

# Kokkuvõte Uus-Meremaa Statistikaameti külastusest

Statistika nõukogu koosolek 20.06.2018

Maiki Ilves

# Uus-Meremaast

Maoori keeles: **Aotearoa** ('pika valge pilve maa')

- Rahvaarv: 4,7 mln (maoore u. 650 000)
- Pindala: 270 534 km<sup>2</sup>
- Ametlikud keeled: inglise, maori (ida-polüneesia juurtega) ja Uus-Meremaan viipekeel

NZ Statistikaamet

- töötajaid: u. 600
- aasta eelarve 68 mln eurot

# Külastuse eesmärgid

Delegatsioonis 2 Eesti SA, 1 RMIT, 1 TEHIK, 3 Soome SA,  
2 Austraalia SA.

Saada ülevaade ja õppida

- pakutavatest andmeteenustest
- andmehalduse rakendamisest riigi tasandil
- nende tehnilistest lahendustest (andmejärv)
- metaandmete haldusest

# Stats NZ strateegia

- Tunnustatud kui „Data Agency“
- Stat NZ juht istub valituskabinetis (ametipost „Chief Government Data Steward“)
- Ei opereeri GSBPM järgi (fookus mitte statistikal, vaid andmetel)
  - 4 valdkonda: Data System Leadership, Insights and Statistics, Data Services, Organisation Capability and Services, (Data Ventures)

<http://archive.stats.govt.nz/~/media/Statistics/about-us/strategic-direction/strat-direction-2016.pdf>

# Our vision

Unleashing the power of data to change lives

# Our purpose

Empowering decisions by adding value to New Zealand's most important data

# Our goals

**To help improve outcomes for all New Zealanders, our goals are to:**

- double the value of data provided by Statistics NZ to New Zealand by 2018
- create a tenfold increase in value of the data provided to New Zealand by 2030

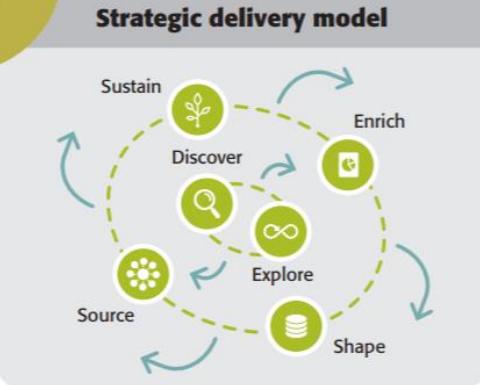
# Our future state

**Who we are and how we act**



**What and how we deliver**

**Creating customer value**



**What drives us**

**IDARE character**

|   |             |
|---|-------------|
| I | INQUISITIVE |
| D | DRIVEN      |
| A | AGILE       |
| R | RESILIENT   |
| E | ENGAGING    |

