Archive Analytics Solutions is presenting an archive system that embodies best practice for long-term, high integrity bit preservation.

But firstly ... many thanks for invaluable help and advice from

- Chris Burnet et al. from netapp.
- James Coomer from DDN
- Paul Watry et al. at the University of Liverpool, UK
**Bit Preservation**

- Immediate and unrelenting attention to operations
- Hardware / IT oriented.
- Unambiguous – do we still have the bits?

**Function Preservation**

- Deferrable (up to a point)
- Software oriented.
- Rules, workflows &c
- Ambiguous – what are we preserving, for whom?

Different skills, different technologies, different time-scales, different impacts, different players ... in fact

... just different
Alloy is about bit-preservation
## Some fundamentals of bit-preservation

| 1) Archives belong to the organization | Contain “finished” objects ... Access via “Roles” ... No “personal” users, |
| 2) Integrity. | Replication, Checksums, signing &c |
| 3) Traceable / tamper resistant | Audit logs on storage, signed &c. |
| 4) Properties ( aka user-metadata) | Finding via metadata ( c.f. search ). |
| 5) Reconstruct just from storage | No dependence on software No encryption or obfuscation Self-evident organization on storage |
| 6) Disaster resistant | Geographical Dispersion, placement constraints, + Documented, intuitive structures & naming |
| 7) “Vault” mentality | Limited access, highly controlled, ‘certifiable” No foreign objects inside the shell. |
| 8) Support for functional preservation | Versions, alternates, dependencies, |
An Open Source based solution for intelligent storing and manipulation/management of data for long periods of time

Alloy is:

- An archival solution that simplifies the creation and management of a TRAC* compliant archive by ensuring that good archival practice is readily practiced.
- Alloy integrates with existing IT infrastructures where possible, and enables archive applications via simple standards-based API(s).
Architecture

- Web interface
- Archival Applications (including ‘actions’)
- Alloy API (CDMI)
- Services (Storage & Catalog) Isolation Layer
  - iRODS
- Triggers & notifications
  - LDAP Server
  - RFC 5424 Server (syslog)

Layers:
- Authentication & Authorization (Auth’n & Auth’zn)
- Replication & Fixity (Repl’n & Fixity)
- "metadata"
- Logs & Audit
- Collections
- Objects
- Manifests
Using Alloy

- Single Front door
- RESTful (CDMI) base API for data access and admin
- Authentication via LDAP
- ‘out of the box’ Browser interface for humans
- Intuitive Replication & Tiering Policy specification
- User defined “properties” (metadata) and search on both names and values.
- IETF RFC-5424 (syslog) external trigger/rules system
- Authorization via roles and ACLs (authentication is LDAP)
- Both hierarchical (c.f. Posix / iRODS &c) & OID access.
Alloy also provides:

- Policy driven replication, with healing/rebalancing.
- Simple to scale out (appliance).
- Config-file driven deployment.
- Slaves manage storage and user access – scalable performance.
- Archive can be reconstructed just from storage – state is not stored only in Alloy.
- Rules can be attached to the IETF RFC 5424 listener – e.g. logstash, syslog-NG, rsyslog (or custom).
- All ‘provenance’ relevant logs are stored to archive.
Storage Tiering

4 - Fastest  0
3 - Standard  1
2 - Cloud     0
1 - Offline   0

Storage Sites

Only where all
- EU
- Scotland
- UK
- ed.ac.uk

Unassigned Labels
- .uk
- EEA
- Midlothian
- ac.uk

Prohibited where
- obsolete
Using Alloy – scripting and automation

• All interaction is via http(s) / CDMI 1.0.2
• Command line tool to perform add, retrieve, delete, list, of collections, objects and metadata.
  e.g. cd, mkdir, ls, get, put, rm, pwd ...

• Can also use `curl’, or scripting-languages via http.
Roadmap for the future

Functionality – applications on Alloy

- TRAC “wizard” – facilitate an ISO-16363 audit
- Escrow, retention and disposition manager
- Alternate API support (S3, SWIFT, CIFS(?) etc)
- Anything else that is widely needed.

Under the covers

- Additional storage layer support (Ceph, Propr’y Object stores – as requested)
- Master-less cluster
- Improved inter-operability support
- Alternate authentication (based on demand)
Availability

- Available under Apache 2.0 license.

- A turnkey installation is available from NetApp
Archive Analytics (AA), a UK company, creates application-specific iRods based applications and providers consultancy, operations and support for the same. AA’s products deliver robust archival, content management and analytic capabilities with explicit, transparent and audit-able guarantees of compliance to applicable laws, policies and practices and ease of configuration and use.

AA will be the solution of choice for cost-sensitive, data intensive applications in the higher education, bio-informatics, and media fields.

AA is a key partner in the iRods and related communities, collaboratively contributing to the development of the open-source platform.
A History of Open Source Collaboration

- Linux
- FreeBSD
- iSCSI
- NFS
- oVirt
- OpenStack
- SnapCreator
- RDMA
- iRODS
- OSF
- DCE
- NDMP

© 2013 NetApp, Inc. All rights reserved.
Broker of Services – Risk and Reward

Cloud Service Providers

Private Cloud

Hyperscale Cloud Providers

Performance
Availability
Protection
Governance

Risk

Access

Security

©2014 NetApp, Inc. All rights reserved.
Alloy - Archive as a Service

Data Centre 1

Alloy Appliance providing replicated storage from an on-site Archive

Replication

Replication of data (files) and catalog (SQL) from DC1 to DC2

Replication of data (files) and catalog (SQL) from DC2 to DC1

Data Centre 2

Alloy Appliance providing replicated storage from an on-site Archive

Client A

Alloy Appliance

File Servers

Network Shares

Client B

Alloy Appliance

File Servers

Network Shares

©2014 NetApp, Inc. All rights reserved.
Thank you.....