

The logo for EMBL Australia Bioinformatics Resource is located in the top left corner. It features the text "EMBL" in a large, bold, sans-serif font, with "Australia Bioinformatics Resource" in a smaller font below it. To the right of the text is a circular emblem composed of green dots, with one red dot in the center. The background of the slide is a light grey with a network of green and orange nodes and lines, and a faint grid pattern.

EMBL

Australia  
Bioinformatics Resource

# EMBL Australia Bioinformatics Resource (EMBL-ABR)

*Network Building Strategy & Scope*  
20<sup>th</sup> June 2016

Andrew Lonie, EMBL-ABR Director  
Vicky Schneider, EMBL-ABR Deputy Director

## Table of Contents

<b>Introduction</b> .....	<b>3</b>
<b>Mission</b> .....	<b>3</b>
<b>Vision</b> .....	<b>3</b>
<b>Objectives</b> .....	<b>3</b>
<b>How</b> .....	<b>3</b>
<b>EMBL-ABR modus operandi</b> .....	<b>4</b>
EMBL-ABR HUB.....	4
EMBL-ABR nodes .....	5
<b>Communication</b> .....	<b>5</b>
Annual EMBL-ABR All Hands meeting.....	5
EMBL-ABR Domain driven workshops.....	5
Dissemination.....	5
Key Documents.....	6
<b>Key Events 2016</b> .....	<b>7</b>
<b>Organisational structure and personnel</b> .....	<b>8</b>
<b>EMBL-ABR Structure</b> .....	<b>8</b>
<b>EMBL-ABR Hub Team</b> .....	<b>8</b>
<b>Key Areas</b> .....	<b>9</b>
<b>Data</b> .....	<b>9</b>
EMBL-ABR Annotation Activities .....	9
<b>Tools</b> .....	<b>10</b>
<b>Training</b> .....	<b>10</b>
<b>Platforms</b> .....	<b>11</b>
Genomics Virtual Lab.....	11
<b>Compute</b> .....	<b>11</b>
<b>International</b> .....	<b>11</b>
Activities .....	11
<b>Appendix 1: Node Description Template</b> .....	<b>12</b>
<b>Appendix 2: EMBL-ABR EoI Template</b> .....	<b>12</b>

## Introduction

EMBL-ABR (Australian Bioinformatics Resource) is a distributed national research infrastructure set up with support from the University of Melbourne and BPA. The EMBL-ABR Hub is hosted by The University of Melbourne and co-located with the VLSCI node.

This Strategic Plan formulates the long-term general goals for EMBL-ABR, the model for the process of network building and defines the EMBL-ABR Hub current and future activities. The current version of the Plan addresses specific deliverables for 2016 and 2017. The Plan will be revised annually, considering input from the International Scientific Advisory Board, the Heads of the EMBL-ABR nodes and the Community Lead Group. It describes the major activities within EMBL-ABR Key Areas of Data, Tools, Platforms, Compute, Training and International.

## Mission

To provide national, coordinated bioinformatics infrastructure and support for life science and medical researchers in Australia.

## Vision

Our vision is that EMBL-ABR should be a sustainable, truly national infrastructure for bioinformatics, supporting life science and medical research in Australia. EMBL-ABR is the liaison towards key international partnerships including ELIXIR, CyVerse and BD2K. In facilitating high-quality, best-practice data driven science with the necessary infrastructure such as competence provision in access to data and analysis methods, EMBL-ABR will be of fundamental importance. EMBL-ABR will coordinate bioinformatics support within Australia and make bioinformatics easily accessible for Australian researchers as well as bringing international recognition to Australian research, data, bioinformatics tools and platforms.

