

# Quobyte and Digital Bedrock

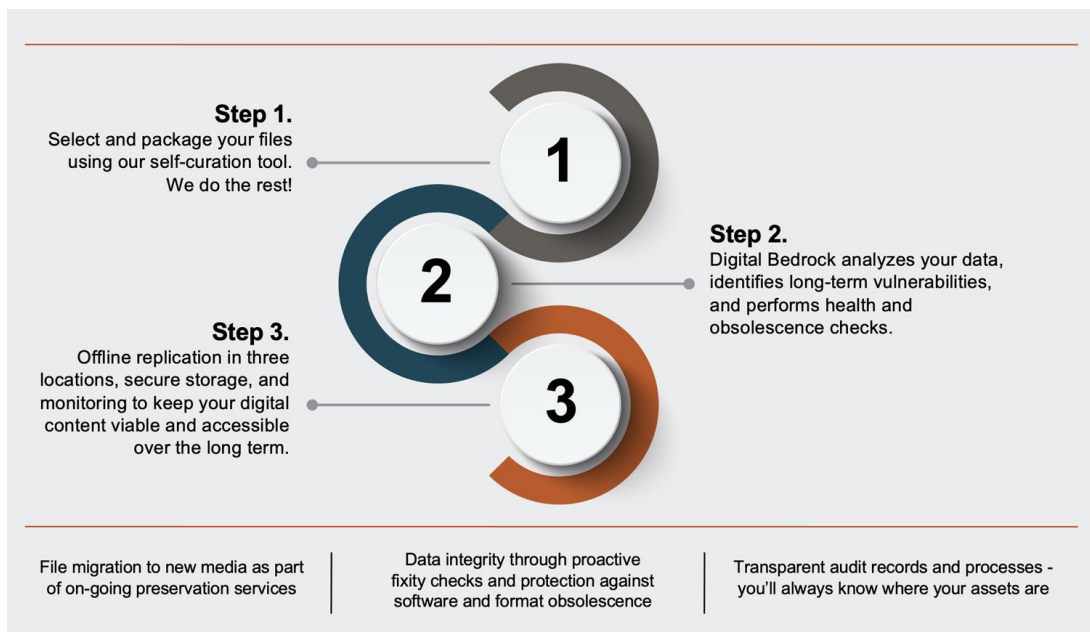
## Putting High-Performance Storage at the Heart of Digital Preservation

Keeping one's valuable digital holdings usable over time requires active digital preservation. The umbrella of digital preservation actions includes content appraisal (should this data be retained long-term?), file format verification through metadata analysis, ongoing "bit health" and obsolescence monitoring, and migrating files from one storage medium to another as physical storage media reaches its end of life. All digital files must be in a secure environment while undergoing preservation processing. Preserved data can be ultimately stored in the cloud, on premises, or on offline media such as LTO. However, before it is written to target storage, it is most effectively processed on a drive-based storage platform.

Linda Tadic founded Los Angeles-based Digital Bedrock to meet the specific needs of digital asset preservation. Her career spans over 30 years serving organizations such as HBO, the Getty Research Institute, ARTstor, and the Media Archive and Peabody Awards Collection at the University of Georgia. When it comes to archiving and preserving large digital projects, few organizations can match Digital Bedrock's expertise and care. Preservation processing – from moving customer data to deep LTO storage and all the steps in-between – comes with many challenges. Having the right storage platform in the center of that process is critical to Digital Bedrock's model and growth. That's why Tadic and Digital Bedrock rely on Quobyte.

More Than an Upload

“Many people think preservation is just putting their data on servers or in the cloud,” says Tadic. “But it’s much more, and that’s where we add value. People don’t need to preserve everything, so we help at the front end with selection and de-duping. We bring it into our digital preservation application, which we created, where it’s scrubbed through a virus check and verified with a checksum – a digital fingerprint. Because our clients include studios and government agencies, we use a SHA512 cryptographic hash algorithm, as recommended by the Department of Defense.”



Meticulous metadata extraction from files is key to Digital Bedrock’s process. For example, metadata allows an archivist to query across a preserved collection for media assets shot with a certain camera with a certain lens within a given date range. According to Tadic, there are a range of metadata extraction tools, most of them open source, and Digital Bedrock uses “all of them together,” because each captures different information or expresses the same information in a different way. Collectively, Digital Bedrock has amassed tens of millions of files, and each file can generate hundreds of metadata points. The database and storage platform handling this analysis must be very robust.

Otherwise, storage becomes the bottleneck blocking the ingestion and processing of customer projects.

Once data processing completes, Digital Bedrock writes the file collection to at least three LTO tape copies simultaneously, as opposed to making one master and cloning it. Each tape is separately verified after writing. Only at this point is the source data deleted from Digital Bedrock's Quobyte servers. Moreover, all data receives an annual SHA512 checksum verification to verify the files' bit health. One LTO copy is kept in a locked safe within Digital Bedrock's downtown Los Angeles data center, protected behind nine levels of biometric security. The other two copies are geographically dispersed for disaster recovery, also in secure locked environments. Because of this high level of detailed processing and security, Digital Bedrock can guarantee that no bad actors can access its clients' data.