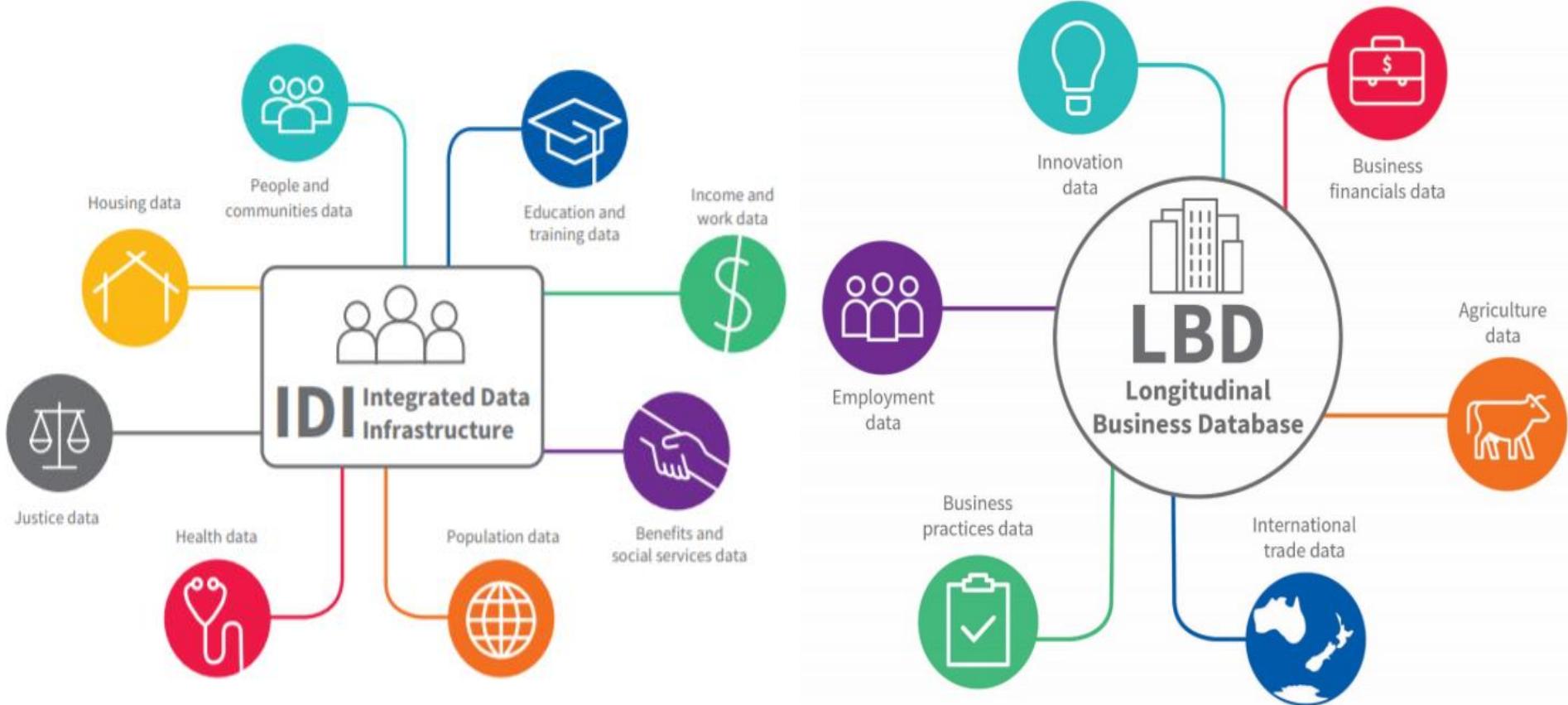
# Strateegia - tuleviku toimimismudel

| Outside In Approach   | Instil customer centricity: Process, capability and culture shift towards human centred design, co-design and collaboration. | Develop an accelerator: A physical space and mind set for rapid development, testing and unleashing.                | Product management and retirement: Active management of products' lifecycles.   | Design physical space: Make physical space available for more collaborative, open and human centred design culture.   | People and capability gaps: Gaps include co-design, collaboration and most viable product development and testing.                      | Develop performance metrics, governance structure and tools: Identify information and governance structure for human centred agility.         | Relationship management systems and tools: Identify the relationship management needed for effective customer relationships. |
|-----------------------|--|---|---|---|---|---|--|
| Stewardship           | Develop a Stewardship role: Define, establish and embed a Stewardship function.  | Brand development: Define the brand and actively work to manage public perception.                                  | Legislation change, lobbying and regulation development: Develop and provide the regulatory environment for the wider ecosystem.          | Fill Stewardship capability gaps: Focus on leadership capabilities for growth and change, and data and information management.                                  | Develop performance metrics, governance structure and tools: Identify information and governance structure for human centred agility.   | Relationship management systems and tools: Identify relationships needed to act as an effective stewards.                                     |  |
| Consultative Approach | Develop a Consulting role: Define, establish and embed a Statistics and Government consultative approach.                    | Design physical space: Make physical space available for more collaborative, open and human centred design culture. | Legislation change, lobbying and regulation development: Mitigate legislative and regulatory issues which prevent open consultation.      | Fill Consulting capability gaps: Focus on capabilities for insights, data and information management, agile, and product management.                            | Develop performance metrics, governance structure and tools: Identify information and governance structure for a consultative approach. | Relationship management systems and tools: Identify relationships needed to deliver effective and valuable consultative services.             |  |
| Data Offering         | Big data environment development: Develop necessary infrastructure for an integrated and robust big data environment.        | Digital data channel development: Provide a suite of digital channels for sourcing and unleashing data.             | Legislation change, lobbying and regulation development: Mitigate legislative and regulatory issues around receiving and unleashing data. | Develop a secure technology environment: Develop environments to safely explore and experiment with new tools and technologies.                                 | Identify and fill capability gaps: Focus on analytical insights, data and information management, and statistical methods.              | Develop performance metrics, governance structure and tools: Identify information and governance structure for receiving and unleashing data. | Relationship management systems and tools: Identify relationships needed to effectively source and unleash data.             |
| Processes             | Value model process re-engineering: Modify processes so data is easier to use for developing new products or unleashing.     | Deconstructing the statistical production workflow: Unpick legacy approaches and modularise data and applications.  | Develop performance metrics, governance structure and tools: Identify the right information and governance structure for process change.  |   |   |   |  |
|                       |  |   |   | <div style="border: 1px solid black; padding: 5px; display: inline-block;">           Key<br/>           Change stream<br/>           Workstream         </div> |   |   |  |

