The Australian BioCommons Community Engagement Strategy: Understanding community-scale challenges to inform solution delivery

Tiff Nelson[1], Andrew Lonie[1], Johan Gustafsson[1,2], and Jeff Christiansen[1]

[1]Australian BioCommons, [2]Bioplatforms Australia

Australian BioCommons

Contents

- Introduction to the Australian BioCommons
- Engagement Strategy development
- Our Community Engagement Strategy
- Communities and their status
- Key lessons learned
- Conclusion

Australian BioCommons

The Australian BioCommons Mission

- To **support life science research communities with community scale digital infrastructure** developed and maintained in concert with international peer infrastructures,
- To enable **access to computational services, and**
- To provide **training and support** solutions that enable the rapid and broad based adoption of the above.

View complete Mission here: https://www.biocommons.org.au/about

Australian BioCommons

The Australian BioCommons Principles

- Observe a **national focus** on capabilities and communities,
- **Partner internationally** to participate in and contribute to larger critical mass efforts,
- Build a software and expertise capability that will **reduce duplication of infrastructure management** in Australia
- Promote and build on, **high throughput cloud infrastructure** that is interoperable with international equivalents, using established, well supported software platforms
- **Streamline the exchange of tools, workflows, data and training and expertise** both nationally and internationally

View complete Principles here: https://www.biocommons.org.au/about

Australian BioCommons

Who is Australia's Life Science Research Community?

100,000

Total pool of Australian publicly funded researchers

Who is Australia's Life Science Research Community?

30,000

We estimate 30% are bioscience related

+200,000 students

Australian BioCommons

Who is Australia’s Life Science Research Community?

50%

15,000

30%

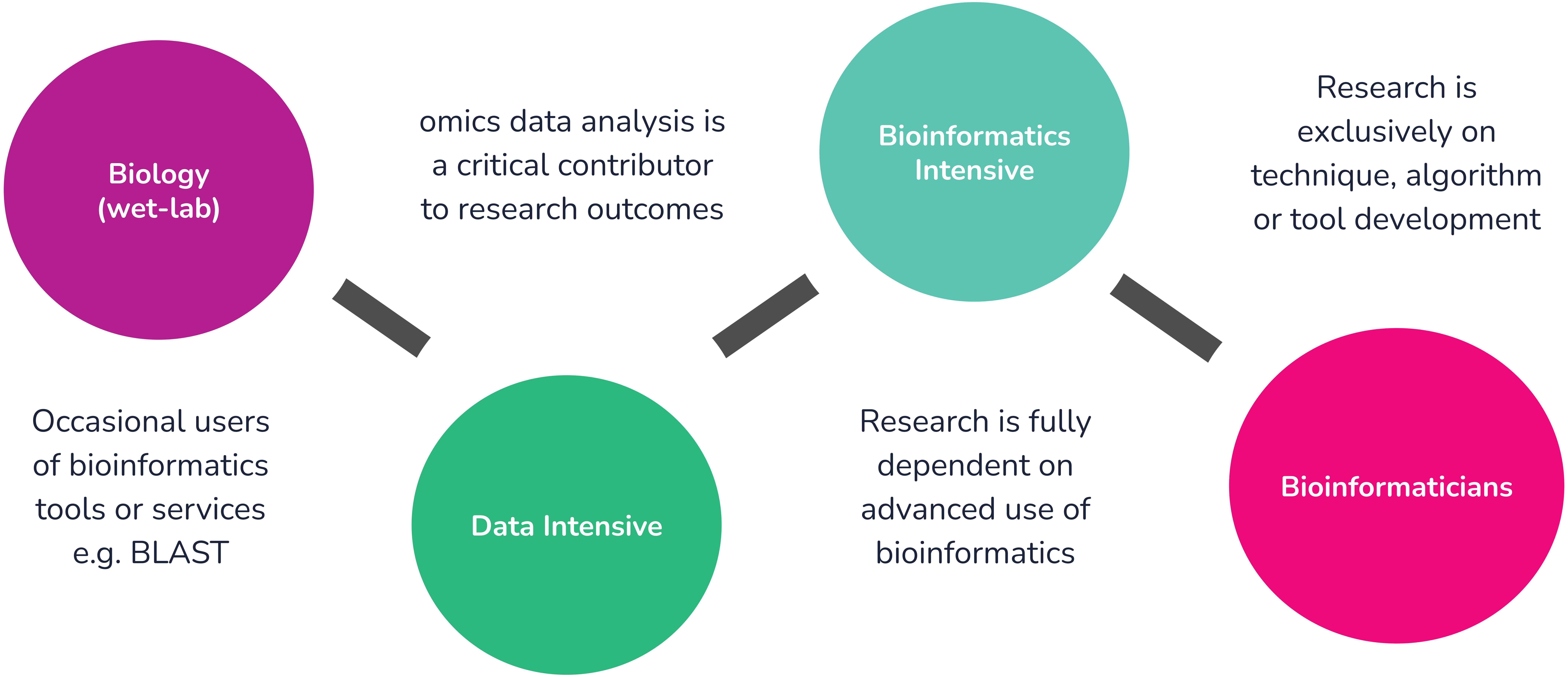
9,000

20%

6,000

What are their skill levels in bioinformatics?

Four broad expertise types



Biology (wet-lab)

