Capturing experimental detail in a paperless environment – Scarab, an Electronic Lab Notebook developed and used at the Structural Genomics Consortium

FERREIRA, L[2]; MASSIRER, K[1,2]; GILEADI, O[2,3]; MARSDEN, B[3].

[1]Center for Molecular Biology and Genetic Engineering, University of Campinas, Brazil; [2]The Strucutural Genomics Consortium, University of Campinas, Brazil; [3]The Strucutural Genomics Consortium, University Oxford, UK

The amount of data generated by a biological research team can overcome the ability of that team in preserve the necessary information for the research in an easy way to retrieve it. Developing a data warehouse to store and manipulate the data can help the researchers to improve the quality of the results and also to perform faster researches. In practice, electronic data warehouse allows to store and retrieve all the data relevant to a project, including historical data from past lab members or from collaborators, in a way that is not possible with paper-based notebooks. The development of Scarab by Molsoft L.L.C. in collaboration with the Structural Genomics Consortium – SGC, aims to satisfy this demand for an electronic manner to support research. Scarab incorporates many features centred around structured and unstructured data without losing mining flexibility. Firstly Electronic Lab Notebook – ELN, where the researcher can manipulate electronic pages to embed tables, images, PDF files and also other files used during the research. Those pages also can be used, shared with and modified by researchers who work together on the same project. A unique feature of Scarab is the ability to flexibly mine and export the data, whereby the researcher, without the assistance of an informatics team, is able to build queries to the database and export the result to use the results for another search or attach the result on a desired ELN page or other files. The development of this data warehouse started in 2002 by Molsoft L.L.C. in partnership with SGC located in Oxford University, since then it has been used as a successful case of a bioinformatics tool supporting all activities developed by SGC laboratories around the world. In 2016 the same model was implemented at SGC-Unicamp, proving the modularity of the data warehouse, changing the location and the activities developed at SGC-Unicamp, but using the same data warehouse schema, implementing only minor changes on the database. At SGC-Unicamp we are currently studying the human kinome composed of 500 proteins and we are collecting data from about 10 different sources of experiments. Scarab has proven to present a solid data warehouse solution to support all activities of biological laboratories and the flexibility to adapt to changing science outputs over time.

FUNDING: FAPESP, SGC-Global, Unicamp.