# Andmeteenuse võimekus

- Alustasid 2014, tunnustatud liider maailmas („D7 nations“)
- Esimene äpp oli karjääri planeerimise lahendus
- Eesmärk jagada andmeid institutsioonide vahel (nn. “open government data“) ja kasvatada andmehalduse võimekust institutsioonides
- Kliendisegmendid – Institutsioonid, teadlased, KOV
- Kasutuslood näited
- Andmed on nii nagu nad saabusid „as-is“
- Andmestikud otsest tuvastamist välistavad
- Algne ülesande hindamine („we can, but should we do it ?“)

Stats NZ's Integrated Data Infrastructure (IDI) is a large research database containing de-identified microdata about people and households.



## Health data

- B4 School Checks – from 2011
- Cancer registrations – from 1995
- Chronic conditions – from 2007
- General medical services claims – from 2002
- Health tracker – 2006-13
- Laboratory claims – from 2003
- Mortality – from 1988
- Immunisation – from 2006
- National non-admitted patient collection – from 2007
- Pharmaceuticals – from 2005
- PHO enrolments – from 2003
- Population cohort demographics and addresses – from 2004
- Mental health and addiction – from 2008
- Publicly funded hospital discharges – from 1988
- National Needs Assessment and Service Coordination Information System (SOCRATES)

## Justice data

- Recorded crime: offenders – from 2009
- Recorded crime: victims – from 2014
- Court charges – from 1992
- Sentencing and remand – from 1998

## People and communities data

- Auckland City Mission – from 1996
- Migrant Survey – from 2012
- Driver licence and motor vehicle registers
- Longitudinal Immigration Survey of NZ – 2005-09
- General Social Survey – 2008-2014

## Population data

- Border movements – from 1997
- Visa applications – from 1997
- Departure and arrival cards – from 1997
- 2013 Census
- Births, deaths, marriages, and civil unions – from 1840

## Income and work data

- Tax and income – from 1999
- NZ Income Survey – from 2006
- Household Labour Force Survey – from 2006
- Survey of Family, Income and Employment – 2002-10
- Household Economic Survey – from 2006

## Housing data

- Tenancy – from 2000
- Social housing – from 1980

## Education and training data

- Early childhood education participation – from 2008
- Primary education – from 2007
- Secondary education – from 2004
- Tertiary education – from 1994
- Industry training – from 2001
- Targeted training – from 2001

## Benefits and social services data

- Benefits – from 1990
- Youth services – from 2004
- Children's Action Plan – from 1996
- Working for Families – from 2003
- Child, Youth and Family – from 1991
- Student loans and allowances – from 1992
- ACC injury claims – from 1994

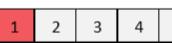
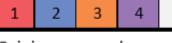
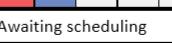
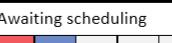
# Uued andmed IDIsse