Occasional users of bioinformatics tools or services e.g. BLAST

Data Intensive

omics data analysis is a critical contributor to research outcomes

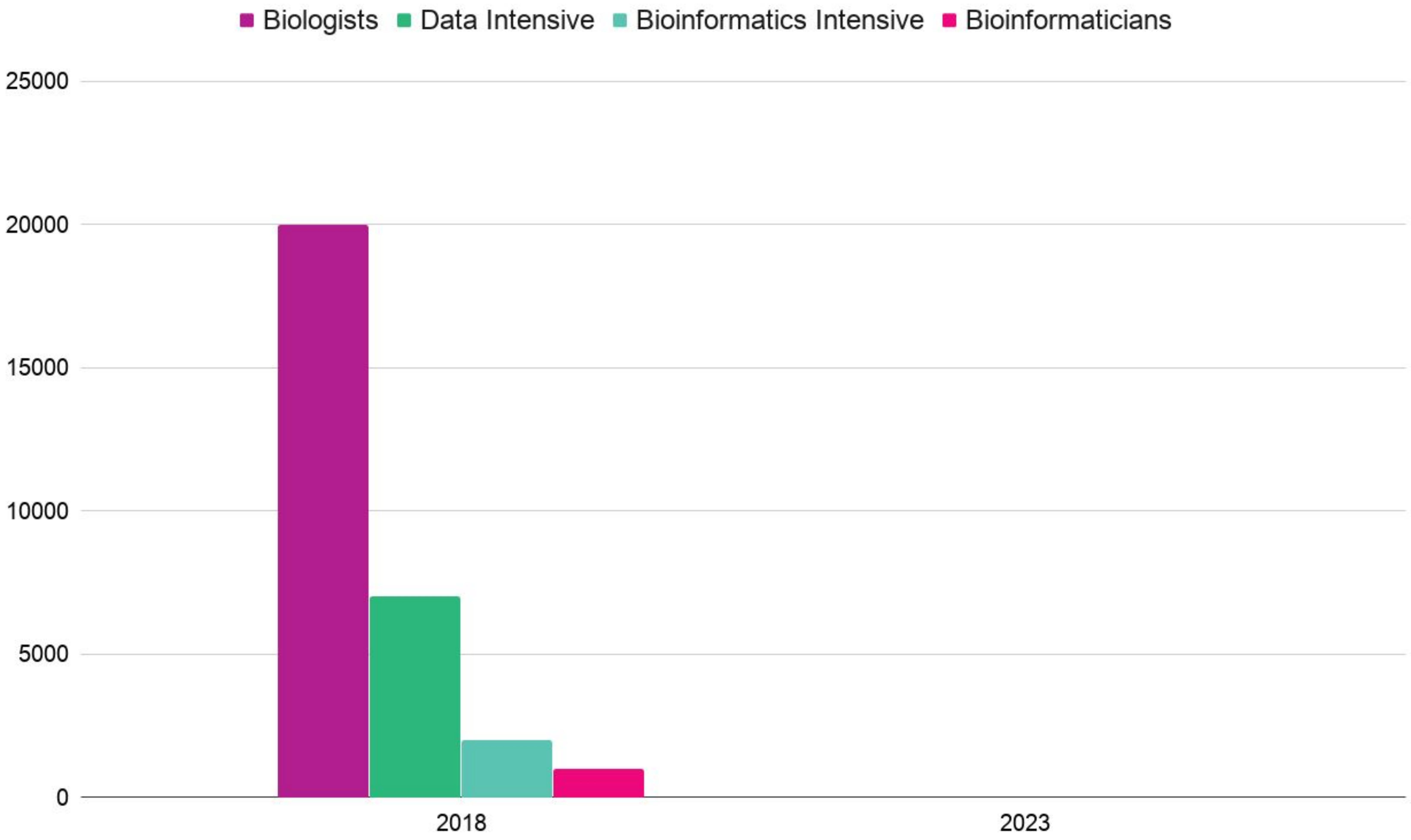
Bioinformatics Intensive

Research is fully dependent on advanced use of bioinformatics

Bioinformaticians

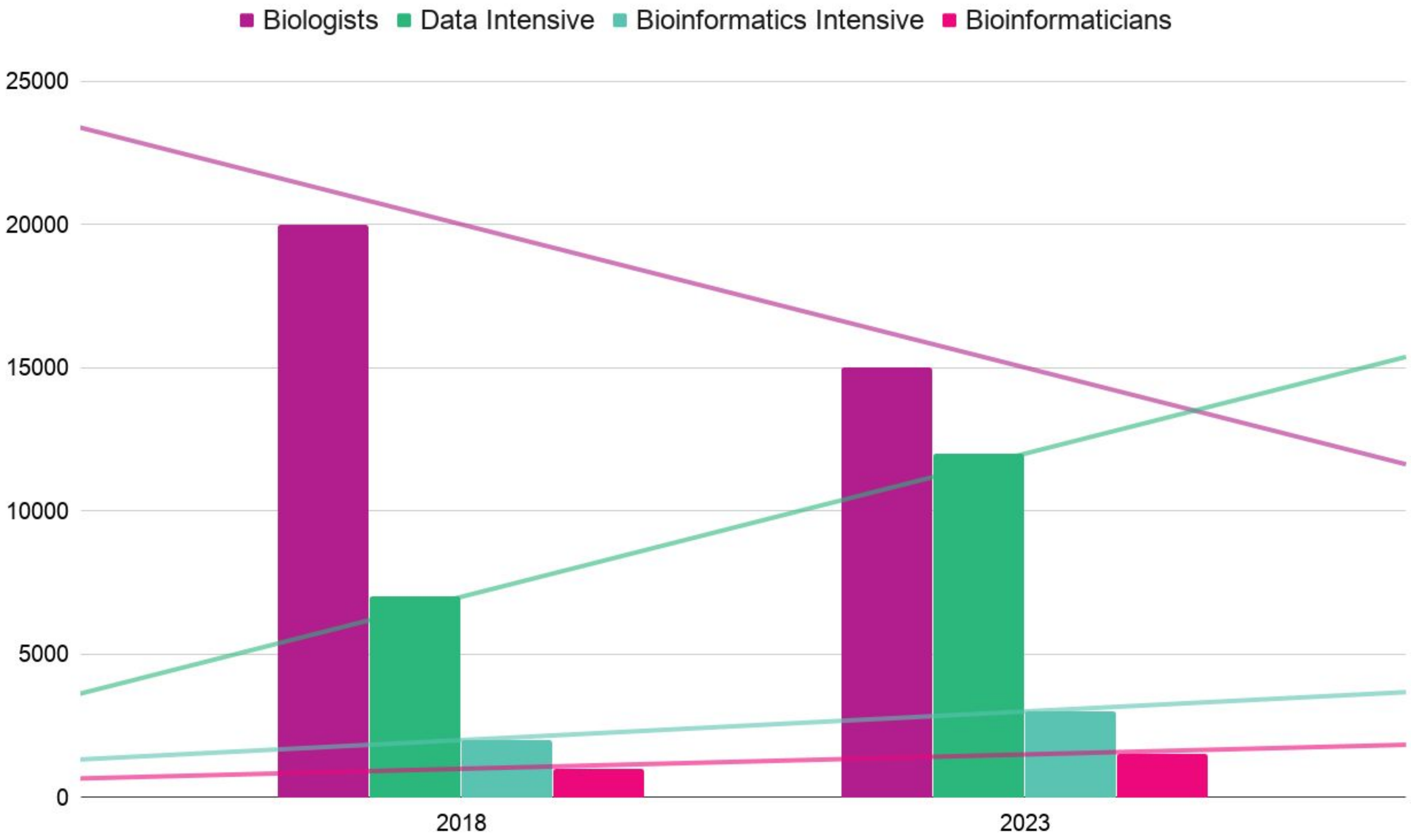
Research is exclusively on technique, algorithm or tool development

How many researchers are in each group?



Australian BioCommons

Expected change over the next 5 years



Australian BioCommons

What does the Australian life science research community require with respect to bioinformatics infrastructure in order to fully participate in, and benefit from an era of data-driven science?

Australian BioCommons

Team

Jeff - 0.4FTE

Primary focus:
Australia-wide community engagements - engagement direction and operations.

Tiff - 0.6FTE

Primary focus:
Australia-wide community engagements - currently around common bioinformatics methods

Johan - 1.0FTE

Primary focus:
Supporting bioinformatics efforts within and across Bioplatforms Australia Framework Data Initiatives (e.g. OMG, GAP, ARG, AM)

