

# Environmental Big Data → Evidence for Policy?

Landcare Research Link Seminar 15

Feb 26<sup>th</sup>, 2013



Landcare Research  
Manaaki Whenua

**Robert Gibb**  
Informatics Science Leader

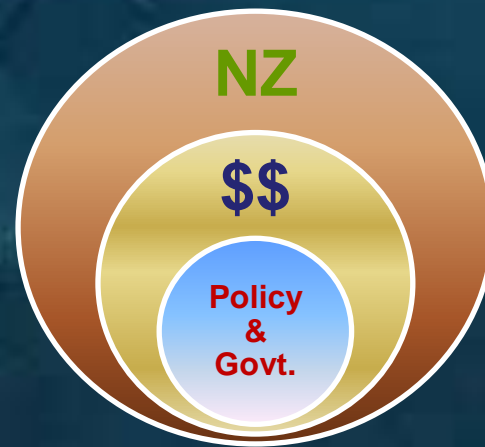
“scientists overestimate the utility  
of what they know ...

while

... policy-makers underestimate  
what they do not know”

Sir Peter Gluckman,  
IPANZ, 2013-2-21

# Global & Local Environment



# **PART 1 .. Policy & Data 2010**

**aspiration meets cold reality**

## **PART 2 .. Fast forward to 2013**

## **PART 3 .. Future perspective**







MathCad ?..

R? ..

Excel? ..

“Data publication? – you mean the scientific paper”

.. Word !

“Archive Processing log is: file notebook or email diary”

Field notebook

“ARCHIVE is a CD in a drawer”

“Don't forget to take one home”

“Master Copy is My Copy”

“Sharing is email & USB sticks at a meeting”

“More disk storage is: a bigger USB drive”

“Backup is IT Dept responsibility”

Is this your scientist's Data Management Reality?



“Email is for communication,  
& SPAM is under control”

“We have a  
Disaster Recovery Plan”

“Scientific Publications  
are held by the library”

“The annual IT budget  
constrains network  
storage SAN size”

“Our procedures meet the  
requirements of the Auditor  
and the Public Records Act”

“HSM is the solution to total  
file volume management”

# Is this your IT Dept's Data Management Reality?

“All network drives  
are backed up to a  
well defined  
schedule”

“Archive is, each  
scientists' responsibility”

“Off site and on-site  
backup tape storage”

“We can't control data if it is on  
another organisation's systems”

“International Best Practice”

“Metadata & Catalogues”

“Scientific Quality is measured in Publications & Reports”

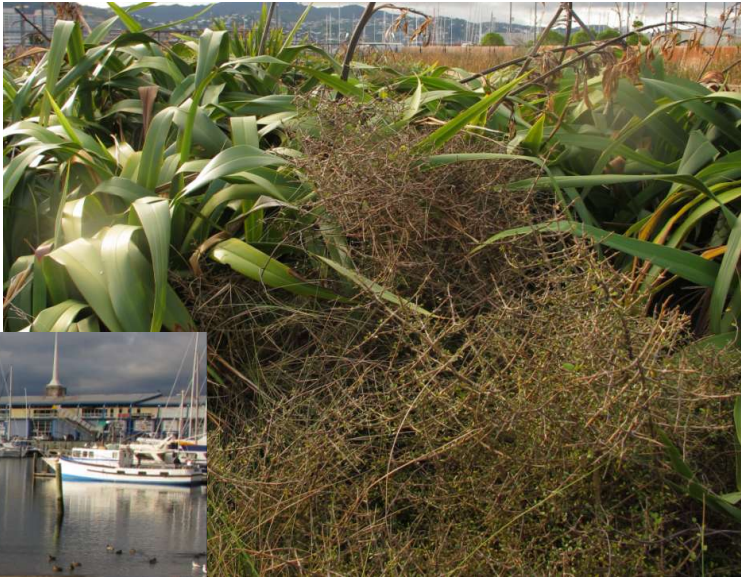
“International Data Standards”

“We expect Best-of-breed Multi-agency, Collaborative teams”

“We fund results, Data Management is presumed”

Is this Your Funder's View?







MathCad &..

grid-R &..

K-Excel &.

