

# Era7 Bioinformatics' Bio4j project team attends final meeting of Google Summer of Code 2014 in California

Granada, 30<sup>th</sup> October 2014

Bio4J was selected this year as a project for being funded by the Google Summer of Code 2014. After months of work, the program ended with a really successful participation of Era7 Bioinformatics team.

Google Summer of Code is a 10-year-old global program that offers funding to leading Open Source projects from all kind of knowledge areas. This is done by means of funding directly students all over the world to help in some new functionalities or improvements of selected Open Source projects. Google tries to identify and fund exciting projects like Era7 Bioinformatics' Bio4j.

To celebrate the success of this program, Google organised a meeting in its headquarters from 23<sup>rd</sup> to 26<sup>th</sup> October and invited delegates from each successfully participating organisation to greet and collaborate. Two **Era7 Bioinformatics delegates** attended the event at Mountain View Google's offices and participated actively in the different activities organised by Google.

"This project has been a great opportunity to make our Bio4j platform an even more useful and valuable tool that we use under the hood of many of our pipelines and services, like BG7 and Genome7", said Eduardo Pareja, CEO of Era7 Bioinformatics. "Based, in part, in these improvements, we can offer now tailored Bio4j based services to be used by other parties in their bioinformatics solutions", added **Dr. Pareja**.

## About Bio4j

Bio4j is a high-performance cloud-enabled graph based open source bioinformatics data platform, integrating the data available in the most representative open data sources around protein information. It integrates the data available in UniProt KB (SwissProt + TrEMBL), Gene Ontology (GO), UniRef (50, 90, 100), RefSeq, NCBI taxonomy, and ExPASy Enzyme DB. The current version has more than 2.000.000.000 relationships, 400.000.000 nodes and 1.000.000.000 properties.

Bio4j provides a completely new and powerful framework for protein related information querying and management. Since it relies on a high-performance graph engine, data is stored in a way that semantically represents its own structure. On the contrary, traditional relational databases must flatten the data they represent into tables, creating *artificial* ids in order to connect the different tuples; which can in some cases eventually lead to domain models that have almost nothing to do with the actual structure of data.

## About Era7 Bioinformatics

Era7 Bioinformatics ([www.era7bioinformatics.com](http://www.era7bioinformatics.com)) is a leader company in Next Generation Sequencing (NGS) and Bioinformatics for Bacterial Genomics and Metagenomics. Era7 was founded in Spain and it has a subsidiary in Cambridge, Massachusetts, USA since 2012. Era7's business model is based on helping its clients in all the steps of a genomics project, from the design of the project and the sequencing, to an expert bioinformatics analysis, specialised bioinformatics consultancy and the development of applications and customised software. This is possible thanks to the strong research-oriented activity that Era7 carries out, the use of Cloud Computing and the advantages of developing software under an Open Source model.

## Contact information

Rosa Martin  
NGS  
[info@era7.com](mailto:info@era7.com)  
34 958 256771