

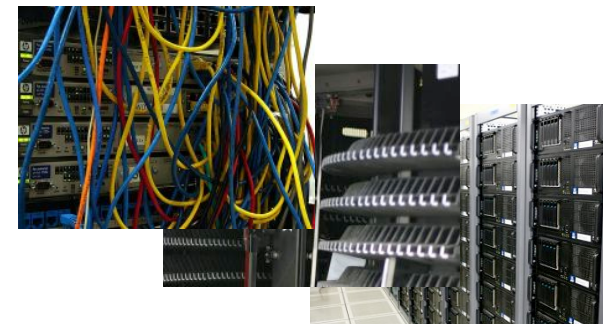


# NZ eResearch 2020: it's all about the data

Dr Andrew Treloar, Director of  
Technology

!

?



CC-BY @atreloar



## What Business are we in?

Theodore Levitt, *The Changing Character of Capitalism*,  
*Harvard Business Review*, July–August 1956

“The railroads did not stop growing because the need for passenger and freight transportation declined. That grew. The railroads are in trouble today not because that need was filled by others (cars, trucks, airplanes, and even telephones) but because it was not filled by the railroads themselves. They let others take customers away from them because **they assumed themselves to be in the railroad business rather than in the transportation business.**”



CC-BY @arelouar

Photo CC-BY [www.flickr.com/photos/jerryjohn/63351338/](http://www.flickr.com/photos/jerryjohn/63351338/)



CC-BY @atreloar

Photo CC-BY [www.flickr.com/photos/stiefkind/6454784607/](http://www.flickr.com/photos/stiefkind/6454784607/)



CC-BY @atreloar  
Photo CC-BY [www.flickr.com/photos/torkildr/3462607995/](http://www.flickr.com/photos/torkildr/3462607995/)

Tools Options

- search tools
- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- Human Genome Variation
- Genome Diversity
- EMBOSS

Edit Attributes

Name:

Join two Queries on data 3 and data 1

Info:

Database/Build:

Click to Search or Select

Number of comment lines:

save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

Change data type

New Type:

tabular

This will change the datatype of the existing dataset but not modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.

save

History Options

- 1.8 Ch
- 14: Draw phyloeny on data 12
- 13: Draw phyloeny on data 11
- 12: find lowest diagnostic rank on data 10
- 11: find lowest diagnostic rank on data 9
- 10: fetch taxonomic representation on data 8
- 9: fetch taxonomic representation on data 7
- 8: a234 within 5% of max
- 7: a1 within 5% of max
- 6: join two Queries on data 4 and data 2
- 5: join two Queries on data 3 and data 1
- 4: a234 max bit score



