Population estimates for Ireland using administrative data sources

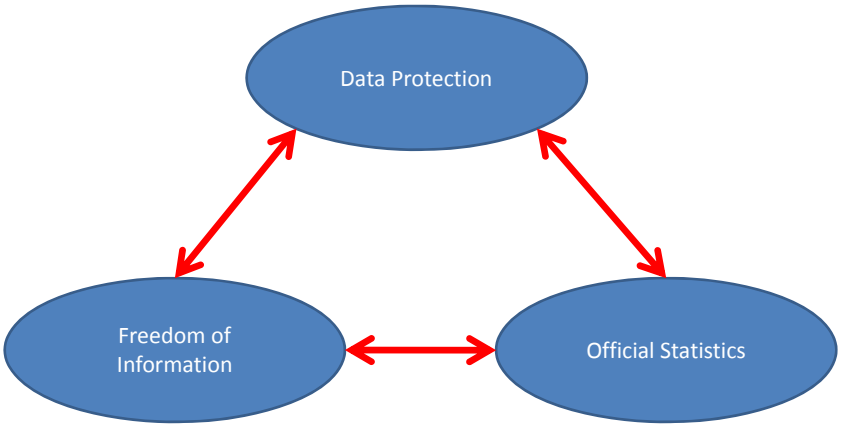
….. first steps

International Conference on Census Methods
National Academy of Sciences
July 31 – August 1, 2014
John.Dunne@cso.ie

Overview

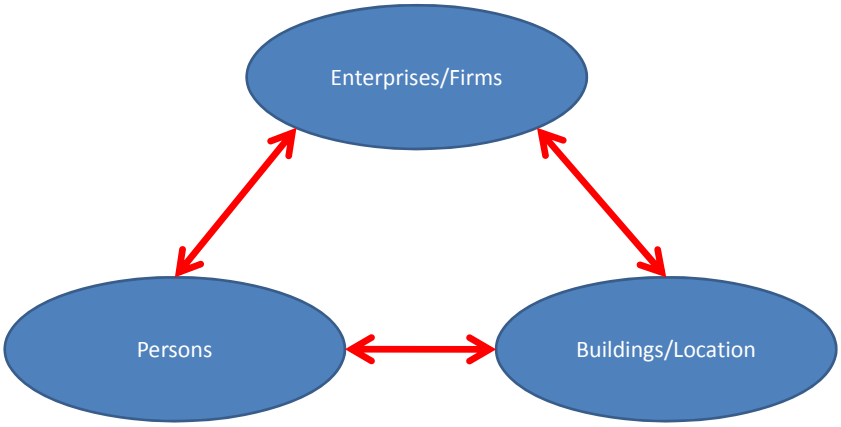
- Legal environment
- Administrative Data Centre ADC @ CSO
- National Data Infrastructure
  - 3 pillars (Persons, Business, Property)
- Overview of Person Activity Register
- Population estimates from admin data
  - Some initial exploratory work