## Objectives

EMBL-ABR aims to:

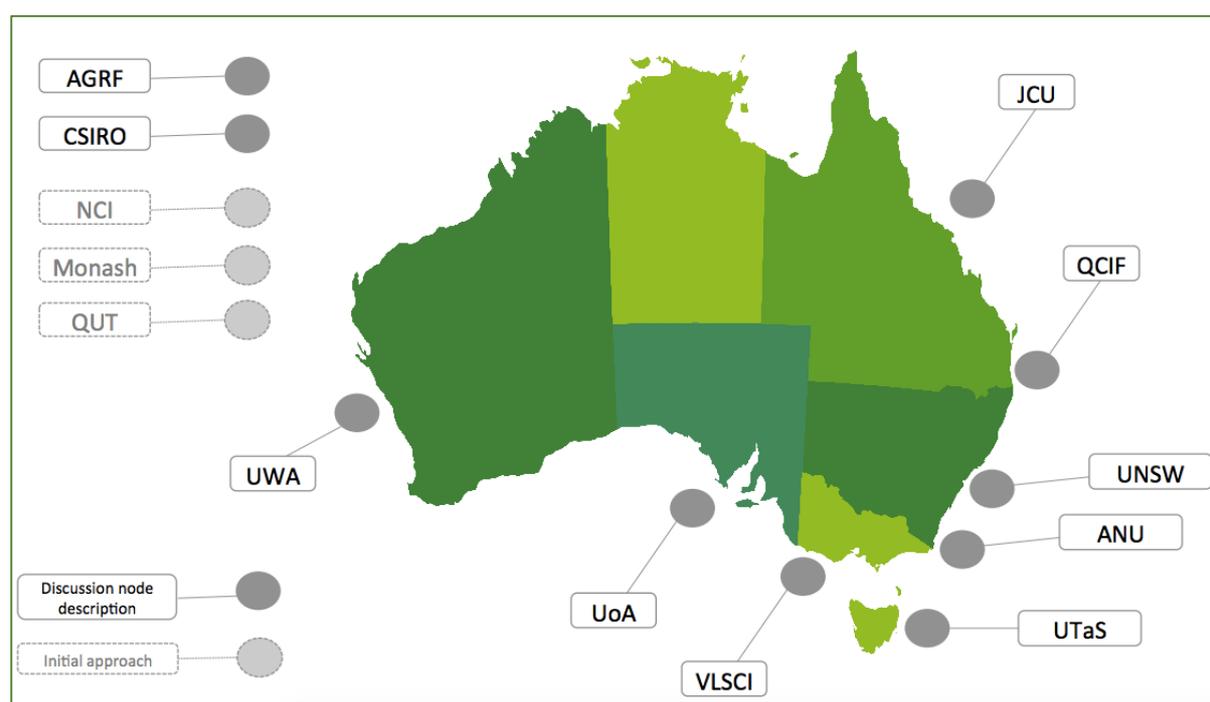
1. increase Australia's capacity to collect, integrate, analyse, exploit, share and archive the large heterogeneous data sets now part of modern life sciences research
2. contribute to the development of and provide training in data, tools and platforms to enable Australia's life science researchers to undertake research in the age of big data
3. showcase Australian research and datasets at an international level
4. enable engagement in international programs that create, deploy and develop best practice approaches to data management, software tools and methods, computational platforms and bioinformatics services.

## How

EMBL-ABR provides research infrastructure to facilitate bioinformatics for life and medical scientists as well as bioinformaticians. This research community is geographically dispersed. The national EMBL-ABR network provides connectivity by harnessing the expertise that already exists within local groups and Institutions across Australia, coordinating it at the national level. This facilitates the sharing of data, tools, and platforms. The network also drives Australian involvement in international initiatives around data standards and research best practice, preventing unnecessary duplication of effort. EMBL-

ABR also improves data openness by facilitating data publishing. Because it is in such high demand from the research community, a high-quality federated bioinformatics training program is a key component of EMBL-ABR. In 2017, EMBL-ABR will also develop a programme with industry to create an interface between bioinformatics in academia and industry and identify the mutual benefits.