Engagement Strategy

01 Identify

02 Research

03 Communicate

04 Document

05 Deploy

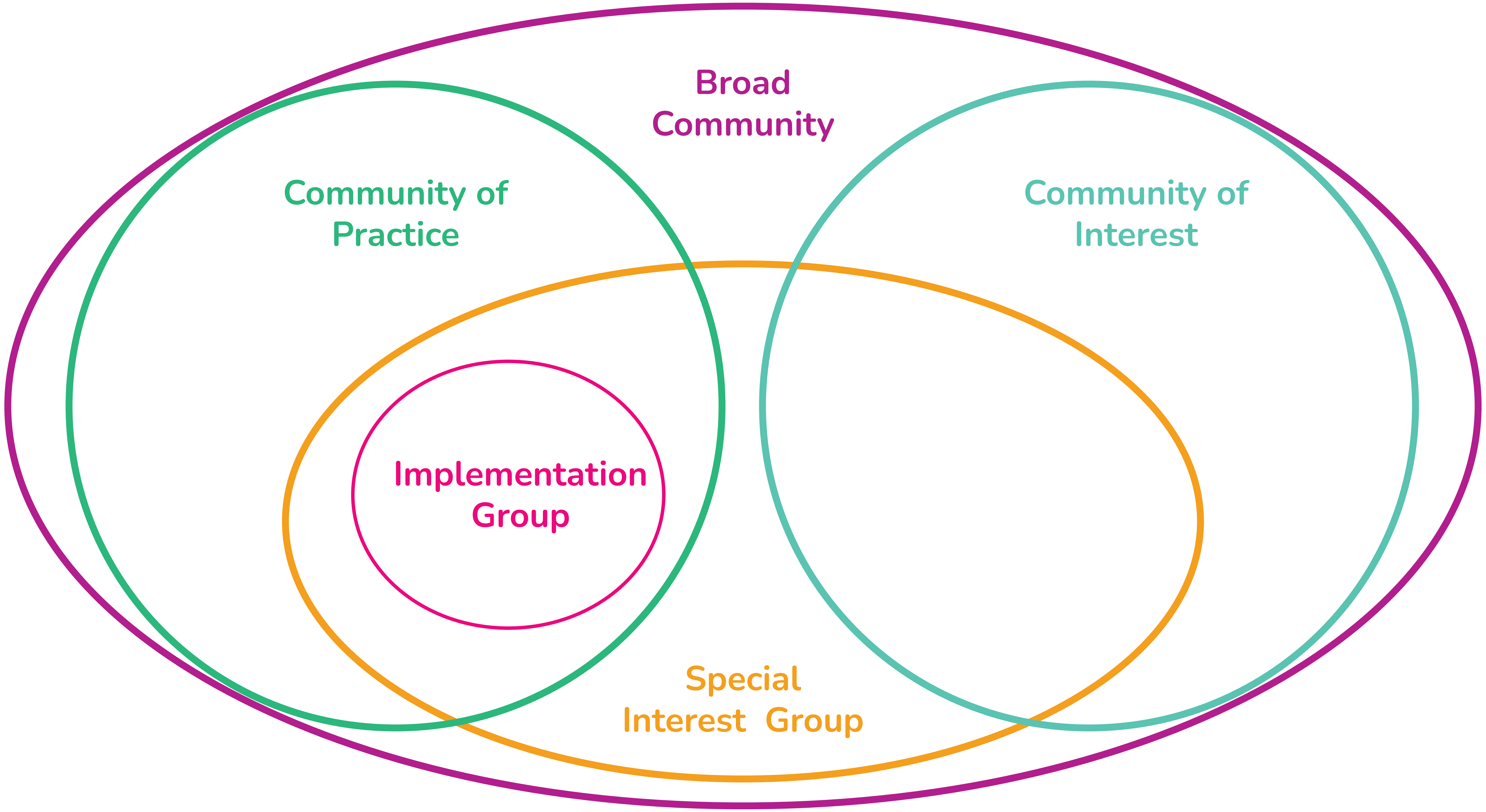
Engagement Strategy: 1/ Identify

Identify meaningful communities of manageable scope around focus areas with known infrastructure challenges, such as genome assembly and annotation of non-model organisms, and microbiome analysis;

metabolomics
microbiome analysis
cancer genomics
genome annotation
genome assembly
proteomics

Australian BioCommons

Engagement Strategy: 1/ Identify



Engagement Strategy: 2/ Research

Research the community topic area to understand broad needs and challenges to engage members

Australian BioCommons

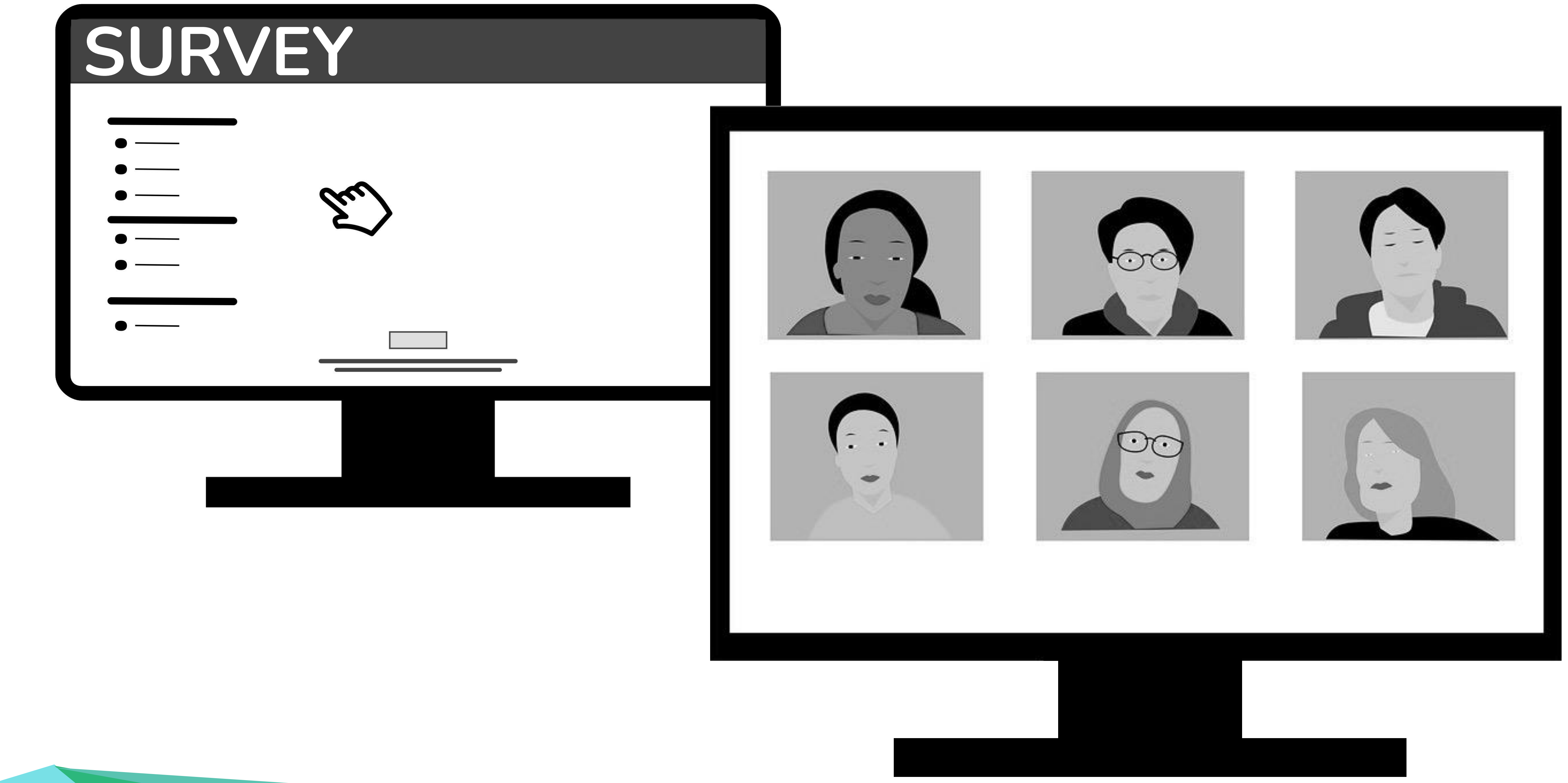
Engagement Strategy: 3/ Communicate

Communicate with the broad community, inclusive of everyone from any expertise level or any institution, to identify issues, roadblocks and solutions / suggestions through electronic surveys, shared discussion boards and virtual meetings.