## Summary of applications to add data to the IDI and LBD

Note: suitable for printing on A3 paper

| Status:                       |  |
|-------------------------------|--|
| <b>1. Initial discussions</b> | Discussions are underway between Integrated Data and the applicant about the feasibility of including this data. |
| <b>2. Gaining approvals</b>   | Completing a business case, privacy impact assessment (PIA) and/or MoU, where applicable.                        |
| <b>3. Scheduled</b>           | Data has been scheduled for loading.   |
| <b>4. Build in progress</b>   | The system build is in progress in preparation for linking.  |
| <b>5. Load in progress</b>    | Data is being loaded as part of the current refresh (full integration) or loaded separately (ad hoc load).       |
| <b>6. Complete</b>            | Data is available for researchers to use.  |

### Full integration (IDI)

| Priority | Dataset  | Status   | Scheduled date | Requested by                                    | Comments  |
|----------|--|--|----------------|---|---|
| 1        | Family Start                                   | Build in progress<br>     | Mar-18         | Oranga Tamariki                                 | Expected to be available by the end of April  |
| 2        | CoreLogic                                      | Initial discussions<br>   | TBC            | MBIE and academic researchers                   | Legal barrier preventing integration  |
| 3        | Social housing data from MSD                   | Build in progress<br>     | Mar-18         | MSD   | Expected to be available by the end of April  |
| 4        | Te Kupenga and Disability Survey data          | Build in progress<br>     | ASAP           | Various government agencies and academic groups | Both datasets will be ad hoc loaded as soon as the IDI ad hoc load process can load wide datasets             |
| 5        | MOE special education data                     | Gaining approvals<br>     | TBC            | Ministry of Education                           | No longer to be loaded in December due to cancellation of the December refresh. Will be re-scheduled for 2018 |
| 6        | Early Start                                    | Gaining approvals<br>     | TBC            | Early Start Project Ltd                         | Early Start will be ad hoc loaded once the appropriate approvals have been gained.                            |
| 7        | Parents as first teachers                      | Gaining approvals<br>     | TBC            | Oranga Tamariki                                 | No longer to be loaded in December due to cancellation of the December refresh. Will be re-scheduled for 2018 |
| 8        | Programme for assessment of adult competencies | Awaiting scheduling<br>  | TBC            | Ministry of Education                           |   |
| 9        | Crash analysis                                 | Awaiting scheduling<br> | TBC            | New Zealand Transport Agency                    |   |

# **IDI toetavad vahendid**

- **Wikit**, mis sisaldab kasulikku teavet ja ressursse (avalik)
- **IDI-koodide jagamise ruumi** (juurdepääs Wiki kaudu), kus teadlased saavad talletada koodi, mida nad soovivad jagada kõigi IDI kasutajatega
- **Metaandmete andmebaasi**, mis sisaldab üksikasjalikke metaandmeid konkreetsete kogude kohta (juurdepääs Wiki kaudu – avalik)
- Online koostöö keskkond **MeetaData**, kus kasutajad saavad üksteisega kontakteeruda, jagada teadmisi, küsida küsimusi ja jagada koodi (avalik)
- Juurdepääs IDI-le **Microsoft SQL Server Management Studio** kaudu

# Andmehaldus – Data Leadership Hub

- Data Standards (kuidas ja mida koguda)
- Data Stewardship (kuidas hallata, enesehindamine)
- Data Knowledge Centre (kust leida? (*easy-to-find*))
- Data Management/Consultancy (nn.“andmete“ mentorlus)

(<https://www.stats.govt.nz/assets/Uploads/Data-leadership-fact-sheets/fact-sheet-data-leadership-hub-Mar-2018.pdf>)

