Collections – From Local to Global

Building strength in an interconnected world

Joanne Daly |
5 May 2014

Donald Hobern, John La Salle, Beth Mantle and others

Structure of the Talk

1. Context of digital collections
2. National and global aggregators
3. Towards 2021 - challenges in data flow

Conclusion

From the perspective of someone outside NH collections:

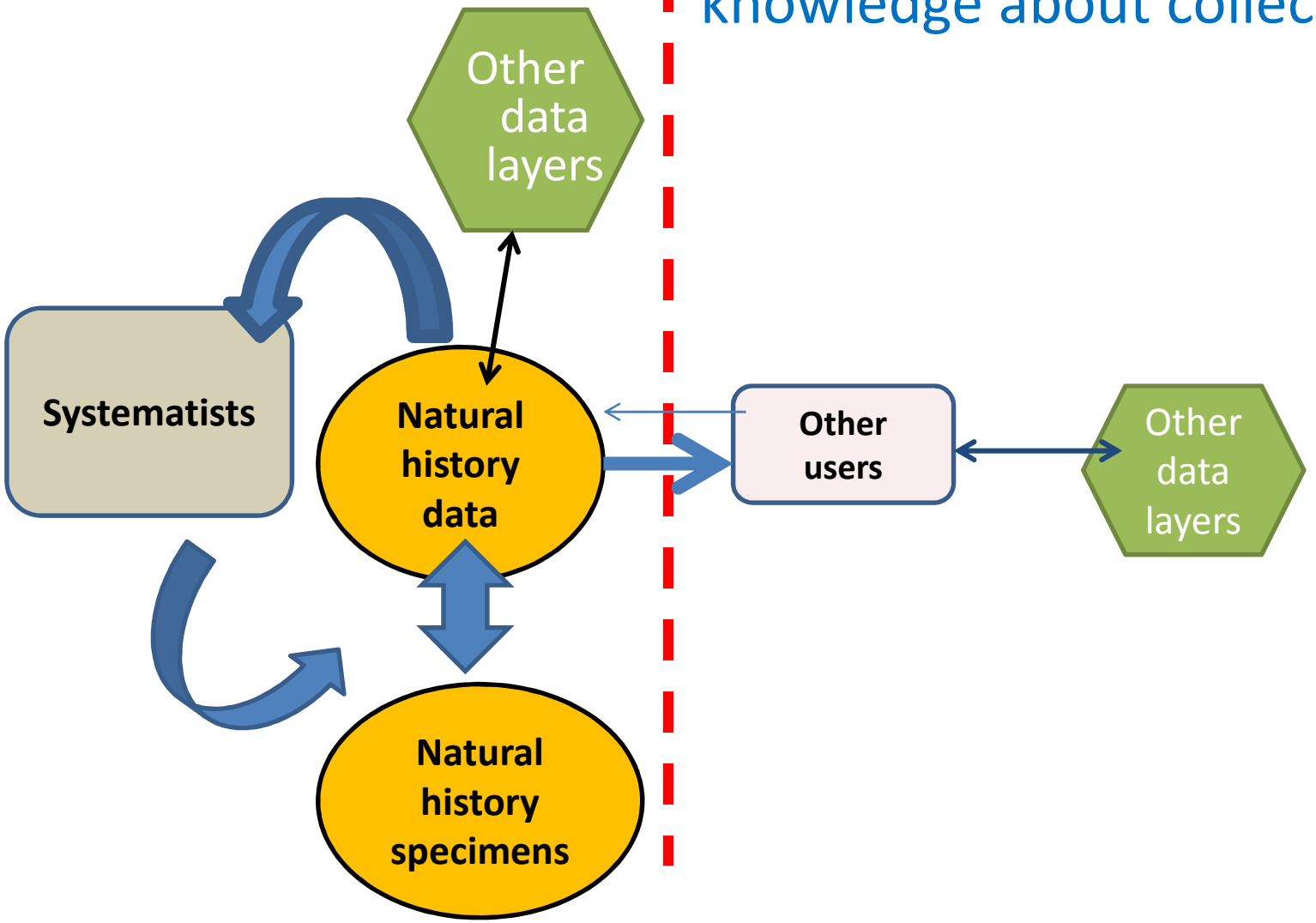
In 20thC natural history collections were seen as stewards of physical specimens and provided species identification

In 21st C they are valued because they provide knowledge in digital form to a wide range of users and applications.

This can only be achieved by aggregating data across collections, across national boundaries and across data types

1. Context for collections as knowledge providers

Collections 20th Century – creators of knowledge about collections



Infrastructure : The IT World since 2001

Socialising information and communication

2001	**GBIF and Catalogue of Life**- commenced **Wikipedia** – socialising knowledge **Creative Commons** – licence agreements for sharing
2003	**Skype** – enabling face to face discussions globally **PLoS Biology** – commences
2004/05	**Facebook and Twitter-** social networking **Cloud computing**–commoditising file storage
2007	**iPhones** **GBIF portal** launched **Atlas of Living Australia** (2007/08) – funded **EOL** launched – major institutions and private funders

Infrastructure : The IT World since 2001

Socialising information and communication

2010	**IPADS - merged comms, computing and cameras**
2011/12	**Tablet Market - Expansion of android based tablets**
	Smartphones - all embracing
	Print media - Continuing decline
2013/14	**2nd Gen portals Launch of new portals for GBIF and EOL**
	Instant publishing of species data

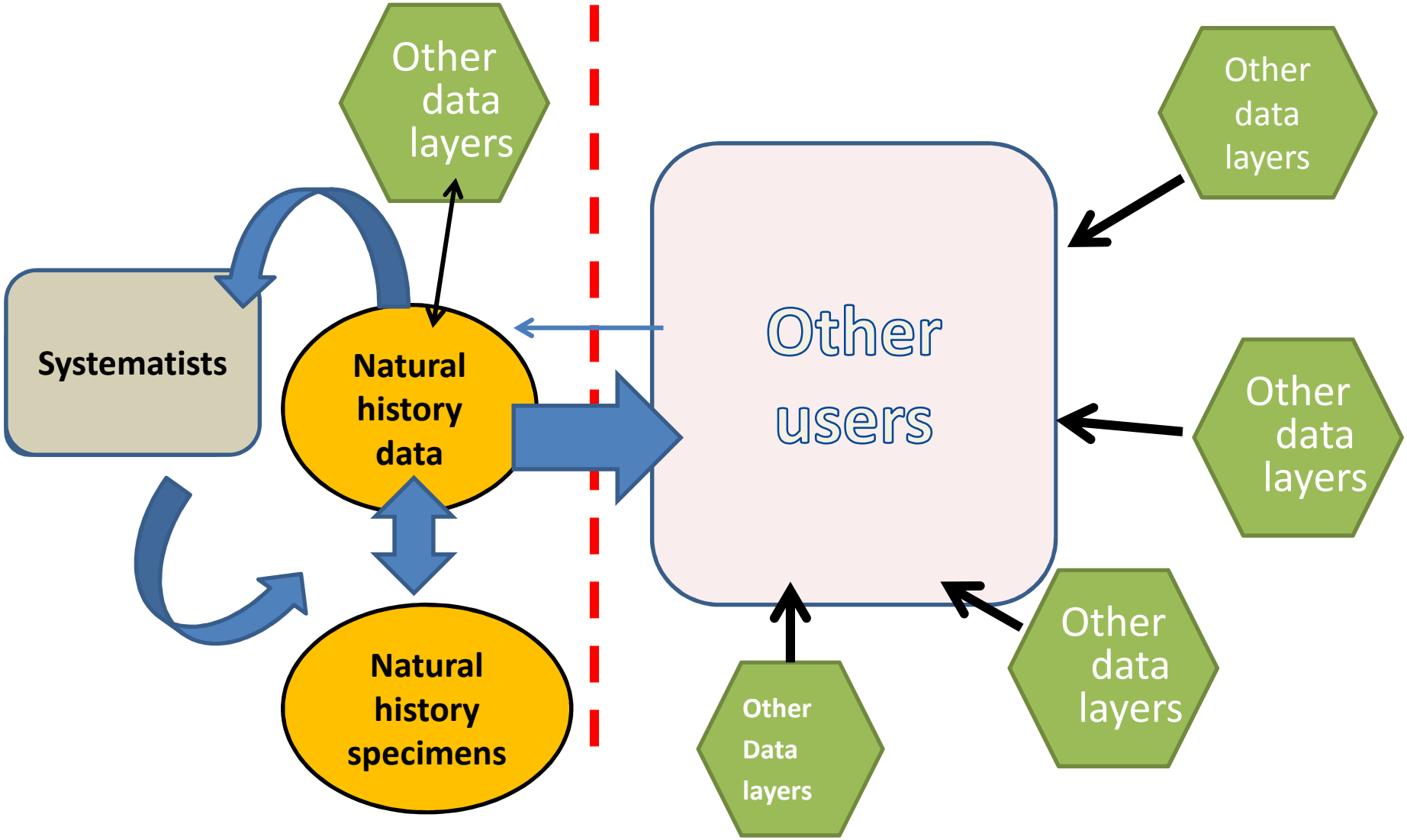
The Research World since 2001

Explosion of computing power and biotechnology

- **Petascale computing** – visualisation and global scale models; 'the grid'
- **Wireless** and **broadband networks – driving the desire for mobility**
- **Major national investments in infrastructure globally** – repositories, shared solutions and workspaces
- Analytical tools enabling **integration of data** across multiple **data layers**
- **Genome projects** from years to months to days? From tens of millions of $$ to KK's of $

All this drives and is driven by a hunger for data of all kinds and an expectation that it will be available

Collections in 21st Century – creators of knowledge about the world around

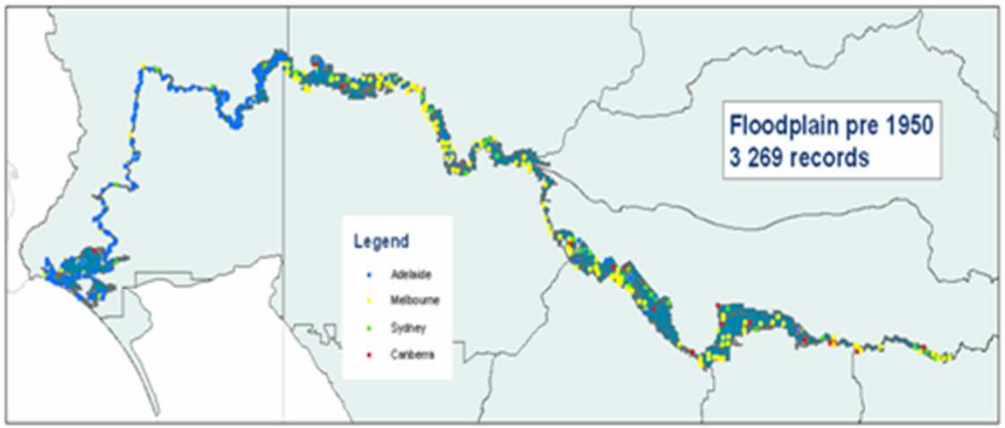


Challenge in the 21st Century is how to convert data about physical collections into...

Bird Collection at Naturalis, Leiden

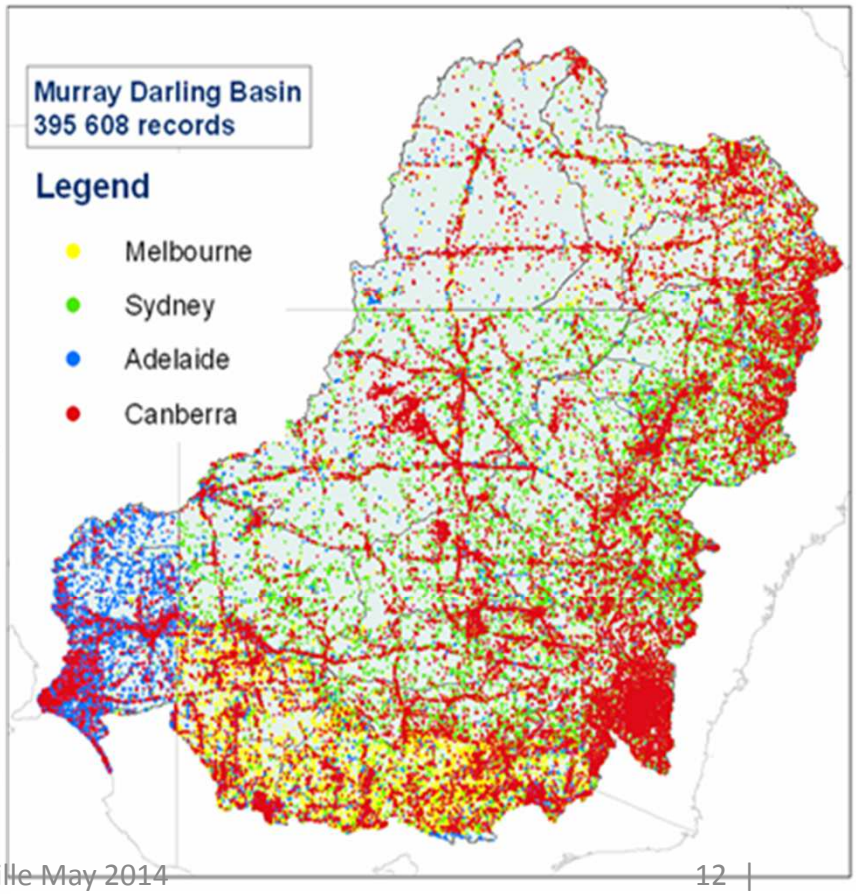
Knowledge products that can deliver dynamic web-based products while...

