests, planning outputs (to name a few considerations).
- ⦿ Learning about the art and science of data integration increases data literacy (esp. 'black-box' tools, linkage error and bias).
- ⦿ It's necessary for methodologists who are working on surveys / censuses / registers to have this working knowledge to guide interpretation of the findings.

# Specific methodological role

- ⦿ Provide advice and assistance on, and develop methodology where needed for:
  - ⦿ Quality Assessment.
  - ⦿ Preparation for linkage such as cleaning, editing, standardisation and imputation
  - ⦿ Linking the source data files together incl. information gathering; use of integration software; new methodologies; and assessment of linking outcomes
  - ⦿ Analysis of linked data
  - ⦿ Data Privacy, Data Sovereignty, Confidentiality & Security
- ⦿ Documenting methodological work, formally (for publishing) or informally as appropriate.
- ⦿ Ensuring a quality plan for the statistical methodologies exists and signed-off & peer-reviewed.

Any questions?

Gary Dunnet  
[gary.dunnet@stats.govt.nz](mailto:gary.dunnet@stats.govt.nz)