BioCommons - challenges and blockers

PAWSEY supercomputing centre

NCI

BIOPLATFORMS AUSTRALIA

aarnet Australia's Academic and Research Network

ARDC Australian Research Data C

Challenge: a large, complex research community

30,000 health/biosciences researchers

18,000 health/biosciences RHD students

48,000 health/biosciences PG course work students

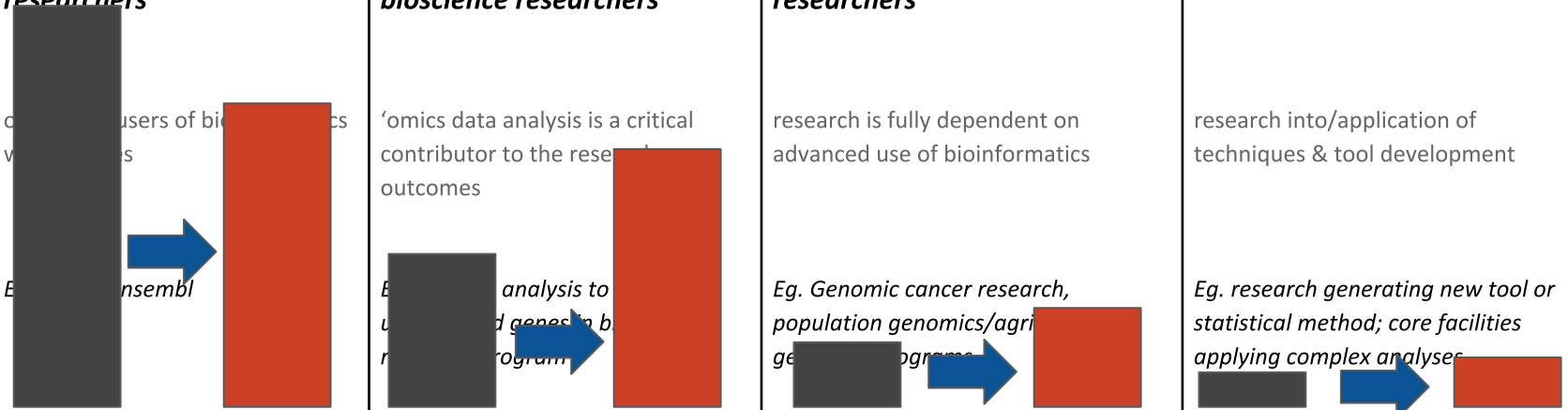
(163,000 + 40,000 =) 200,000 health/biosciences UG students

1,000 to 1,500 bioinformatician/computational biologists

Challenge: New Methods and a Growing User base

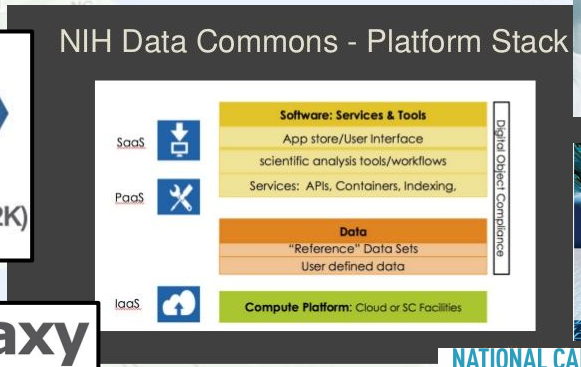
Estimated # Australian biology researchers in 2018: **30,000**

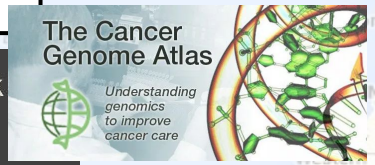
20,000 *(→ 15,000)* ***biology-focussed bioscience researchers***	**7,000** *(→ 12,000)* ***data-intensive bioscience researchers***	**2,000** *(→ 3,000)* ***bioinf-intensive bioscience researchers***	Estimated #: **1,000** *(In 5 years → 1,500)* ***bioinformaticians***
users of bi cs s	'omics data analysis is a critical contributor to the rese outcomes	research is fully dependent on advanced use of bioinformatics	research into/application of techniques & tool development
E nsembl	*E analysis to u d genes in b r rogram*	*Eg. Genomic cancer research, population genomics/agri ge ogram*	*Eg. research generating new tool or statistical method; core facilities applying complex analyses*