It is the actual sum of activities and tasks performed by the nodes (which we are currently collecting through the Node description and Activity EoI) that will form EMBL-ABR Programme of work. Hence, 2016 will showcase pilot projects and prototypes of efforts with more input from the Hub that we expect to have as the project develop, where activities and tasks in the key area are done by the nodes and the hub retains the coordination, communication and impact monitoring role. Currently, the nodes that are working on a node description form and/or engaging with the hub are illustrated below:



## EMBL-ABR modus operandi

EMBL-ABR is organized as a network, with a central coordinating hub and multiple nodes at the institution level. This forms a federated system spanning the highly dispersed geographical locations of Australian life science and medical research that also has the advantages of centralised core activities. The EMBL-ABR Hub accommodates the EMBL-ABR Coordination Hub as well as EMBL-ABR Core Activities.

## EMBL-ABR HUB

The Hub is responsible for:

- the coordination and integration of services run from the EMBL-ABR nodes.
- developing and delivering the EMBL-ABR Core Programme and overall management and coordination of EMBL-ABR node projects.
- providing oversight of Australia biological data by creating data registries and coordinating the adoption of data standards as well as driving participation in international data standards development.
- liaising with other national and international science infrastructure initiatives.

## EMBL-ABR nodes

EMBL-ABR nodes are set up based on specific criteria, where prospective node members complete an EMBL-ABR node description form (see Appendix 1). This requires them to describe what their node will provide in terms of expertise in bioinformatics experimental design, analysis and interpretation, training and other specific local expertise. Head of nodes will be responsible for the monitoring and documentation of the activities across their nodes by adopting the monitoring and tracking processes provided by EMBL-ABR Hub.

EMBL-ABR also provides a vehicle for individuals or group of individuals and other initiatives to approach and fill in an EMBL-ABR Expression of Interest Activity form (EMBL-ABR EoI), which also has specific criteria relevant to the vision and mission of EMBL-ABR.

## Communication

Communication is key for the success and benefits of EMBL-ABR within and externally. Hence EMBL-ABR has an active web site (<https://www.embl-abr.org.au/>) which provides up-to-date information both to our users and for internal purposes. EMBL-ABR has also a monthly newsletter, which first issue was released in June 2016: <http://bit.ly/1UEZDrU>. These are both created and maintained by EMBL-ABR Hub.

EMBL-ABR Hub is currently exploring project management systems which aim to facilitate tracking of projects and allowing for EMBL-ABR Hub staff and nodes staff to stay informed and updated on the ongoing projects. Outreach activities are key to ensure the network involved stays informed and also key partners and stakeholders are up to date with the outcomes and projects under EMBL-ABR.

These include the following activities:

### Annual EMBL-ABR All Hands meeting

This is crucial when dealing with a distributed infrastructure so that all involved (nodes and hub) are aware of the spread of expertise, competencies and projects updates. This meeting also will include key infrastructure collaborators (e.g. Nectar and RDS). The first EMBL-ABR All Hands meeting is scheduled to take place on 8 December 2016.

### EMBL-ABR Domain driven workshops

(online and f2f): these meetings are focused on a particular domain or bioinformatics topic: e.g. clinical genomics, agriculture or large-scale data management. These workshops facilitate networking and knowledge transfer bringing in-depth discussions. In 2016 we will be focusing on the key area of Data and running 5 one-day workshops. We will also run a Hackathon on Bioschemas which is relevant for the key area of Tools.

### Dissemination

Presentations at different institutions and universities, nationally and internationally, including key conferences and events. We have a full agenda for 2016, which includes the trips from the Deputy Director around Australia (Hobart, Sydney, Brisbane and Adelaide) with remote chats with relevant people in Perth, Canberra and Melbourne, as well as some training activities and two major international Bioinformatics conferences (ISMB and ECCB), where EMBL-ABR is attending with a booth and engaging directly with the attending to generate awareness about Bioinformatics in Australia and specifically EMBL-ABR. Attendees will be asked specific questions regarding Australian Bioinformatics and training as shown in the figure below that will be used to inform our future priorities.

