

2009 NGDA Development Plans

Original aspirations

- Formulate foundational principles
 - standard operating procedures (SOPs)
 - to guarantee base level of preservation
 - analogy: electrical wiring conventions
 - SOP: black=hot, white=neutral
 - to avoid shocks, miswiring
- Demonstrate principles in action

Principles

- Capture & archival of semantics
- Uniform representation of data & semantics
- Survivable physical representations
- Maintenance of dependencies
- Independence of layers
 - library/repository
 - repository/storage

Where are we now?

- Capture & archival of semantics ★★
- Uniform representation of data & semantics ★★
- Survivable physical representations ★★
- Maintenance of dependencies ★★
- Independence of layers
 - library/repository ★★
 - repository/storage ★★

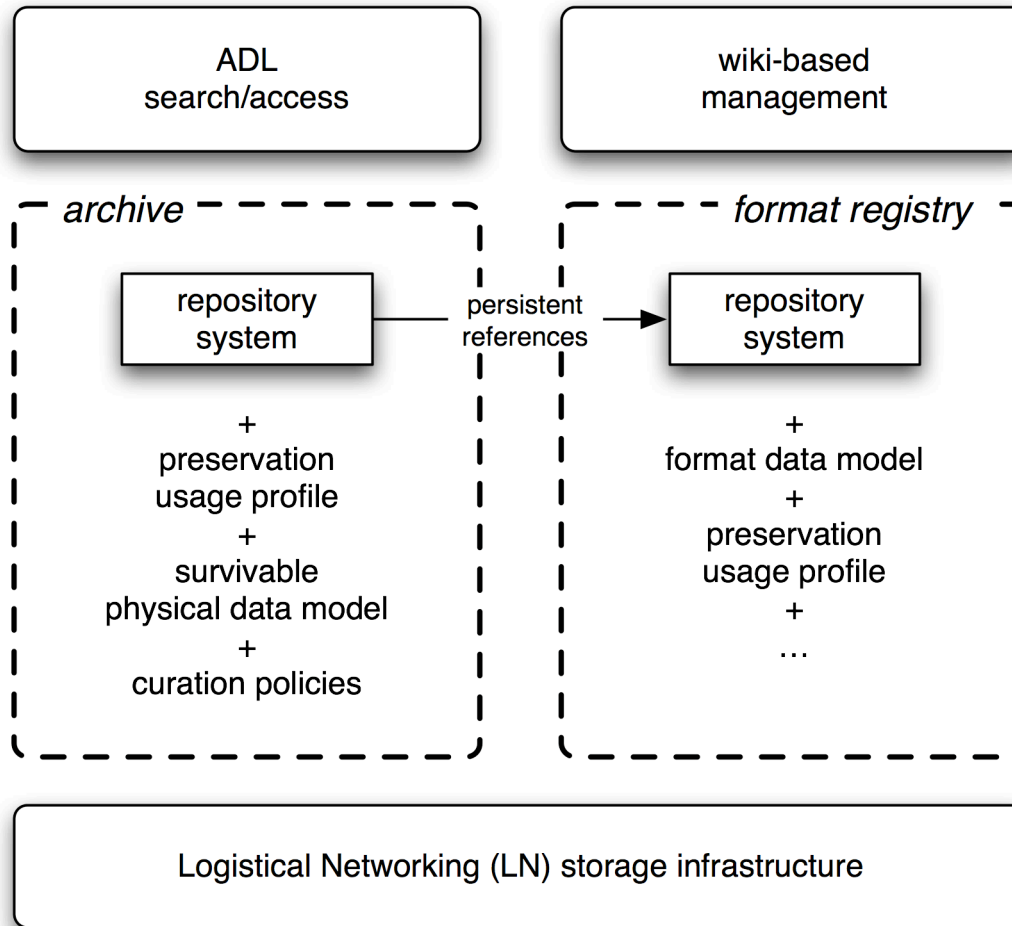
Development (I)

- Move to Fedora-based system
 - (research) to delineate principles
 - (practice) to facilitate federation, ingest
- Extensions needed:
 - storage abstraction
 - physical representation of archive metadata
 - standard representation of formats, other SBOs (semantics-bearing objects)
 - maintenance of dependencies
 - inter-object
 - inter-archive
 - archive/repository descriptor

Development (II)

- ADL on Fedora
 - to provide meaningful access to data
- Modules needed:
 - OAI crawler integrated with metadata mapper
 - universal collection driver (MINT)

1,000 words



Development (III)

- Federation documentation: how to join
 - contractual
 - as depositor
 - as node
 - technical
 - ADL
 - repository
 - storage (?)

Development (IV)

- Population
 - of archive
 - of format registry

Will we have succeeded?

- ✓ Will have a copy of CaSIL
 - ✗ but CaSIL is a website
 - ✗ but CaSIL is living entity; ours is a snapshot
- ✓ Will provide data access via ADL
 - ✗ but not to CaSIL's index maps
 - ✗ but differently organized
- ✓ Will have distributed copies of data
 - ✗ but no (end-to-end) proactive integrity checking
 - ✓ NGDA archive data model includes checksums
 - ✗ but in Fedora?

Will we have succeeded?

- ✓ Will have survivable representations of data & semantics
 - ✗ but use, usability are best guarantors of successful preservation
- ✓ Will have semantics (formats, etc.)
 - ✗ but there's Wikipedia and the larger world in general
 - ✗ but not needed (now)
 - Rosenthal: "format obsolescence is the prostate cancer of preservation"