# „Data system leadership“ printsiibid

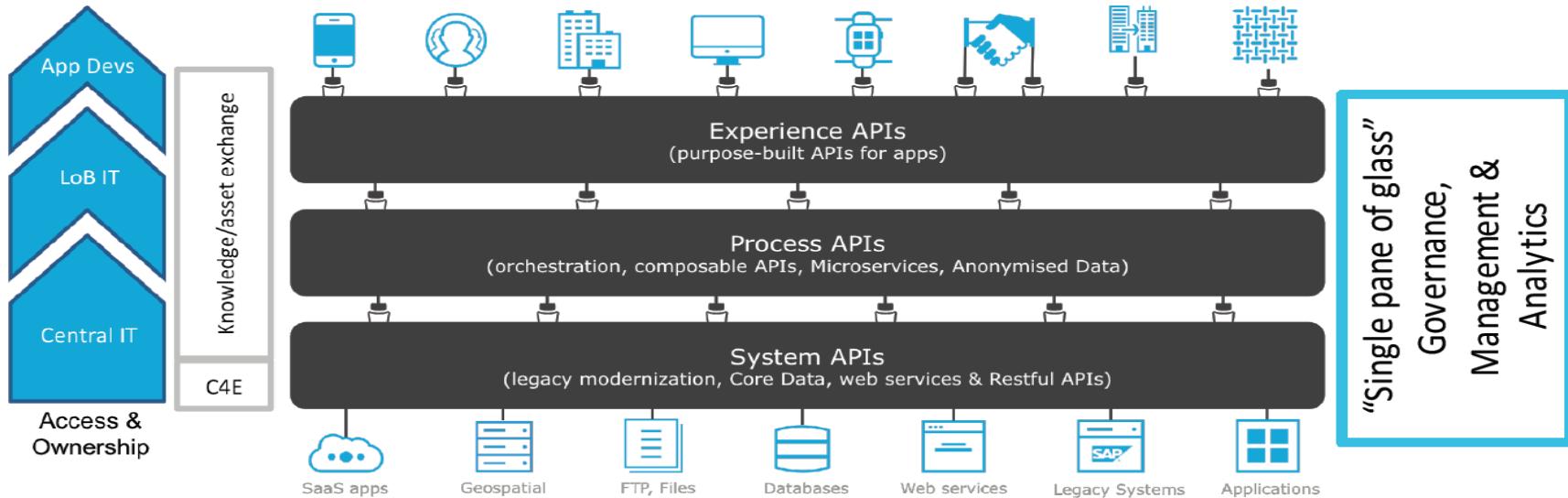
- Seadusandlus, seadusandlus, seadusandlus!
- Väiksel maal on suured võimalused (NZ)
- Ole usaldusväärne – andmed ei leki, kasutus reguleeritud, privaatsus kaitstud
- Ole vastutustundlik (5 SAFE Framework  
[https://en.wikipedia.org/wiki/Five\\_safes](https://en.wikipedia.org/wiki/Five_safes))
- Ole võimekas – **juhi ootusi, tee mida lubasid**
- Tegutse reputatsiooni nimel kahel suunal – valitus(asutused) ja avalikkus
- Erinevate allikate andmete linkimine suurendab kõikide osapoolte võimalusi – nad saavad sellest (lõpuks) aru
- Vajalik valitsusala ministeeriumi toetus

## The 'five safes' framework for the IDI and LBD



1. **Safe people** – researchers can be trusted to use data appropriately and follow procedures.  
Researchers must pass referee checks before we allow them to work with data. We require them to sign a declaration of secrecy under the Statistics Act 1975 and follow our rules and protocols. Researchers who break our protocols can be banned, blacklisted, or prosecuted.
2. **Safe projects** – the project has a statistical purpose and is in the public interest.  
Research is restricted to the analysis of groups, not individuals, and must be in the public interest. This means that the research is focused on finding solutions to issues that are likely to have a wide public benefit. The Government Statistician or delegated authorised person signs off all research proposals.
3. **Safe settings** – security arrangements prevent unauthorised access to the data.  
Data can only be accessed through a secure Data Lab environment. Computers are not connected to a network and only Stats NZ staff can release data to researchers.
4. **Safe data** – the data inherently limits the risk of disclosure.  
We de-identify data, which means we remove personal identifying information such as names and addresses, and encrypt (ie replace with another number) identifiers such as IRD and NHI numbers. See our de-identified data fact sheet for more information about the benefits, risks, and possible uses of de-identified data. Researchers in the IDI get access only to the data relating to their research; researchers in the LBD get access to all LBD data.
5. **Safe output** – the statistical results produced do not contain any identifying results.  
Researchers must confidentialise output before it can be released from the Data Lab, and Stats NZ staff double-check results to ensure individuals cannot be identified. See Microdata output guide for the methods and rules researchers must use for confidentialising output produced from Stats NZ's microdata.