(Murray Darling Basin research)



Floodplain: plant collections

MDB: Plant collections



Many global and regional biodiversity initiatives emerged

Conabio	**1992**
GenBank	**1992**
TDWG	**1985/1994**
ITIS	**1996**
Species 2000	**1997**
GBIF	**2001**
Catalogue of Life	**2001**
Morphbank	**2000-2002**
Barcode of Life	**2004**
Atlas of Living Australia	**2007**
Encyclopedia of Life	**2007**
Biodiversity Heritage Library	**2007**

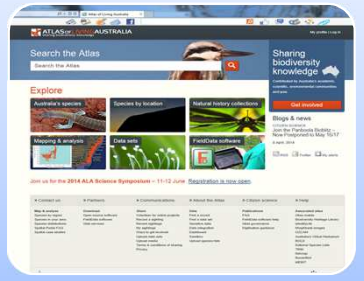
GeoBon
IPBES
Lifewatch

Drivers for Federated Systems for natural history data

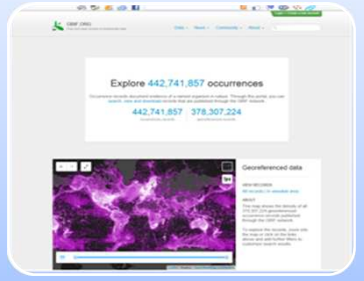
Past effort was inadequate to meet the challenge in biodiversity conservation and management

- Data aggregation is essential to address questions
- Many national efforts in biodiversity had been fragmented, underfunded and uncoordinated
- Species level work needs to move from a cottage industry to industrial scale
- Imminent 'data deluge'

2. Why different initiatives?



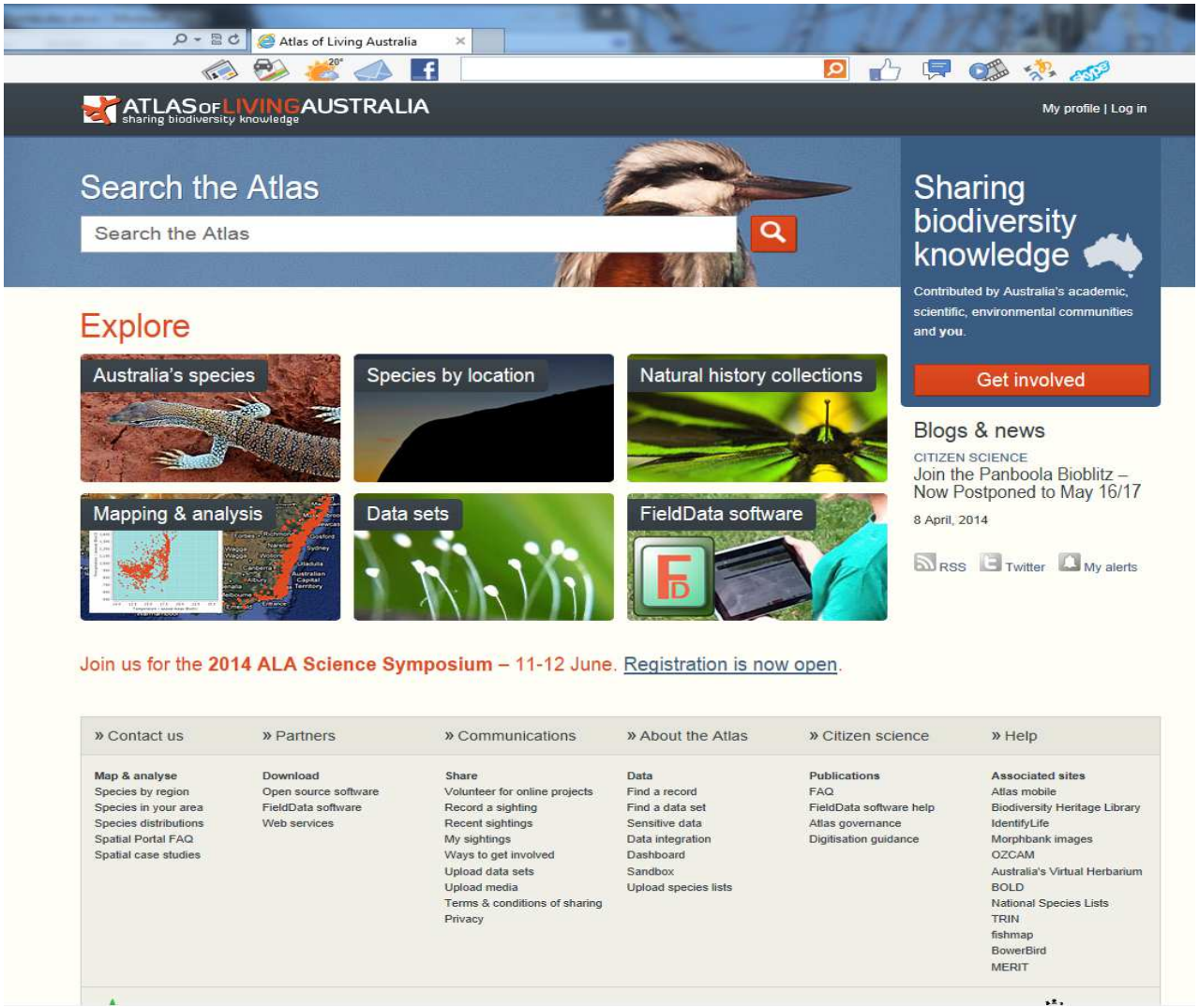
National Site – ALA – partnership among collections



Global aggregator – GBIF – intergovernmental agreement

Global aggregator - EoL - partnership of major institutions

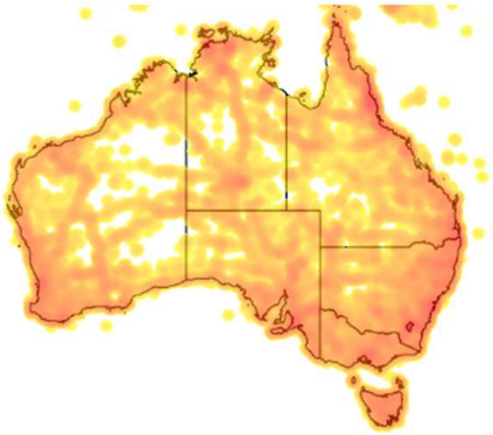
www.ala.org

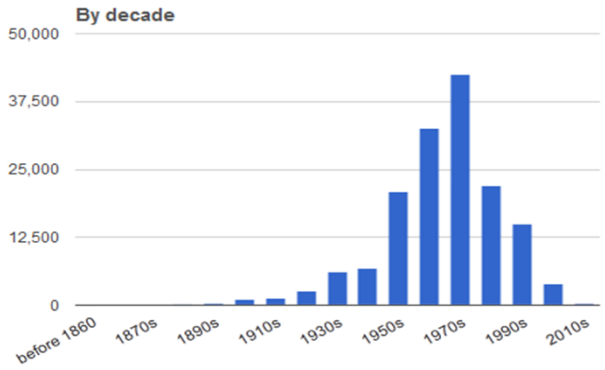

Atlas of Living Australia
ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge
My profile | Log in
Search the Atlas
Search the Atlas
Sharing biodiversity knowledge
Contributed by Australia's academic, scientific, environmental communities and you.
Get involved
Explore
Australia's species
Species by location
Natural history collections
Mapping & analysis
Data sets
FieldData software
Blogs & news
CITIZEN SCIENCE
Join the Panboola Bioblitz – Now Postponed to May 16/17
8 April, 2014
RSS
Twitter
My alerts
Join us for the 2014 ALA Science Symposium – 11-12 June. Registration is now open.
» Contact us
» Partners
» Communications
» About the Atlas
» Citizen science
» Help
Map & analyse
Species by region
Species in your area
Species distributions
Spatial Portal FAQ
Spatial case studies
Download
Open source software
FieldData software
Web services
Share
Volunteer for online projects
Record a sighting
Recent sightings
My sightings
Ways to get involved
Upload data sets
Upload media
Terms & conditions of sharing
Privacy
Data
Find a record
Find a data set
Sensitive data
Data integration
Dashboard
Sandbox
Upload species lists
Publications
FAQ
FieldData software help
Atlas governance
Digitisation guidance
Associated sites
Atlas mobile
Biodiversity Heritage Library
IdentifyLife
Morphbank images
OZCAM
Australia's Virtual Herbarium
BOLD
National Species Lists
TRIN
fishmap
BowerBird
MERIT

ALA and natural history collections

- 147 NH collections in Australia
- 45 M records of which 9.4 M preserved specimens
- National species lists
- 409 spatial layers
- Next 12 months: adding phylogenetic capacity; e-flora for Australia; virtual collections (explore through images);

Use of Insect data in ANIC - 2013/14 from ALA



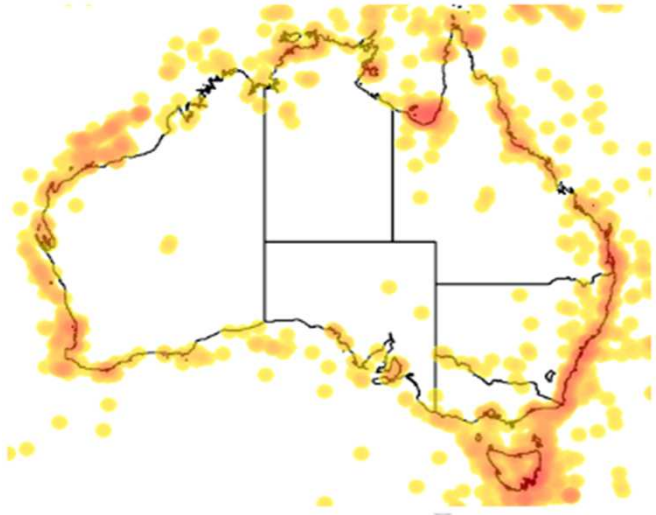


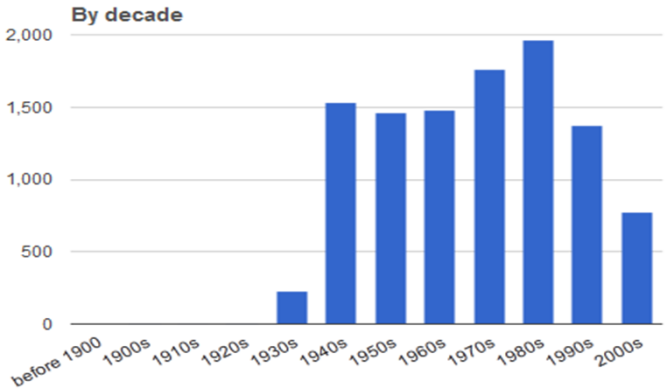
		ANIC (#rank)	Best or #2 in country (#rank)
Number of specimens (estimate)		12,000,000 (#1)	2,500,000 MV (#2)
Number of records in ALA		193,591 (#3)	230,000 MV (#1)
% of specimens in ALA		2% (#4)	12% AM (#1)
Downloads per record in 12 months		**18 (#1)**	**14 QM (#2)**
Species in ALA		4,621 (#3)	12,700 MV (#1)
Types	**Primary**	**16,656 (#1)**	2,816 QM (#2)
	Total		20,000 MV (#1)
Specimens with conservation status		966 (#2)	2,600 MV (#1)

Breakdown by reason of occurrence record downloads 2013 in ANIC through ALA

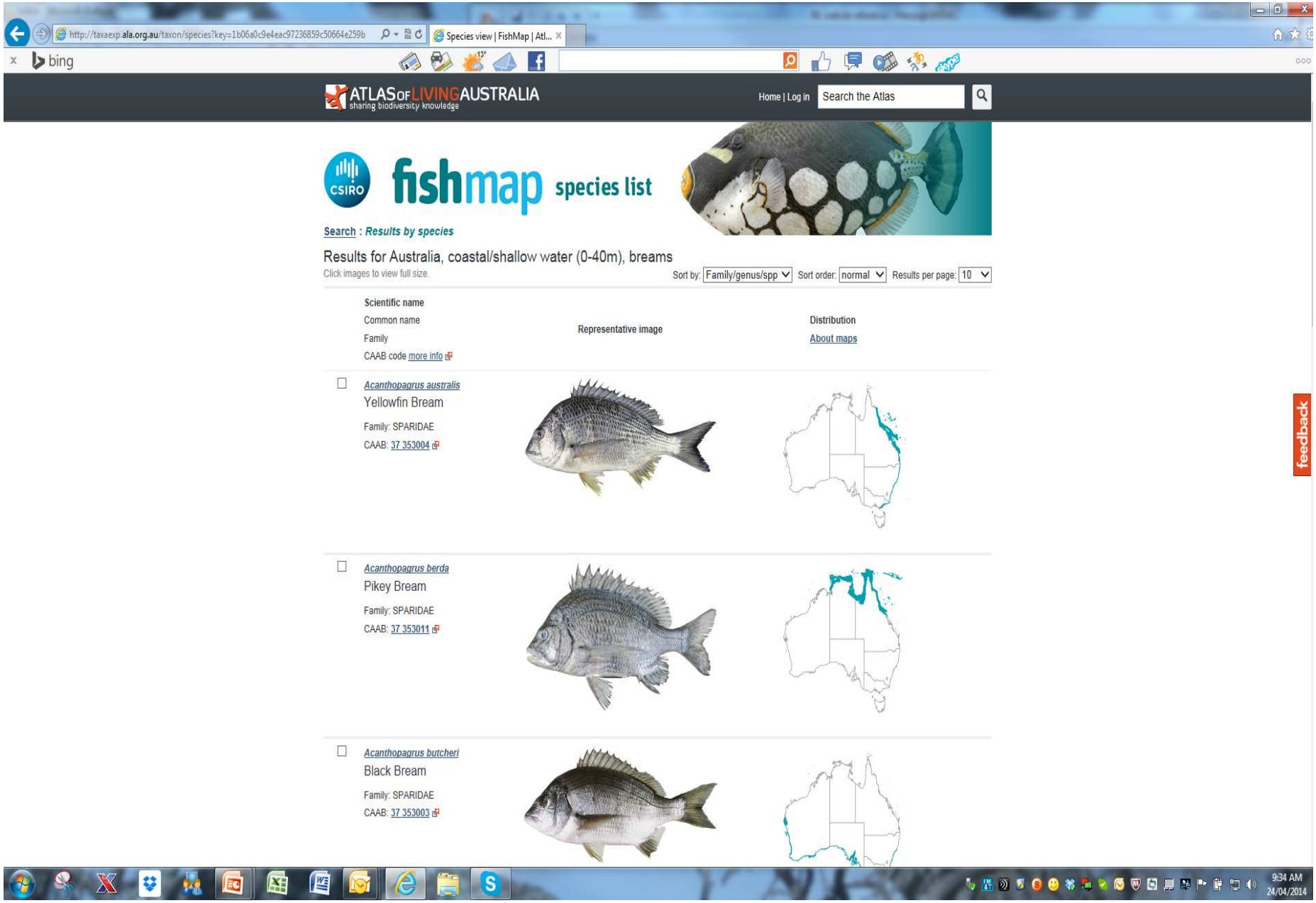
Category	Prop of total	Downloads	Events
Ecological Research	34%	1,184,843	148
Scientific Research	33%	1,155,809	153
Environmental Impact	12%	408,348	236
Education	9%	323,006	191
Collection Management	6%	193,669	3
Conservation	3%	117,023	221
Testing	2%	57,570	239
Other	0%	12,153	52
Other scientific Research	0%	3,859	17
Systematic Research	0%	3,563	5
Biosecurity	0%	1,357	9
TOTAL	100%	**3,461,200**	**1274**