Legal environment



Key : 3 Legislative pillars

NDI - Joined up data

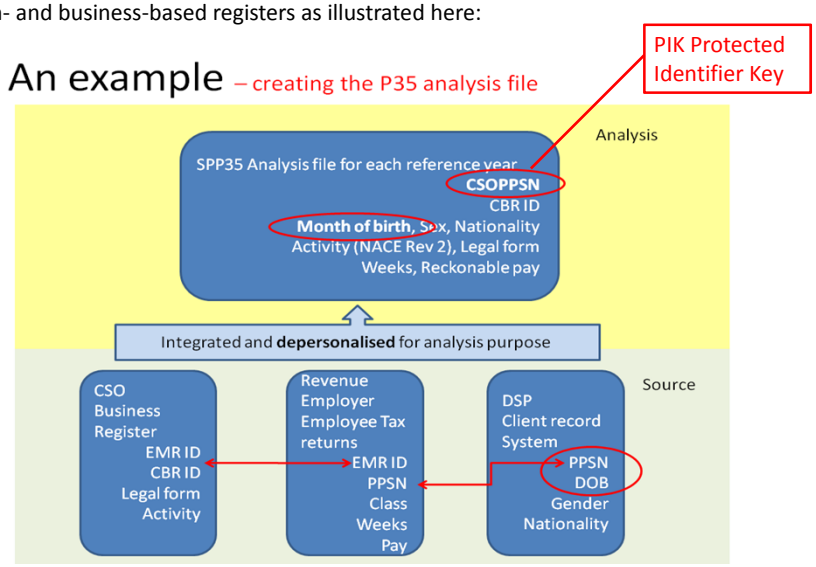


Key : Permanent official identification

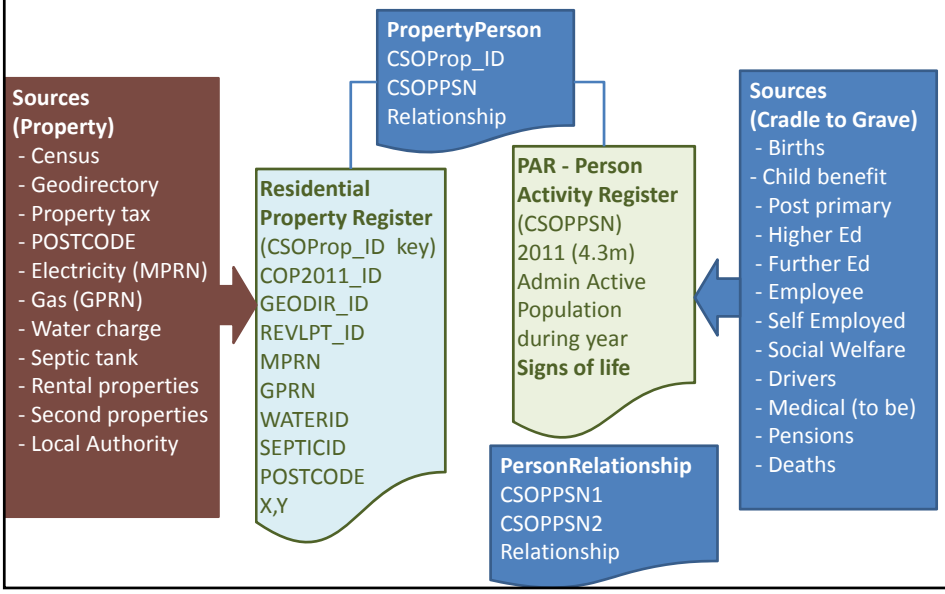
Rationale for a National Data Infrastructure
http://ipa.ie/pdf/Forum_Vol_61_4.pdf

ADC @ CSO

An example of an Analysis tier data flow is the P35 (employee) dataset, which links person- and business-based registers as illustrated here:



Statistical view of NDI – Persons and Property



Purpose of PAR Person Activity Register

'Cradle to Grave'

Purpose :

To track specific cohorts within the population

To act as a master key to administrative data in identifying feasibility of specific analyses

To investigate population structures over time

- Population estimates?

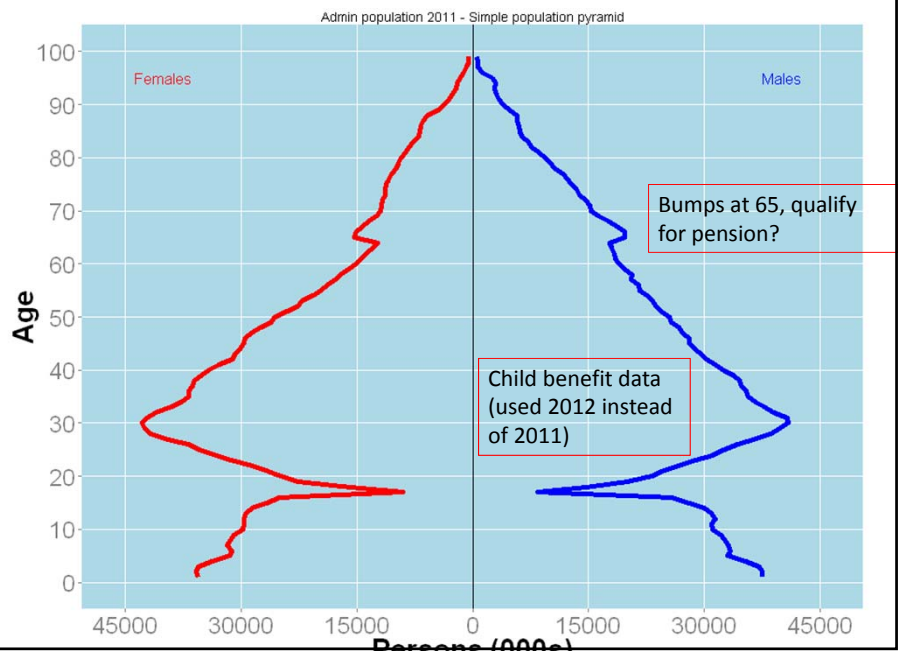
Population estimates using admin data?

For the year **2011**,
taking a "signs of life" approach
the Person Activity Register identified activity from

4.35m persons

on Public Administration Systems.

