**Big Data Use Cases: Progress and Futures**

Geoffrey Fox Indiana University; Ilkay Altintas, UC San Diego and SDSC

Under the auspices of the NIST Big Data Public Working Group, we formed a community of interest from industry, academia, and government, with the goal of developing a consensus set of Big Data requirements across all stakeholders. The major activities were gathering various use cases from diversified application domains and extracting requirements. Initially we developed a use case form with 26 fields and these were completed for 51 areas. They were spread over broad areas: Government Operations (4), Commercial (8), Defense (3), Healthcare and Life Sciences (10), Deep Learning and Social Media (6), The Ecosystem for Research (4), Astronomy and Physics (5), Earth, Environmental and Polar Science (10), Energy (1). These are of course only representative and miss many important cases but they form an interesting set.

After gathering the use cases, a multi-step process was used to extract requirements. First specific requirements were extracted from each use case and then they were mapped to broad characteristics motivated by structure of the reference architecture. These characteristics were data sources (data size, file formats, rate of grow, at rest or in motion, etc.); data lifecycle management (curation, conversion, quality check, pre-analytic processing, etc.); data transformation (data fusion/mashup, analytics); capability infrastructure (software tools, platform tools, hardware resources such as storage and networking); Security & Privacy; and data usage (processed results in text, table, visual, and other formats). Then we aggregated all the 437 specific requirements into 35 high-level generalized requirements.

For future work, we need to feedback lessons from the requirements analysis into use case template and consider the consequences for standards. In particular this would require the “complete application” to be broken up into components of the workflow. Building a set of demonstrations and benchmarks based on use case components is also attractive.