Australian National Fish Collection 13/14 - ALA





		ANFC (#rank)	Best in country (#rank)
Number of specimens (estimate)		148,000 (#4)	350,000 MV (#1)
Number of records in ALA		32,425 (#2)	163,167 AM (#1)
% of specimens in ALA		20% (#4)	77% QVMAG & AM (#1)
Downloads per record in 12 months		**12.4 (#1)**	**10.1 SAMF (#2)**
Species in ALA		2,926 (#2)	6,062 AM (#1)
Types	Primary	44 (#4)	844 AM (#1)
	Total	481 (#2)	4,234 AM (#1)
Specimens with conserv. status		21 (#7)	1,656 AM (#1)


http://taxaexp.ala.org.au/taxon/species?key=1b06a0c9e4eac97236859c50664e259b
Species view | FishMap | Atl...
bing
ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge
Home | Log in
Search the Atlas
CSIRO
fishmap species list
Search : Results by species
Results for Australia, coastal/shallow water (0-40m), breams
Click images to view full size.
Sort by: Family/genus/spp
Sort order: normal
Results per page: 10
Scientific name
Common name
Family
CAAB code more info
Representative image
Distribution
About maps
Acanthopagrus australis
Yellowfin Bream
Family: SPARIDAE
CAAB: 37 353004
Acanthopagrus berda
Pikey Bream
Family: SPARIDAE
CAAB: 37 353011
Acanthopagrus butcheri
Black Bream
Family: SPARIDAE
CAAB: 37 353003
feedback
9:34 AM
24/04/2014

Breakdown by reason of occurrence record downloads 2013 in ANFC through ALA

Category	Prop (%)	Downloads	Events
Testing	40	163,260	34
Scientific research	17	70,734	371
Environmental impact	16	64,078	40
Education	8	34,401	49
Other	8	32,918	14
Collection management	8	32,624	10
Conservation	1	5,378	90
Ecological research	1	3,441	104
Systematic research	0	455	6
Other scientific research	0	106	3
Biosecurity	0	11	2
Total	100%	407406	723

Occurrence downloads by reason ALA 23/4/14

Category	Prop of Total (%)	Downloads	Events
Other	27	398 M	17,537
Ecological Research	24	356 M	20,793
Scientific Research	22	327 M	32,815
Conservation	10	143 M	6,673
Environment impact	7	108 M	3,773
Education	5	77 M	6,596
Collection Management	3	41 M	438
Systematic Research	1	19 M	1,495
Biosecurity	1	9 M	403
Other Scientific Research	0.3	4 M	555
Total	**100%**	**1.48 B**	**91,078**

Caution

- Downloads do not equate to usage
- It will underestimate the usage for the collections data as reference material
- This is high in taxonomic and general species identification work

Examples of Utility of ALA to NH collections in Australia

1. Collections:

- **Searches on data** to locate specimens, images, drawers etc
- Allow data to be contextualise with **enriched data** – e.g. altitude, depth, salinity
- Provide **image** galleries
- Allows data to be **visualised and mapped** which can be used to detect outliers
- **High visibility of collections** through field apps

2. External users:

- Assist with or reduce **external enquiries**
- **Reduce** the number and size of **loans**
- Can assist in **data curation** by allowing remote input

Examples of Utility ALA, cont'd

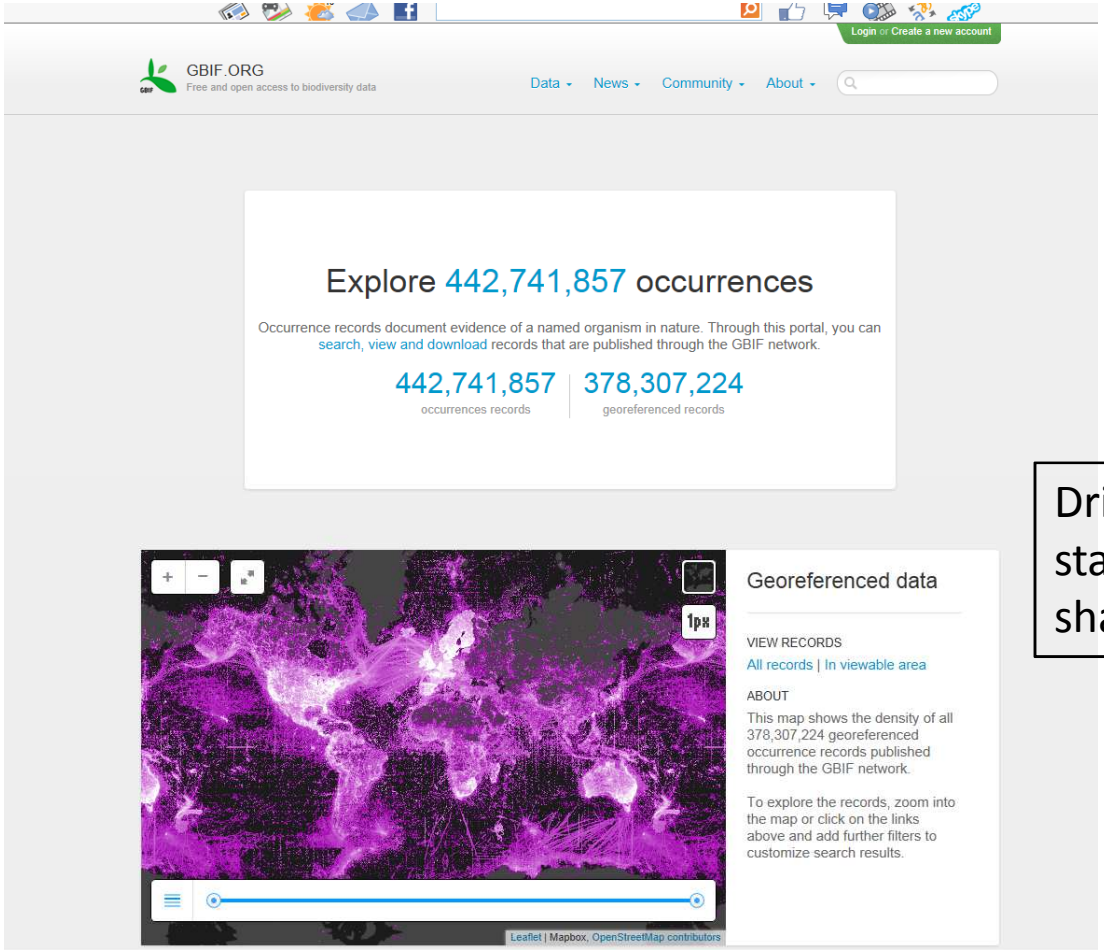
3. Community

- Can lead to **agreement on national names list** and consistence of names
- Led to **unity among disparate collections** to focus on shared problems and solutions
- Given a real **sense of community** in faunal collections

So what do global initiatives do?

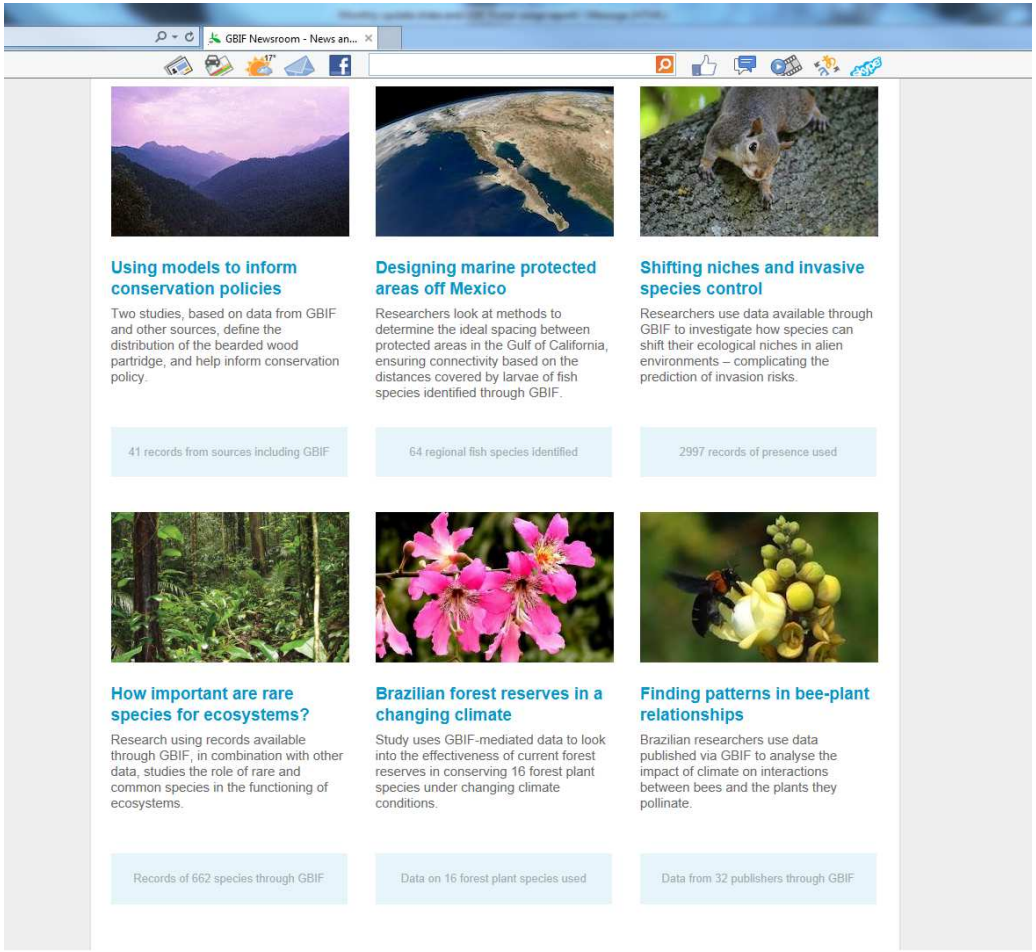
Data accessible through GBIF: 97 M are specimen records, 210 M human observation records

Global solutions to informatics architecture

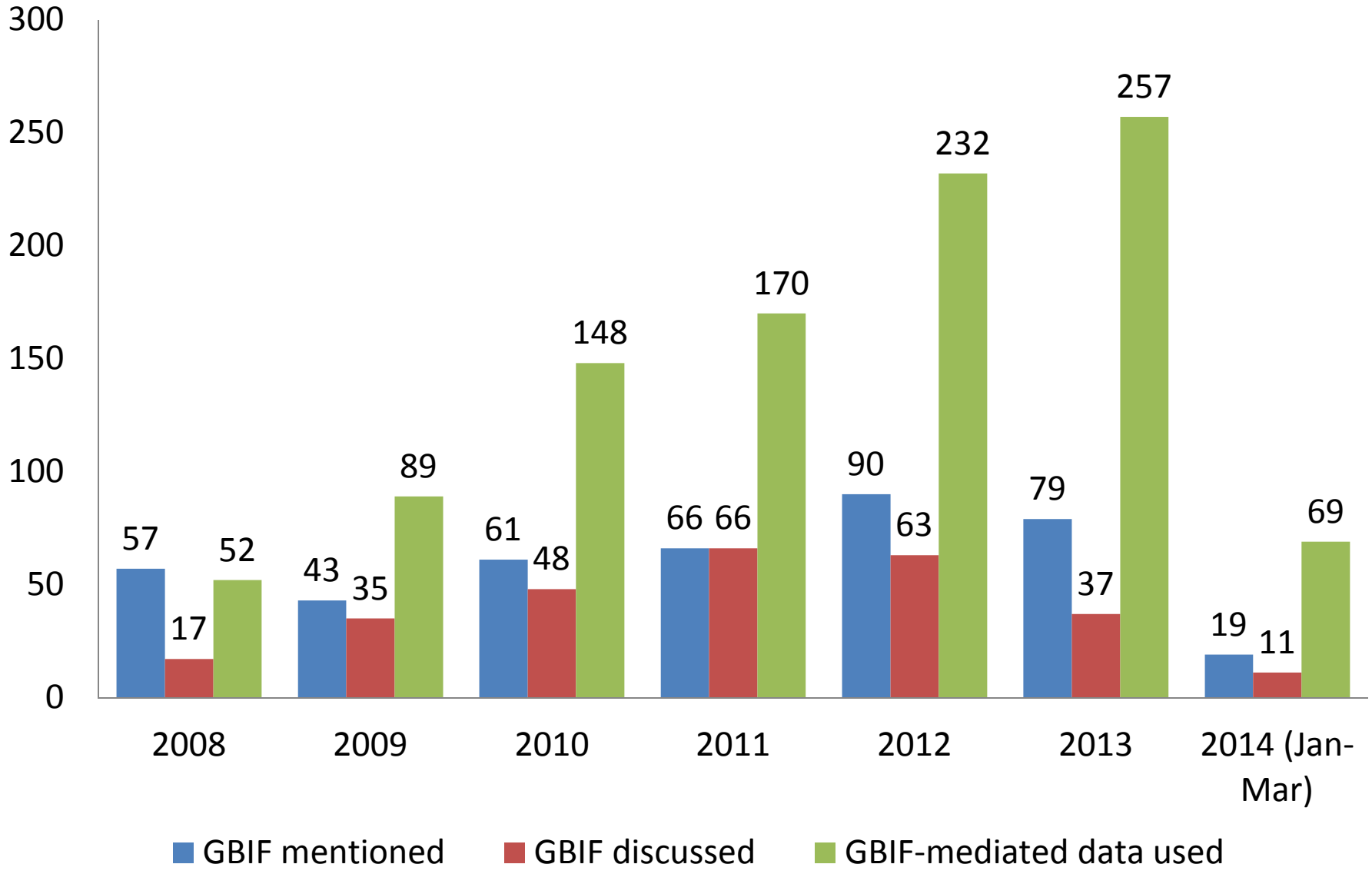


Drive global standards for sharing

Featured uses of data accessed through GBIF http://www.gbif.org/newsroom/uses



GBIF citation in research



New Developments in GBIF

- **Expanded** number of fields of information
- Searches possible for records designated at **holotypes**
- Release of ability to view **images, access audio and video** support (early May)
- **Instant** access to data

Provide global solutions: Instant access to data in publications (with Pensoft Publishers)