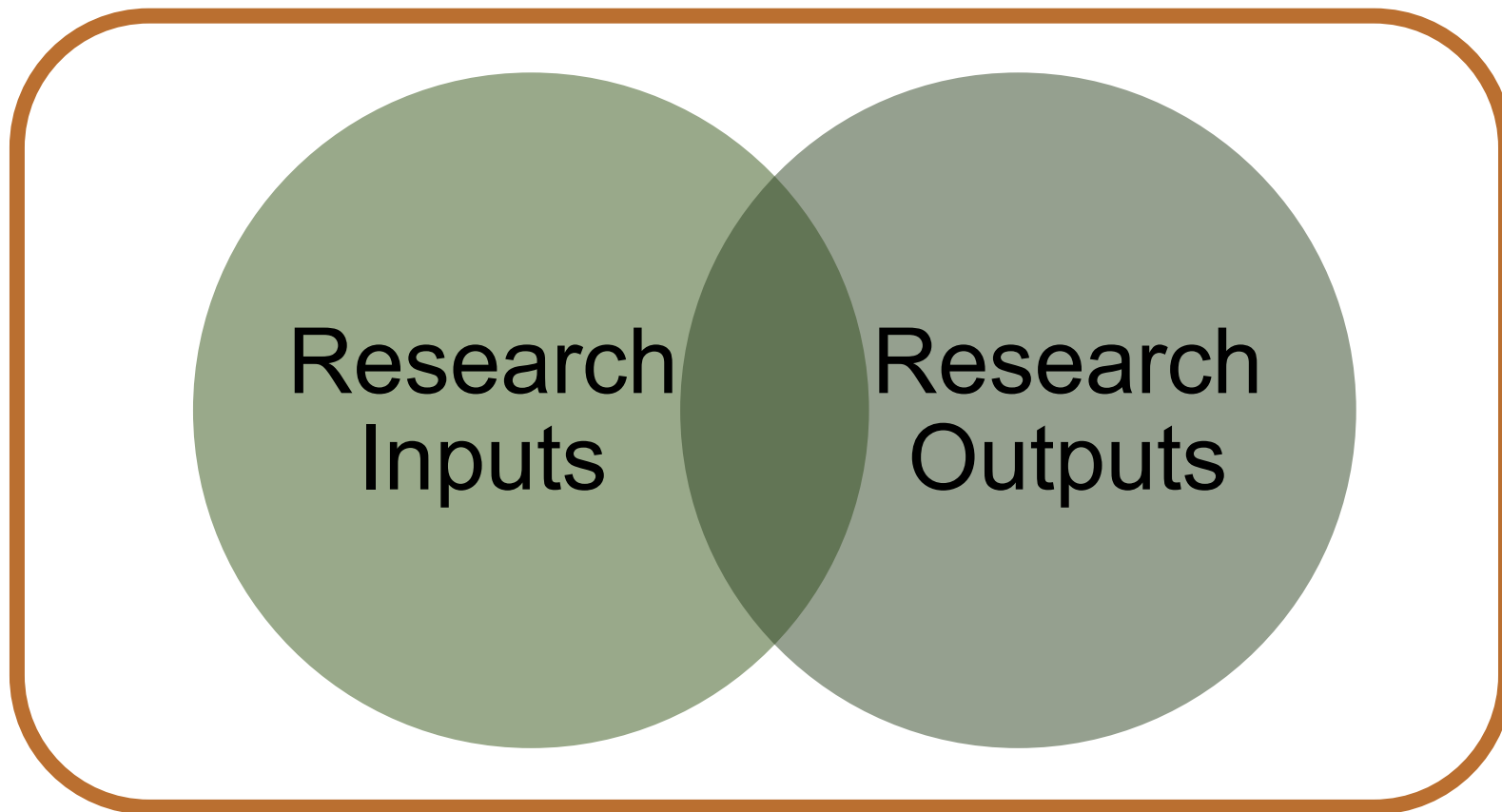


Photo: Dave Allen, NIWA

# Data **is** infrastructure

- Analyses
  - Human genome
  - Subduction boundary 3D visualisations
- Observations
  - Climate change records
  - Southern Sky astronomy surveys
- Cultural organisations holdings
  - Language records
  - Polynesian current maps

# What is research data?



## ANDS enables transformation of:

Data that are:

- ❑ Unmanaged
- ❑ Disconnected
- ❑ Invisible
- ❑ Single use

To Structured Collections that are:

- ❑ Managed
- ❑ Connected
- ❑ Findable
- ❑ Reusable

so that Australian researchers can easily publish, discover, access and use/re-use research data.

## ANDS' distinguishing characteristics

- Engages with institutions not researchers
- Provides range of services
- Contributes to policy development
- Nationally resourced organisation
- Active internationally

# The Research Data Alliance (RDA)

- Global community-driven organization launched in March 2013 to accelerate data-driven innovation
- RDA focus is on building the **social, organizational and technical infrastructure** to
  - *reduce barriers to data sharing and exchange*
  - *accelerate the development of coordinated global data infrastructure*



# RDA Approach: CREATE → ADOPT → USE

15



## RDA Members come together as

- **Working Groups** – 12-18 month efforts to build, adopt, and use specific pieces of infrastructure
- **Interest Groups** – longer-lived discussion forums that spawn Working Groups as specific pieces of needed infrastructure are identified.

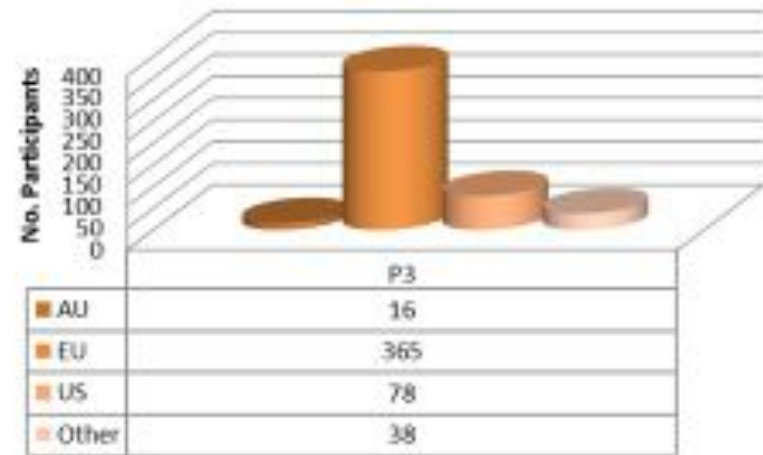
## Working Group efforts focus on the development and use of data sharing infrastructure

- **Code, policy, infrastructure, standards, or best practices that are adopted and used** by communities to enable data sharing
- **“Harvestable” efforts** for which 12-18 months of work can eliminate a roadblock
- **Efforts that have substantive applicability** to groups within the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today**

### *Final participant breakdown by region*

REGION	P3	%
AU	16	3%
EU	365	73%
US	78	16%
Other	38	8%
<b>Total</b>	<b>497</b>	<b>100%</b>

### Participants per region



497 participants from 35 countries



## Questions?

- [andrew.treloar@ands.org.au](mailto:andrew.treloar@ands.org.au)
- <https://speakerdeck.com/atreloar/ereseearch-2020-its-all-about-the-data>
- @atreloar