“Data publication is a in web-repository, with documentation, reports & processing scripts”

“Processing is K-wave scripted, se grid-enabled, shareable & publishable for access anywhere”

.. K-Word !

real-time sensors

PDA  
e-notebook

“ARCHIVE is Google-able ..

“ .. in the field and from home”

“Sharing is via secure managed web-service”

“Master is Versioned”

“More disk storage is: available everywhere”

“Backup is IT Dept responsibility”

**The scientist in the complex real world ...**



“Science Collaboration tools pervade the science process”

“The annual budget constrains network grid-storage pool”

“Disaster Recovery Plan: storage is replicated globally:”

“Data publication is in a web-repository, with documentation, reports & processing scripts”

“Our procedures define the standards for the Auditor and the Public Records Act”

“the pool is like cloud storage on steroids for NZ science,”

“All storage is replicated, & distributed”

“Data is lifecycle managed on inter-generational timescales”

“Archive is: persistently referenced repository”

“Staff in other organisations have controlled access to our in-house systems”

## The IT Dept's NZ-science Club





# Stakeholders as on-line participants in NZ Science

The IT  
Dept's  
NZ-science  
Club

The Scientist in the  
complex real world ...



**PART 1 .. Data & Policy 2010**

**PART 2 .. Fast forward to 2013**

**POLICY+ → DATA+ MAPPING+**

**PART 3 .. Future perspective**

**POLICY+**



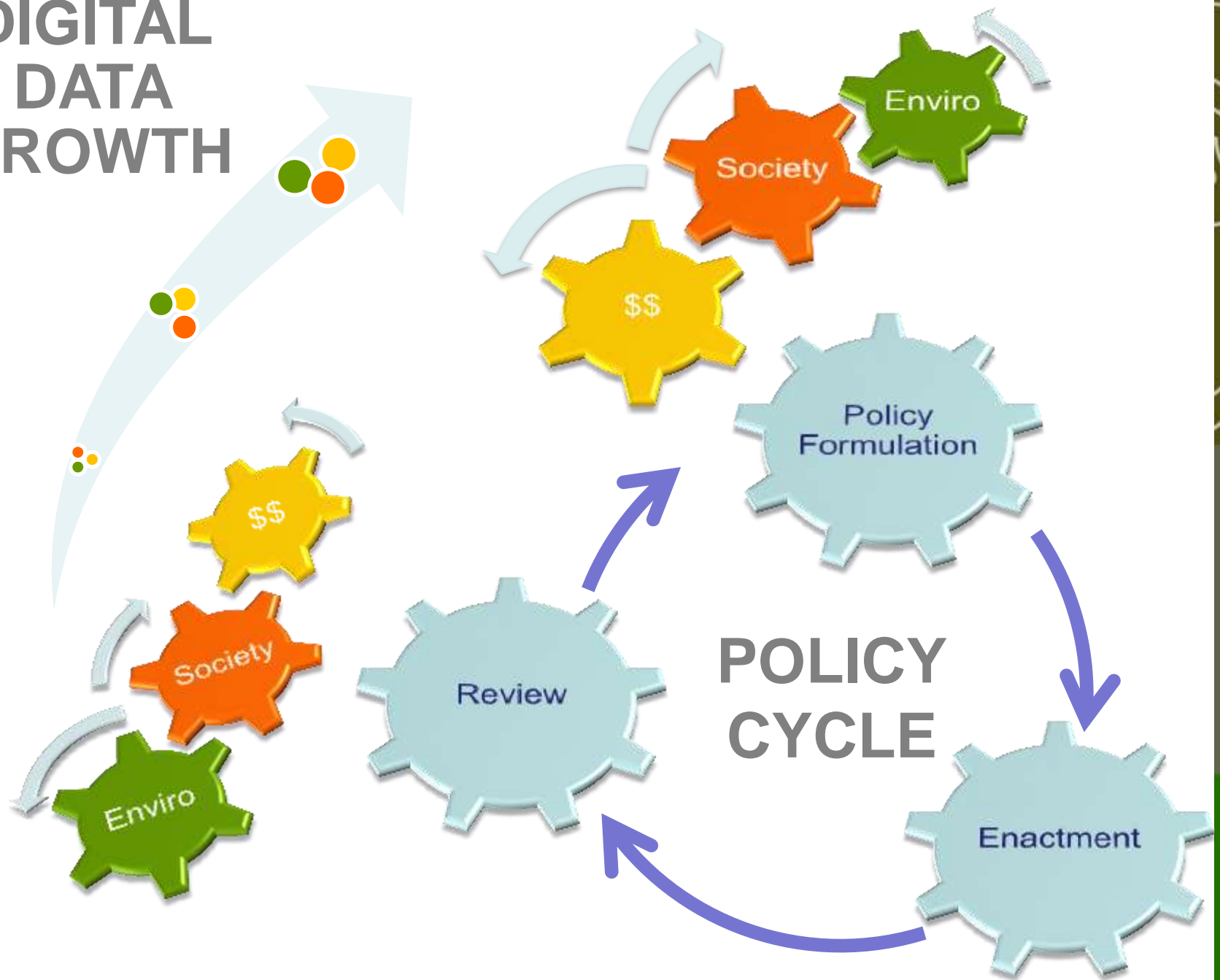


- New Zealand Government Open Access and Licensing (NZGOAL)
- MoRST Environmental Data Management Policy Statement
- Significant \$\$s moved from vote **Environment** to vote **Science**  
to do targeted research to enhance LCDB data-product &  
deliver LCDB3 & LCDB4 as by-products of Kyoto imagery

## **POLICY+** pivotal events in 2010

**... that influenced data investment behaviour**

**DIGITAL  
DATA  
GROWTH**



**DATA+**







- Launched August 2010
- Data download focus
- Minimal visualisation capability
- Aimed at GIS professionals
- Requires registration
- Uses existing open source technologies
- Strong standards-based metadata component
- Many formats supported and different projections
- Web services
  - Catalogue Service - linked to data.govt.nz, GEOSS
  - WFS – Simple features
- Potential for user uploads and user comments

<http://iris.scinfo.org.nz>

not just Landcare Research

... but also **linz** data service,