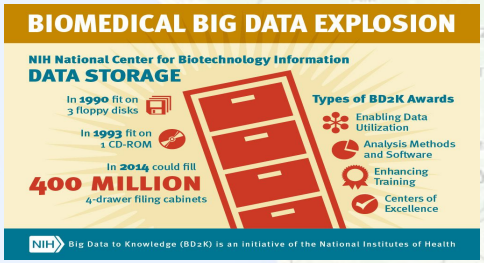


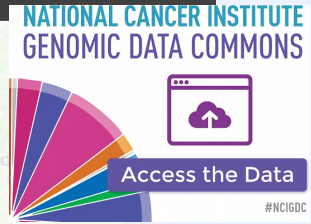
Important Transitions

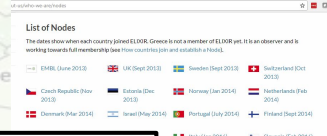
Challenge: a global wave of investment

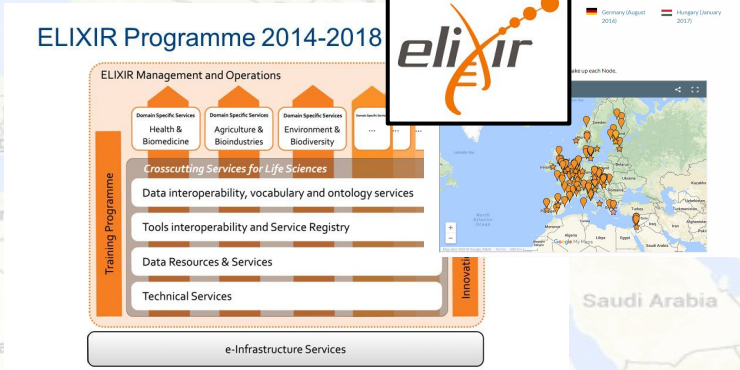


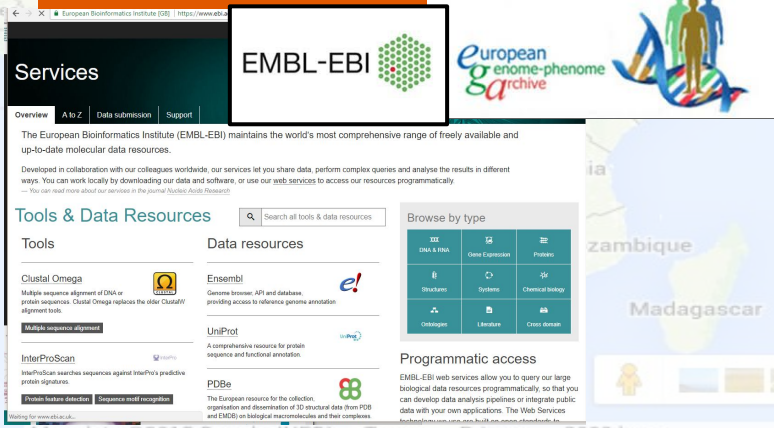












US National Institutes of Health: Data Commons

Data Commons Platform

Use

- User Portals & Workspaces
- Scientific Analysis Tools & Workflows
- Services: Identifiers, Search, Authentication

Data

- NIH-funded Data Sets

Cloud

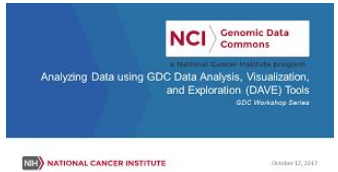
- Leverage Cloud Service Providers
- Store large biomedical data sets

FAIR

The Data Commons Platform describes the components needed to store and make NIH data accessible in the cloud

Pilot Phase

Galaxy



TOPMed: WGS data from 120,000 individuals funded by National Heart, Lung, and Blood Institute.

