

Strategic opportunities through applying semantic technologies to modernising official statistics

Dr Siu-Ming Tam

SemStats 2013

22 October 2013

Welcome to SemStats 2013!

The goal of the SemStats workshop is to explore and strengthen the relationship between the Semantic Web and statistical communities...

...to provide better access to the data held by statistical offices.

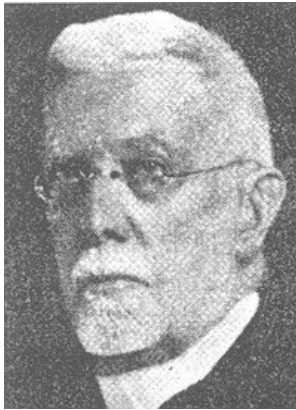
...and to assist in local and global Modernisation of Statistical Production and Services

Statistics

- *Statistics is the science of learning from data, and of measuring, controlling, and communicating uncertainty* – American Statistical Association
- Origin of the term Statistics
 - Statistics 1.0
 - Statistics 2.0

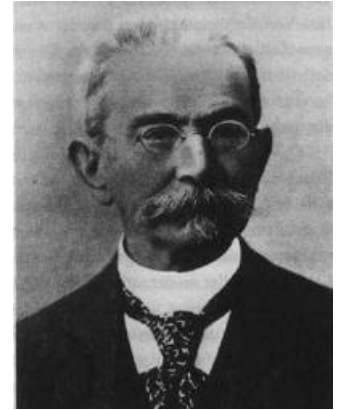
Change brings uncertainty....

- Statistics help government, business and the community understand and shape change
 - Evidence based decision making
- But what happens when statisticians face change?
 - In technology
 - In meeting increasing and users expectations
 - In scientific methods for producing statistics
 - In budgets!



Kaier

Georg von Mayr in response to Anders Kaier's
*On the Method of Representative
Enumerations Serving as a Type of the Entire
Population of a State*
(ISI 1895)



Von Mayr

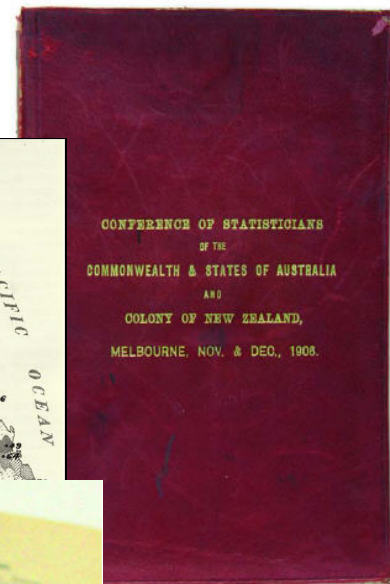
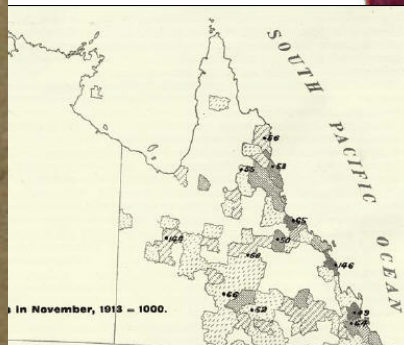
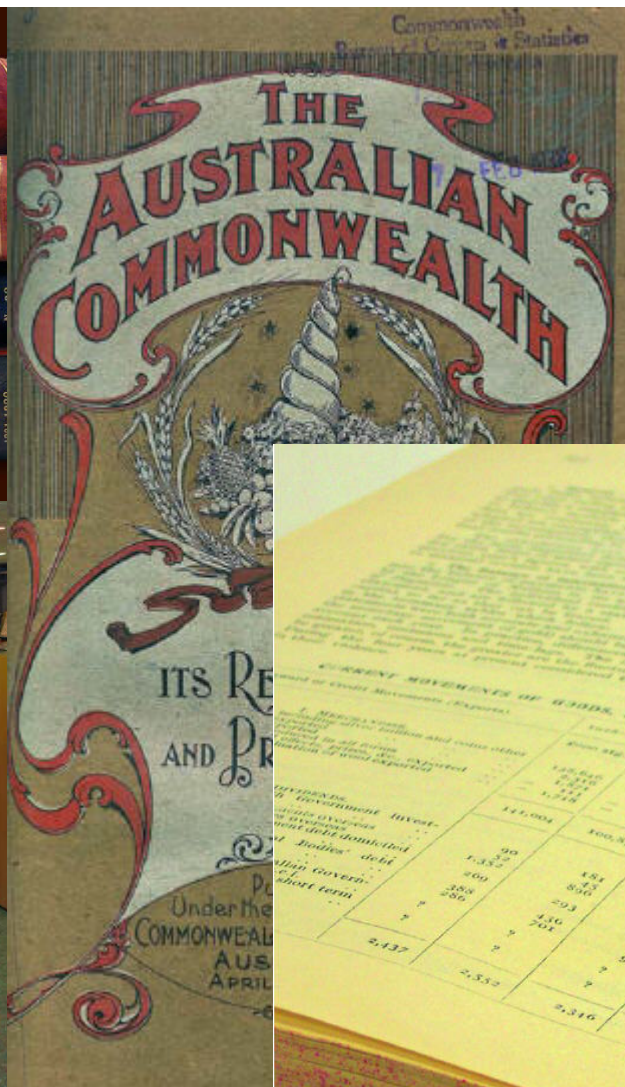
“...I regard as most dangerous the point of view found in his work. I understand that representative samples can have some value, but it is a value restricted to the domain already illuminated by full coverage. **One cannot replace by calculation the real observation of facts.** A sample provides statistics for the units actually observed, but not true statistics for the entire terrain.