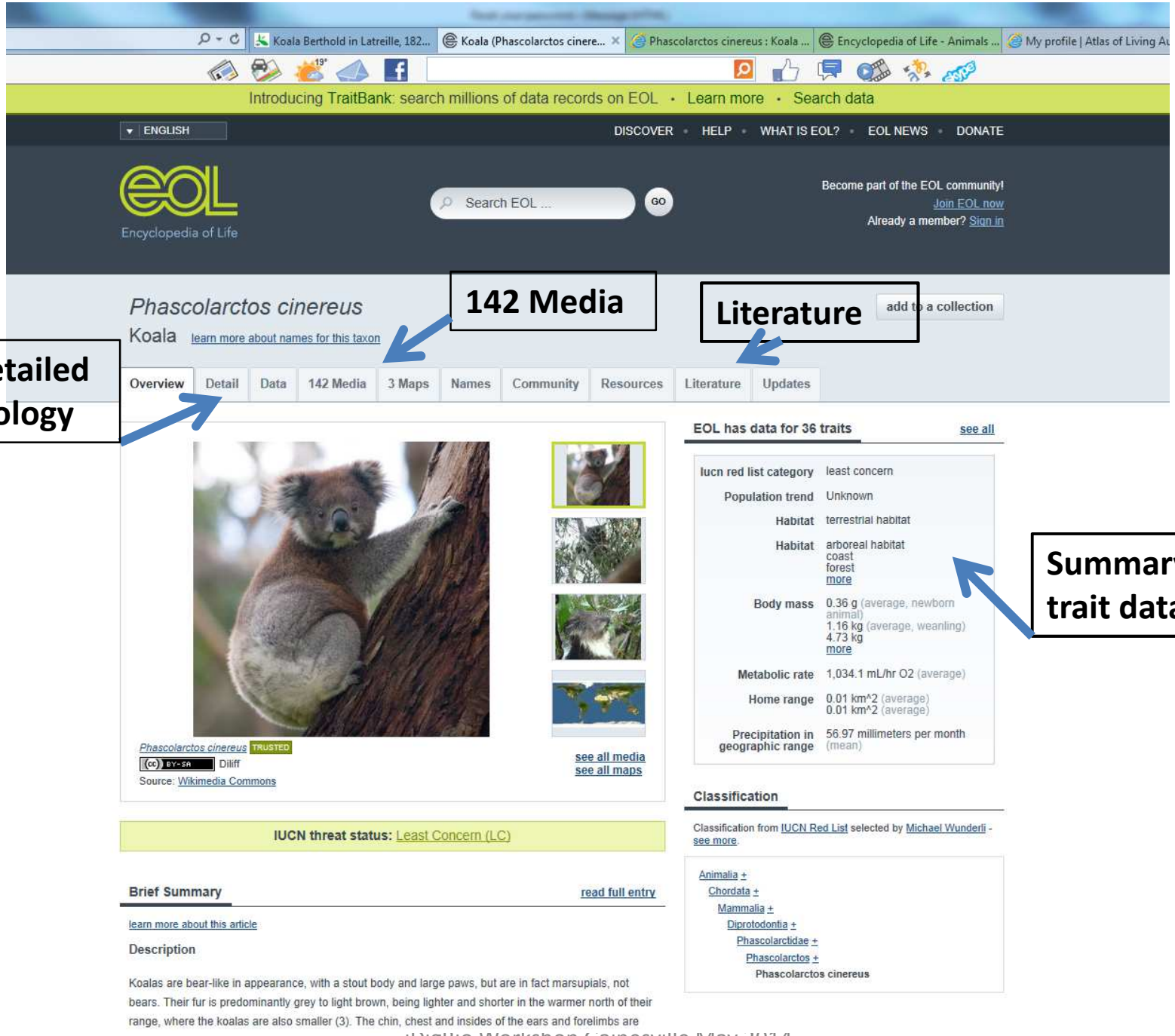
Biodiversity Data Journal

Supervised by Jeremy Miller and Menno Schilthuizen, from Naturalis Biodiversity Center, Leiden

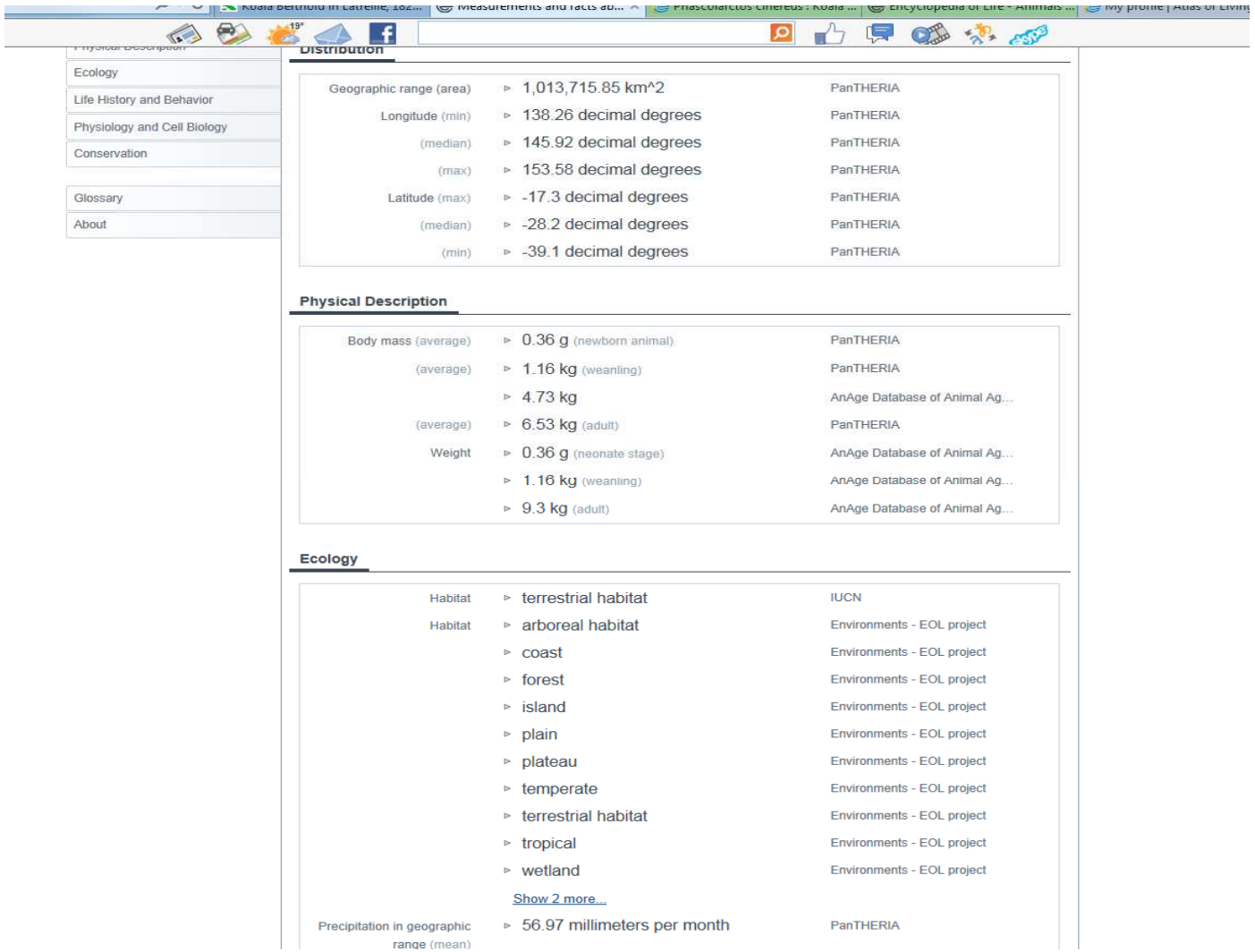
Occurrence data harvested by GBIF

EoL shows rich data with images and species description


http://eol.org/
Encyclopedia of Life - Anim...
bing
Introducing TraitBank: search millions of data records on EOL • Learn more • Search data
ENGLISH
DISCOVER • HELP • WHAT IS EOL? • EOL NEWS • DONATE
eol
Encyclopedia of Life
Become part of the EOL community!
Join EOL now
Already a member? Sign in
Global access to knowledge about life on Earth
Search EOL ...
GO
Paralonchurus goodei
Croaker
EOL News
see more
Edward O. Wilson Biodiversity Symposium
Marie Studer of the Encyclopedia of Life and other international biodiversity experts will join Dr. Edward O. Wilson for the Edward O. Wilson Biodiversity Symposium at the University of Alabama from April 22-24, 2014.... more
APRIL 23, 2014 13:12
Happy Earth Day!
Celebrate Earth Day by exploring some of our favorite EOL collections!
Best Images on EOL
... more
APRIL 22, 2014 12:56
One Species at a Time Podcast: Riftia
Did life on earth have started in deep sea vents? Scientists at Woods Hole Oceanographic Institution (WHOI) are testing this hypothesis and their current work is reviewed in this ... more
APRIL 16, 2014 14:01
Community Activity
see more
Katja Schulz marked the classification from "Species 2000 & ITIS Catalogue of Life: April 2013" as preferred for "Stanhopea".
ABOUT 1 HOUR AGO
reply
Eric Peterson commented on "File:NIEdot315.jpg": #4 in this figure would be what is currently called Letharia columbiana.
ABOUT 3 HOURS AGO
reply
Peter Zaia added text to "Venom toxicity" on "Dendroaspis polylepis GÜNTHER 1864".
D. polylepis has a varied venom LD50 ranging from 0.05 mg/kg...
ABOUT 4 HOURS AGO
reply


Introducing TraitBank: search millions of data records on EOL • Learn more • Search data
ENGLISH
DISCOVER • HELP • WHAT IS EOL? • EOL NEWS • DONATE
Encyclopedia of Life
Search EOL ...
GO
Become part of the EOL community!
Join EOL now
Already a member? Sign in
Phascolarctos cinereus
Koala
learn more about names for this taxon
add to a collection
Overview
Detail
Data
142 Media
3 Maps
Names
Community
Resources
Literature
Updates
142 Media
Literature
Detailed biology
Summary trait data
Phascolarctos cinereus TRUSTED
Diliff
Source: Wikimedia Commons
see all media
see all maps
EOL has data for 36 traits
see all
iucn red list category least concern
Population trend Unknown
Habitat terrestrial habitat
Habitat arboreal habitat
coast
forest
more
Body mass 0.36 g (average, newborn animal)
1.16 kg (average, weanling)
4.73 kg
more
Metabolic rate 1,034.1 mL/hr O2 (average)
Home range 0.01 km^2 (average)
0.01 km^2 (average)
Precipitation in geographic range 56.97 millimeters per month (mean)
IUCN threat status: Least Concern (LC)
Classification
Classification from IUCN Red List selected by Michael Wunderli - see more.
Animalia +
Chordata +
Mammalia +
Diprotodontia +
Phascolarctidae +
Phascolarctos +
Phascolarctos cinereus
Brief Summary
read full entry
learn more about this article
Description
Koalas are bear-like in appearance, with a stout body and large paws, but are in fact marsupials, not bears. Their fur is predominantly grey to light brown, being lighter and shorter in the warmer north of their range, where the koalas are also smaller (3). The chin, chest and insides of the ears and forelimbs are

Trait Bank is a searchable, open digital repository of organism traits with 3 million records, 330 attributes and 384,000 taxa

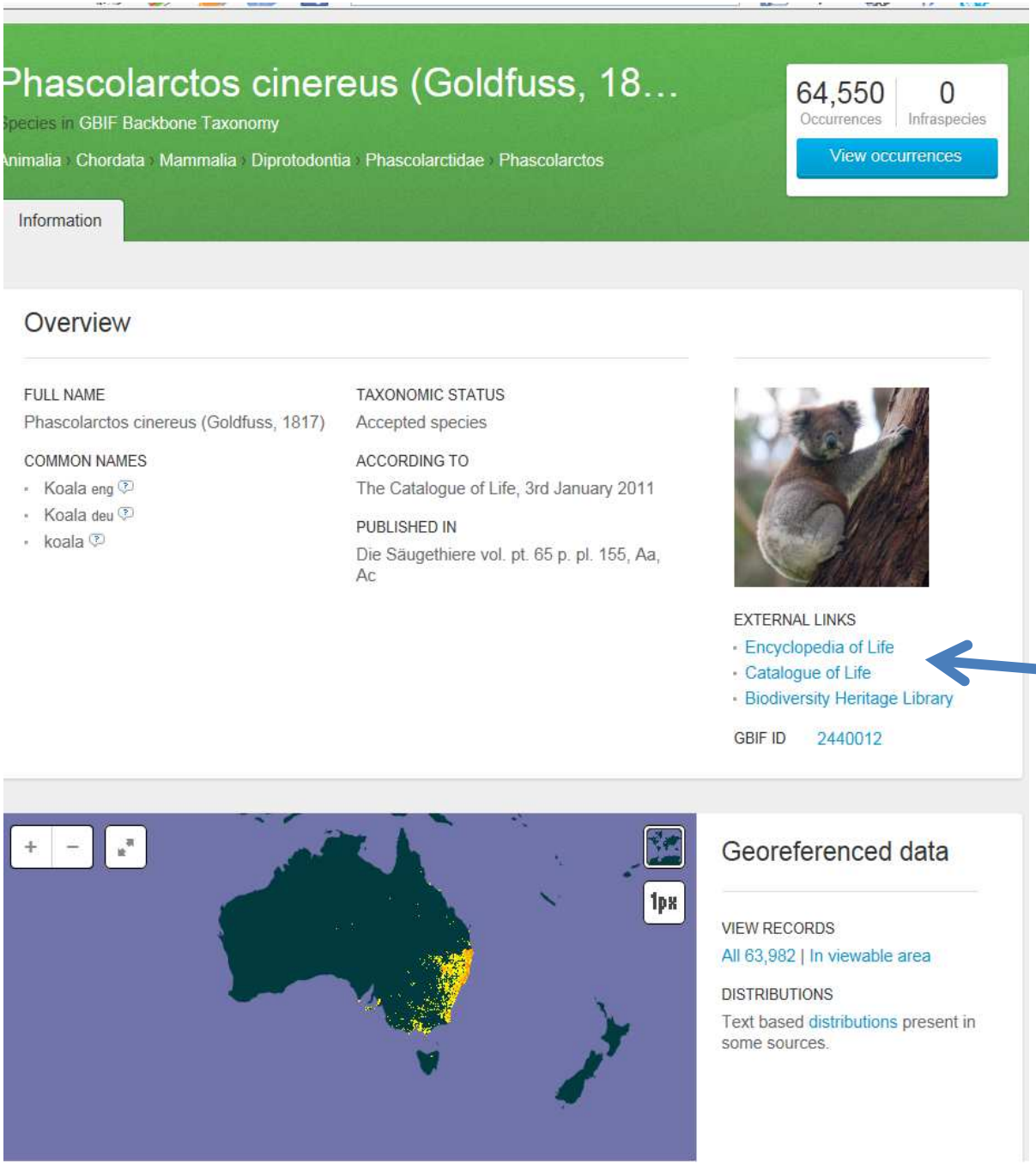


Why have these different initiatives?