Genotype Tissue expression project: 714 donors and 11688 RNA-seq samples across 53 tissue sites and 2 cell lines

GTEx

ALLIANCE of GENOME RESOURCES

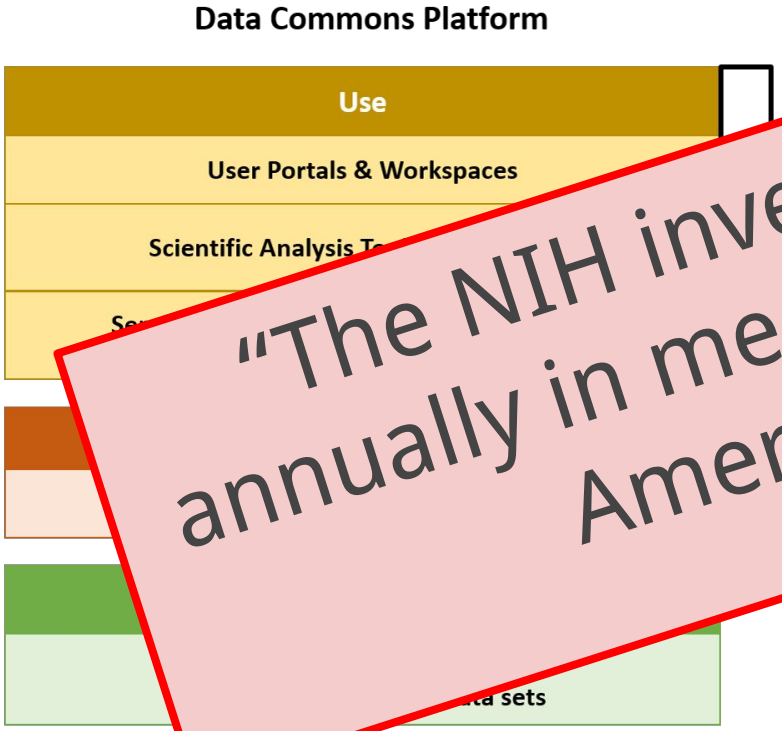
Model Organism Databases: WormBase, FlyBase, Zebrafish Information Network, Saccharomyces , Mouse and Rat Genome Databases.

amazon web services™

Azure

Google Cloud

US National Institutes of Health: Data Co



The Data Comm latform describes the components needed to store and make NIH data accessible in the cloud

"The NIH invests US$39 billion annually in medical research for the American people."

project: 714
-seq samples across
53 tissue sites and 2 cell lines

GTEx

ALLIANCE of GENOME RESOURCES

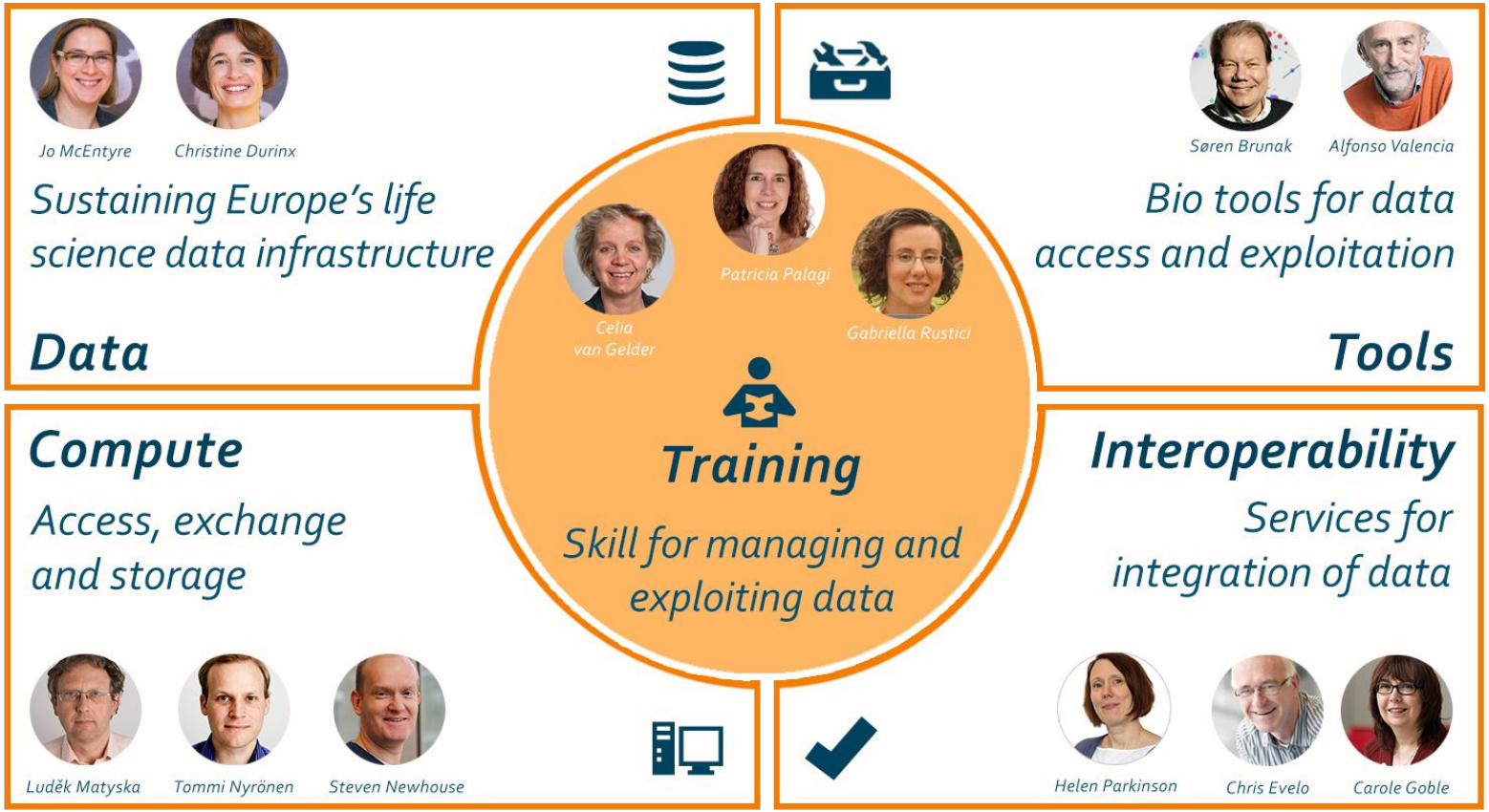
Model Organism Databases:
WormBase, FlyBase, Zebrafish Information Network, Saccharomyces , Mouse and Rat Genome Databases.

