inktank

Digital preservation with open source

Sage Weil
Founder and CTO, Inktank

What is Ceph?

- Originated from DOE grant to UCSC
- Scale-out distributed storage
- Fault tolerant – no single point of failure
- Storage hardware agnostic
- Commodity components
- Single cluster, multiple storage protocols
  - Object, block, and file
- Free and open source

The role of open source in archival storage architectures

- Cost
  - Software and hardware
- Vendor independence
- Transparency
- Insulation from vendor business decisions, health

How can large archives expand their use of open software?

- IT departments conditioned to purchase proprietary platforms
- Robust, scalable open solutions now exist
  - Ceph
  - GlusterFS
  - Swift
- Value proposition is compelling
- Invest in budding open source projects and communities

What interfaces will be used in archival storage in 1, 5, 10 years?

- Near-term
  - File
  - Object
- Long term
  - Probably not file-based POSIX (NFS, CIFS)

- HTTP-based REST protocols are simple
  - Standards like CDMI are appealing, but overrated

Final thought

- Every tax dollar spent on proprietary solutions
  - funds development of proprietary technology
  - funds marketing activities
  - benefits only a single vendor's products
- Every tax dollar spent on open source solutions
  - funds software that can benefit all users
  - builds solutions that can out-survive any vendor

Thank you!