



Contribution ID: 58

Type: Poster

## Improving replica placement strategies using information from existing communication infrastructures

In highly data-driven environments such as the LHC experiments a reliable and high-performance distributed data management system is a primary requirement. Existing work shows that intelligent data replication is the key to achieving such a system, but current distributed middleware replication strategies rely mostly on computing, network and storage properties when deciding how to replicate data-sets across a global set of data centres.

While the distributed nature of such data management systems reduces the requirement for co-location of data and users interested in specific data, reliability and performance considerations mean that where possible co- or close-location are preferred.

We present an approach for improving existing replication strategies based on geographical data available in existing communication infrastructures. Information on the geographical distribution of interested users is extracted from an existing communication infrastructure using automated analysis of locational expressions in research documentation, operational logbooks, e-mail correspondence or web presences. Combined with the linking of data-sets to interested users this allows for an intelligent, anticipatory data replication strategy for data placement at locations close to the interested users.

### Summary

Replica placement strategies based on feature extraction by natural language parsing.

**Primary authors:** Mr LASSNIG, Mario (CERN & University of Innsbruck, Austria); Mr HALL, Mark Michael (Cardiff University, Wales, UK)

**Presenters:** Mr LASSNIG, Mario (CERN & University of Innsbruck, Austria); Mr HALL, Mark Michael (Cardiff University, Wales, UK)

**Track Classification:** 1. Computing Technology