amazon web services™

Azure

Google Cloud

ELIXIR : 5 platforms of shared services led by leading European scientists



ELIXIR : 5 platforms of shared services led by leading European scientists


Jo McEntyre


Christine Durinx

Sustaining Euro
science data inf

Data

ren Brunak


Alfonso Valencia

tools for data
d exploitation

Tools

Compute

Access, exchan
and storage


Luděk Matyska


Tommi Nyrönen

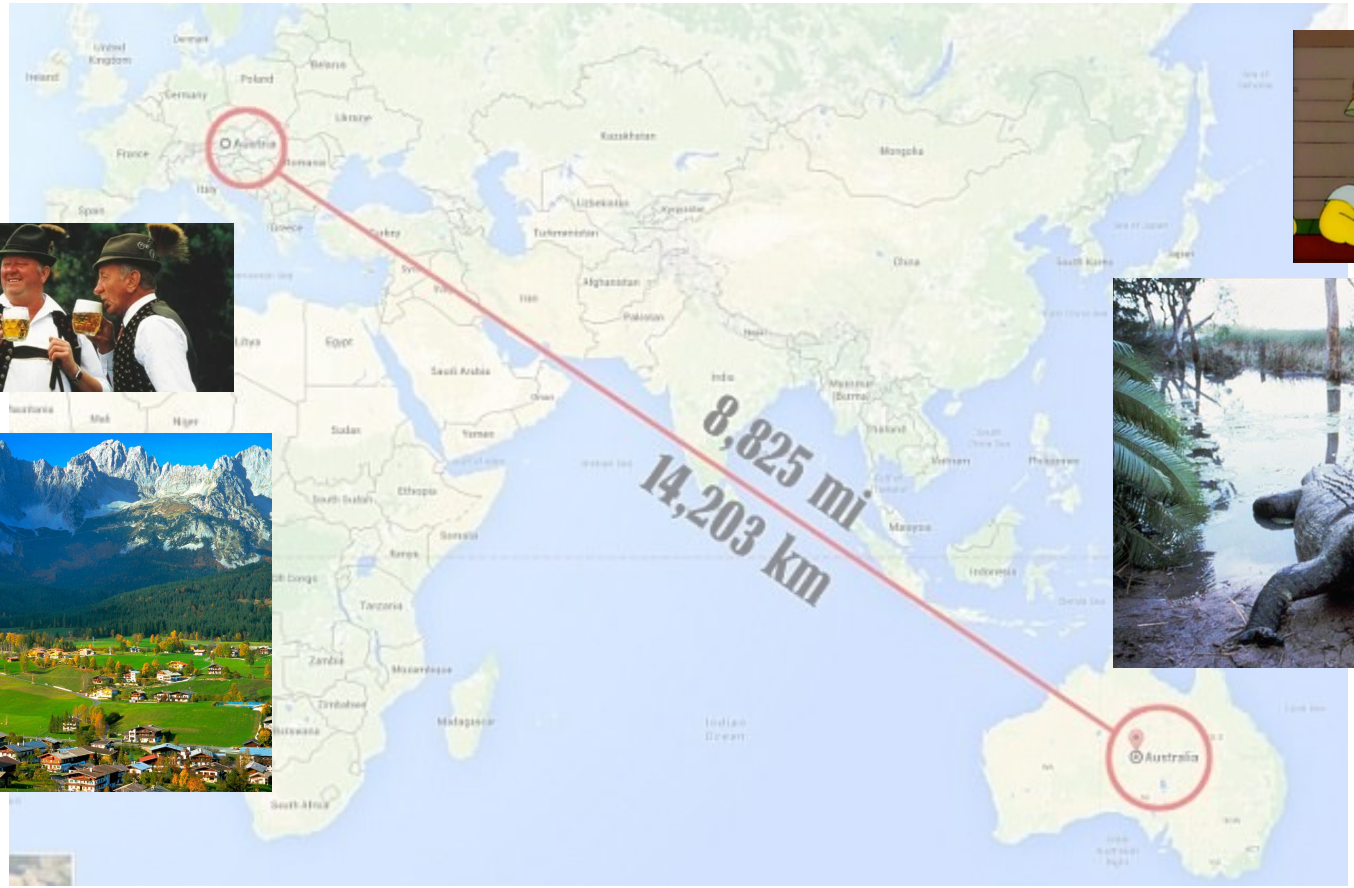
operability

Services for
ration of data

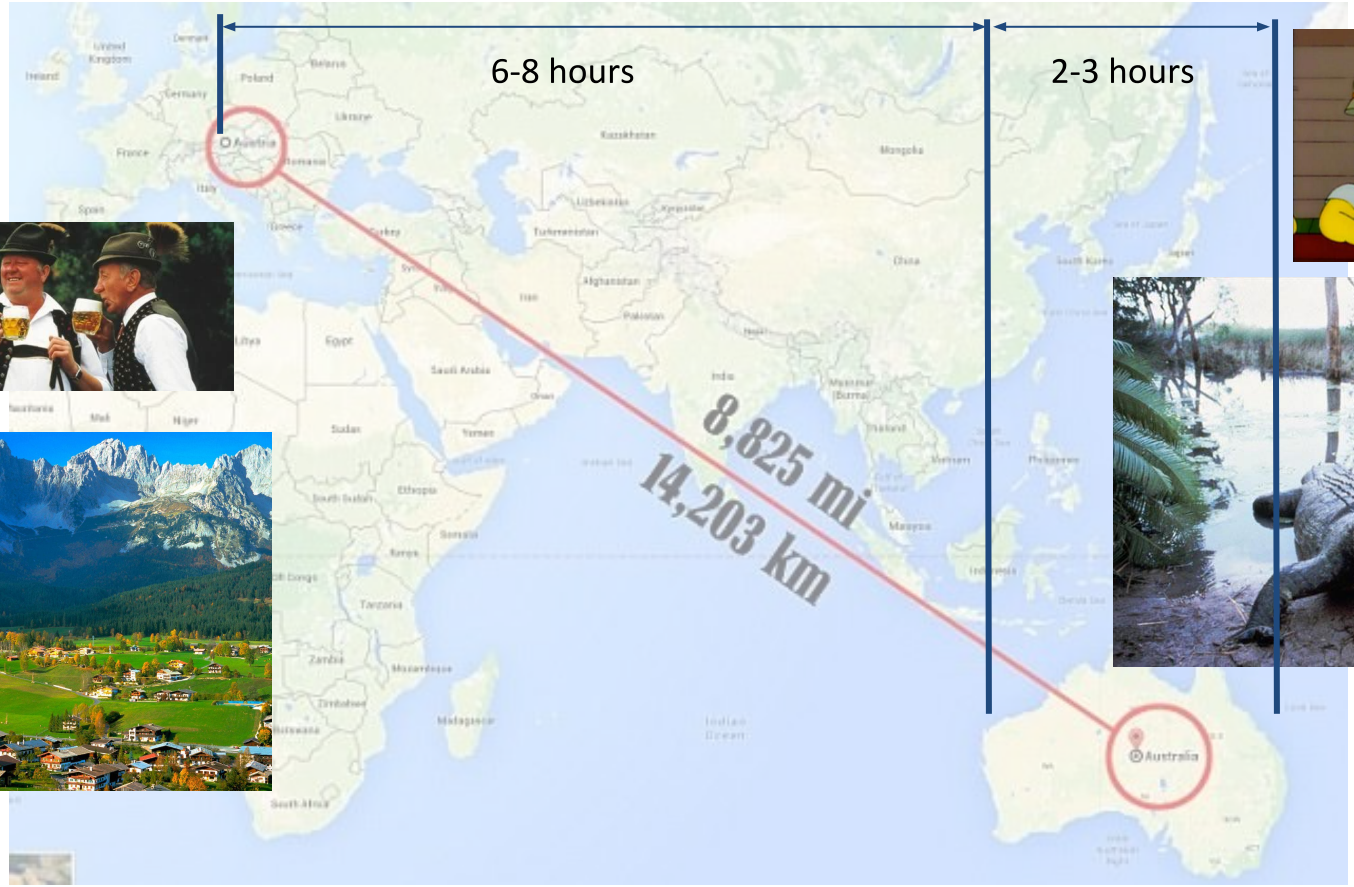
Chris Evelo

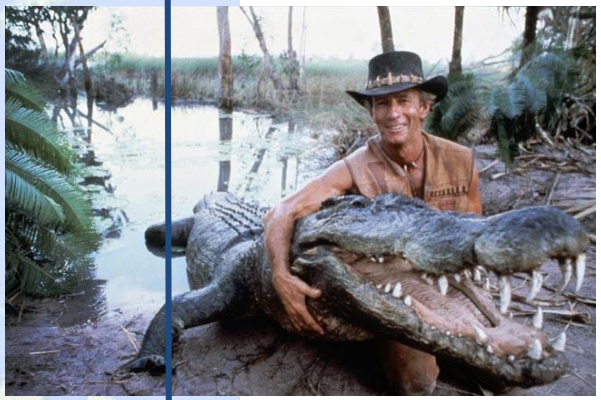

Carole Goble

Challenge: Australia != Austria

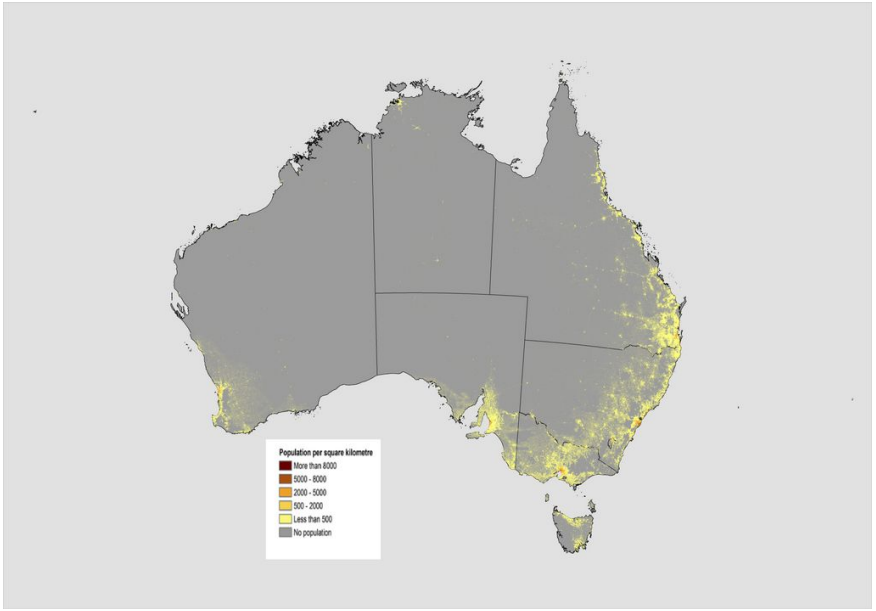


Australia != Austria





Challenge: Australia is big, sparse and distant



BIOPLATFORMS AUSTRALIA

Australia: big and sparse

51	Madagascar	26,969,307	26,262,313	2.69%	587,041 km²	46/km²
52	Cameroon	25,876,380	25,216,267	2.62%	475,442 km²	54/km²
53	Ivory Coast	25,716,544	25,069,230	2.58%	322,463 km²	80/km²
54	North Korea	25,666,161	25,549,604	0.46%	120,538 km²	213/km²
55	Australia	25,203,198	24,898,152	1.23%	7,692,024 km²	3/km²