**niwa** environmental information catalogue

and **data.govt.nz** – which harvests these catalogues

# DATA+



**MAPPING+**

# Principles

- K.I.S.S.
- Focus more on presenting knowledge and information rather than data(sets)
  - Inform users, assist in their understanding
  - Not “just” maps
- Goal-based design – users with questions they want answered
- Re-usable – both software and data/map layers
- Extensible
- Implemented using open source software
- Built on best practice and open standards





## First of the new portals

- Home page with good explanatory information
- Google maps-like navigation
- Reusable BASE layers from LINZ topographic data
- High quality on-screen cartography
- Speed of map redraws/zooms/pans
- Search by coordinates/location/address
- Use of metadata, legends and explanatory information to inform user
- Reuses existing link to soil fact sheets generator
- High quality hard copy cartography (PDF)



OurEnvironment  
Landcare Research  
Mataaki Whenua

Home | Getting Started | About | Map | Glossary | Terms of Use | Data Provenance

## New Zealand's Environments

OurEnvironment  
Landcare Research  
Mataaki Whenua

Home | Getting Started | About | Map | Glossary | Terms of Use | Data Provenance

### Create Your Map

Tick the topics of interest to you & then press this button



thematic focus, it has been general user although even the coming months additional added to Our Environment can be answered.




What is the surface rock like?



What is the land capable of being used for?



What vegetation would I find?



Tools: Location Search, Layers, Layer Metadata, Legend, Report, Suitability for Production Forestry

Search by place (e.g. Te Anau), address or by entering a [UTM](#) or a [Whole lot code](#) coordinates (e.g. 7590671, 5819114)

Layers:

- Permanent Snow & Ice
- Direction of Slope
- Steepness of Slope
- Erosion Severity (Observed)
- Surface Rock Type
- Land Suitability
  - Suitability for Arable Cropping
  - General Suitability
  - Multiple Use Land
  - Pastoral or Forestry Land
  - Conservation Land
  - Suitability for Production Forestry
    - Very High
    - High
    - Moderate
    - Low
    - Very Low
    - Unsuitable
  - Most Intensive, sustainable land use
  - Land Use Capability
- Ecosystems & habitats
- Drainage
- Simple Coastal Outline
- Monochrome Terrain Map
- Colour Terrain Map

Layer Metadata: Suitability for Production Forestry

Class	Description	Area
2	High	1,500 ha
3	Moderate	70,600 ha
		83,500 ha
		81,500 ha
		91,600 ha
		+ 100 ha