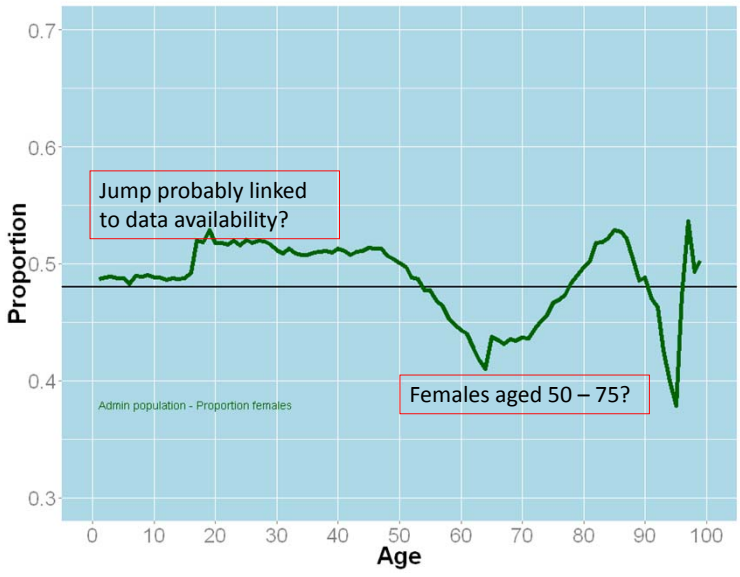
2011 Admin count by age and sex



Bumps at 65, qualify for pension?

Child benefit data (used 2012 instead of 2011)

Look at proportion female

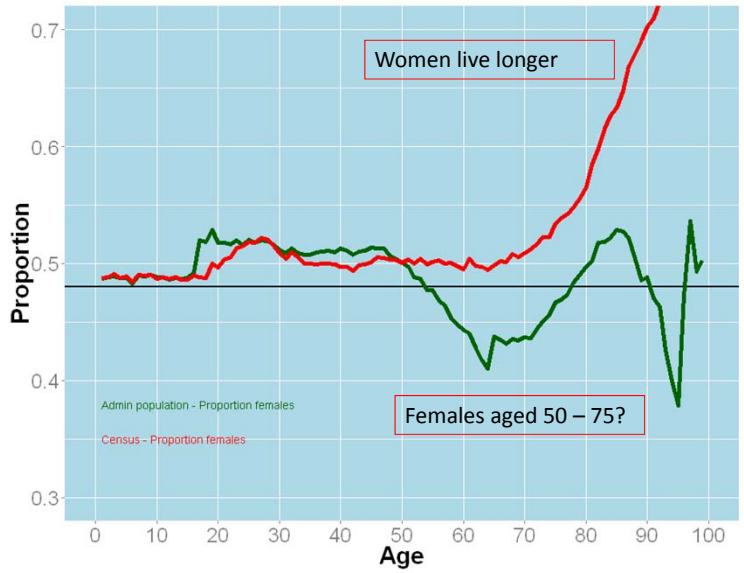


Jump probably linked to data availability?

Females aged 50 – 75?

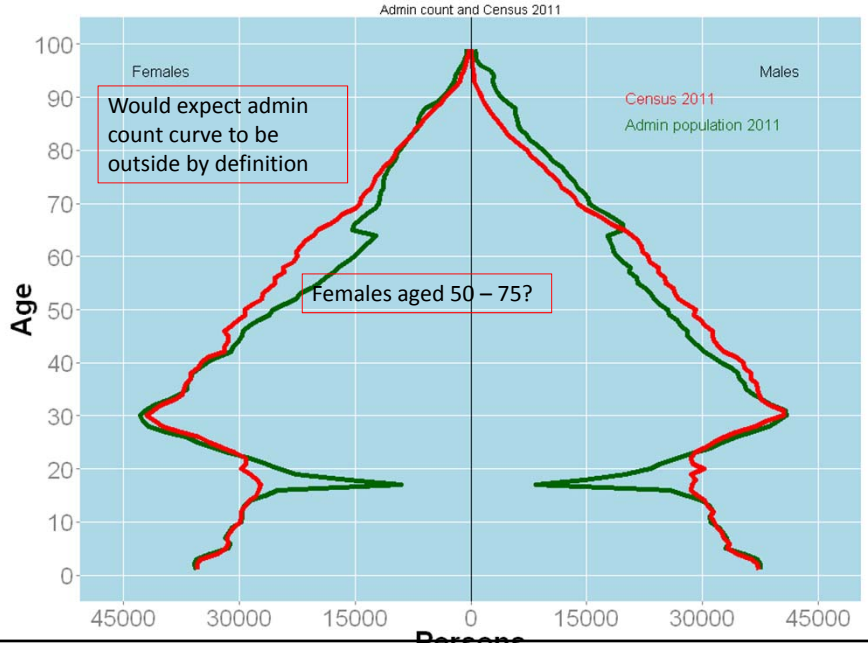
0.48

Look at proportion female
- admin V census



0.48

Admin Count V Census 2011



Where to now?

