# Understanding Storage Intermediaries

Helen Hockx-Yu, University of Notre Dame

# Background

- Based on a paper co-authored with Don Brower, Computational Scientist at the Center of Research Computing, University of Notre Dame
- Accepted and presented at iPres 2021
- Build on the learning obtained from enterprise data migration projects, necessitated by switching storage intermediary products.
- Full paper

# What are storage intermediaries

- Device and/or software that sits between your computer and storage media.
- Translate between different protocols and APIs
- Access Control
- Perform useful functions, e.g.
  - Caching
  - Encryption
  - Compression
  - Deduplication
- Examples: AWS Storage Gateway; Spectra Logic's BlackPearl Converged Storage System

# File transformations = dependencies

- Underlying storage organization may not match logical organization
- Files might be renamed -
  - e.g. UUID, Hashed value (Content Addressable Storage)
- Files might be broken up, compressed and stored as many separate chunks
- Additional metadata
  - file systems metadata versus in-file metadata
- "Deleted" files
- "Phantom" files

# Objects stored in AWS S3 by Avere



| Name |
| --- |
| 00000A9C_2C29E7108AF746E19785195A754BDD70.00000000009.FFFFFBE1.0001 |
| 00000AB6_41F9724940000000F64732132000000001F3EFB957AE8F0C4E98804840000000F6473200.00000000016.FFFFF24B. |
| 00000AF2_1AB8FD1A963A4824BA0BFD54FA7F3FA3.00000000009.FFFFFDE9.0001 |
| 00000B04_A12F2B1340000000DDF8B30F20000000013E4621929A8D016045646840000000DDF8B300.00000000016.FFFFFCD1 |
| 00000B13_96D56C67CE7F44AF808C63FAC1F3F939.00000000009.FFFFFE5D3.0001 |
| 00000B7C_E658EA9D5B724D05BCBEE4B011671DED.00000000009.FFFFFB25.0001 |
| 00000B93_5A67F5F7B1A64CDE81BA008BEABD9741.00000000009.FFFFFBE1.0001 |
| 00000B9F_977E346B4C924B3287AFE5DA731F0174.00000000009.FFFFFB43.0001 |
| 00000BA3_77E19765CA0049D5A547B16C1BB99163.00000000009.FFFFFDE9.0001 |
| 00000C04_79F754888B484479A154ED7CE0CA9624.00000000009.FFFFFDE1.0001 |
| 00000C10_4A9F37251D34474090610C4E7467C39A.00000000009.FFFFFB5D.0001 |
| 00000C21_91DB72CD02020100CBA031000000000005898F69A207D078.00000000011.FFFFDC9F.0001 |
| 00000C22_55FE16F18A3B4A4FB52E11EB3C67F5F5.00000000009.FFFFF11B.0001 |
| 00000CB5_B18ACDEB8E3A4E178C7794465AE2CB01.00000000009.FFFFF6A7.0001 |

# Recommendations for digital preservation

- Be aware of reduced redundancy and diversification
- Trade-off between efficiency in data transfer and storage, and file integrity and independence.
- Rethink exit strategy
- Plan for data migration
- Think about access
- Call out the dependencies; give the same level of attention as checksums and file formats