It is **especially dangerous** to propose representative sampling in the midst of an assembly of statisticians. Perhaps for legislative or administrative goals sampling may have uses - but one must never forget that it cannot replace a complete survey. It is necessary to add that there is among us these days a current in the mind of mathematicians that would, in many ways, have us calculate rather than observe. **We must remain firm and say: no calculation when observation can be made.**”

100+ years later...

- Sample surveys for statistical units (persons, businesses) are now “orthodox”
- but face to face interviews are expensive (and have other limits, e.g. timeliness, provide load)
 - Can we use Web surveys?
 - Can we use administrative registers instead?
 - Can we sample big data?
 - Can we integrate data from multiple sources to form a well rounded perspective?

From paper based products



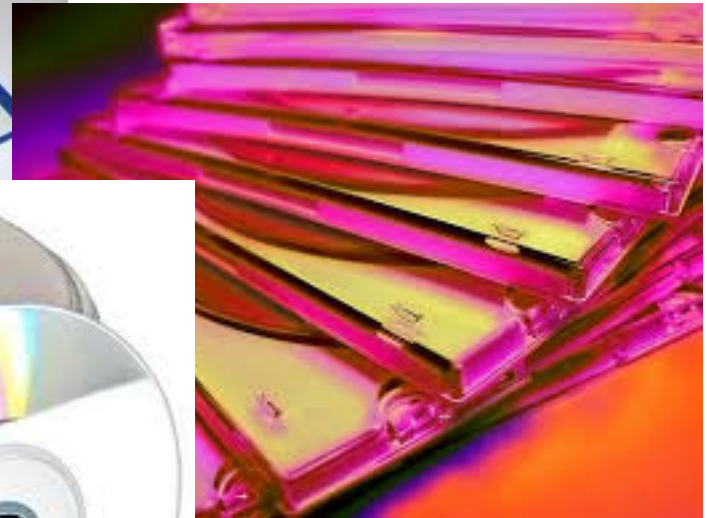
1906 Conference of Statisticians.



	1900-01	1901-02	1902-03	1903-04
Government expenditure	1,000,000	1,100,000	1,200,000	1,300,000
Investment	500,000	600,000	700,000	800,000
Revenue	1,500,000	1,600,000	1,700,000	1,800,000
Services	200,000	250,000	300,000	350,000
Gold	100,000	150,000	200,000	250,000
Total	2,500,000	2,650,000	2,800,000	2,950,000



To electronic products



To web-based products



Australian Bureau of Statistics



Australian Bureau of Statistics
Celebrating the International Year of Statistics 2013