56	Taiwan	23,773,876	23,726,460	0.20%	36,193 km²	657/km²
57	Niger	23,310,715	22,442,822	3.87%	1,267,000 km²	18/km²
58	Sri Lanka	21,323,733	21,228,763	0.45%	65,610 km²	325/km²
59	Burkina Faso	20,321,378	19,751,466	2.89%	272,967 km²	74/km²
60	Mali	19,658,031	19,077,749	3.04%	1,240,192 km²	16/km²
61	Romania	19,364,557	19,506,114	-0.73%	238,391 km²	81/km²
62	Chile	18,952,038	18,729,160	1.19%	756,102 km²	25/km²

BIOPLATFORMS AUSTRALIA

Challenge: Australia is wealthy

52	Cameroon		475,442 km²	54/km²
53	Ivory Coast		322,463 km²	80/km²
54	North Korea		120,538 km²	213/km²
55	Australia		7,692,024 km²	3/km²
56	Taiwan		36,193 km²	657/km²
57	Niger		1,267,000 km²	18/km²
58	Sri Lanka		65,610 km²	325/km²
59	Burkina Faso		272,967 km²	74/km²
60	Mali		1,240,192 km²	16/km²
61	Romania		238,391 km²	81/km²
62	Chile		756,102 km²	25/km²

	World	**80,683,787**
1	United States	19,390,604
2	China[n 5]	12,237,700
3	Japan	4,872,137
4	Germany	3,677,439
5	United Kingdom	2,622,434
6	India	2,597,491
7	France	2,582,501
8	Brazil	2,055,506
9	Italy	1,934,798
10	Canada	1,653,043
11	Korea, South	1,577,524