Suitability for Production Forestry

High  
Moderate  
Low  
Very Low  
Unsuitable

## Second of the new portals

- S-map like features +
- Targeted at interested lay person but of value to scientist also
- Information pop-ups instead of factsheets
- Where available, links to LRIS data layers
- Reuses BASE layers and context layers
- More extensive metadata, legends and explanatory information to inform user
- Starts to introduce data visualisation & reporting
- Roadmap for further development







# Earth Beneath Our Feet

- Project
- Staff
- Graphics
- Get Geological Profile
- Get Geological Cross-Section

## Geological Cross Section

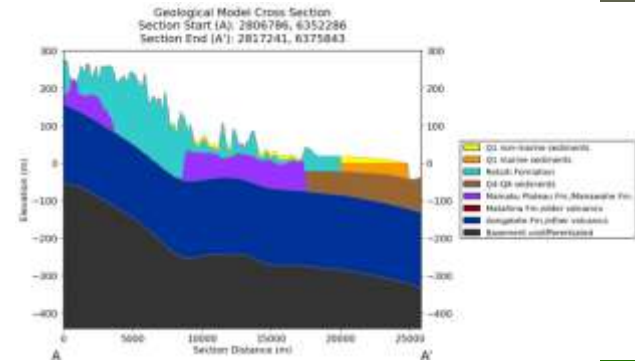
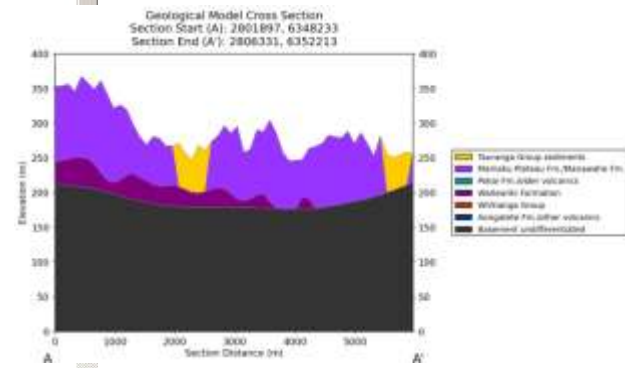
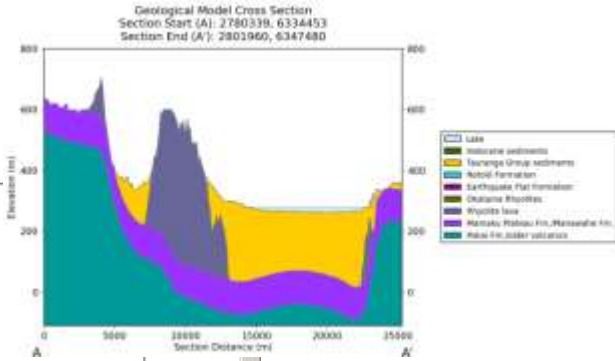
Section name:

Start of section  
Enter your location, either street address or coordinates

Please enter your street address below:

or if you know your location enter it here

NZMG	NZTM	WGS84
<b>NZMG</b>		<input type="button" value="Reset"/>
Easting	<input type="text"/>	
Northing	<input type="text"/>	



each portal is tailored for a specific audience  
... are **your needs** catered for?  
come and talk to us