## EMBL Australia Bioinformatics Resources would like to know more about you.

### Do you:

- know anyone working in life sciences in Australia?
- have a general perception of Australian science?
- know of any Australian bioinformatics tools?
- currently have any collaborations with Australian scientists / bioinformaticians?

### NO YES

- Which field? \_\_\_\_\_
- What is it (in a few words)? \_\_\_\_\_
- Which one/s? \_\_\_\_\_
- Who or which institution? \_\_\_\_\_

### Your Details:

Institution \_\_\_\_\_

Current position \_\_\_\_\_

Most valuable bioinformatics training ever \_\_\_\_\_

Most needed bioinformatics training (for yourself) \_\_\_\_\_

Most needed bioinformatics training for life scientists \_\_\_\_\_

**Optional:** Stay up to date with future courses and events by subscribing here to EMBL-ABR news:

Name \_\_\_\_\_ Email \_\_\_\_\_

Thank you!

## Key Documents

In order to maximize transparency and communication, there is a dedicated section on ENBL-ABR website for key documents, here we will also make available the actual node descriptions and EOI from the various Head of nodes and Activity Leads as this become finalized. Key Documents page can be found here: <https://www.embl-abr.org.au/key-documents/>. We also prepared a list containing the portfolio of documents we have been creating during the past 3 months in order to support the various levels of communication, internal and external to the emerging network we have the following portfolio of documents, marketing elements and events.

# EMBL

## Australia

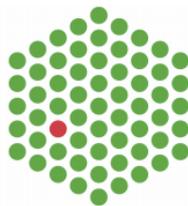
### Bioinformatics Resource

## portfolio

during the past 3 months and a half we have seen a transformation of EMBL-ABR. In order to support the various levels of communication, internal and external to the emerging network we have the following portfolio of documents, marketing elements and events.

### We categorize these below as in:

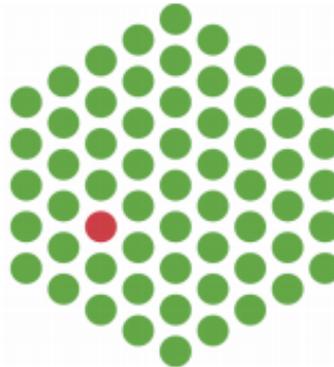
- EMBL-ABR Key Docs\_HUB\_Secretariat  
Lead: Fiona Kerr
- EMBL-ABR Comms\_HUB\_Comms & Training  
Lead: Helen Gardiner
- EMBL-ABR in progress\_HUB  
Lead: Vicky Schneider + Pip + Christina



EMBL-ABR ISAG TOR
EMBL-ABR Node description
EMBL-ABR Eoi Activity
EMBL-ABR Structure slide
EMBL-ABR 2-pager_general
EMBL-ABR Slide deck_general
EMBL-ABR Website & twitter
EMBL-ABR Comms guidelines
EMBL-ABR Training Template
EMBL-ABR Current projects & activities
EMBL-ABR surveys
EMBL-ABR BP guidelines (e.g. data chaperoning)
EMBL-ABR Training monitoring