Let’s look at koalas

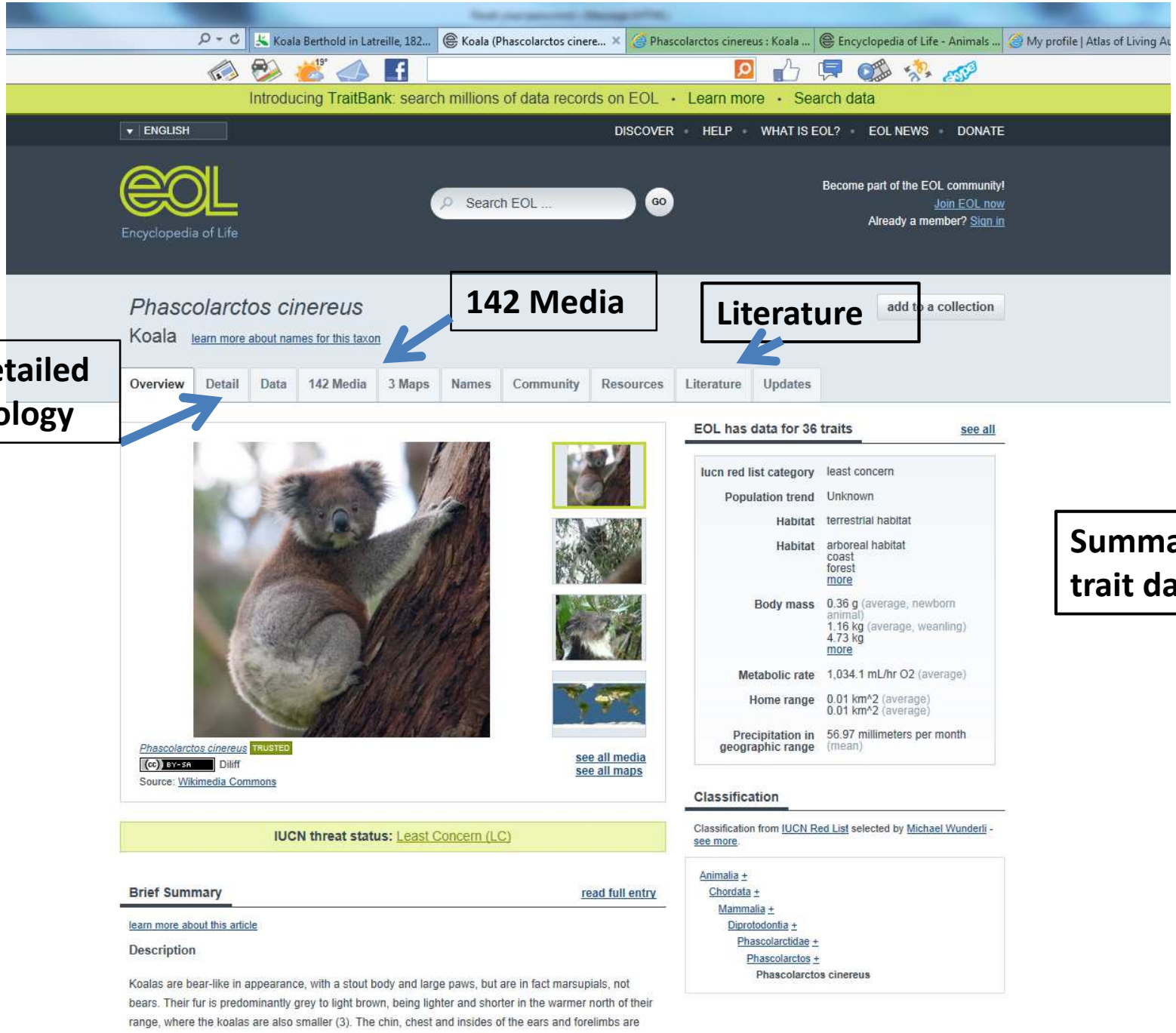
Phascolarctos cinereus

View of Koala in GBIF



64,550 Records, duplicates have been removed and is now around 37,000

Links to Other sites:
EOL
COL
BHL


Introducing TraitBank: search millions of data records on EOL • Learn more • Search data
ENGLISH
DISCOVER • HELP • WHAT IS EOL? • EOL NEWS • DONATE
Encyclopedia of Life
Search EOL ...
GO
Become part of the EOL community!
Join EOL now
Already a member? Sign in
Phascolarctos cinereus
Koala
learn more about names for this taxon
add to a collection
Overview
Detail
Data
142 Media
3 Maps
Names
Community
Resources
Literature
Updates
142 Media
Literature
Detailed biology
Phascolarctos cinereus TRUSTED
Diliff
Source: Wikimedia Commons
see all media
see all maps
EOL has data for 36 traits
see all
iucn red list category least concern
Population trend Unknown
Habitat terrestrial habitat
Habitat arboreal habitat
coast
forest
more
Body mass 0.36 g (average, newborn animal)
1.16 kg (average, weanling)
4.73 kg
more
Metabolic rate 1,034.1 mL/hr O2 (average)
Home range 0.01 km^2 (average)
0.01 km^2 (average)
Precipitation in geographic range 56.97 millimeters per month (mean)
Summary trait data
IUCN threat status: Least Concern (LC)
Classification
Classification from IUCN Red List selected by Michael Wunderli - see more.
Animalia +
Chordata +
Mammalia +
Diprotodontia +
Phascolarctidae +
Phascolarctos +
Phascolarctos cinereus
Brief Summary
read full entry
learn more about this article
Description
Koalas are bear-like in appearance, with a stout body and large paws, but are in fact marsupials, not bears. Their fur is predominantly grey to light brown, being lighter and shorter in the warmer north of their range, where the koalas are also smaller (3). The chin, chest and insides of the ears and forelimbs are

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

My profile Logout Search the Atlas

Species | Locations | Collections | Mapping & analysis | Data sets | Blogs | Get involved | About the Atlas

Home → Australia's species → *Phascolarctos cinereus*

Phascolarctos cinereus (Goldfuss, 1817)

Record a sighting | Alerts

Koala

Name source
Australian Faunal Directory

Rank
Species

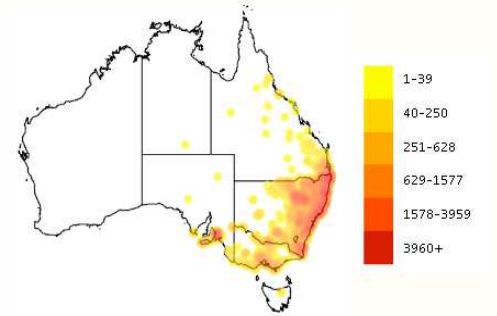
Data links
LSID | JSON / WMS /RDF

Species presence
Recorded In Australia
Terrestrial Habitats

Conservation status
IUCN Least Concern
AU Vulnerable
NSW Vulnerable
NSW Endangered Populations
QLD Vulnerable

Overview | Gallery | Names | Classification | Records | Literature | Sequences

Occurrence records map




Source: Wikipedia

View records list | Map & analyse records

Description

The Koala is a unique Australian marsupial, often incorrectly called a Koala Bear. The males are larger than females. Koalas from southern areas are about 30% larger than the Queensland koalas Koala's fur is thick soft. Ears have long white hairs on the tips. Koalas can live as long as 17 years, however life expectancy is usually less than 10 years due to disease, attacks by dogs, road kills. ... source: OZ Animals

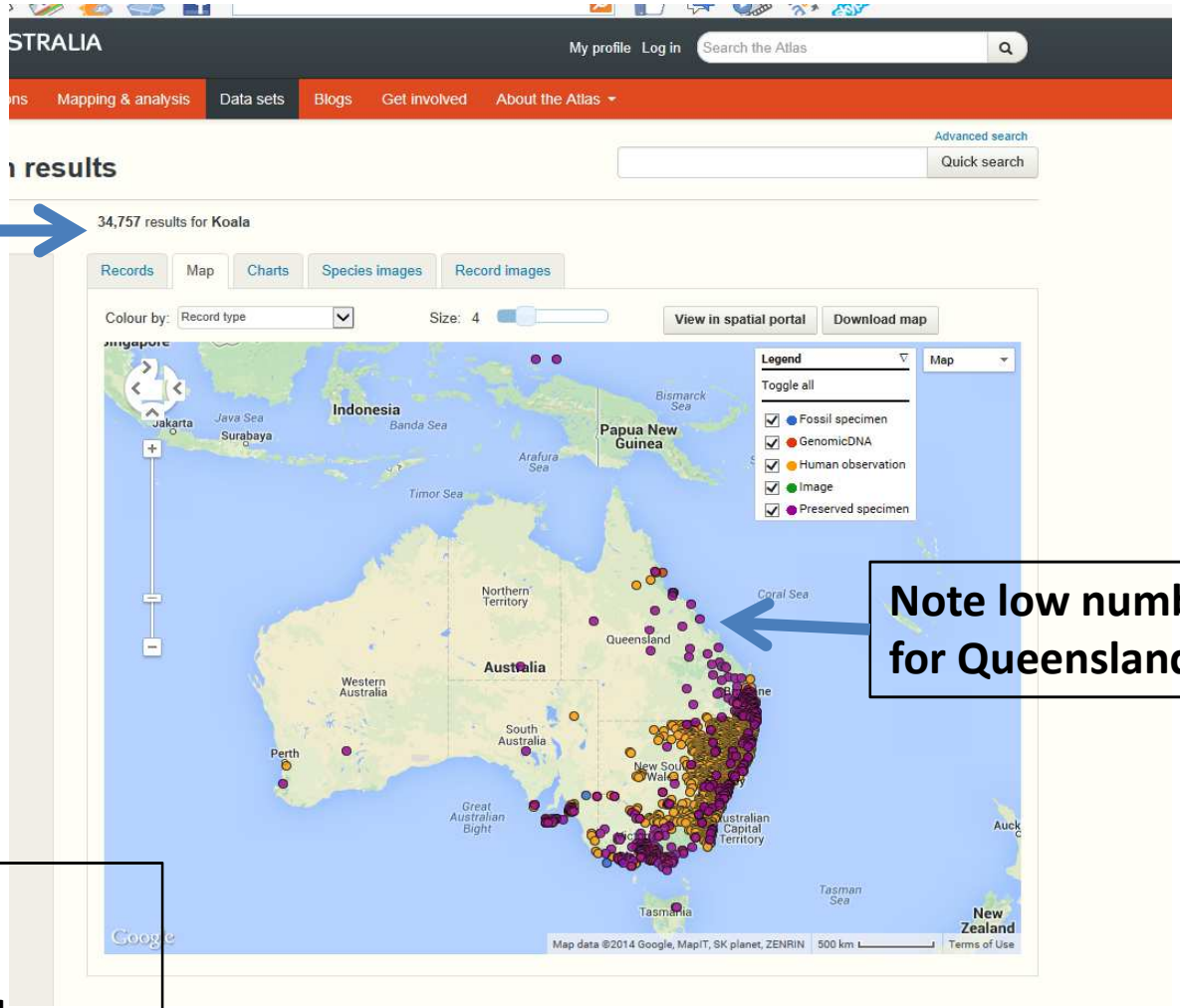
The Koala is an arboreal marsupial with fur ranging from grey to brown above, and is white below. It has large furry ears, a prominent black nose and no tail. It spends most of its time in trees and has long, sharp claws, adapted for climbing. Adult males weigh 6 - 12 kg and adult females weigh 5 - 8 kg. ... source: Office of Environment and Heritage