Mobile Survey Participant Information About Us Careers Contact Us

Rate the ABS website

Home Complete Survey Statistics Services Census Topics @ a Glance Classifications News & Media Education Links Help



search

- [What's New](#) NEW Catch up on the latest
- [Australia At A Glance](#) Statistical summary
- [Key National Indicators](#) Australia's current
- [ABS Catalogue](#) Publications available from
- [About ABS Products and Services](#)
- [How to Contact the ABS](#)
- [Media Releases](#)
- [Release Advice](#) Shows details of the process
- [ABS Library Extension Program](#) Content
- [ABS Education Services](#) Information re
- [Links to other Statistical Agencies](#)
- [ABS Annual Report 1995-96](#)
- [ABS Papers, Classifications and sele](#)

All Statistics
Statistics
Understanding Statistics
Statistical Geography
Statistical Quality Management
National Statistics
Australia's Population
Key Economic Indicators
Census Data
Consumer Price Index
Labour Force
National Accounts
Measures of Australia's Progress
Regional Statistics
Selected information with a regional focus
National Regional Profile
Key Products
Australian Social Trends
Australian Year Books
Australian Economic Indicators
Papers and Articles
Selected information & statistics on a range of topics
Economy
Environment & Energy
Industry
People

All Headlines Labour Force: Unemployment rate decreased 0.1 pts to 6.2% in September 2013 (seasonally adjusted) (released 10 October 2013)

Spotlight on ...



Do you run or own a small business?

Would you like to know more about where your target market lives or how your industry is doing? Find out this and more in our [Small Business Portal](#).

Product Releases

[Previous](#) | [Upcoming](#) | [All Updates](#)
[ABS Release Calendar](#)
[Main Economic Indicators](#)

No Releases for Friday 18 October 2013 at 11.30am (Canberra time)

Releases for Thursday 17 October 2013



The value of merchandise goods imported in September 2013 fell \$0.4b (2%) to \$21.3b - [International Merchandise Imports, Australia](#) (cat. no. 5439.0)



Detailed labour force data available today including hours worked - [Labour Force, Australia, Detailed - Electronic Delivery](#) (cat. no. 6291.0.55.001)

Media Releases

[View All](#)
[Subscribe](#)

10/10/2013 [Labour Force, Australia](#)

09/10/2013 [Young women lag behind young men on numeracy skills, but perform well on literacy](#)

08/10/2013 [Internet Activity, Australia](#)

03/10/2013 [Value of irrigated production in 2011-12 at 10 year high](#)

03/10/2013 [Directory of Family and Domestic Violence Statistics](#)

02/10/2013 [Building Approvals, Australia](#)

