

TECHNOLOGY NETWORKS

Date: 18 November 2016

Headline: sbv IMPROVER Datathon

Byline: Anna MacDonald

Article online at: <http://www.technologynetworks.com/news.aspx?id=198229>

New approach to connecting data to analytics stimulates multi-disciplinary collaboration and innovation at first sbv IMPROVER Datathon.

How do you share a large dataset among a diverse group of scientists, encourage them to expand computational analytics strategies, create customizable analytical tools, and demonstrate reproducibility in a complex, multi-endpoint toxicological study? A diverse group of scientists have come together at the first sbv IMPROVER Datathon in order to do just that and explore new ways of collaborating across disciplines and approaching large datasets.

The Datathon was a working session in which participants collaborated to turn datasets into scientific insight. Data was provided to teams who then developed research questions and preliminary findings. In advance of the event, the sbv IMPROVER team prepared protocols to explore the dataset, which enabled participants to access the data as well as to understand and design novel workflows which connect the data to analytics.

Custom analytics modules (called “gadgets”) for data exploration, visualization and comparison, were built for the Datathon, enabled by an open connectivity platform technology called Garuda. This community-built open platform provides a unique framework to dynamically connect data with publicly available tools or participant developed algorithms.

One month prior to the Datathon, participants were able to download data from a dedicated sbv IMPROVER webportal. The data comprised a range of ‘omics and functional measurements taken from a seven-month inhalation study of heated tobacco,1,2 supplemented with unpublished lung and blood DNA methylation data. Classical toxicological end points were assessed in the study together with additional transcriptomics, proteomics and lipidomics measurements, all generated through high-end profiling technologies and creating a substantial, rich and varied dataset.

The Datathon attracted participation from computational biologists, bioinformaticians and data scientists specializing in predictive analytics, text analytics, data mining and statistical analysis. Drivers for participation included the depth and diversity of the dataset on offer, and the ability to explore that dataset within an open, transparent environment that facilitates innovation and discovery. Participants also had the opportunity to have their methodologies made available as gadgets on the platform.

“The customized workflows based on an open platform technology have enabled Datathon participants to dive into the data from one of our most ambitious and comprehensive studies to date,” said Dr Nicolas Sierro, Manager Genomics, Biological Systems Research, Philip

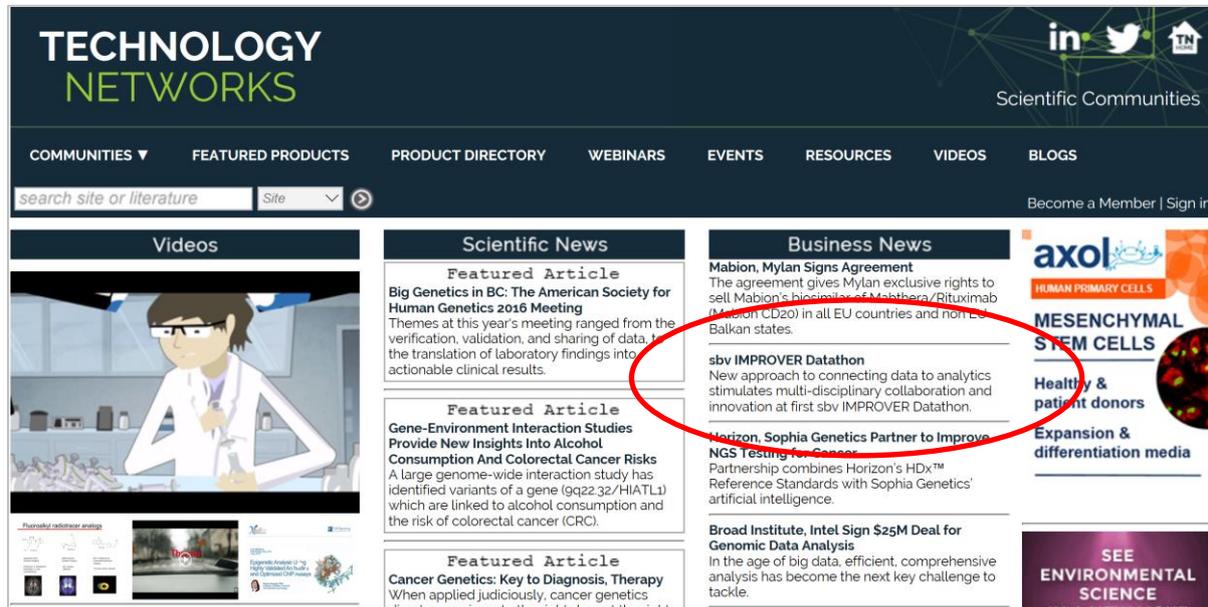
Morris International (PMI). “sbv IMPROVER is about driving open innovation in scientific discovery, facilitating transparent research frameworks, and enhancing dialogue within and between different scientific specialisms. The open platform has proved to be an innovative partner, helping us realize these goals.”

Dr. Samik Ghosh, CTO, SBX Corporation, Tokyo, Japan, commented: “The Datathon has established an exciting new model for biological investigation. It provides a blue-print for the creation of verifiable workflows that are capable of dealing with the complexity and diversity of biomedical data, and ultimately the ability to turn this data into knowledge and understanding. We are delighted that PMI chose the open platform based technology to empower the first sbv IMPROVER Datathon through our purpose-built gadgets, and also that the participants so clearly demonstrated the depth of the data and the ability to develop multi-modal analytics to verify and reproduce their study findings.”

Garuda (www.garuda-alliance.org) is a community-built platform providing an open framework to connect, discover and navigate through different applications, databases and services in biology and medicine. Powered by language-independent Application Programming Interfaces (APIs), it can connect various software packages and combine them into bespoke ‘gadgets’ customized for specific data analysis.

sbv IMPROVER is a collaborative initiative, led and funded by PMI, which has run a series of open science challenges, projects and events since 2012. It aims to develop a robust methodology for verifying scientific methods and results in the context of industrial and academic research. As the community has grown, its focus has expanded across a range of topics in biomedical research.

Technology Networks HomePage Screengrab (18 November 2016):



-- ENDS --