Online resources

Barry Armstead Photography
Images
BowerBird

Records of Koalas on ALA

34,757 records

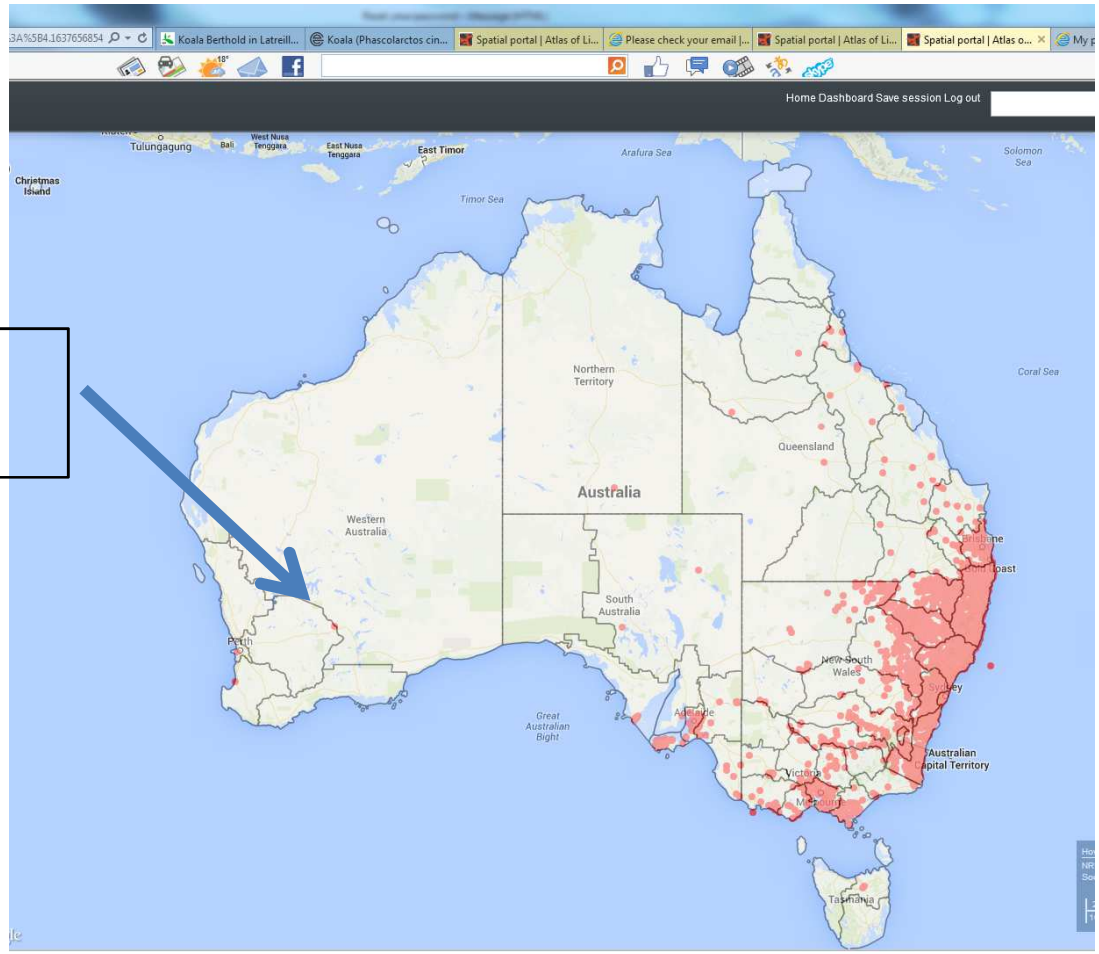


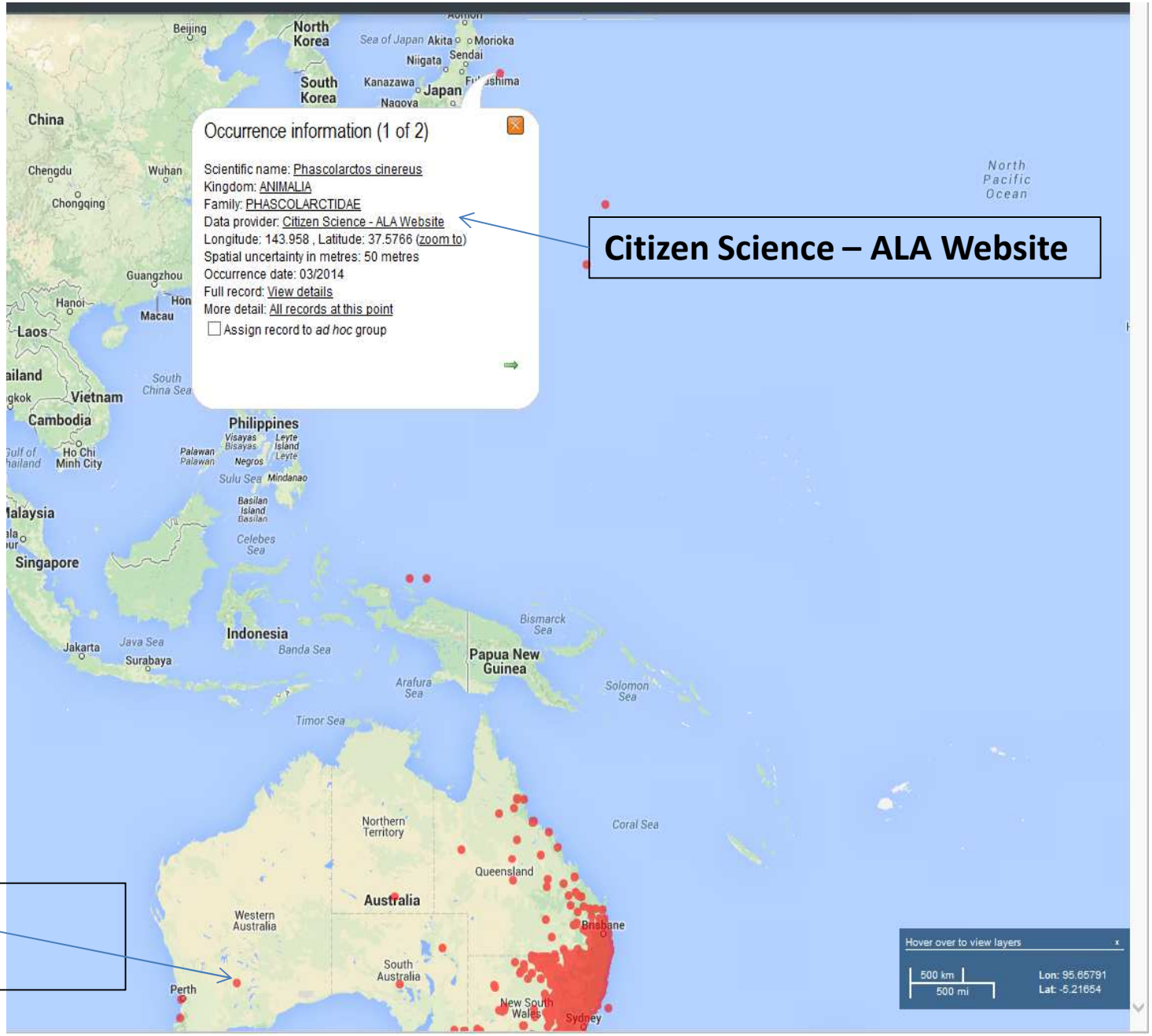
Note low numbers for Queensland

Orange: Human Observation
Purple: Preserved specimens

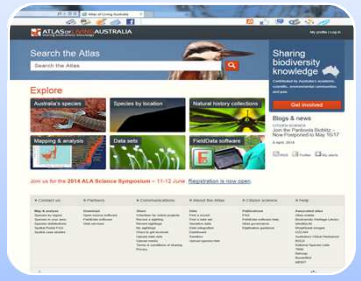
Koala records in ALA with overlay of NRM regions

NRM boundaries



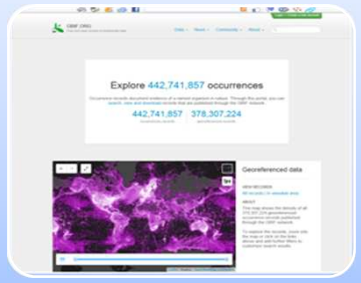

Occurrence information (1 of 2)
Scientific name: Phascolarctos cinereus
Kingdom: ANIMALIA
Family: PHASCOLARCTIDAE
Data provider: Citizen Science - ALA Website
Longitude: 143.958 , Latitude: 37.5766 (zoom to)
Spatial uncertainty in metres: 50 metres
Occurrence date: 03/2014
Full record: View details
More detail: All records at this point
Assign record to ad hoc group
Citizen Science – ALA Website
Agrilus koala
coleopteran
Hover over to view layers
500 km
500 mi
Lon: 95.65791
Lat: -5.21654

Advantages of different initiatives



National Site – ALA – partnership among collections

- Engage with local initiatives – Biodiversity surveys; local collections
- Multiple layers of other kinds of data
- Customised to national and regional needs
- Sense of community



Global aggregator – GBIF – intergovernmental agreement

- Get agreement among countries to share data
- Provide large amounts of primary data
- Frame standards for data sharing
- Provide Global solutions to data sharing

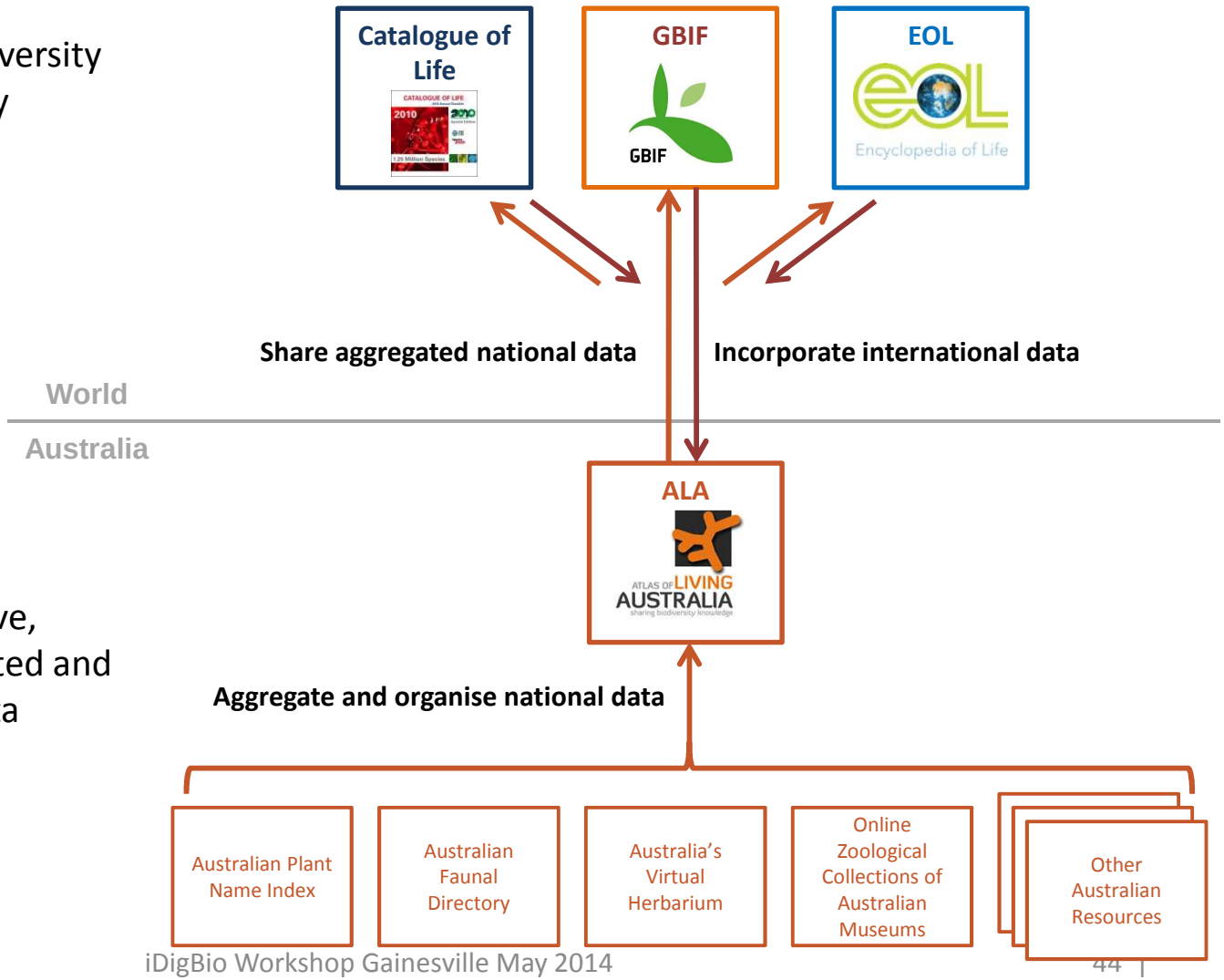
Global aggregator - EoL - partnership of major institutions

- Presents wide range of knowledge about species
- Different kinds of data – images, trait data, detailed biology
- Strong educational aspect
- Different languages and different scripts

National to Global

- GBIF mission
  - To make the world's biodiversity data freely and universally available via the Internet

- ALA Mission
  - To develop an authoritative, freely accessible, distributed and federated biodiversity data management system



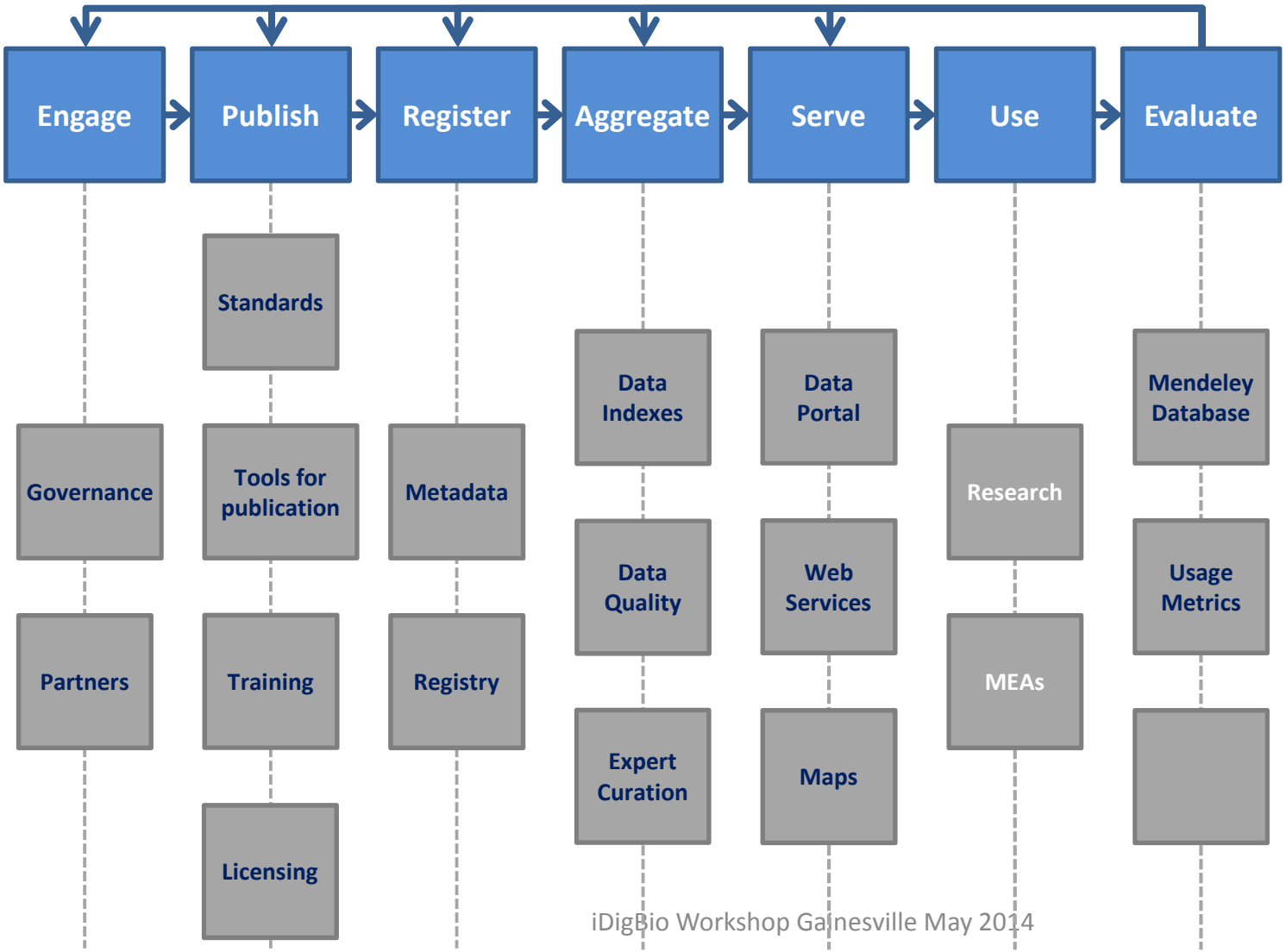
3. Towards 2021

Looking forward

- **At beginning of 21st Century the focus was on global solutions**
- **Global virtual infrastructure must grow/build on national infrastructure** – because the infrastructure is distributed and relies on strong national efforts
- **Virtual infrastructure at different levels (regional, national, global) serves different purposes and communities**
- **Specialisation allows tailoring to different communities while sharing solutions for all**
- **Advanced infrastructure must assist those countries that are less able to muster national effort - developing countries**
- **Infrastructure needs sustainable funding**

Structured Data Pathway

(after D. Hobern, GBIF)



Key Data Sharing Challenges

1. Aggregating and integrating data that were never meant to be aggregated is difficult!

- **Data management principles: data sharing protocols, IPR, 'ownership'…**
- **Data standards: compatibility, integration;**
- **Data types – what is needed? By whom?**
- **Data quality – fitness-for-use? By whom?**
- **Data volumes – how much is enough?**
- **Benefits/incentives for data sharing? Attribution etc;**
- **Data security - Open Access vs treatment of 'Sensitive Data':**
- **Data download and use tracking – who is using data, for what, where are the gaps and needs of different groups?**
- **Analytical tools: web services – what good is access without the means to interrogate?**
- **Who pays for the 'pipes & plumbing' for data to 'flow'?**