Australian BioCommons

Connecting Virtually

For geographically dispersed communities with a large number of potential members (i.e. > 50), virtual and electronic methods for knowledge transfer is efficient.



Meeting with the Community

Meeting with community members facilitates relationships and provides the opportunity for:

1. the community to interact and question the goals of the project; and,
2. the engagement team to clarify details and ensure information representing community needs is accurate.

Australian BioCommons

Documentation of Issues

Item Number	User Story Category	Assigned to (Roadmap) or community	Person / Group	BioCommons Community (relevant to this user story)	As a...	I require/would like...	I do not require / would not like...	So / because...	Priority	STATUS	..
#0030	Compute Infrastructure	Genome Annotation	JB	Genome Annotation	Researcher	to have access to appropriately resourced working data storage and compute for genome annotation with common tools and pipelines, e.g. ENSEMBL, iTasses annotation pipelines, Comparative Annotation Toolkit		that I can annotate reference genomes	High - Essential	Assigned	..
#0025	Software & Database Access	Genome Annotation	JB	Genome Annotation	Researcher	to have access to appropriately resourced web accessible working data storage and compute for manual curation and correction for automated genome annotation with common tools and pipelines, e.g. Apollo.		that I can annotate reference genomes	High - Essential	Assigned	..
#0115	Compute Infrastructure	Genome Assembly	AL	Genome Assembly	Researcher	access to between 0.5 to 1.5 TB RAM with 64-128 CPUs and 10TB free disk space		to assemble and annotate non model eukaryotic organisms genomes	High - Essential	Assigned	..
#0119	Compute Infrastructure	Genome Assembly	AL	Genome Assembly	Researcher	access to computational resources that don't require a clear prediction of required time and compute upfront		much of the genome software hasn't been refactored to work in a multi threaded environment and this is problematic, especially with new species, because it is very difficult to predict without first trialling.	High - Essential	Assigned	..
#0135	Compute Infrastructure	Genome Assembly	AL	Genome Assembly	Researcher	alternative access schemes to NCMAS to gain access to computational infrastrucre.		NCMAS caters to larger established groups rather than smaller niche working groups.	High - Essential	Assigned	..
#0152	Compute Infrastructure	Genome Assembly	AK	Genome Assembly	Researcher	to have access to computational capability to run software for 5 days on 30 cores (short)		I can assemble plant genomes completely	High - Essential	Assigned	..
..	..	..	..	..	..	..	..	..	..	..	..

Australian
BioCommons

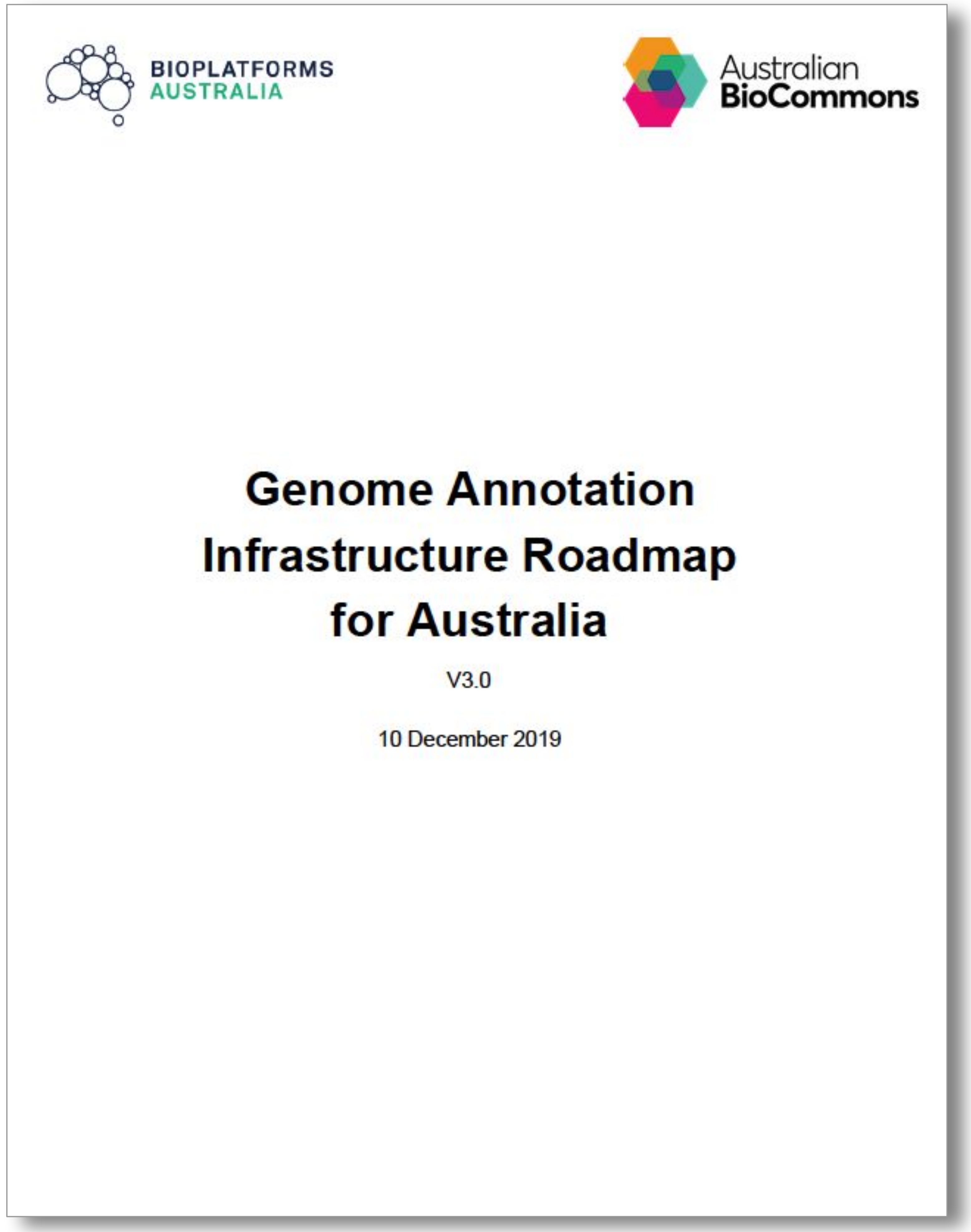
Engagement Strategy: 4/ Document

Document the challenges and detail conceptual solutions (in discussion with infrastructure specialists).

Iterate to achieve endorsement from a subset of practitioners from the community.

Formalise in an Infrastructure Roadmap: a blueprint for solutions that can be deployed to address the community challenges.