- Different measures
- Dependency on admin data sources over time
- A coverage survey ? (Can be costly)

Is there another way?

DSE Dual System Estimate or Capture Recapture

A second independent source such that

	In 2nd Source	Out 2nd source
In PAR	**A**	**B**
Out PAR	**C**	**D**

Population estimate = A + B + C + D

where D = C * (B/A)

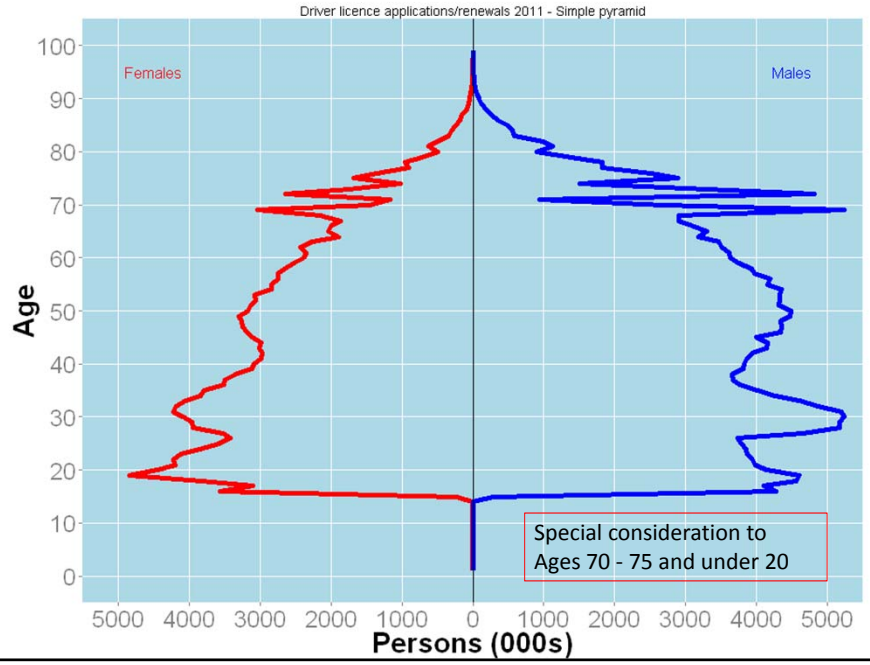
An example

	In 2nd Source	Out 2nd source
In PAR	**900**	**4.5m**
Out PAR	**100**	**0.5m**

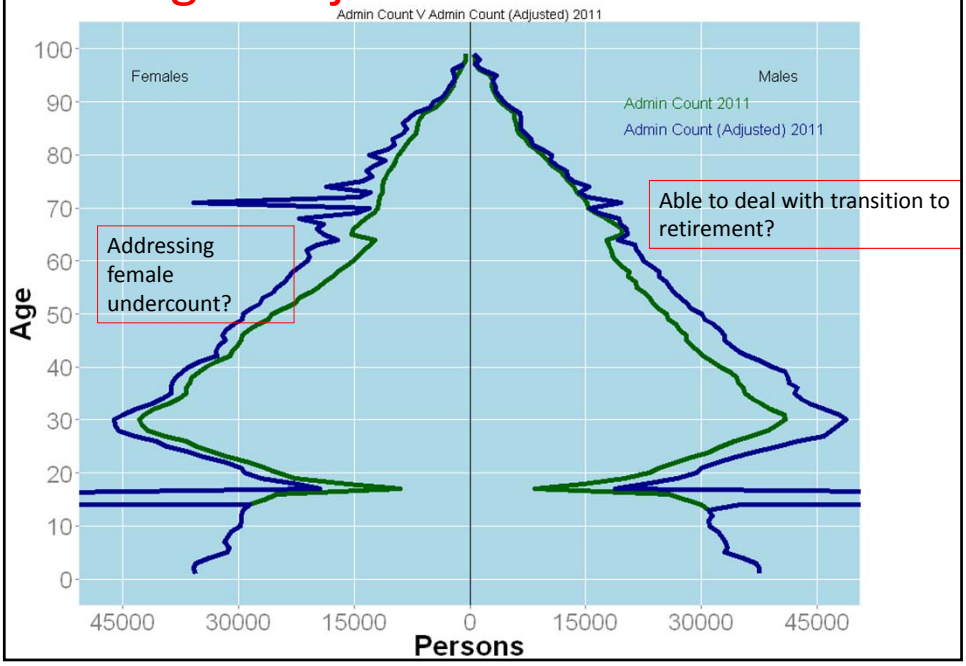
Population estimate = 5m + 1000

2 Questions
Is there such a source?
What are we estimating?

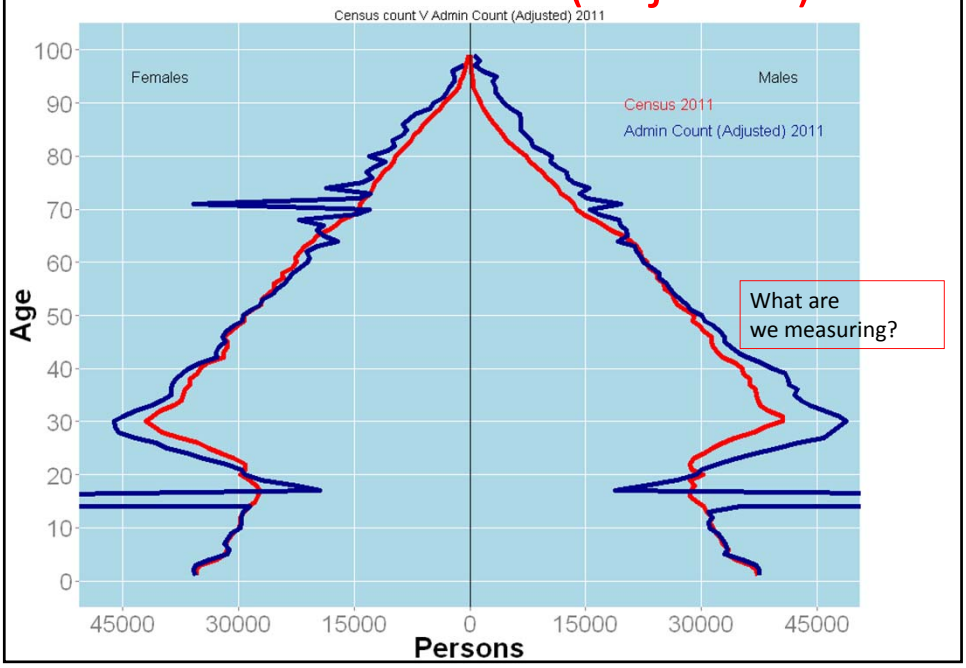
What about driver licence renewals?



Looking at adjusted admin counts



Census V Admin Count (Adjusted)



What are we measuring?

Some considerations

	In DL Renewals	Out DL Renewals
In PAR	**A**	**B**
Out PAR	**C**	**D**

Population estimate = A + B + C + D

where D = C * (B/A)