Responsibilities for Digital infrastructure - towards 2021

Owners

- Create all new data in digital form
- Digitise and the publish priority legacy data
- Store the data in:
  - » an enduring database structure
  - » an enduring repository
  - » a format that can be used in different devices
- Provide Metadata about the data
- Updating and cleaning the records, or
- Accept annotated versions of their data

Responsibilities for Digital infrastructure - towards 2021

User Responsibility

1. Accept **password access**
2. Say what they are **using** the data for
3. **share** their solutions and cleaned data
4. **acknowledge** their usage

Responsibilities for Digital infrastructure - towards 2021

Data portals

- must continue to provide an **enduring** IT backbone
- be able to **evolve with IT** – relaunched every 5-8 yrs
- Is **user friendly** and well integrated into contemporary IT solutions

Responsibilities for Digital infrastructure - towards 2021

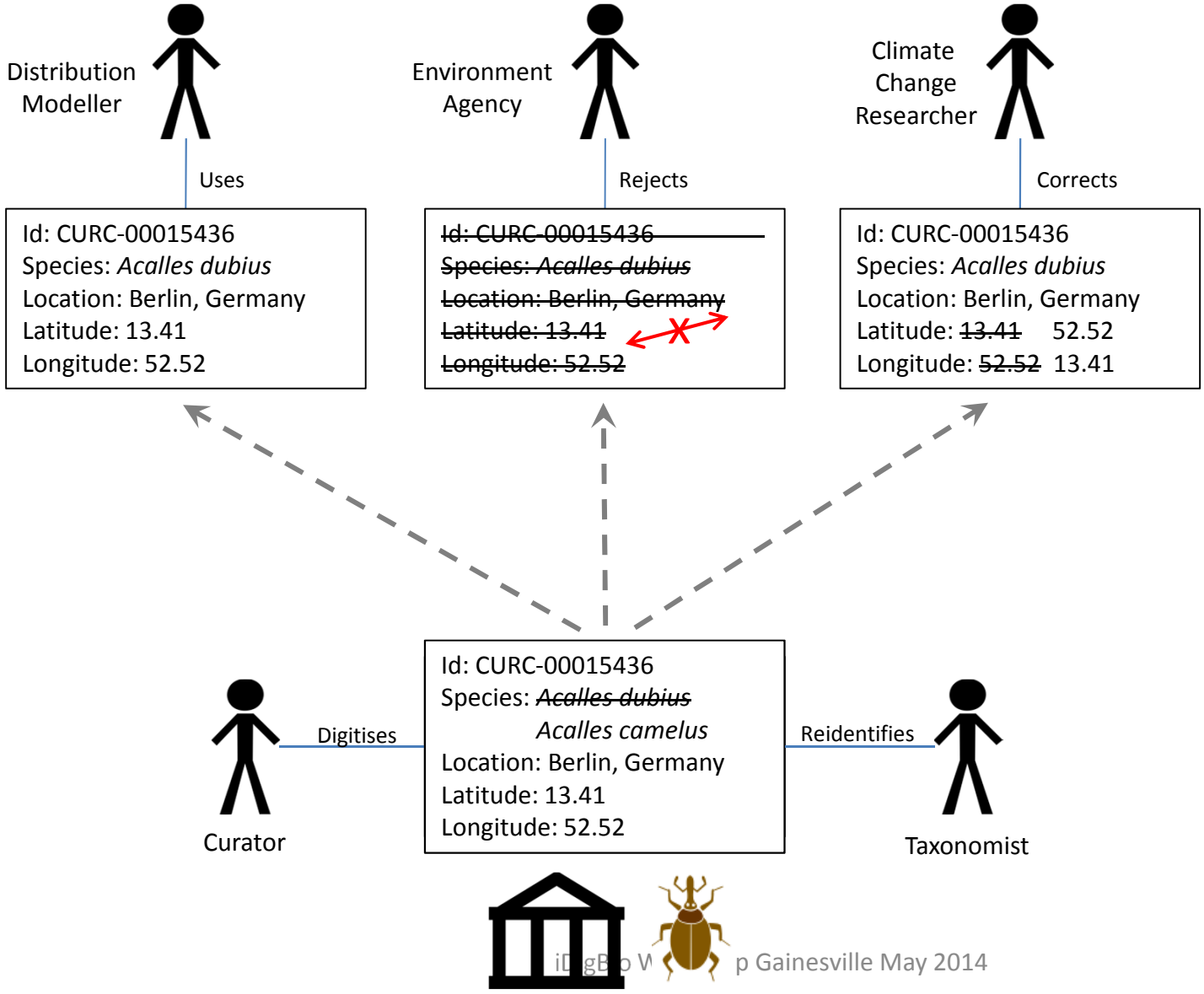
All parties must move to a new modality for data sharing, data quality

...a wiki based approach to improving data

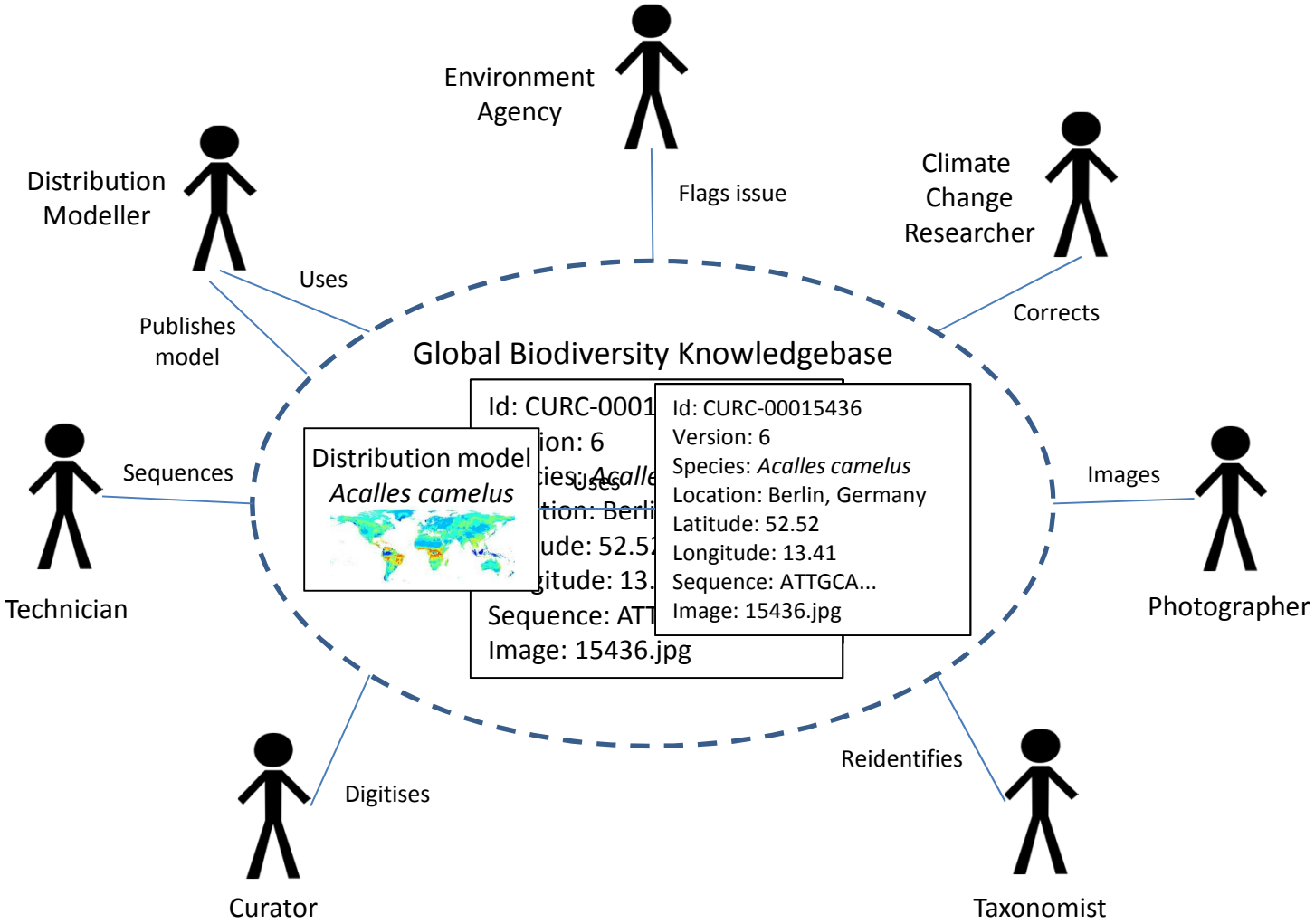
(after Donald Hobern, GBIF)

Weaknesses with distributed data

(from D. Hobern)



Toward a shared knowledgebase
(from D. Hobern)



4. Conclusions

Conclusion

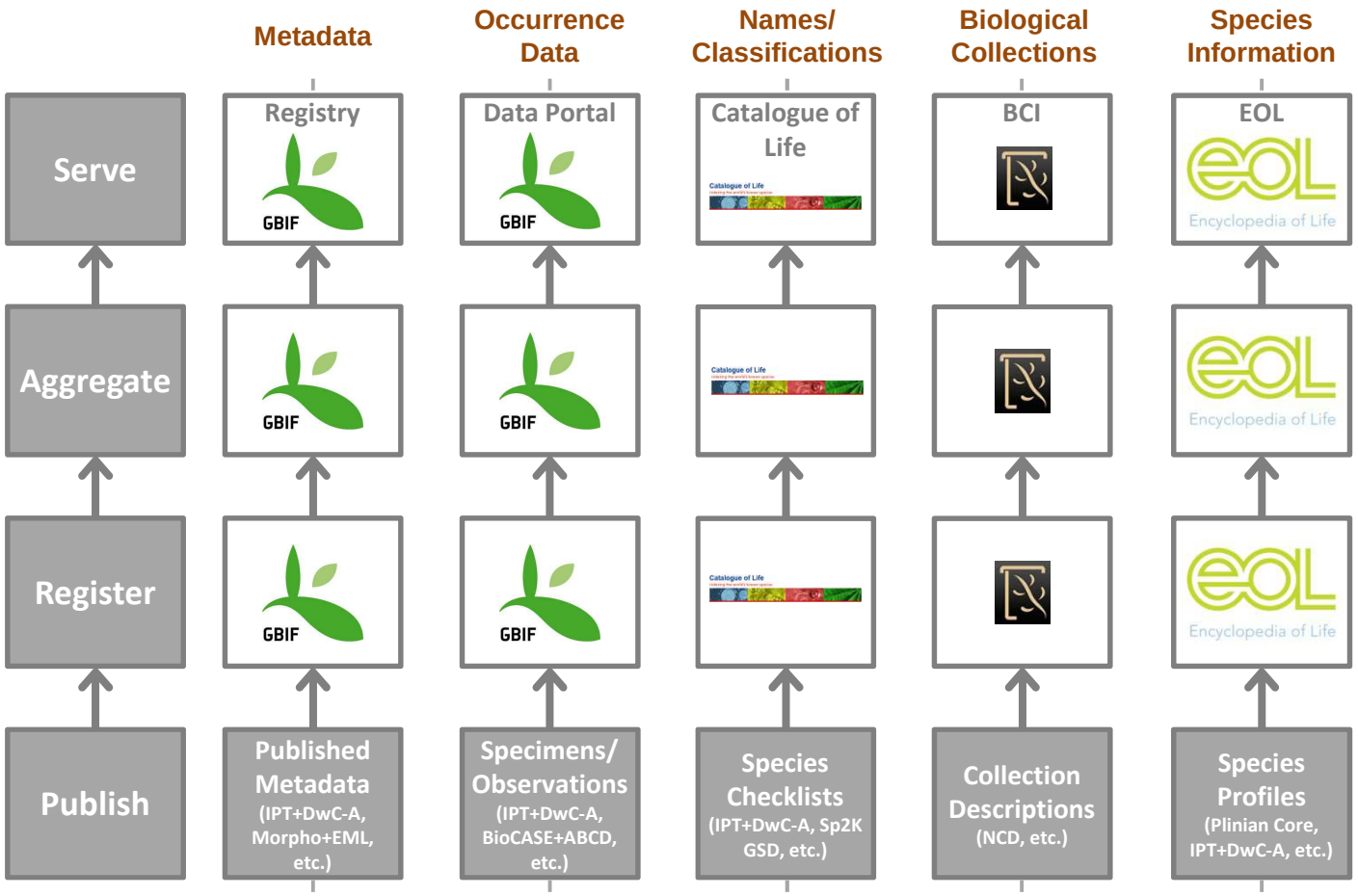
In 20thC natural history collections were seen as stewards of physical specimens

and provided species identification

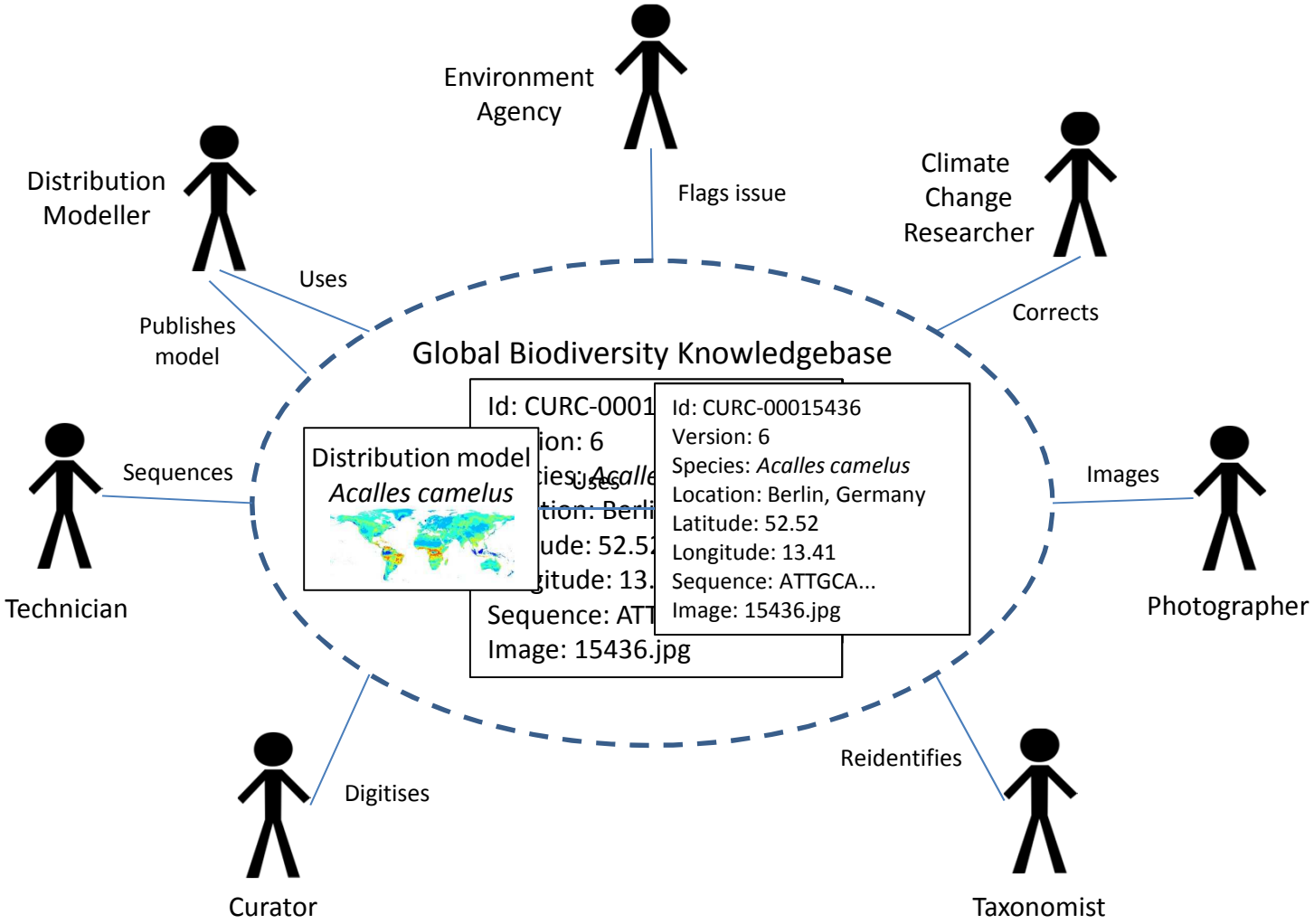
In 21st C they are valued because they provide knowledge in digital form to a wide range of users and for a wide range of applications.

