

The IMAGE unified data portal to integrate and represent European gene bank data

P. Cozzi¹, J. Fan², A. Sokolov², O. Selmoni³, E. Vajana³, S. Joost³, E. Groeneveld¹, P. Flicek², P. Harrison² and A. Stella^{1,4}

¹UCSC, Via Emilia Parmense, 84, 29100 Piacenza, Italy, ²EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge, CB10, United Kingdom, ³EPFL, Route Cantonale, 1015 Lausanne, Switzerland, ⁴CNR, IBBA, Via A. Corti 12, 20133 Milano, Italy; paolo.cozzi@ibba.cnr.it

The Innovative Management of Animal Genetic Resources (IMAGE) project was established to improve animal gene bank management within Europe. One key challenge that this project addresses is the integration and transparent use of the vast information stored within more than 60 gene banks/genetic collections spanning 20 European countries, together with the collection of newly generated data. This imposes a major strategic challenge for the management and conservation of animal genetic resources due to the huge amount of heterogeneous data distributed across gene banks in different locations, storage formats and languages. The solution we present here comprises: (1) a well-defined metadata rule set ensuring high quality and comparable data across the diverse collections originating in different storage formats and languages; (2) development of a single Inject tool helping gene bank managers to enhance, standardise, tag and submit their gene bank data to a common data pool that integrates all gene bank records from across Europe; (3) the sustainability offered by archiving of this data within the EBI BioSamples public archive; and (4) a bespoke data portal that integrates gene bank metadata with generated 'omic datasets from within IMAGE and cross referencing to other gene bank and breeding database resources from across Europe such as those hosted by the FAO. Within the data portal, a Geographic Information System tool is included to assist the user in identifying/ storing the geographical origin of the samples as well as displaying individual/population genetic parameters and biological attributes through interactive maps. Querying across all types of data is also expected to facilitate targeted search to identify genetic material of interest residing somewhere in the partner gene banks and collections.