User Guide

Background

The European Bioinformatics Institute provides international access to data in molecular bioscience generated by researchers worldwide, including Australia. In its present state, Australian specific data is difficult to isolate within the EBI databases, particularly for the non-domain user. This software allows for Australian data to be discoverable in the form of collections of nucleotide and protein sequences. These collections can be defined as Australian data of two types: a) data submitted from Australian-based researchers; b) data associated with sets (and subsets thereof) of Australian species.

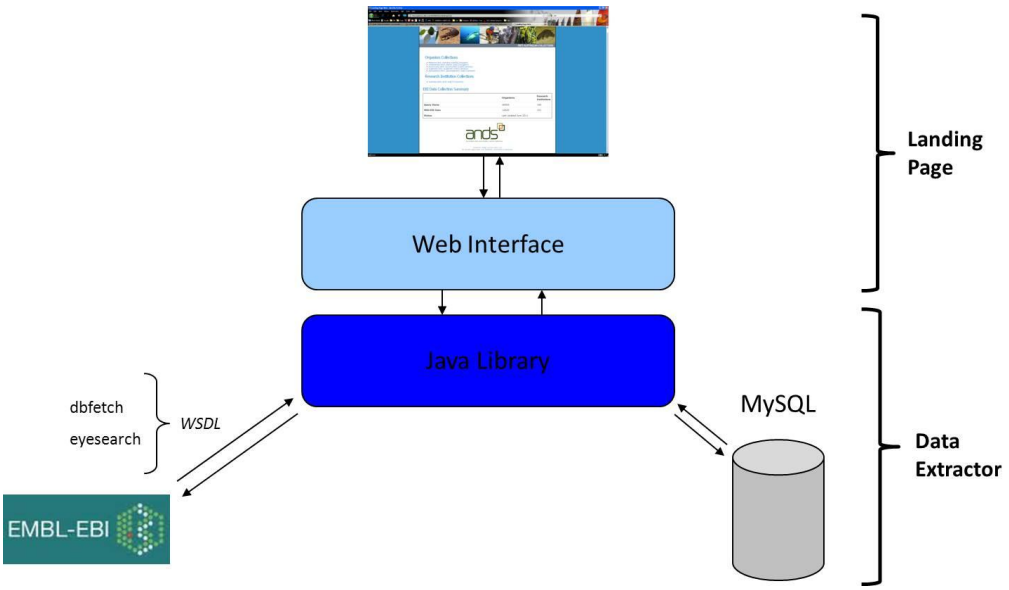
The software provides the following:

- The software queries EBI databases using Australian 'query items' (researchers or species) and stores data attained in a database.
- The software automatically generates collections which can be submitted to Research Data Australia. These collections are then discoverable by interested parties.
- The collections provide a link back to a landing page, which allows a visual display of collections of 'Australian' data found at the EBI. The landing page also the ability to navigate back to the primary data at EBI.

Overview of the product

The EBI-RDA linkage system is comprised of two main parts (Figure 1):

1. ***Data extractor***: this system is a standalone Java Program (executable jar) that pulls query items from a local data store (MySQL database) and queries them (using web services) against the Uniprot and ENA databases made available by EBI.
2. ***Landing page***: a website to allow visual display of the data. This was developed used JSP/Servlets and displays the relevant data that is stored in the MySQL database.



Installation and reuse

To install the software, see the installation guide: http://ebi-rda-linkage.sourceforge.net/ebi-rda-linkage-installation-guide.pdf

To modify the software, see the developer's guide: http://ebi-rda-linkage.sourceforge.net/ebi-rda-linkage-developers-guide.pdf

Landing Page Navigation

The landing page provides the following information:

- Metadata for the collection such as synonyms and a description of the collection

Australian Organism Collection

Name	Phascolarctos cinereus
Synonyms	Koala
Rank	species
Description	This data collection contains all currently published nucleotide (DNA/RNA) and protein sequences from the Australian dwelling organism Phascolarctos cinereus, commonly known as Koala. The nucleotide (DNA/RNA) and protein sequences have been sourced through the European Nucleotide Archive (ENA) and Universal Protein Resource (UniProt), databases that contains comprehensive sets of nucleotide (DNA/RNA) and protein sequences from all organisms that have been published by the International Research Community. The identification of the species Phascolarctos cinereus as an Australian dwelling organism has been achieved by accessing the Australian Plant Census (APC) or Australian Faunal Directory (AFD) through the Atlas of Living Australia.
Collection Content Last Updated	June 2011
Number of EBI Records	240
Aggregation by Contact	QFAB on behalf of the EMBL Australia Mirror of the EMBL-EBI Email
Search all data in EBI	Search EBI for any reference using the search term: **Phascolarctos cinereus**

- A subset of relevant results eg protein or DNA sequences. These results also provide a link back to the primary data at the EBI

EBI Data for Species

All Records | DNA Records | Protein Records

- **DNA Records:** The first 100 DNA records for this species are displayed. Other records (total: 149) can be viewed at EBI
- All available protein records are displayed

172 items found, displaying 1 to 20.

[First/Prev] **1**, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Accession Number	Sequence Header	EBI Database	View in EBI
AB011223	Phascolarctos cinereus mitochondrial ND1 gene for NADH dehydrogenase subunit 1 , complete cds.	ENA	Go to EBI
AB241053	Phascolarctos cinereus mitochondrial DNA, nearly complete genome.	ENA	Go to EBI
AF252263	Phascolarctos cinereus cytochrome P450-4A15 (CYP4A15) mRNA, partial cds.	ENA	Go to EBI
AF262696	Phascolarctos cinereus olfactory receptor koa2 gene, partial cds.	ENA	Go to EBI
AF262697	Phascolarctos cinereus olfactory receptor koa3 gene, partial cds.	ENA	Go to EBI

- The ability to navigate to related collections eg. For a species, the ability to view collections of related species.