# MAPPING+



**PART 1 .. Data & Policy 2010**

**PART 2 .. Fast forward to 2013**

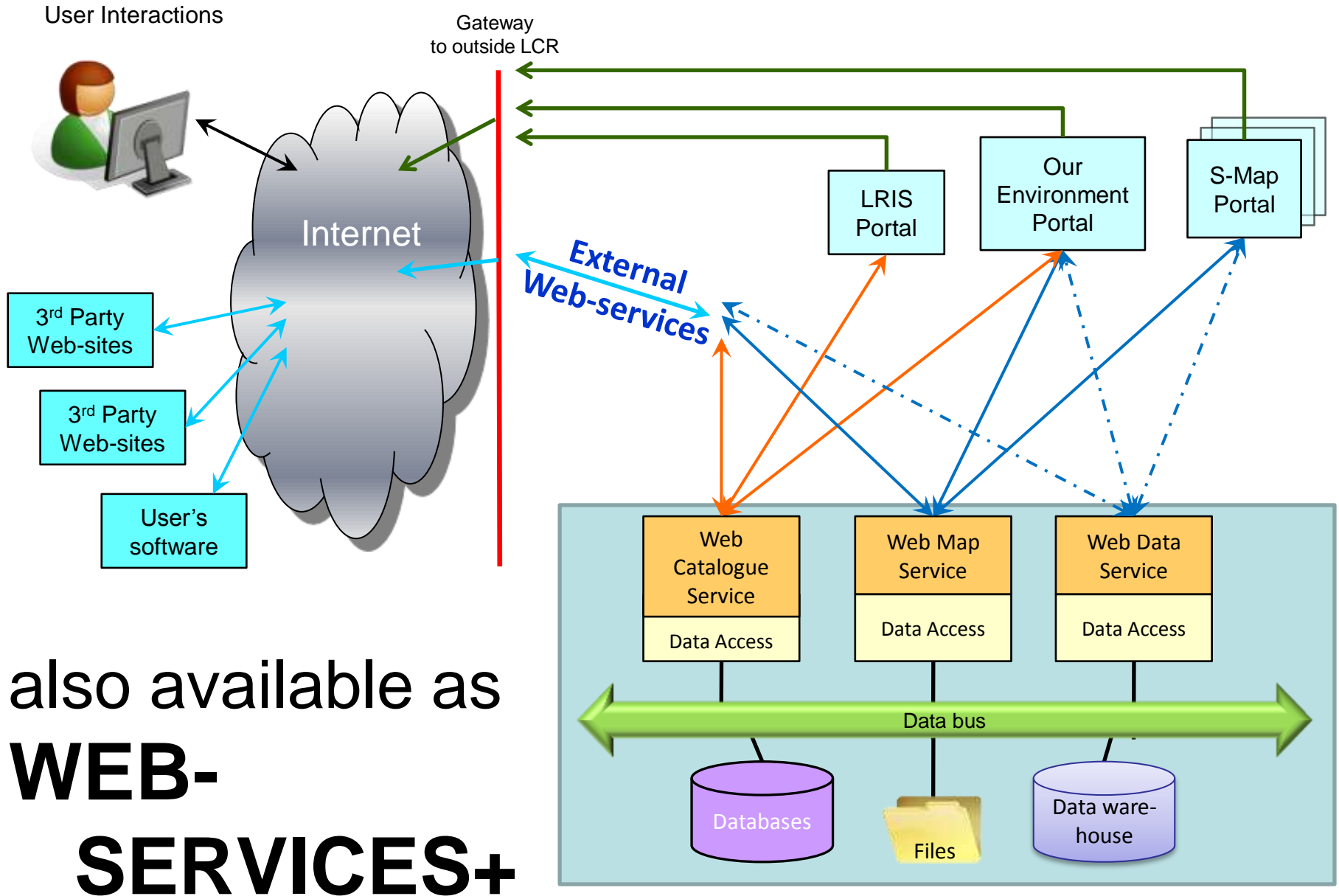
**PART 3 .. Future perspective**

**WEB-SERVICES+**

**+ BIG-DATA → POLICY++**

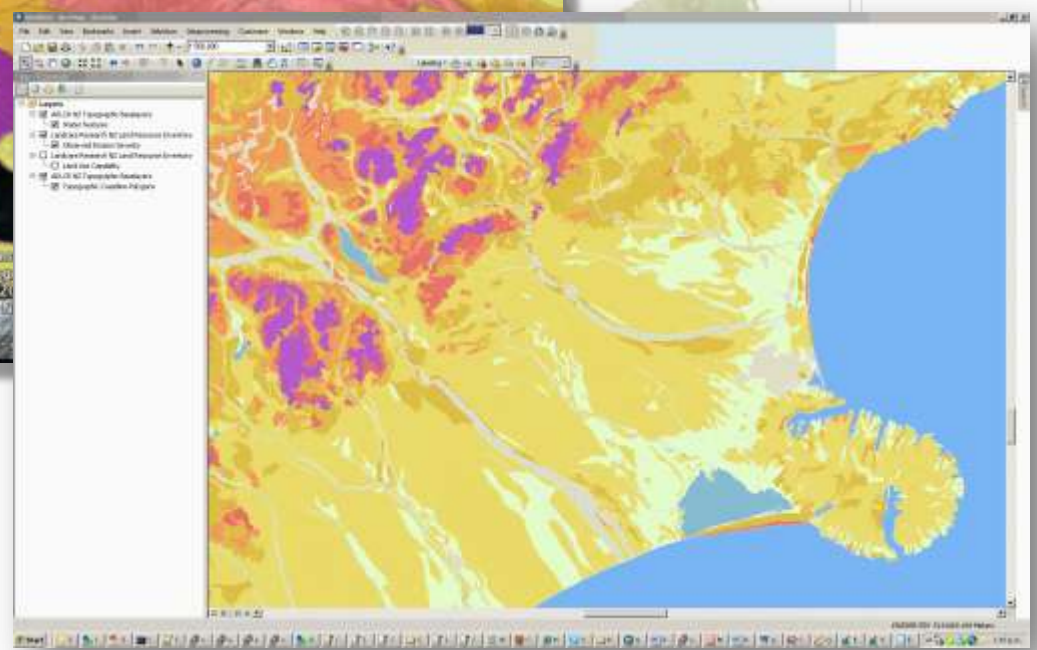
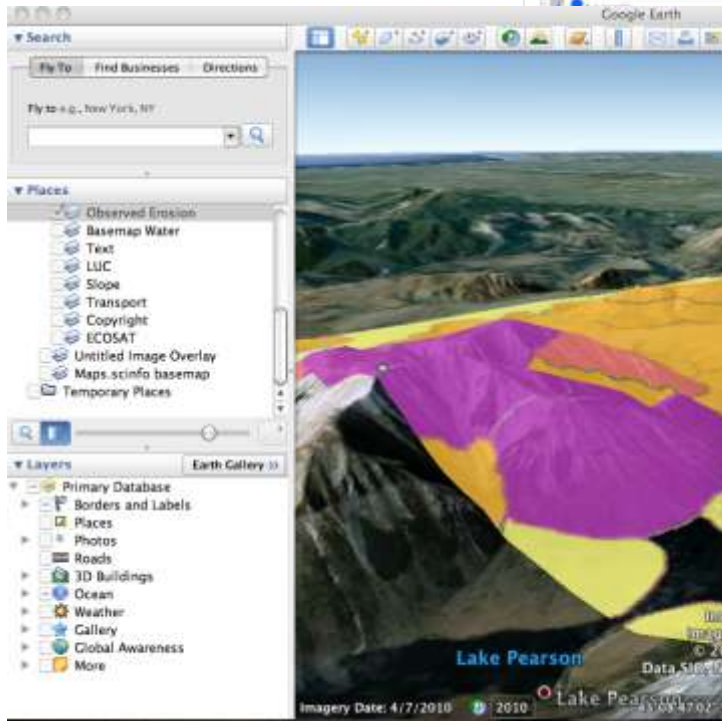
**WEB-SERVICES+**

# DATA+ and MAPPING+ content ...



also available as  
**WEB-SERVICES+**





not just **Landcare Research**

... but also **gns** web-services & **niwa** web-services  
and **horizons.govt.nz** with other RCs

# WEB-SERVICES+

**+ BIG DATA**



# BIG-DATA ... what is it?

Not large homogeneous data sets

Rather a fire-hose of diverse data from  
**real-time** internet usage &  
**web-service** data feeds

With potential for real-time feedback to  
influence individual decision-making  
eg via smart-phone