# Tehnoloogia - uus arhitektuur



- Teenuse põhiselt arhitektuurilt andmepõhisele arhitektuurile
- Rakenduse põhiselt tarbija põhisele arhitektuurile
- Teeninduselt iseteenindusele
- „Käsitsi“ ja „poolkäsitsi“ tehnoloogiatelt automaatsetele tehnoloogiatele

# Tehnoloogia detailsemalt

- Küsitlused – SalesForce
- Metaandmete süsteem – Colectica, Aria
- Andmeladu – MS-SQL (kasutavad privaatpilve!)
- Andmehõive – Pentaho
- Andmejärv – Hadoop, Cloudera (eraldid andmelaost!), Hive, Impala
- Analüüs – SAS / R
- Visualiseerimine – R Shiny
- Avaldamine - .STAT (versioon 5), PC-Axis
- Sisemine tööprotsess – Microsoft Teams

# Metoodika arendused

## ■ Admin.andmete laialdasem kasutuselevõtt

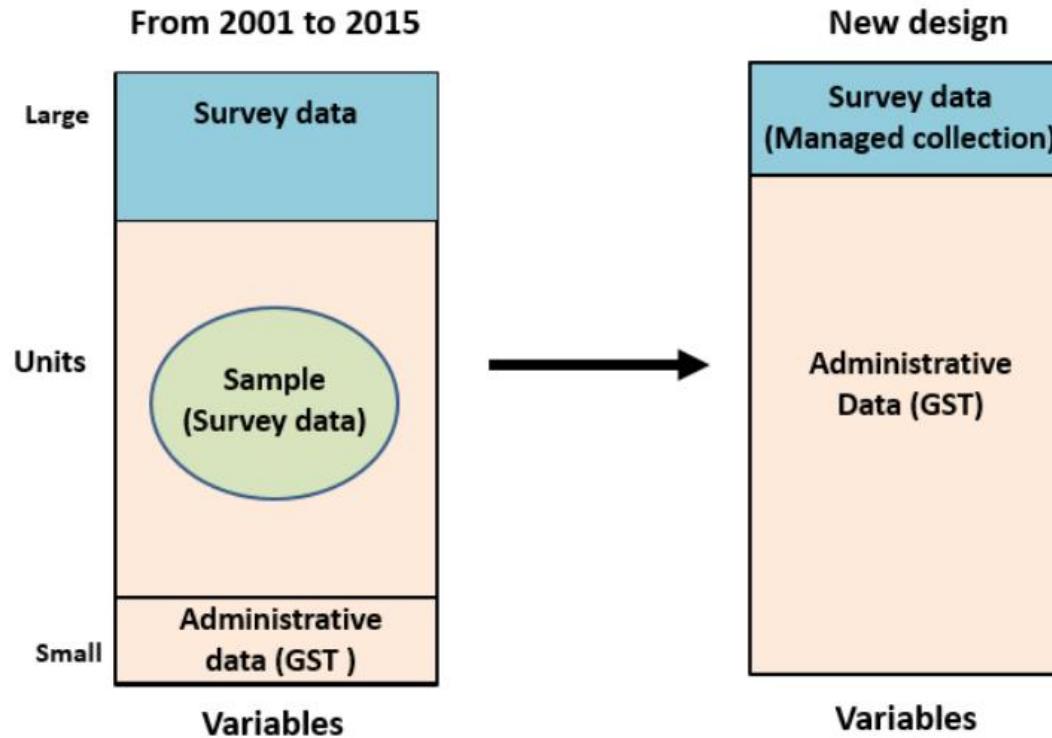
Statistics NZ's vision: "to transform the collection and production of statistics" -> Administrative data will be the primary source of information, supplemented where necessary by direct collection.