This page last updated 18 October 2013

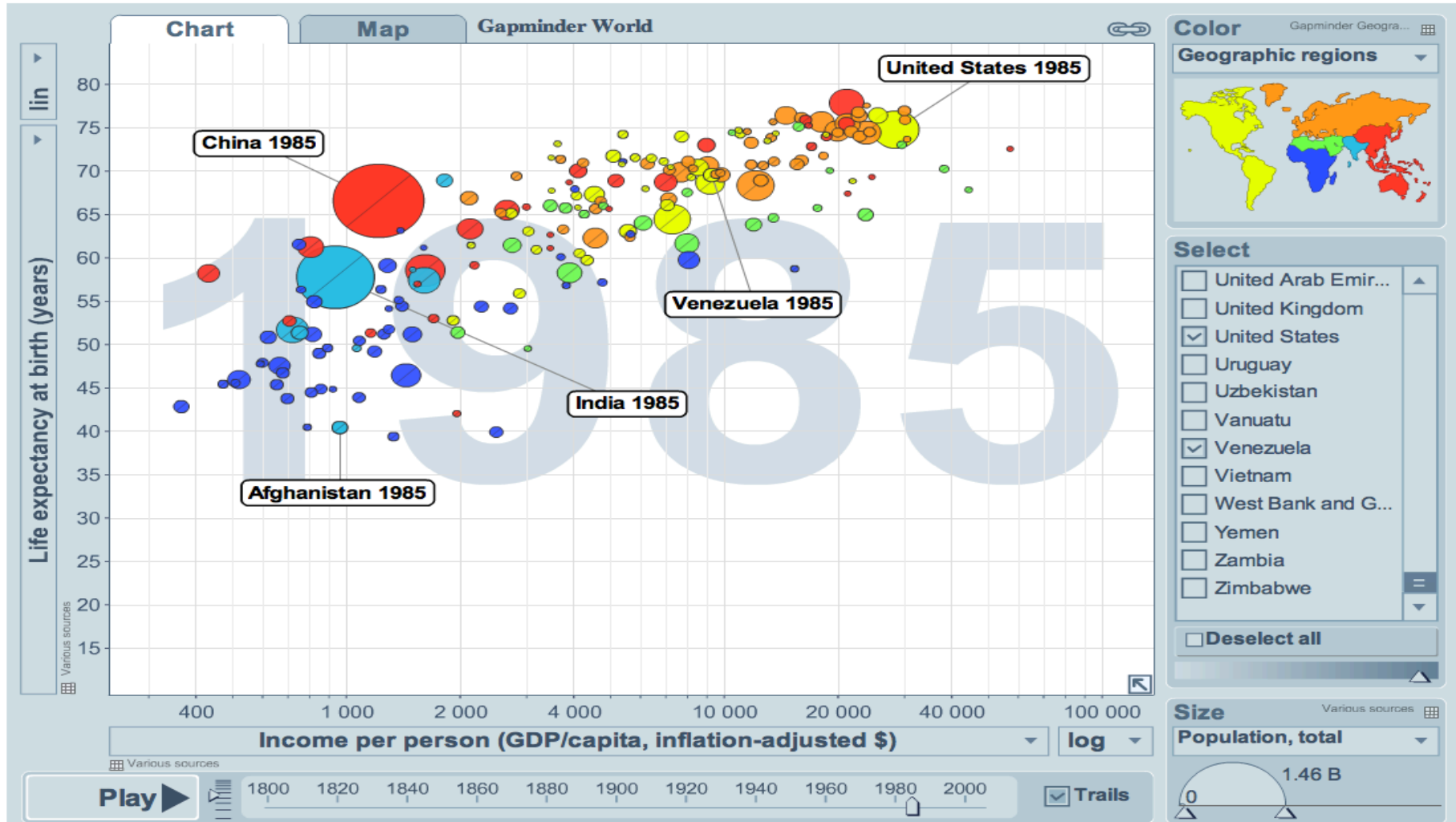


AUSTRALIAN POPULATION CLOCK
23,251,001
1 new person: 1 min 19 sec
[How does this work?](#)

Complete my survey

AUSTRALIAN Social Trends
JULY 2013

Gapminder World by Hans Rosling



Australian House Price Indexes - Static

6416.0 - House Price Indexes: Eight Capital Cities, Jun 2013 [Quality Declaration](#)

LATEST ISSUE Released at 11:30 AM (CANBERRA TIME) 06/08/2013

[Summary](#) | [Downloads](#) | [Explanatory Notes](#) | [Related Information](#) | [Past & Future Releases](#)

[Page tools:](#) [Print Page](#) | [Print All](#) | [Email Notification](#) | [RSS](#) | [BOOKMARK](#) | [SEARCH](#) | [Search this Product](#)

[Main Features](#)
[Analysis](#)
[ABS House Price Index](#)
[Methodology](#)
[Use of Price Indexes in Contracts](#)
[About this Release](#)

JUNE KEY FIGURES

Established house prices	Mar Qtr 13 to Jun Qtr 13 % change	Jun Qtr 12 to Jun Qtr 13 % change
Weighted average of eight capital cities	2.4	5.1
Sydney	2.7	6.1
Melbourne	2.4	3.3
Brisbane	1.9	3.7
Adelaide	0.3	0.6
Perth	3.4	11.0
Hobart	-1.0	1.2
Darwin	2.9	7.7
Canberra	1.0	2.6

Established house prices, Weighted average of eight capital cities - Quarterly % change