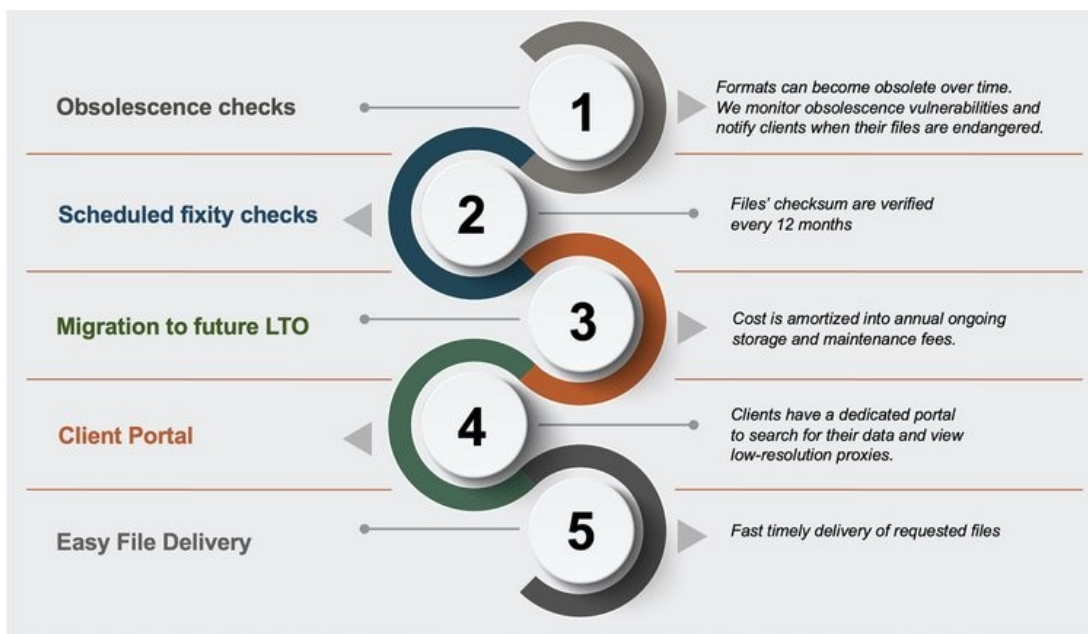
"We have a petabyte of active storage where we're safekeeping client data while it's being processed at different stages," says Tadic. "But that storage keeps getting refreshed. This is why the Quobyte system is really important, to help us be able to manage all of that data – to bring it in, allow us to run processes, and then delete that data to make room for the next client. We have clients spending millions of dollars to produce or acquire works we preserve, and they don't want it kept online. So, managing how that data moves from ingestion to preservation is key."

### The Drawbacks of Earlier Open Options

Digital Bedrock performs its data processing across six on-prem storage servers, each with its own hard disk storage. The cluster typically processes between 30 and 50 files concurrently. Exporting can span up to 18 different tape drives. Not surprisingly, storage throughput plays a key role in determining the balance struck between job completion time and how many files can process concurrently.

Digital Bedrock had originally used an open source storage platform. But as data being stored and processed increased as the company grew, it presented serious challenges to overall performance.

“We wanted to use our own hardware, so we worked with Linux,” says Digital Bedrock project manager/developer Diana Eppstein. “We installed new 18TB drives in the servers, but we were wasting a lot of time copying data from one machine to another, like pre-processing on one system and then having to move it to another machine. Then time was spent loading files to check the checksums and verify the files were good. We were stressing our systems and always worrying about the data, because file movement is when you make mistakes and have errors pop up. One of the biggest reasons we wanted Quobyte was because it lets us keep files in one location throughout processing.”



Part of the Linux platforms' shortcomings could be addressed by adding significantly more disk storage, but Digital Bedrock wanted to avoid unnecessary infrastructure and heavy capital investments.

Additionally, Digital Bedrock system administrator, Marco Cova, notes that prior storage platforms required significant effort for configuration and management. Simply getting the platform up and running took several days, and ongoing maintenance presented unique and at times vexing challenges.

"We had so many issues," he says. "We'd get bottlenecks nobody could explain, and it took forever for a single person — me, in this case — to troubleshoot problems when they arose."

#### Quobyte: Easy Operation With Outsized Results

When we spoke to Cova, it had been over nine months since he first installed Quobyte for Digital Bedrock. The process was so simple and easy, he says, that he hardly remembers doing it. Quobyte aggregated the performance of all storage servers and the performance it delivered out of the box more than met Digital Bedrock's performance criteria, making any tuning or maintenance unnecessary.

"You don't even notice that you're running a storage platform," he adds.

According to Diana Eppstein, Digital Bedrock typically runs project sets composed of 40 to 50 terabytes of data. Quobyte allows the company to process a project in a quarter of the time required by earlier platforms. Part of this improvement stems from inefficiencies in how the prior solution (FreeNAS/TrueNAS) went about data deletion.

Consider that a project might entail running five metadata extraction tools simultaneously across 80,000 project files. Digital Bedrock typically runs eight to ten such projects in tandem. As soon as one project finishes, the data needs to

be deleted from hot storage to make room for the next project. On the original storage solution, Eppstein had to write code that would intentionally slow the deletion process to prevent errors in the system performance.

This problem vanished when Digital Bedrock adopted Quobyte. The company could run its processing and archiving workflow without the previous bottlenecks. The backlogs that hindered prior operations went away, allowing the company to process more jobs and thus increase revenue.

Beyond performance, Digital Bedrock's adoption of Quobyte has yielded a range of secondary benefits.

"We love the speed and that it's very open," says Linda Tadic. "But we also like working with other startup-type companies. It's easier to build better relationships with people. The lower cost was helpful, but who we partner with, having the right relationships, that's just critical."