EMBL-ABR ISAG TOR	PDF to be found on dedicated webpage for ISAG members	FIONA
EMBL-ABR Node description	<a href="https://www.embl-abr.org.au/key-documents/">https://www.embl-abr.org.au/key-documents/</a>	FIONA
EMBL-ABR Eoi Activity form	<a href="https://www.embl-abr.org.au/key-documents/">https://www.embl-abr.org.au/key-documents/</a>	FIONA
EMBL-ABR Structure slide	<a href="https://www.embl-abr.org.au/key-documents/">https://www.embl-abr.org.au/key-documents/</a>	FIONA
EMBL-ABR 2-pager_general	by request to Helen Gardiner	HELEN
EMBL-ABR Slide deck_general	by request to Helen Gardiner	HELEN
EMBL-ABR Website & twitter	<a href="https://www.embl-abr.org.au/">https://www.embl-abr.org.au/</a> @EMBL_ABR	HELEN
EMBL-ABR Comms guidelines	<a href="https://www.embl-abr.org.au/styleguides/">https://www.embl-abr.org.au/styleguides/</a>	HELEN
EMBL-ABR Training Template	<a href="https://www.embl-abr.org.au/styleguides/">https://www.embl-abr.org.au/styleguides/</a>	HELEN
EMBL-ABR Current projects & activities	<a href="https://www.embl-abr.org.au/">https://www.embl-abr.org.au/</a>	HELEN
EMBL-ABR Monthly Newsletter	<a href="https://www.embl-abr.org.au/news/">https://www.embl-abr.org.au/news/</a>	HELEN
EMBL-ABR BP guidelines (e.g. data chaperoning)	In development	PIP
EMBL-ABR Training monitoring	In development	CHRISTINA
EMBL-ABR surveys	In development	VICKY

# EMBL



## Australia

## Bioinformatics Resource

# Key Events 2016

**Join EMBL Australia Bioinformatics Resource in a variety of key events on Bioinformatics with a particular focus on Data life-cycle for 2016!**

*Where: Lab-14, 700 Swanston St, Carlton VIC 3053, Melbourne, Australia*

*Hosted by EMBL-ABR Hub and EMBL-ABR@VLSCI node*

### External Faculty

Rafael Jimenez, ELIXIR, UK

Sandra Orchard, EMBL-EBI, UK

Dan Bolser, EMBL-EBI, UK

Jioty Khadake, NIHR BioResource, UK

EMBL-ABR Nodes from across Australia will be taking part in this activities, according to nodes expertise and the life scientists community needs.