Established house prices, Quarterly % change - June quarter 2013

Australian House Price Indexes - Interactive

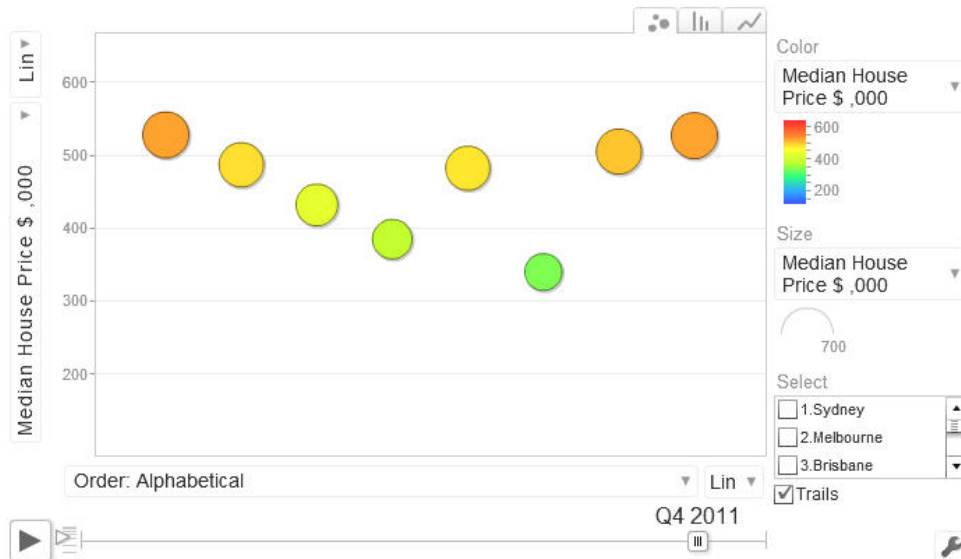
6416.0 - House Price Indexes: Eight Capital Cities

Find out how your state compares using House Price Index data in Google Motion Charts!

Features of Google Motion Charts

- Adjust the speed of the animation by dragging the small white triangle on the right next to the Play button
- Change the colour and size, and select variables (capital city) on the Chart


To view the animation, the free plug-in '[Adobe Flash Player](#)' may need to be installed on your computer.



Each Capital City is represented by a bubble and its colour and size is related to the Median House Price (unstratified) at that point in time.

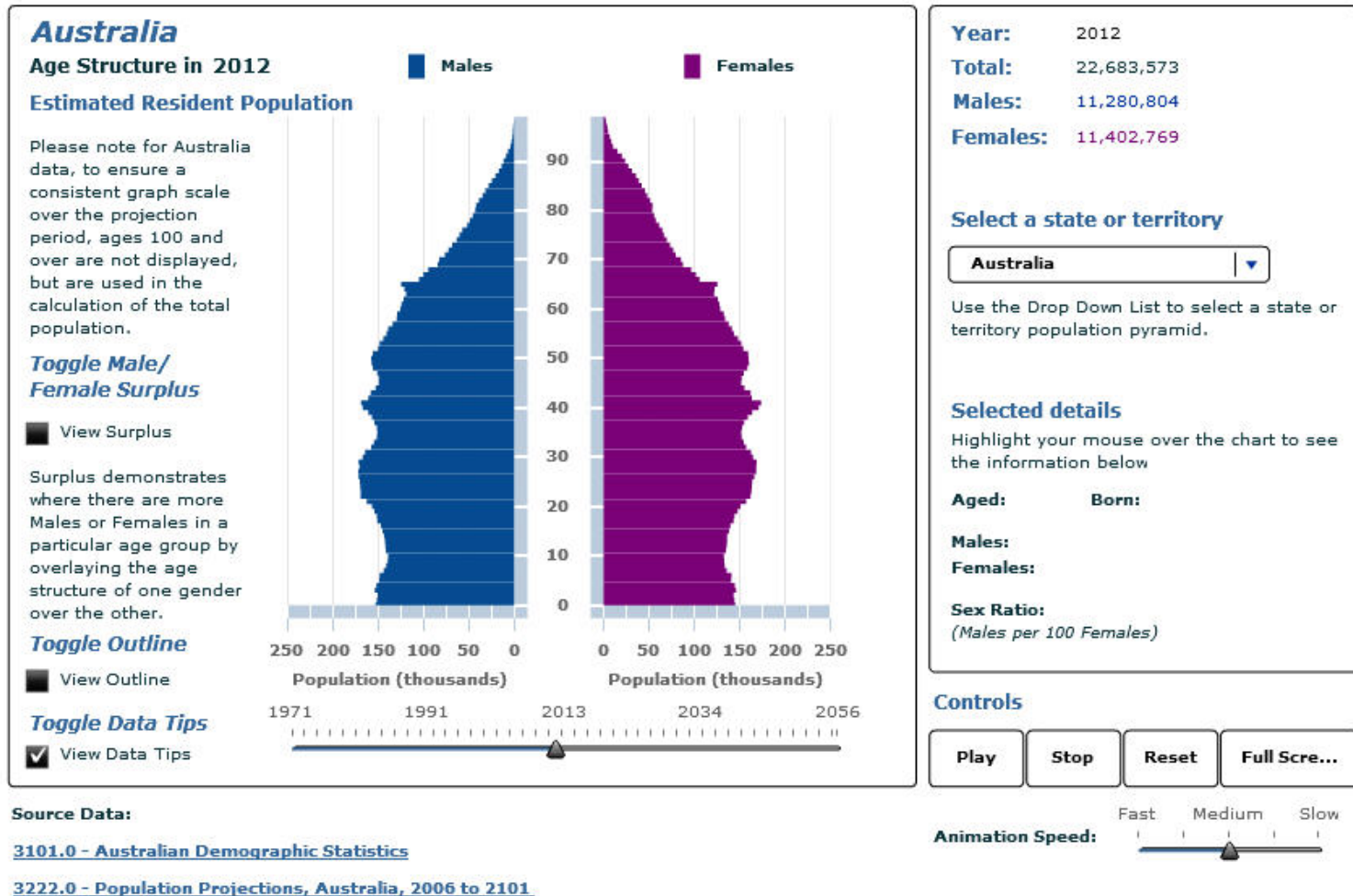
The default view is:

- X-axis: Eight Capital Cities
- Y-axis: Median House Price (unstratified)
- Time: Press Play and the chart is animated from March Quarter 2002 onwards

Click Play  to begin the animation.

Source: Cat. No. [6416.0 - House Price Indexes: Eight Capital Cities, June 2013](#)
Table 7. Median House Price (unstratified) from March Qtr 2002 to December Qtr 2012

Australian Population Pyramid



Official Statistics

Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation – UN Fundamental Principle

OGD: Oil, Gold, Democracy?

Data is the new gold

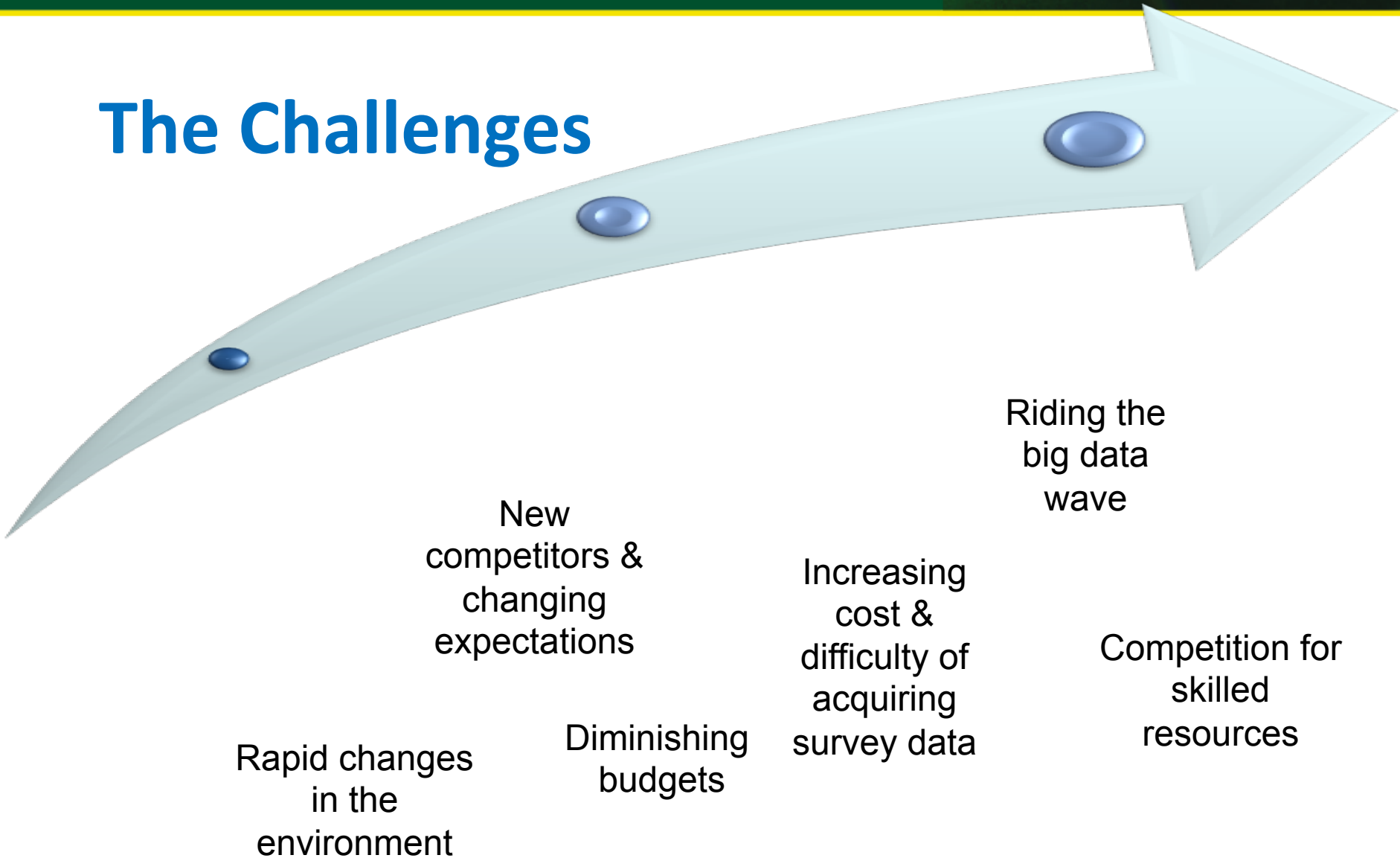
- *Just as oil was likened to black gold, data takes on a new important and value in the digital age...*
 - Neelie Kroes, VP of EC for Digital Agenda
- Public sector information already generates 32 billion € (est) of economic activity each year
- The Open Data package – for data to be able to be re-used released by the EU – will double that

