



Contribution ID: 246

Type: **Talk**

High Throughput Computing in bioinformatics: workflows, containers and emerging paradigms

Tuesday, 4 December 2018 11:00 (30 minutes)

Next Generation Sequencing has brought genomic analysis within the range of a great number of laboratories, while increasing the demand for bioinformatic analysis. These typically comprise workflows composed out of chains of analyses with data flowing between workflow steps. Such analysis is amenable to High Throughput Computing, a form of high performance computing characterised by a focus on overall analysis throughput rather than optimisation of a single application. In recent years workflow languages and container technologies have become a key part in composing efficient, reproducible and re-usable bioinformatic workflows. These technologies, however, pose a challenge for High Performance Computing providers as they require different characteristics from an execution environment to that provided by traditional HPC clusters. These challenges will be discussed and some approaches to solving them will be discussed.

Presenter Biography

Peter van Heusden is a researcher at the South African National Bioinformatics Institute, where he has been developing research computing infrastructure since the 1990s. His research focus includes scientific workflow languages and workflow management systems, research software and systems engineering and biological sequence analysis with a specific focus on the pathogen *Mycobacterium tuberculosis*.

Primary author: Mr VAN HEUSDEN, Peter (South Africa National Bioinformatics Institute)

Presenter: Mr VAN HEUSDEN, Peter (South Africa National Bioinformatics Institute)

Session Classification: HPC Technologies

Track Classification: HPC Technology