For more details visit: <http://www.embl-abr.org.au/about/events/>

---

21st March 2016

EMBL-ABR Training & Learning wk

---

11<sup>th</sup> April 2016

EMBL-ABR Training day: CyVerse

---

22<sup>nd</sup> June 2016

Focus Group EMBL-ABR network mt

---

24<sup>th</sup> October 2016

Annotation & Curation BP wk

---

25<sup>th</sup> October 2016

Data life-cycle Plants workshop

---

26<sup>th</sup> October 2016

Data life-cycle Animals workshop

---

27<sup>th</sup> October 2016

Data life-cycle Microbes workshop

---

28<sup>th</sup> October 2016

Data life-cycle Health workshop

---

8<sup>th</sup> December 2016

EMBL-ABR All Hands Meeting

---

9<sup>th</sup> December 2016

EMBL-ABR Training across the network

---

12<sup>th</sup> December 2016

Bioschemas hackathon

---

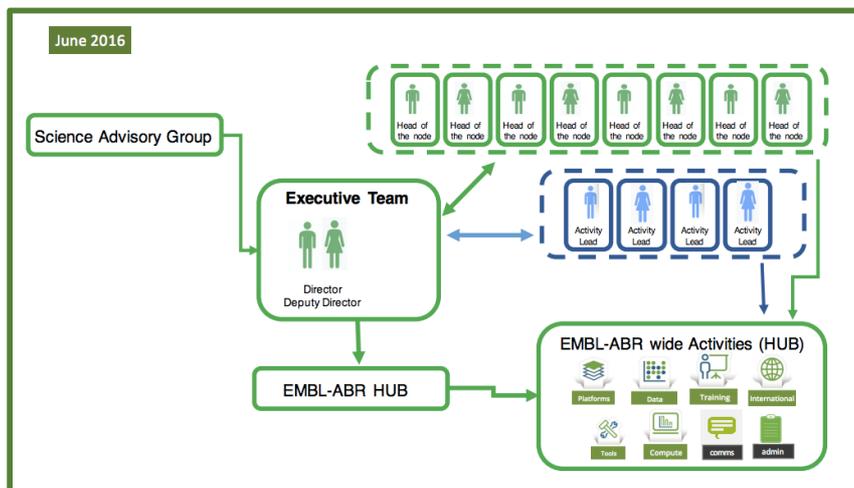
13<sup>th</sup> December 2016

Web resources for biology & Open Source BP wk

## Proposed organisational structure and personnel

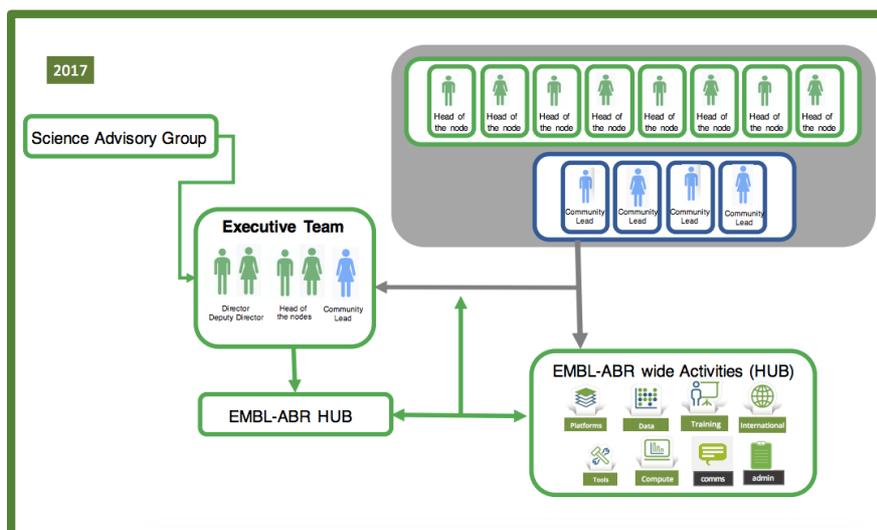
### Proposed EMBL-ABR Structure

EMBL-ABR organisational scheme is shown here. It currently has an Executive Team (Director & Deputy Director); an International Science Advisory Group, both entities are in operation. We are then working on setting up the Head of Nodes Group as we get the node description forms and the commitment from the nodes. We expect to have this third entity in place by the All-Hands meeting in December 2016, as well as



start to see a better picture in terms of Community Leads (that represent the Head of Activities we are also currently collecting through EMBL-ABR EoI).

By 2017 we expect to see the Executive Team composed of the present members plus representatives from the Head of Nodes and Community Lead Group.



### EMBL-ABR Hub Team

Currently EMBL-ABR Hub is managed by Deputy Director V. Schneider, and 50% of a FTE as Open Data Coordinator, Philippa Griffin. The Hub received operational support from core staff at EMBL-ABR@VLSCI node:

Andrew Lonie, Director

Vicky Schneider, Deputy Director and Hub Lead

Philippa Griffin, Open Data Coordinator and Activity Project Management

Fiona Kerr, Executive Management and Secretariat

Helen Gardiner, Communication Manager

Christina Hall, Training Manager

Claudia Curcio, Administration Officer

Ben Moran, Systems Administrator

Madison Flannery, Developer

Resourcing of the Hub is currently under discussion and we foresee a two phased process: 2016-2017 based on what is available and possible from existing skills and resources at the EMBL-ABR@VLSCI Node. From 2018 forward, this resourcing will be re-mapped for the long term, taking into account the abilities to secure national level funding.

