Technologies survey and analysis of current long term archiving systems in the context of Earth Observation

Ruben F. Perez (1), Oscar Perez (1), Oscar Portela (1), Rosemarie Leone (2), Mirko Albani (2), and Vincenzo Beruti (2)

(1) GMV Innovating Solutions, Science and Earth Observing Processing Systems, Spain (rfperez@gmv.com, oopn@gmv.com, ospa@gmv.com), (2) European Spatial Agency, ESRIN, Italy (Rosemarie.Leone@esa.int, Mirko.Albani@esa.int, Vincenzo.Beruti@esa.int)

Long term data preservation includes the activities required to ensure the data integrity and the data access for very long terms, in principle without time limitation. The activities concern therefore a continuous consolidation and technical evolution of archives, archive management systems and the related data access systems to guarantee the basic data preservation and proper data accessibility. The evolution is concerning not only data and format migration, but also the processing chains, the algorithms and the data access technology so that users can always actually receive and process the data products in accordance to the up to date requirements.

In this context, the object of the LAST project (Long term data Archive Study on new Technologies) is to perform an independent study on best practices and assessment of different archiving technologies mature and available on the market for operation in the short and mid-term time frame, or eventually available in the long-term. Emphasis will be put on technologies better suited to satisfy the requirements of ESA, LTDP and other European and Canadian EO partners in terms of digital information preservation and data accessibility and exploitation.

LAST activity can be considered as a due diligence process spanning across three phases:
1) Collection and analysis of the archiving and retrieval requirements from ESA, its partners, and other organizations both from the EO domain and other fields where similar archiving needs do exist. This phase is aimed at the provision of the definition of a common set of requirements, and a survey of the practices and solutions currently in use.
2) An archiving technology survey, where an extensive analysis of technologies already available in the market, or currently in development but expected to be ready in the short/mid-term. This phase includes not only the storage technologies themselves, but also the associated servers, operating systems, databases, network infrastructures and protocols, hardware storage systems, software architecture etc.
3) A testing and benchmarking campaign during which the best three complete solutions previously identified are evaluated as a whole. These solutions are defined through the combination of some of the components previously examined, performing the appropriate analyses and trade-offs in accordance to the requirements identified in the first phase.

The results obtained during phases 1 and 2, and their relationships with other current ESA activities are summarized, focusing on the results of the technological surveys to provide with an overview of the state of the art, and the identification of the different technological areas of interest in relation to those systems.

The evaluation methods that take into account the preferences and specific requirements of long term archiving systems are also described, involving the identification of general technological areas and of the evaluation criteria in each technological area, including the setting of weights for ranking the different evaluation criteria in each technological area. In addition, this method provides the definition of a metric related to the technological gap between a current system and a system in which some of the current technologies may be replaced by the best evaluated ones.