## ■ Hinnaindeksid: skännerandmed ja veebikammimine

## ■ Rahvastikuprognoosid kasutades Bayes'i hinnanguid

## ■ Automaatne konfidentsiaalsuskontroll „Confidentiality on the fly“

# Paradigm shift in sub-annual financial designs

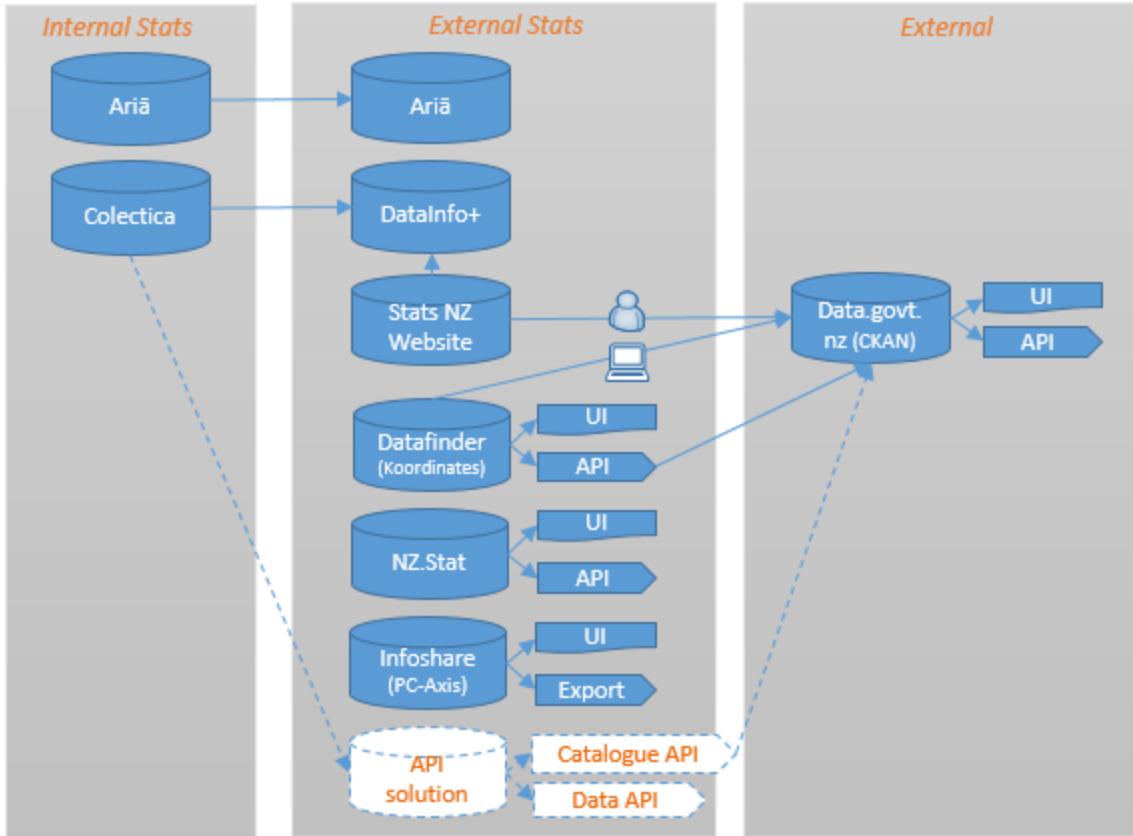


Goal: Maximising the VAT administrative data use in manufacturing, wholesale trade and selected services industries

# Metaandmed ja open data

- Kaks süsteemi metaandmete keskseks hoidmiseks ja haldamiseks
  - Colectica – tunnused, muutujad, struktuurne ja tehniline meta
  - Aria (<http://aria.stats.govt.nz/aria/>) – mõisted, klassifikaatorid
- NZ Open Data Action Plan  
(<https://data.govt.nz/community/news/open-data-action-plan/>)