What are we measuring?
Independence
Closed
D represents those that didn't renew a driver licence but otherwise are similar to C in that there is no record of interaction with the state

- who are they?
- are there other gaps?

Seek to minimise undercoverage in Person Activity Register (max A + B)
Using driver licence renewals as a measure

Future directions

Include admin data in annual population estimates

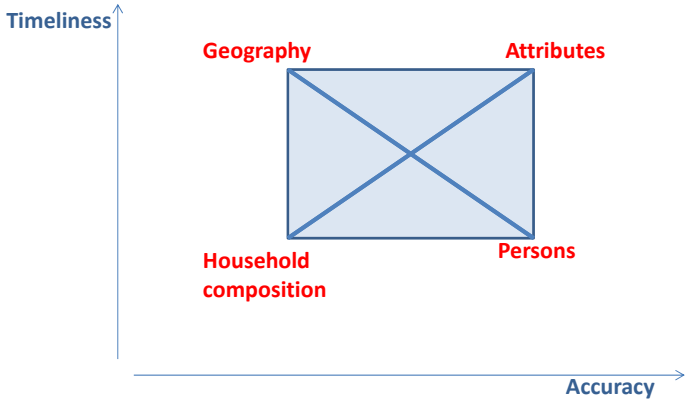
- Sustainability over time but timeliness

Postcodes (letter box) arrive in 2015

- Small area statistics
- Household composition

Are there other independent sources of value?

Fit with Census



Concluding remarks – Going forward

Opportunity for big leaps (Postcodes a game changer?)

Responsible leadership

Develop the statistical system to exploit NDI

Limited capabilities/resources

As the circle of light increases,
so does the circumference of darkness around it.