This can only be achieved by aggregating data across collections, across national boundaries and across data types

Global outlook



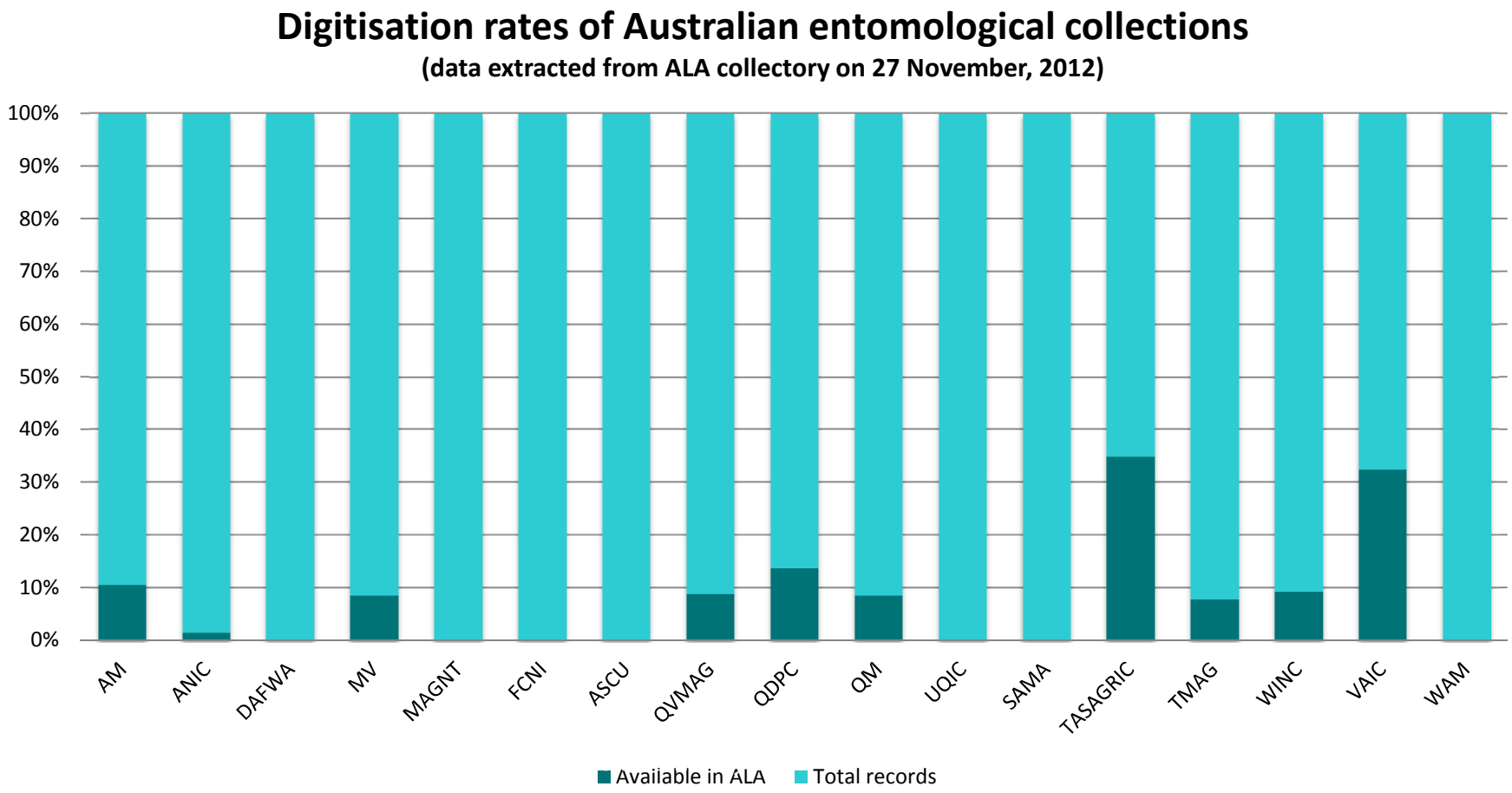
Toward a shared knowledgebase
(from D. Hobern)



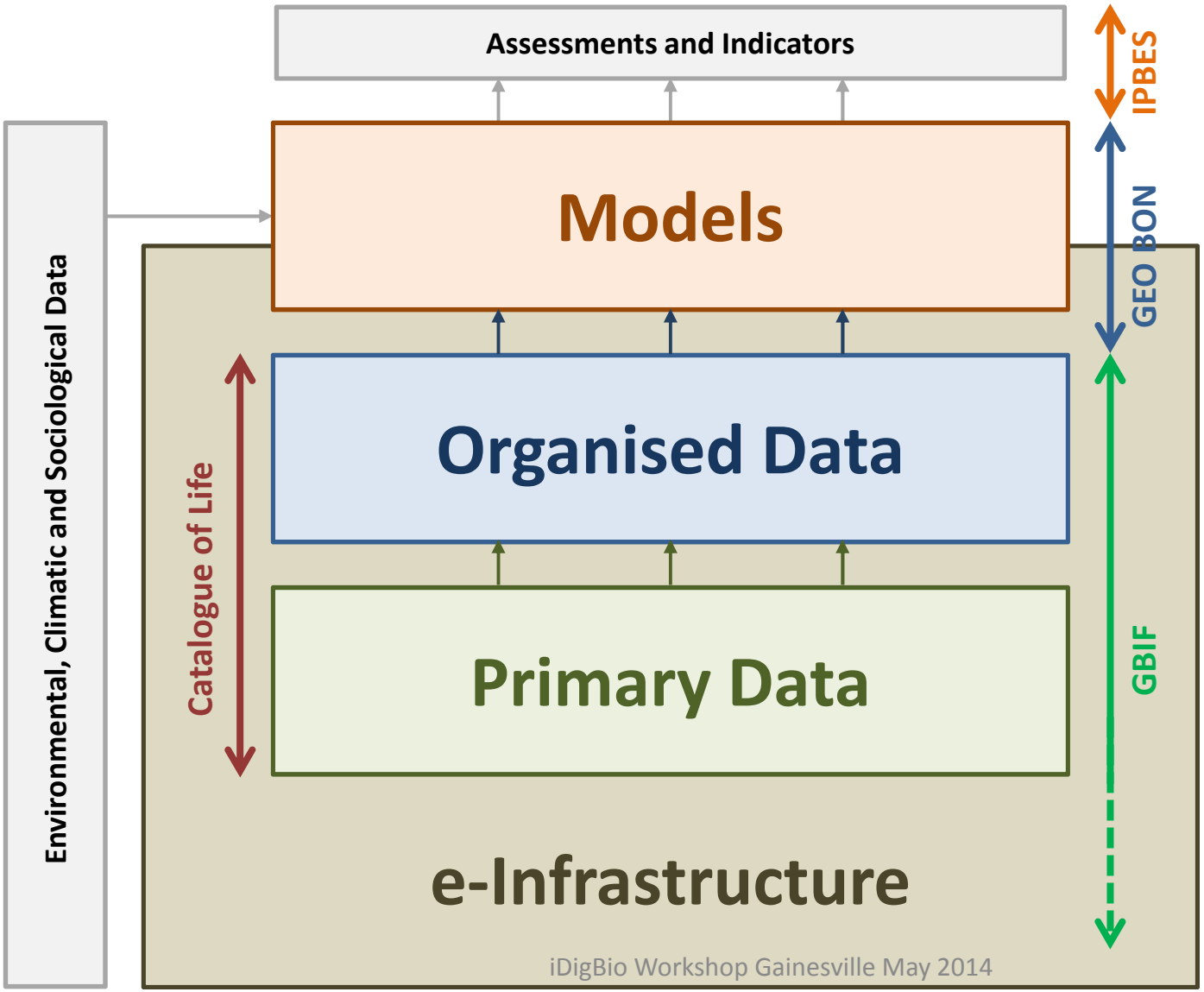
Thank you

Key Data Sharing Challenges

2. Current rates of digitisation



GBIO landscape (from D. Hobern)



Digital infrastructure - towards 2021

Community and Users

- Provide information back to owners about use
- Recognise that data is not clean because this means different things to different people
- Be prepared to clean data and share it
- Build community owned versions of annotated data

Responsibilities for Digital infrastructure - towards 2021

Collections' role

- **Digitally-born data** will be the currency of research but is yet to penetrate beyond fundamental sciences; data volumes are exploding;

  "if it cant be collected digitally, it won't be done/funded"

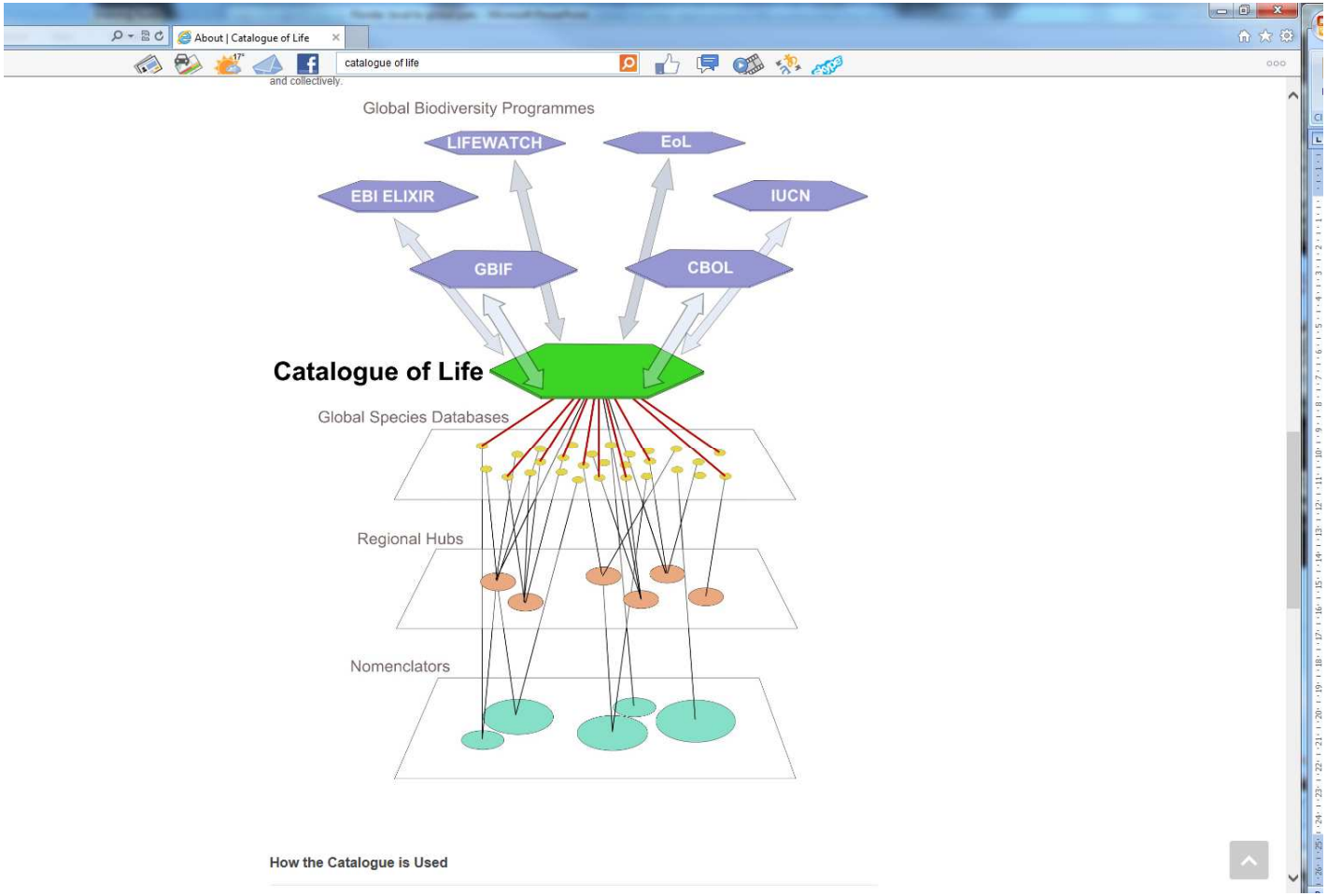
- **Legacy data -** digitise that which is useful

- **Vouchered material needs data rich with images, data etc** and will be a vital link in biodiversity for multi-cellular organisms and will link all levels of biological order

- **Microbial systems** need quite different solutions and vouchered material may not exist and traditonal taxonomy might be irrelevant

- **Training and tools** are essential if this data is to be used (and therefore funded)

- **Look to private sector** for underpinning solutions wherever possible (Wiki, Google, Flickr etc)


About | Catalogue of Life
catalogue of life
and collectively.
Global Biodiversity Programmes
LIFEWATCH
EoL
EBI ELIXIR
IUCN
GBIF
CBOL
Catalogue of Life
Global Species Databases
Regional Hubs
Nomenclators
How the Catalogue is Used