# Metaandmed ja open data- vahendid



## Hetkeseis

- Puudub API, mis kataks ära kogu vajaduse
- Manuaalsete ja automaatsete laadimisprotsesside „segu“

## Ülemineku faas

- Uue API arendus
- Automaatne laadimisprotsess

## Tulevik

- Tooteid ja teenuseid pakutakse API liidese kaudu
- Automatiseritud konfidentsiaalsuse tagamine, mis toetab klientide suuremat iseseisvust (iseteenindus)

## Classifications

### Info

|              |                     |
|--------------|---------------------|
| Abbreviation | ANZSCO              |
| Audience     | OSS                 |
| Lifecycle    | Released            |
| Version      | 1.2.0               |
| Valid from   | 13-May-2013         |
| Last update  | 21-02-2018 04:01:02 |

### Levels

|                 |                |
|-----------------|----------------|
| Levels          | 5 (1548 Codes) |
| Major Group     | 9              |
| Sub-Major Group | 44             |
| Minor Group     | 99             |
| Unit Group      | 361            |

# Australian and New Zealand Standard Classification of Occupations V1.2.0

[Overview](#) [Browse](#) [Advanced](#) [Usage](#) [Discussion](#)

- **221** Accountants, Auditors and Company Secretaries [2]
- **222** Financial Brokers and Dealers, and Investment Advisers [3]
- **223** Human Resource and Training Professionals [3]
- **224** Information and Organisation Professionals [8]
  - **2241** Actuaries, Mathematicians and Statisticians [3]
    - 22411** Actuary
    - 22412** Mathematician
    - 22413** Statistician
  - **2242** Archivists, Curators and Records Managers [4]

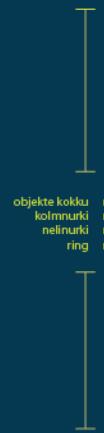
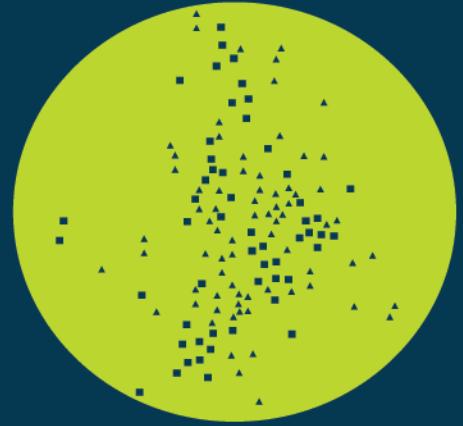
# Mida nemad õppisid?

- Alusta väikselt ja laiene kasutades paindlikku ja iteratiivset protsessi jälgides kasutajate vajadusi
- Juhi nii andmete kasutajate kui ka andmeandjate ootusi vastavalt oma keskkonna võimalustele ja arengutele
- Tee andmete leidmine ja neile ligipääs lihtsaks
- Ole läbipaistev ja kommunikeeri uusi andmeallikaid ning uusi teenuseid
- Ürita ennustada tuleviku muutusi, nõudlust andmete järele, nendele ligipääsu meetodeid ja kasutamise viise

# Mida meie õppisime?

- Jaatavad teemad
  - Andmejärv („data lake“) ja andmete virtualiseerimine
  - Andmehaldus – Kaasamine (trust) vs. Sund
  - Sisesed integratsioonid API tasandil, mitte otse
  - Modulaarne süsteem („built services, not apps“)
- Mõtted / ideed
  - Tiim, kes toetab institutsioone andmehalduses
  - Kommunikatsioon - A4 infolehed, infograafika
  - Eksperimentide tegemise kultuur (s.h keskkond)
  - Kliendid kutsutakse tööde tegemise juurde
- Tõdemus – arhitektuuriliselt ei ole loodud midagi uut vaid on oskuslikult kasutatud olemasolevaid lahendusi





objekte kokku n 129  
kolmnurki n 74  
nelmnurki n 54  
ring n 1

ES eesti  
STATISTIKA