

## **Title: Text mining: the next data frontier. An infrastructural approach.**

### **Abstract:**

Recent years witness an upsurge in the quantities of digital research data, offering new insights and opportunities for improved understanding. Text and data mining is emerging as a powerful tool for harnessing the power of structured and unstructured content and data, by analysing them at multiple levels and in several dimensions to discover hidden and new knowledge. However, text mining solutions are not easy to discover and use, nor are they easily combinable by end users. OpenMinTeD aspires to enable the creation of an infrastructure that fosters and facilitates the discovery and use of text mining technologies and interoperable services. It examines several use cases identified by experts from different scientific areas, ranging from generic scholarly communication to literature related to life sciences, food and agriculture, and social sciences and humanities. OpenMinTeD text mining tools, services and associated resources will run on the cloud, requiring an in-depth optimization of service deployment and execution via scalable VM-based service distribution and use of distributed storage.