Find our completed Roadmaps here: zenodo.org/communities/australianbiocommons/



BIOPLATFORMS AUSTRALIA

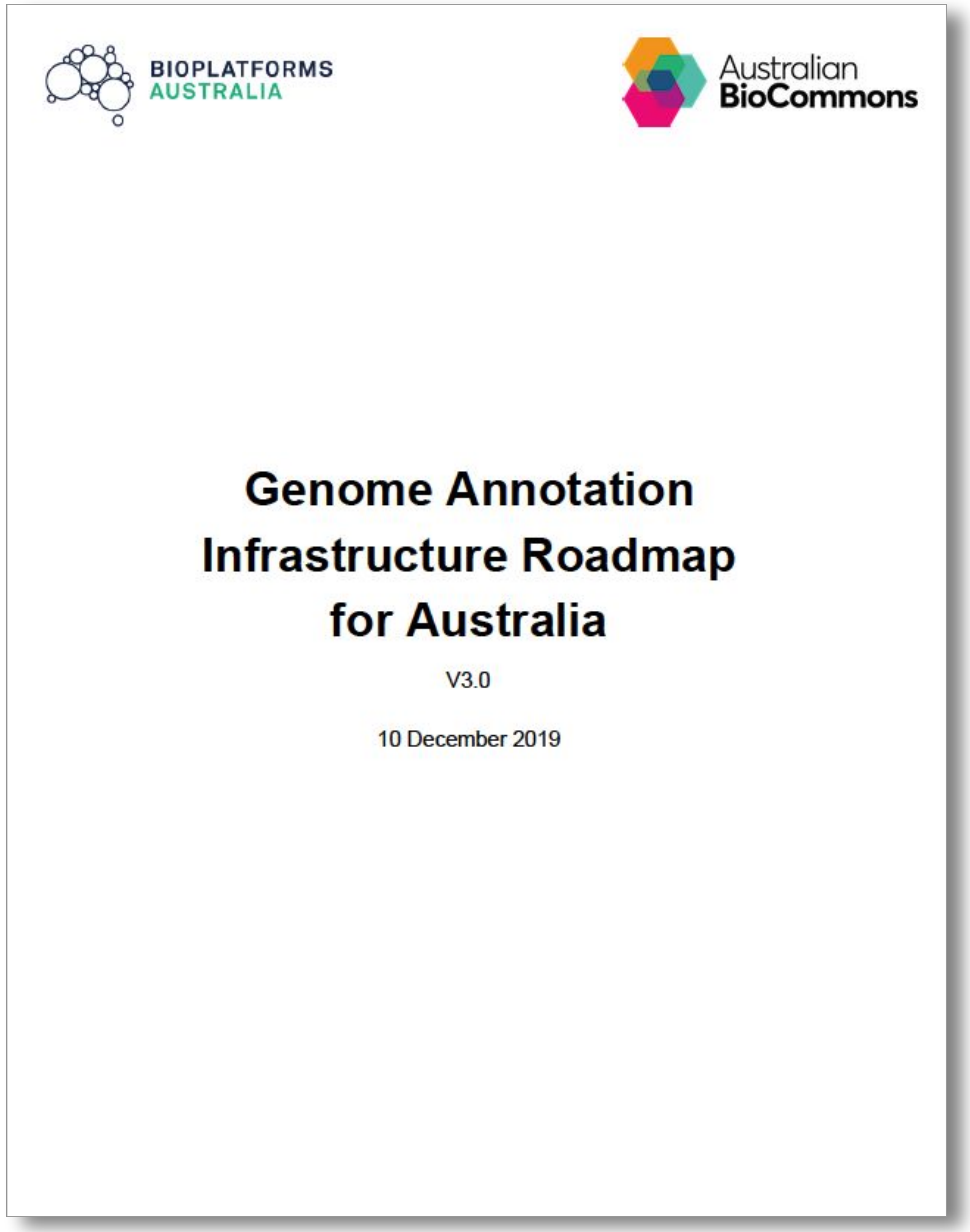
Australian BioCommons

Genome Annotation Infrastructure Roadmap for Australia

V3.0

10 December 2019

Documenting a Roadmap

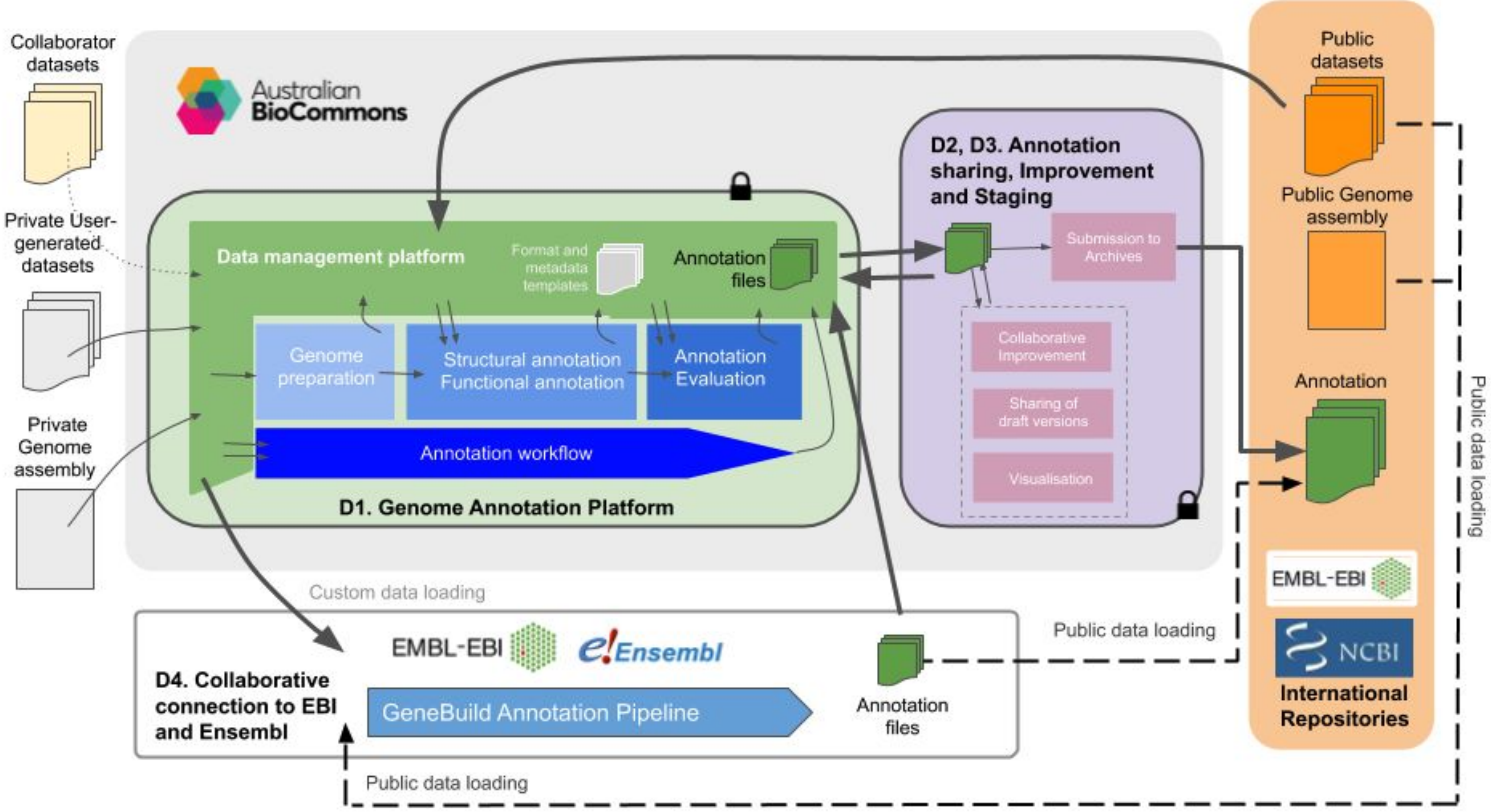

BIOPLATFORMS AUSTRALIA

Australian BioCommons

Genome Annotation Infrastructure Roadmap for Australia

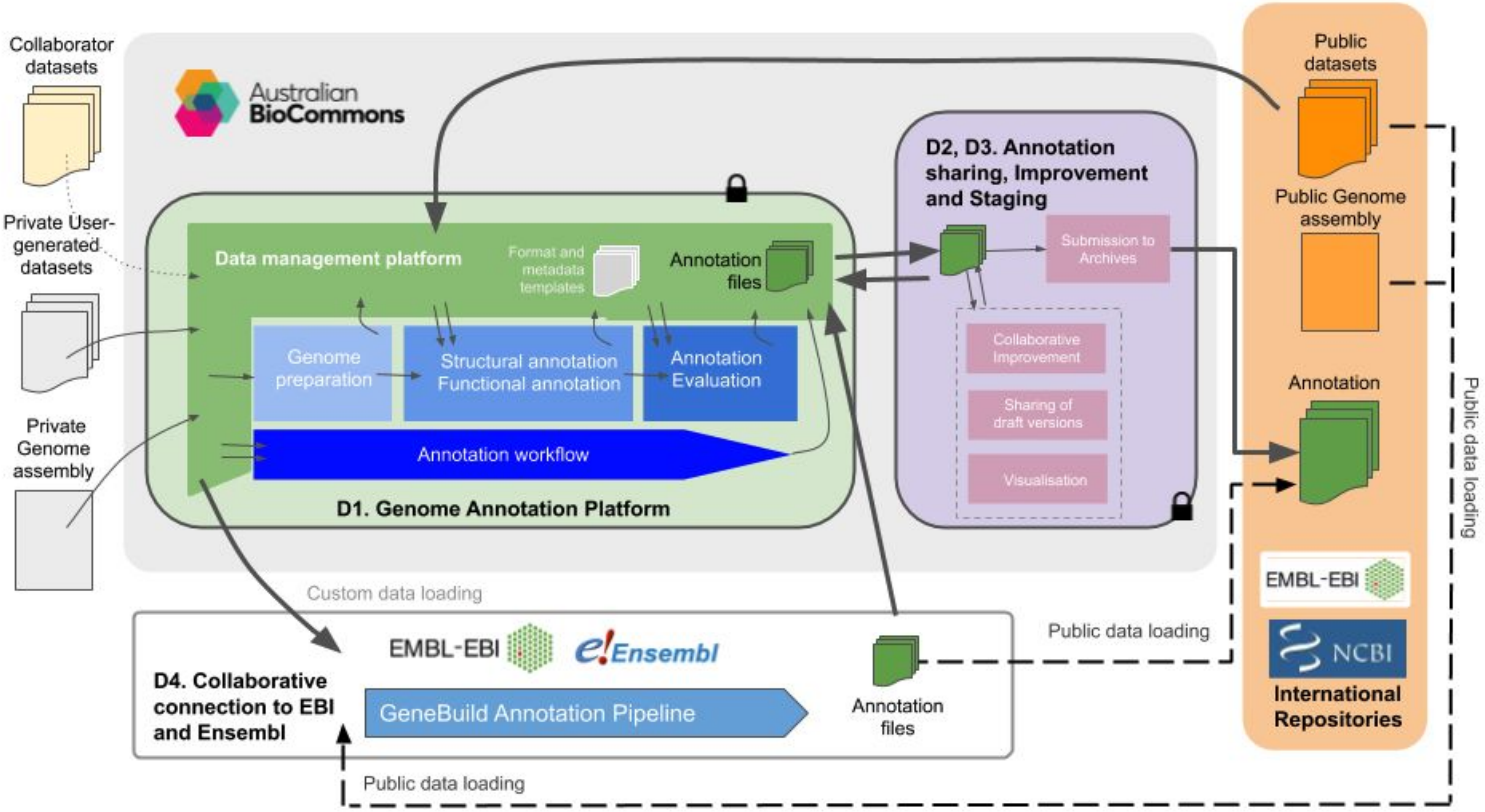
V3.0

10 December 2019



Australian BioCommons

