

# Applications of modern Erasure Coded Big Data paradigms to WLCG data resilience

*Tuesday 11 October 2016 16:30 (15 minutes)*

Previous research has shown that it is relatively easy to apply a simple shim to conventional WLCG storage interfaces, in order to add Erasure coded distributed resilience to data.

One issue with simple EC models is that, while they can recover from losses without needing additional full copies of data, recovery often involves reading the all of the distributed chunks of the file (and their parity chunks). This causes efficiency losses, especially when the chunks are widely distributed on a global level. Facebook, and others, have developed “Locally Repairable Codes” which avoid this issue, by adding additional parity chunks summing over subsets of the total chunk distribution, or by entangling the parity of two stripes to provide additional local information.

Applying these approaches to data distribution on WLCG storage resources, we provide a modified encoding tool, based on our previous approach, to generate LRC encoded files, and distribute them appropriately. We also discuss the potential application to the natural chunking of WLCG style data, with reference to single-event data access models and WAN data placement. In particular, we consider the advantages of mechanisms to distribute load across potentially contended “fat” Tier-2 storage nodes which may need to serve “thin” Tier-2 and Tier-3 resources in their geographical region.

## Tertiary Keyword (Optional)

Object stores

## Secondary Keyword (Optional)

Storage systems

## Primary Keyword (Mandatory)

Distributed data handling

**Primary author:** SKIPSEY, Samuel Cadellin

**Co-author:** EBERT, Marcus (University of Edinburgh (GB))

**Presenter:** EBERT, Marcus (University of Edinburgh (GB))

**Session Classification:** Posters A / Break

**Track Classification:** Track 4: Data Handling