



Meeting the Digital Preservation Needs of Smaller Organizations

Jefferson Bailey, Director of Archiving & Data Services
Internet Archive, jefferson@archive.org

About Internet Archive & Vault

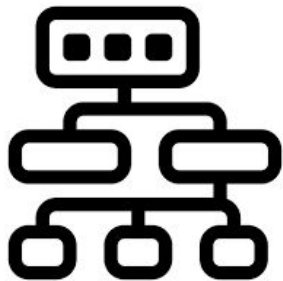


- IA founded 1996
- Non-profit digital library & archive
- Self-owned and operated data centers (7 in 3 countries)
- Digitization, archiving, web, and data products & services used by 2000+ mission aligned organizations in 40+ countries
- 99+ PB archived data (unique)
- Archiving 80-100 TB per day
- Top 150 website, 1.5M users per day
- Vault launched in 2022/2023
- Web application for digital preservation
- Data stored in IA data centers
- Multiple deposit methods, collection management features, fixity operations, integrations, etc
- Around 50 organizations currently using for all types of data and needs
- Majority are 100+ or fewer employees
- MG to TB scale collections

Vault Origins

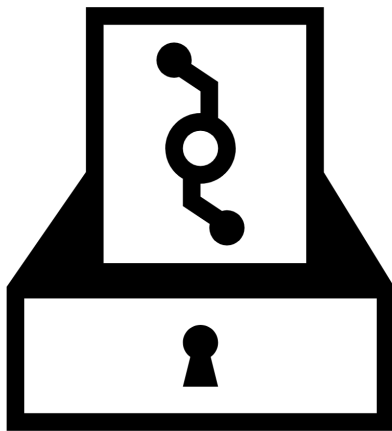


- IA provides a range of storage and preservation services, from very basic to highly customized contracted services
- Vault is a low-cost, extensible service that is more feature-rich than basic archive.org services and less feature-rich and expensive than our enterprise-scale contracted archiving / digital preservation services
- Vault features:
 - Content Agnostic
 - Custom Replication
 - Custom Geolocation (3 countries)
 - Multiple Storage Architectures (block & object)
 - Custom Fixity Audit/Repair Frequencies
 - All the deposit, API, integration, etc bells & whistles



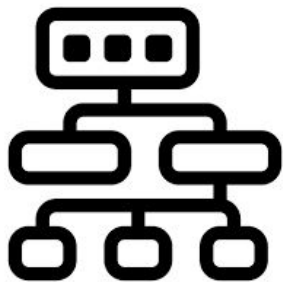
- **Cost, For-Profits, Vendor Lock:** Many services are provided by for-profits that are not necessarily values aligned in practice or pricing models, exit fees, etc
- **OSS:** DP OSS can have a very high barrier of entry as far as install, features, maintenance, dev time, etc
- **Commercial Cloud:** There is a push from “above” to use commercial cloud service for preservation w/o consideration of needs, costs, vendor lock, etc
- **Feature Creep/Retreat:** Lack of clarity on what a DP service should/shouldn’t do vs. in-house actions
- **NoColo:** Many DP systems/services are not colocated with collecting, processing, or access systems and this siloing introduces challenges
- **MoColo:** Many DP systems/services are just apps/layers on top of commercial cloud infrastructure

Value Propositions



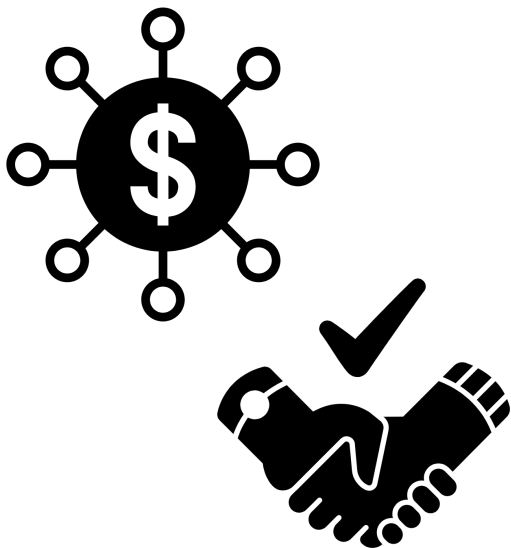
- Mission-aligned non-profit partner
- Self-owned and self-operated data centers in many locations, non-commercial cloud
- Geographic and geopolitical redundancy
- IA's efficiencies of scale and data center ops allows enable preservation services at costs far less than the commercial cloud
- One-time, per-TB cost, pay once preserve forever
- No annual storage or data egress costs
- A la carte pricing by feature
- Tiered costs for “public” vs “private” data
- Access to philanthropic and funder subsidies

Smaller Org Challenges in Digital Preservation



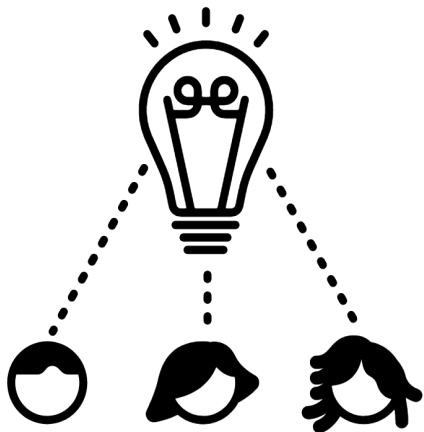
- **Getting out from under the desk:** Many services are provided by for-profits that are not necessarily values aligned in practice or pricing models, exit fees, etc
- **Getting out from under the cloud:** DP OSS can have a very high barrier of entry as far as install, features, maintenance, dev time, etc
- **Getting out from under the accessions:** There is a push from “above” to use commercial cloud service for preservation w/o consideration of specific/community needs
- **Getting out from under the budget:** Lack of clarity on what a DP service should/shouldn’t do vs. in-house actions
- **Getting out from under the bureaucracy:** Many DP systems/services are not colocated with collecting, processing, or access systems and this siloing introduces challenges

Smaller Org Digital Preservation User Stories



- **Nascent Digital Practice:** Early stages of digital preservation strategy; varying technical proficiency
- **Large-Scale, Sudden Acquisitions:** e.g. 100 TB of data from grant-funded project, 500 TB for a new film & media archive
- **Consortial Considerations:** nimble solutions for diverse users
- **Cost Considerations:** high cost of commercial cloud solutions or local implementations
- **Desired Features:** Geolocation, Dashboard tools/Monitoring, Replication functions

Vault Design Considerations for Smaller Orgs



- **Nascent Digital Practice:** Simple, familiar UI, extended features nested away; dashboard-orientation; multiple deposit options
- **Large-Scale, Sudden Acquisitions:** Simple front-end, scalable, complex (hidden) back-end
- **Consortial Considerations:** Variable access and authentication methods + high-impact integrations
- **Cost Considerations:** All-in-one, one-time cost for budgeting realities and core features; extra, ad-hoc costing for special features
- **Desired Features:** Do the core digress things well for everyone (replication, geolocation, fixity monitoring) with user-driven further feature development

THANKS!

Jefferson Bailey
Director of Archiving & Data Services
jefferson@archive.org
vault@archive.org