## Key Areas

EMBL-ABR brings together Australian bioinformatics resources, some of which are incorporated into Activities (e.g. annotation) and Flagship projects (e.g. Sepsis). EMBL-ABR therefore is catalysing the creation and operation of new services in these **Key Areas**: data, tools, platforms, compute and training. EMBL-ABR also classifies “International” as a key area since it aims to be the Australian contact point towards the European infrastructure for biological information, ELIXIR, and other relevant international collaborations such as BD2K and CyVerse. EMBL-ABR will coordinate the Australian contributions to the ELIXIR infrastructure. Here we summarise the aims and activities for each of the Key Areas:

### Data

EMBL-ABR aims to support Australian scientists across all the levels of the data management cycle (from collection to analysis, archiving and data publication). EMBL-ABR will provide the infrastructure and guidance to empower researchers in an understanding of best practice in data storage, data management and data publication. EMBL-ABR will also share best practice in storage of sensitive data gained through the existing expertise in the nodes. This clearly requires a communication effort to translate this knowledge to the Australian life science researchers (training & outreach) and to track the impact on a national and international level.

### EMBL-ABR Annotation Activities

New genomes are being sequenced at an increasing rate, including genomes of unique and valuable Australian species. However, much of the value of a genome assembly comes from identifying where the important features are, like genes, splice variants, repetitive regions, and small RNAs. This process, called ‘annotation’, can be very time-consuming, requiring careful attention. The most successful genome annotation projects have typically involved numerous researchers and multiple lines of evidence, including RNA and protein data obtained from the target species, data from related species, and computational gene model predictions based on the DNA sequence. Data is valuable if it brings insights into the understanding of living systems. Annotation of data is therefore crucial to its value. EMBL-ABR will drive the effective coordination of expertise already present in a dispersed form among Australia Bioinformatics community in the area of annotation.

There is a need for collaborative bioinformatics infrastructure to incorporate multiple sources of information, and support multiple users with the ability to check others’ annotations. The current best-practice approach to this problem is the WebApollo software, which fulfils these criteria. However, WebApollo still requires hosting with sufficient memory and compute resources, which may be inaccessible locally to some Australian researchers. It may also be challenging to set up, especially for biologists without a computing background. This infrastructure-level problem is solved by providing a ‘pre-packaged’ WebApollo setup in the Genome Annotation Platform, accessible to everyone with an Australian research affiliation via the national NeCTAR Research Cloud.

Currently EMBL-ABR is working on a Genome Annotation Platform setup for the Australian research community. This is a collaborative cloud-based platform for Australian researchers (and international collaborators) to annotate a reference genome using one or more lines of evidence. It is set up as a WebApollo server running in a custom Genomics Virtual Lab instance on the NeCTAR cloud.

## Tools

Without appropriate software tools, the full potential for innovation and discovery emerging from the sheer quantity of accessible life science and biomedical data cannot be reached. This means developing software to tackle issues on data management (e.g. data provenance and compression), analysis and visualization for the Australian research community. This is an area where there is a real risk for Australia research to fall behind other countries as there is currently no national mechanism for supporting Australian researcher-created software and useful tools are regularly left un-maintained and un-hosted once the creator moves to a different institution.

EMBL-ABR is currently exploring the need for bioinformatic software hosting, maintenance and support among Australian researchers and bioinformaticians. Ways to contribute to international bioinformatics software standards and tool repositories are also being actively investigated. We would love to hear from you about your experiences, including issues you are facing or future plans for software development. ELIXIR, the European infrastructure for biological information is building a portal to bioinformatics resources world-wide. This project is a community effort coordinated by the Danish node of ELIXIR. The "[Tools & Data Services Registry](#)" provides essential scientific and technical information about analytical tools and data services for bioinformatics.

EMBL-ABR Hub is currently collaborating with ELIXIR and bio.tools in the collection and dissemination of Australian bioinformatics tools as well as tools relevant for the Australia life science researchers.

## Training

The advent of high throughput technologies has greatly increased the demand from life scientists to learn how to use bioinformatics resources and tools. This worldwide phenomenon is also widespread in Australia. EMBL-ABR aims to provide new opportunities and harmonise existing training efforts across Australia to maximise the reach and scope of bioinformatics training, as well as to share experience and best practice from Australia with international initiatives.