What would be the added return from Linked Open Data?

NSO values include...

- Open access to data
 - Free, equal access, and open licensing system
- Support data with metadata
- Linked open data (?)
 - Tim Berners-Lee's 5 star linked open data format
 - Semantic web is not well understood by statistical leaders
 - Web 3.0 vs Web 2.0
 - SDMX vs DDI

The Challenges



Changing expectations?

- Users of statistics expect more data, more quickly
- They also want to:
 - “mash up” with data from other sources
 - explore the statistics visually, interactively, with their own analytical techniques
 - discover their own stories

Riding the Big Data wave

- Challenges and opportunities
 - Business comes first
 - Validity of statistical inferences
 - Scientific methods, **including measuring, controlling and communicating uncertainty**, need to be developed & proven
 - Legitimacy, Legacy and Leveraging
 - Replacing collections
 - Blending
 - Gaps
 - Operational efficiency

NSO responses: Modernisation program

- HLG – Europe, Asia and Pacific
 - Products/Process vision
- Products vision
 - Integrated and better access to micro data
 - More small population group/area data
- Process vision
 - Statistics Production Model/ Industrialisation
 - Metadata standards
- International collaboration

How can semantic technologies help?

Challenges for Semantic “webbers” and official statisticians

- How do Linked Data open the way to new sources of data?
- How can Semantic Web Technologies help better understand and use new sources of data, including web pages, social media, other sources of big data?
- How can the Semantic Web help us better present our statistics, tell statistical stories, and allow users to interact with data cubes, geospatial, time series and graphical visualisation?
- How can the semantic technologies help us to be more efficient and effective?

Could Semantic Web in any sense, also form “part of the challenge”?

- Is there a risk of fundamentally different ontologies emerging for various topics (eg “Occupation”) which compete with relevant statistical classifications that are based on proven conceptual design principles and rigour?
- Is there a risk that poor definition of concepts associated with Linked Data, including incorrect assertions by the provider that their concepts equate semantically with published statistical definitions, lead to misinterpretation and misapplication?

The way forward for Semantic Statistics

- **Articulate the value proposition** of the semantic web for official statistics, particularly to senior leaders of NSOs
- **Raise awareness** of semantic web amongst data managers and those involved in disseminating statistics
- **Continue the dialogue** between the two groups – more Semstats!

Questions?