Documenting a Roadmap



Collaborator datasets

Private User-generated datasets

Private Genome assembly

Australian BioCommons

Data management platform

Format and metadata templates

Annotation files

Genome preparation

Structural annotation
Functional annotation

Annotation Evaluation

Annotation workflow

D1. Genome Annotation Platform

D2, D3. Annotation sharing, Improvement and Staging

Submission to Archives

Collaborative Improvement

Sharing of draft versions

Visualisation

Public datasets

Public Genome assembly

Annotation

EMBL-EBI

NCBI

International Repositories

Public data loading

Custom data loading

D4. Collaborative connection to EBI and Ensembl

EMBL-EBI

e!Ensembl

GeneBuild Annotation Pipeline

Annotation files

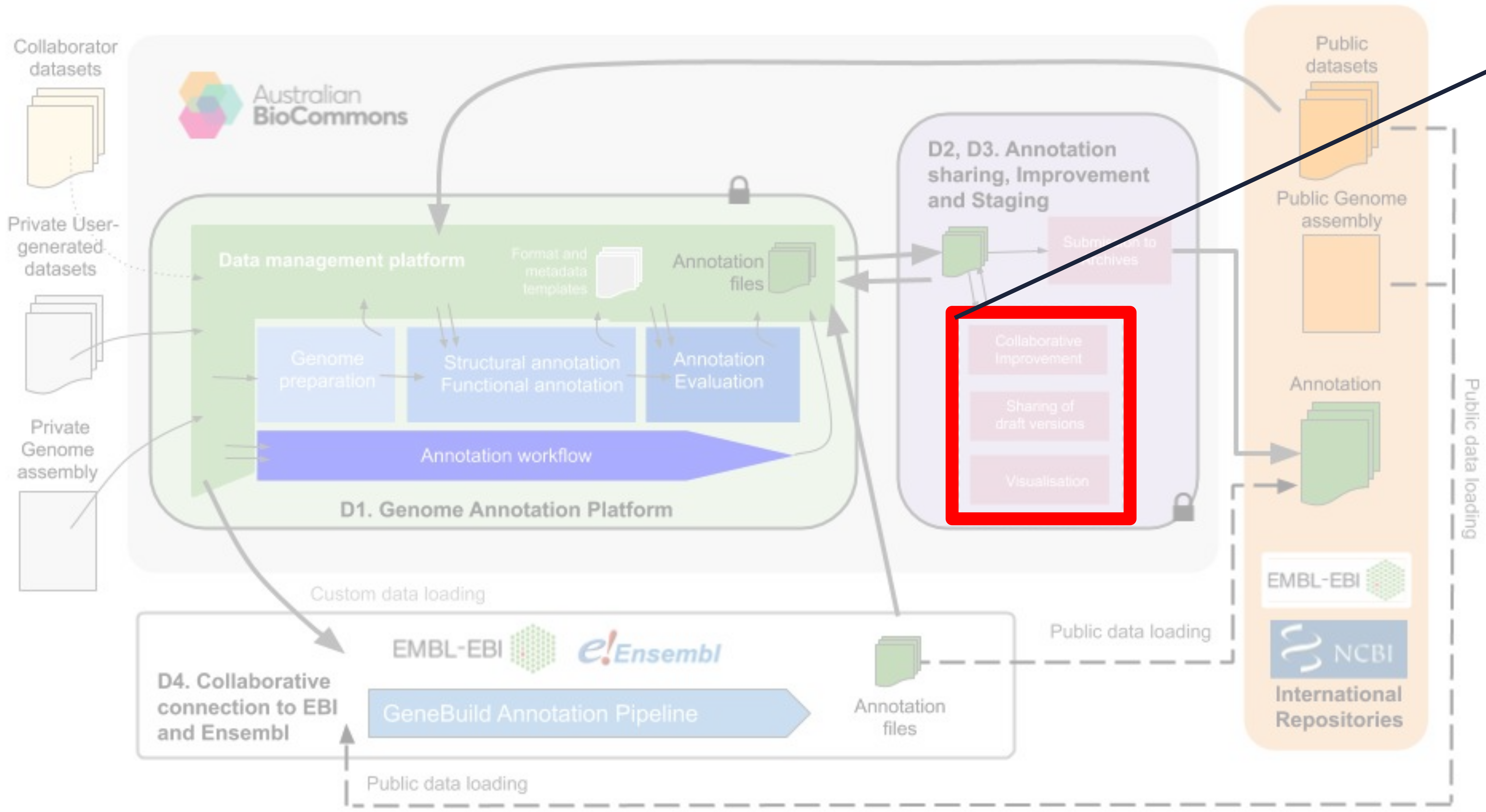
Public data loading

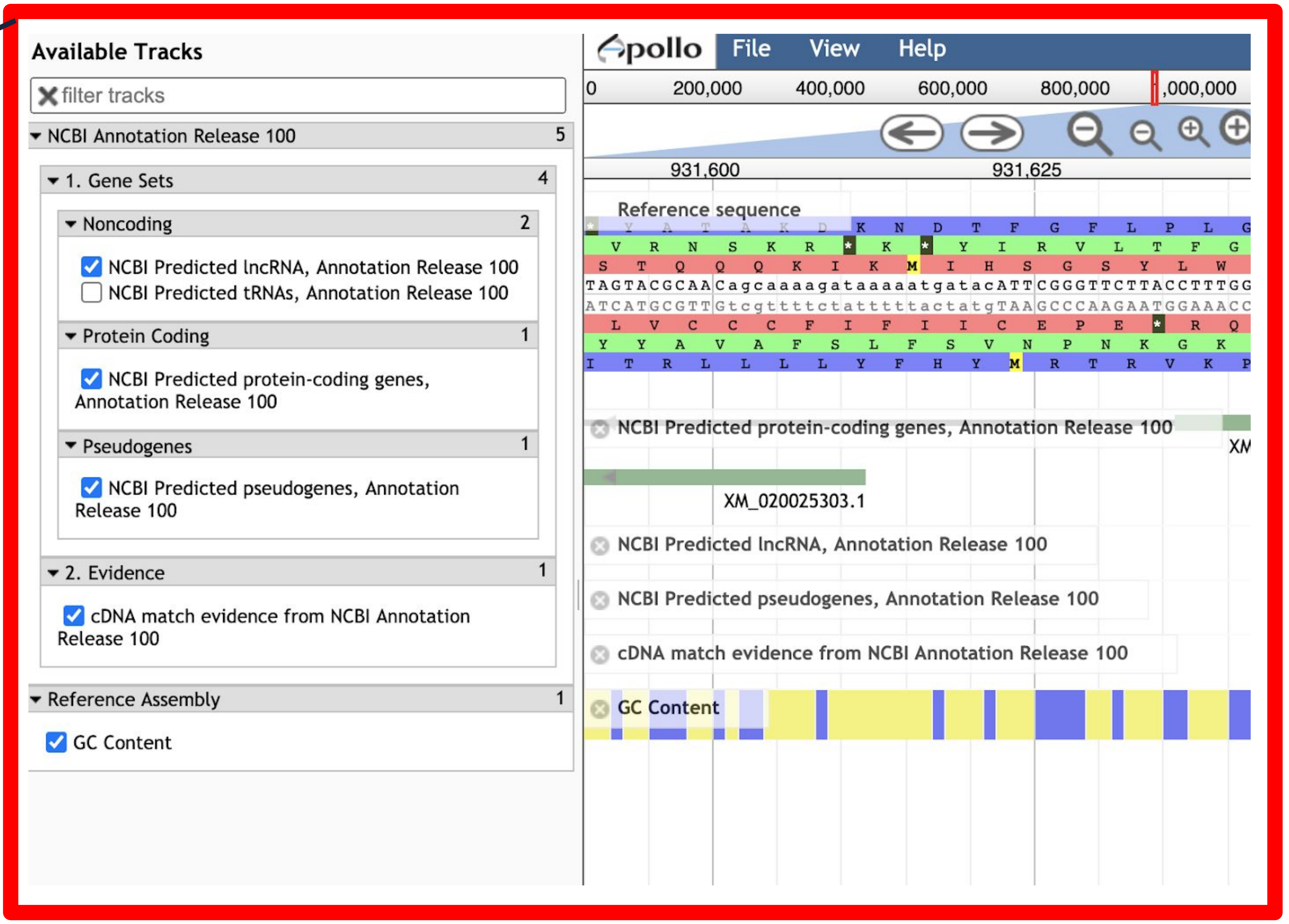
Public data loading

Engagement Strategy: 5/ Deploy

Deploy and implement solutions with our infrastructure partners, with testing and feedback from the community.

