



**Sam Fineberg, Distinguished Technologist, Hewlett Packard Enterprise**

#### **TITLE**

**Long Term Data Preservation using the SIRF Container Format**

#### **ABSTRACT**

If digital artifacts are to serve as the basis for cultural and personal memories, it is essential that their lifetimes exceed those of memories in the human mind, and of the many physical objects found in museums. By definition, this involves timescales that exceed the maximum lifetime of any single storage device or any single storage technology. If the computer industry continues to ignore these issues, it is likely that much of the content generated over recent years will be lost for future generations, leading to a digital dark age. Cloud and object storage help achieve this timelessness, by abstracting storage from physical devices, enabling it to migrate to new technologies as old ones are retired. However, they lack the key metadata needed to preserve meaning of the data into the far future. For example, a standard application level hash can be used to ensure that the bits have not changed, and information about data formats can be used to facilitate migration to new formats. An even bigger issue is to retain the semantic information needed to understand what the metadata means. For example, if I had an email from 20 years ago, would I be able to determine the identity of the sender and recipients. The Storage Networking Industry Association (SNIA) Self-contained Information Retention Format (SIRF) is being developed to define a standard storage container for long term retention. SIRF enables applications to interpret stored data regardless of the application that originally produced it. It includes a catalog in which the key metadata can be stored, enabling the data to be preserved and understood far into the future. SIRF can be deployed as a layer on top of OpenStack Swift, S3, or other object storage interfaces, providing a standardized way to represent preservation objects in the cloud. We have developed a reference implementation of SIRF, OpenSIRF, that encompasses an API and a set of core classes that allow SIRF to be integrated with object storage. More information on SNIA's efforts to address long term retention can be found at <http://www.snia.org/ltr>. More information on OpenSIRF can be found at [www.opensirf.org](http://www.opensirf.org).

#### **BIOGRAPHY**

Dr. Fineberg is a Distinguished Technologist in the Hewlett Packard Enterprise Storage Chief Technologist Office. Dr. Fineberg has over 25 years of experience in areas including storage, cloud computing, archiving, non-volatile memory, and high performance computing. Sam leads cloud and big data technical strategy for HP's multi-billion dollar storage division, driving participation in the OpenStack cloud storage community and solutions for Hadoop. Sam was previously co-chair of the Storage Networking Industry Association (SNIA) Long Term Retention Technical Working Group, and he is an active member of the SNIA Analytics and Big Data Committee.