12	Russia[n 3]	1,530,751
13	Australia	1,323,421
14	Spain	1,311,320
15	Mexico	1,149,919

BIOPLATFORMS AUSTRALIA

Challenge: Australia is wealthy...and productive

52	Cameroon		475,442 km²	54/km²
53	Ivory Coast		322,463 km²	80/km²
54				/km²
55				n²
56				/km²
57				km²
58				/km²
59	Burkina Faso		272,967 km²	74/km²
60	Mali		1,240,192 km²	16/km²
61	Romania		238,391 km²	81/km²
62	Chile		756,102 km²	25/km²

	World	**80,683,787**
1	United States	19,390,604
2	China[n 5]	12,237,700
9	Italy	1,934,798
10	Canada	1,653,043
11	Korea, South	1,577,524
12	Russia[n 3]	1,530,751
13	Australia	1,323,421
14	Spain	1,311,320
15	Mexico	1,149,919

"...In 2013 we contributed to **3.9 per cent** of the world's research output (in terms of publications) from 0.3 per cent of the world's population, **ranking 9th in the OECD**."

BIOPLATFORMS AUSTRALIA

How do we use infrastructure to remain competitive in a global research landscape?

How do we use infrastructure to remain competitive in the global research landscape?

How do we keep up with global $ and critical mass?

15	Mexico	1,149,919

AUSTRALIA

How do we use infrastructure to remain competitive in the global research landscape? How do we keep up with global $?

What do we specialise in, what do we bring to the global table?

How do we use infrastructure to remain competitive in the global research landscape?
How do we keep up with global $?
What do we specialise in, what do we bring to the global table?

How do we deal with distance, internally and externally? Data movement?

How do we use infrastructure to remain competitive in the global research landscape?
How do we keep up with global $?
What do we specialise in, what do we bring to the global table?
How do we deal with distance, internally and externally? Data movement?

How do we keep staff?

How do we use infrastructure to remain competitive in the global research landscape?
How do we keep up with global $?
What do we specialise in, what do we bring to the global table?
How do we deal with distance, internally and externally? Data movement?
How do we keep staff?

What's Australia's place in the biosciences global research infrastructure context?

15	Mexico	1,149,919