Example Deployment





Available Tracks

filter tracks

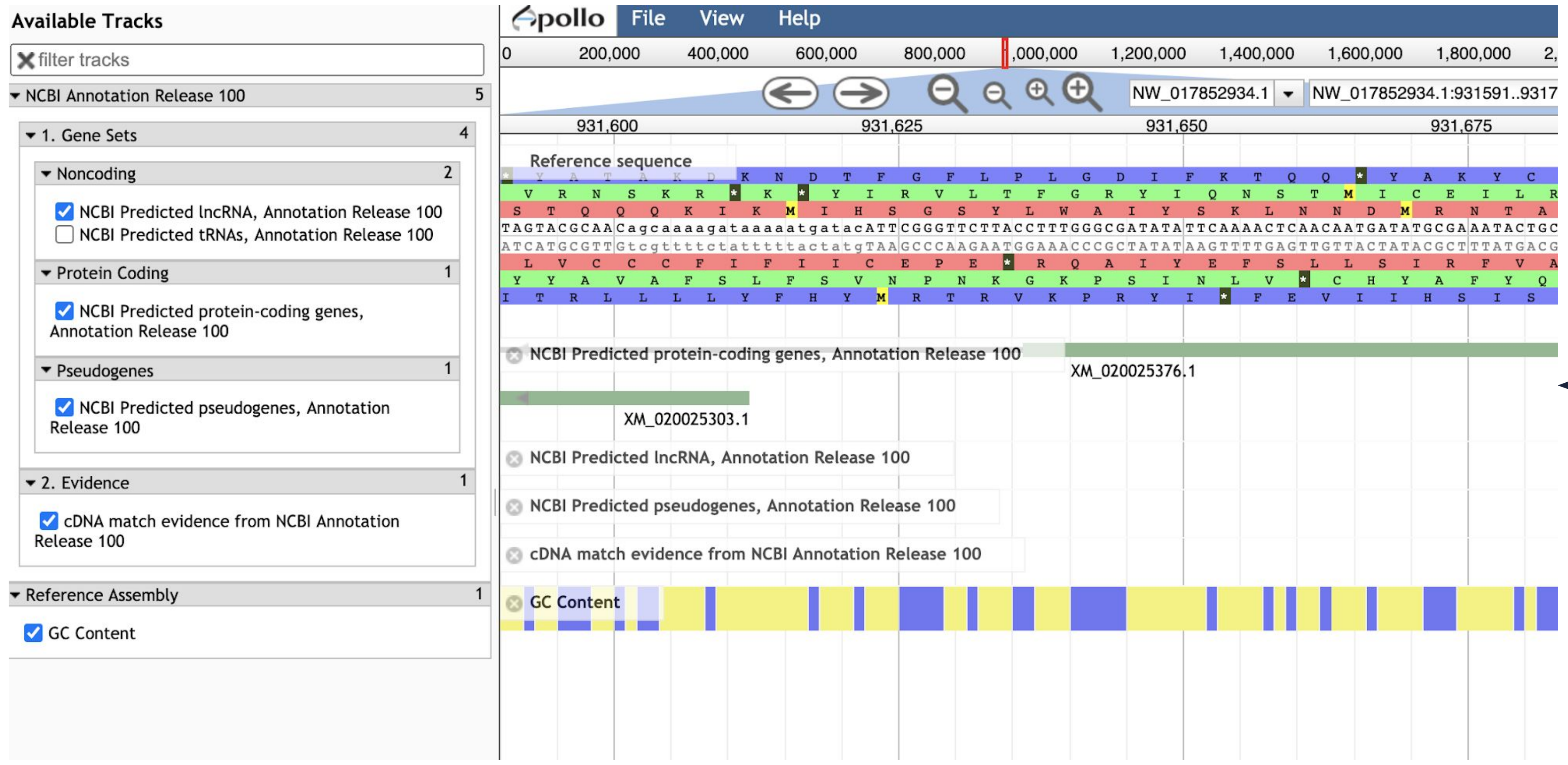
- NCBI Annotation Release 100 5
  - 1. Gene Sets 4
    - Noncoding 2
      - ☑ NCBI Predicted lncRNA, Annotation Release 100
      - ☐ NCBI Predicted tRNAs, Annotation Release 100
    - Protein Coding 1
      - ☑ NCBI Predicted protein-coding genes, Annotation Release 100
    - Pseudogenes 1
      - ☑ NCBI Predicted pseudogenes, Annotation Release 100
  - 2. Evidence 1
    - ☑ cDNA match evidence from NCBI Annotation Release 100
- Reference Assembly 1
  - ☑ GC Content

Apollo for manual curation and sharing of genome annotations

Example Deployment: Apollo

- Apollo for manual curation and sharing of genome annotations

genome annotation



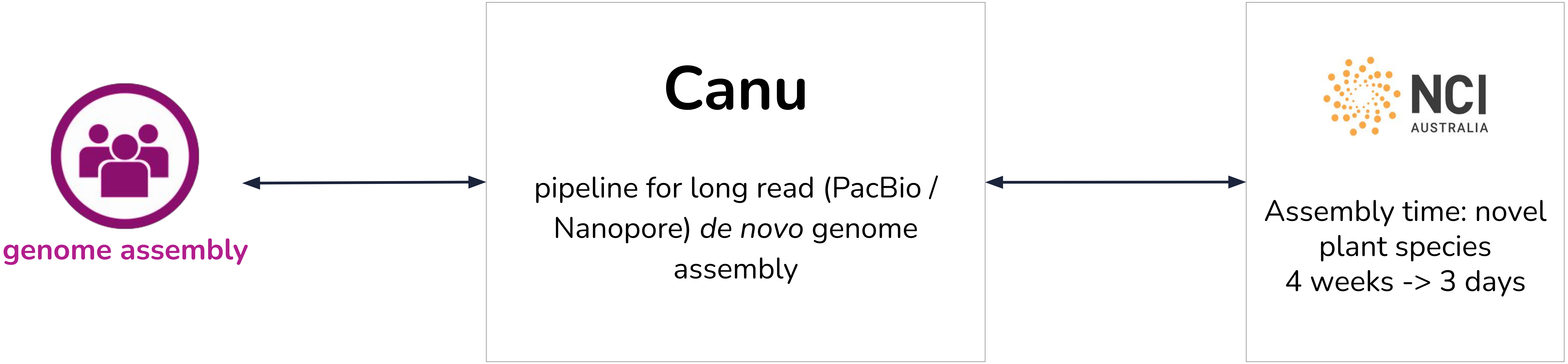
QCIF

PAWSEY supercomputing centre

Australian BioCommons

Example Deployment: Workflow Optimisation

- Workflow optimisation on computational infrastructure for *de novo* genome assembly.



genome assembly

Canu

pipeline for long read (PacBio / Nanopore) *de novo* genome assembly

Assembly time: novel plant species
4 weeks -> 3 days

Now uses a wrapper for Canu that allows job distribution across multiple nodes on Gadi HPC

Communities and their Status

	Lead	Conceptualise Community Group	Develop Survey	Invite participation in SIG	Group Meeting(s)	Draft Infrastructure Roadmap	Finalise Infrastructure Roadmap	Deployment activities	Implementation Group	Communicate deployments to Community
Genome annotation	Tiff	y	y	y	y	y	y	CLI (Fgenesh++ @Pawsey)	LS, PB	
								GUI (Apollo)	LS, TW, AL	
Genome assembly	Tiff	y	y	y	y	y	y	CLI (@NCI & Pawsey))	GP	
Microbiome analysis	Tiff	y	y	y	y	y				
Proteomics	Johan	y	y	y	y					
Metabolomics	Johan	y	y	y	y					
								GUI (Galaxy)	TV	
Phylogenomics / Comparative Genomics	Tiff	y								
								GUI (Galaxy)	JB	
others - TBD										

Australian BioCommons

Key Lessons Learned

- **Clarify project goals**. When engaging it is important to have clear messages about the goals and scope of the project, so as to not unintentionally overpromise.
- **Use a diversity of voices**. During the course of researching a community topic or focus point, get to know the expert practitioners and engage with them to discuss the broad challenges and invite them to give their perspective during meetings and activities.
- **Know your members**. Be clear and upfront about who can participate and at which stage; a community member with an ‘interest’ but no practical skills in running a methodology may provide valuable information on community needs but will not be the best resource for advising on roadblocks or solutions.
- **Be transparent**. Make it easy to join communities; communicate how to be involved and provide accessible documents and records for new members to find.
- **Document with detail and context**. Deployment and implementation teams need quality information to design appropriate solutions; the documentation of this information in various forms, e.g. Roadmap, spreadsheet of ‘user stories’ including questions and comments, detailed minutes and video recordings for reference, facilitates the efficient development of solutions.

Conclusions

- Through this engagement process, the Australian BioCommons has identified and then coordinated work to deploy essential infrastructure that was previously lacking to support critical communities (e.g. those undertaking genome annotation).

- Successful outcomes are measured by positive responses from the community (e.g. turning up in large numbers, actively joining the discussion), active use by early adopters, and uptake of deployed services.

Thanks!

Email me at: **tiff@biocommons.org.au**

View our webpage at: **biocommons.org.au**

Tweet us at: **@AusBiocommons**

View our Roadmaps, Communication Strategy and more at: **zenodo.org/communities/australianbiocommons/**

Australian
BioCommons