# BIG-DATA ... for example

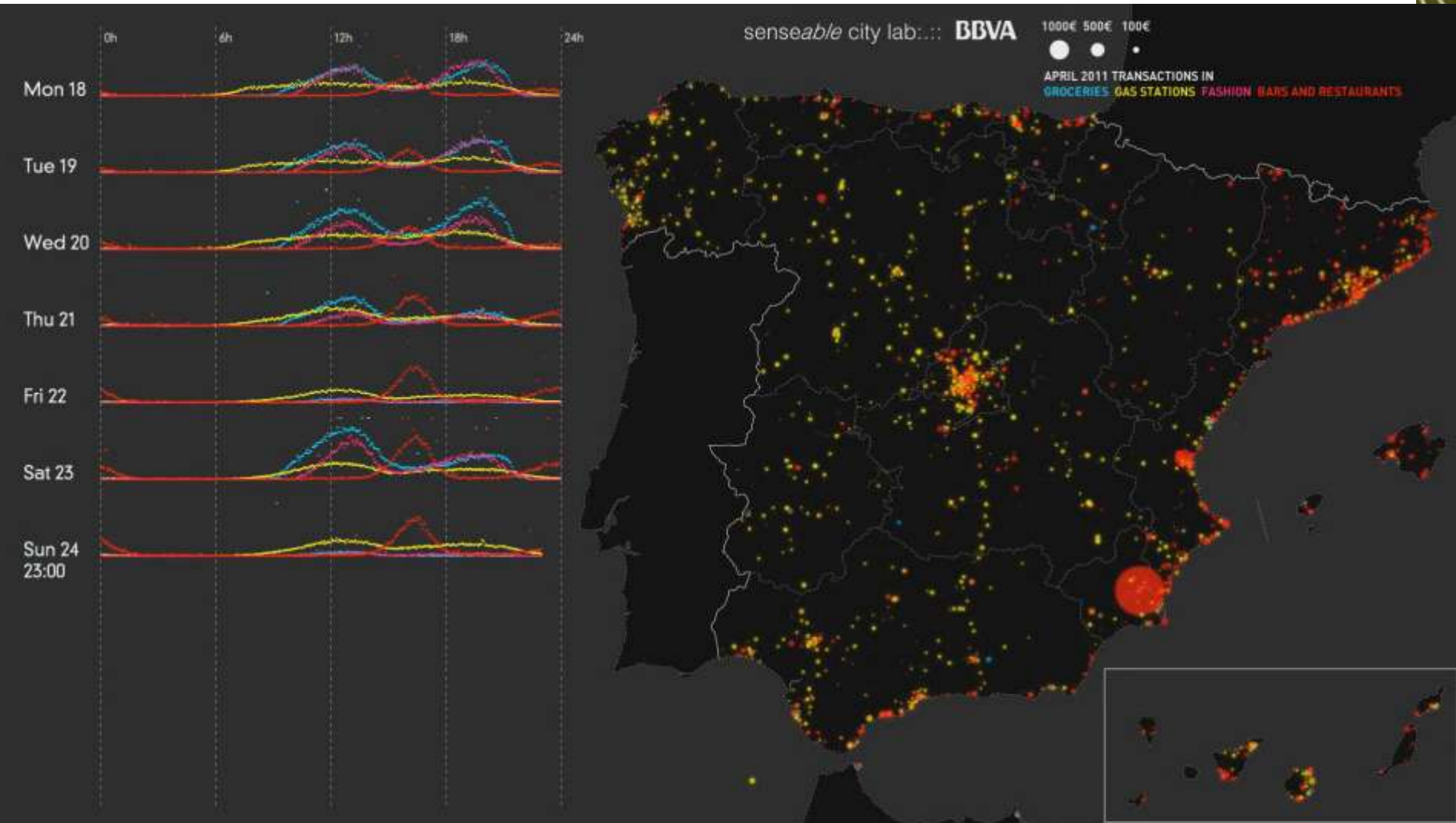
How should I travel to a cross-city appointment?

walk, jog, bicycle, public transport, car or **combination**

web-service analyses routes based on: current gps position; appointment address & time; real-time feeds of traffic flows, public transport bus and train locations, routes & timetables; green-bike usage patterns & availability; congestion statistics during the trip; and ... relative risks of being late