There are three main factors contributing to training gaps among life science researchers:

1. insufficient exposure to the principles and application of bioinformatics
2. the speed at which technologies, data analysis practices and tools develop and the lag in how these are incorporated into the current undergraduate and postgraduate curricula
3. lack of support within some research groups or institutions for students and postdoctoral researchers to learn bioinformatics skills and tools.

These factors currently mean that researchers can find themselves involved with, or even designing, projects that require skills and experience they lack.

The **EMBL-ABR Training Programme** will comprise a series of focal topics to tackle these training gaps, with the aim of enabling sustainable and scalable bioinformatics use by life scientists across Australia.

EMBL-ABR has established collaborations and joined key global efforts in Bioinformatics Training and Education. EMBL-ABR is currently engaging with:

- [The Gulbenkian Training Programme in Bioinformatics \(GTPB\)](#)
- [Bioinformatics and computational biology courses at the University of Cambridge](#)
- [Courses and Workshops from EMBL-HD and outstations](#)
- [ELIXIR IT](#) (Train the Trainer Taskforce)
- [CyVerse](#)

EMBL-ABR is also an active member of The Global Organisation for Bioinformatics Learning, Education & Training [GOBLET](#).

## Platforms

Life scientists benefit greatly from easy access to bioinformatics platforms that link multiple tools, facilitate data sharing and analysis, and trace and record analysis pipelines. Developing, maintaining and extending platforms is not a trivial process. Constant improvement is critical as research fields evolve to maximise the benefits gained from biological data.

EMBL-ABR@VLSCI node is currently providing existing expertise and knowledge in platform development through the Genomics Virtual Lab (GVL) supported by the Nectar cloud.

## Genomics Virtual Lab

The [GVL](#) is a cloud-based platform for bioinformatics research and teaching, used extensively within Australia and internationally. It contains a pre-installed set of bioinformatics tools including [Galaxy](#), [Rstudio](#) and [iPython Notebook](#). Public server instances are available for teaching.

## Compute

The aim of this program of work is to assist in the design, architecture and aspects of the delivery of infrastructure to support EMBL-ABR activities and ensure interoperability across the EMBL-ABR network as well as with international efforts. In addition, this program will identify pathways that respond to the future computation, analysis and storage infrastructure requirements.

## International

Bioinformatics resources and support for life scientist communities around the world are being provided by a range of established or developing centralised, distributed, and mixed model national research infrastructures. Effort at the national scale in several countries has received national level funding and in some cases additional commitment from participating institutions. Such initiatives stem from the necessity to respond to the large demand for e-infrastructure to allow users to efficiently and safely store, share, analyse and publish their data. In this data-driven era, it is vital to participate in international efforts in Data, Standards, Interoperability, Tools, Resources and Training.

EMBL-ABR is working intensively towards establishing a variety of links and partnerships with major International and National Bioinformatics e-infrastructures, such as [ELIXIR](#).

## Activities

Galaxy Australia Community: <https://www.embl-abr.org.au/galaxy-australia-community/>

CyVerse Australia: <https://www.embl-abr.org.au/cyverse-au/>

Bioschema.org: <https://www.embl-abr.org.au/bioschemas/>

## Appendix 1: Node Description Template

can be found here: [https://www.embl-abr.org.au/wp-content/uploads/2016/06/EMBL-ABR-Node-Description\\_Template\\_June2016v2.pdf](https://www.embl-abr.org.au/wp-content/uploads/2016/06/EMBL-ABR-Node-Description_Template_June2016v2.pdf)



## Appendix 2: EMBL-ABR EoI Template

can be found here: [https://www.embl-abr.org.au/wp-content/uploads/2016/06/EMBL\\_ABR-EoI-Activity-Form\\_Template16v2.pdf](https://www.embl-abr.org.au/wp-content/uploads/2016/06/EMBL_ABR-EoI-Activity-Form_Template16v2.pdf)