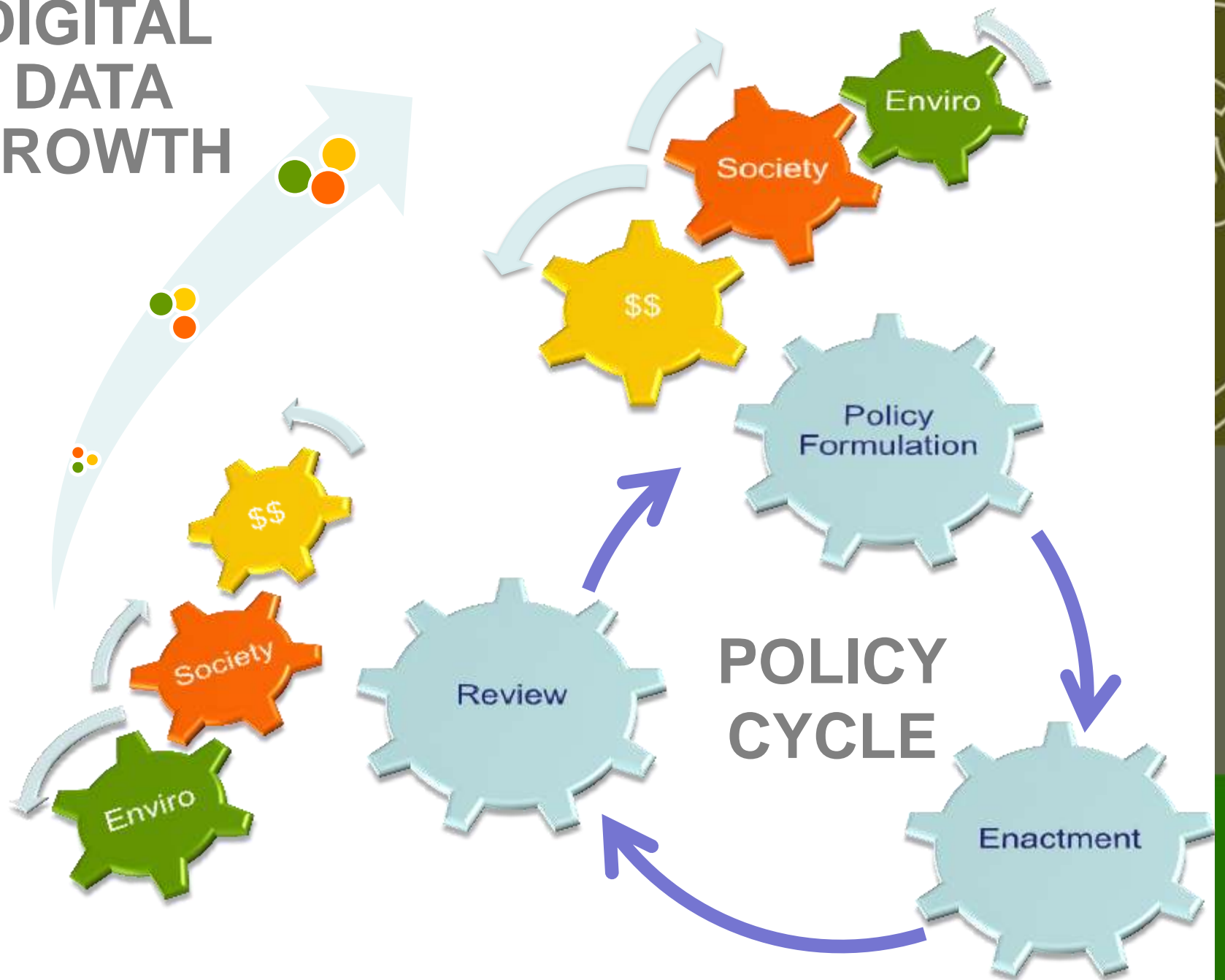
app consumes web-service & advises **fastest**, **cheapest**, **personal fitness target** &/or **greenest** options which update continuously as travel progresses & situation changes

# Social Behavioural Big-Data



Spanish credit-card spending Easter 2011

**DIGITAL  
DATA  
GROWTH**





# Social Behaviour Big-Data with Environmental Impact ... 201?

**What if** data is available real-time & instead of matching \$ transactions with  
**Groceries**, **Fuel**, **Fashion** and **Bars & Restaurants**

... **we match** \$ transactions against map of embodied energy footprint, or map of employment impact, or health outcome alternatives, or balance of trade impact ... or ?



**How would** ... this impact policy formulation ? or policy review ?

# Environmental Big-Data ... 20??

**There are many private sensors in the environment**

eg soil moisture for irrigation control or  
heavy farm equipment monitoring to schedule maintenance  
but **is the data available?**

- ... should it be as **proof of 'license to operate'**
- ... what policy would encourage or mandate availability
- ... how would this influence world perception of

**'NZ 100% Pure' ?**