Complex Data Sets, Software Preservation and Emulation

A distributed approach to long-term care

RDAP Summit 2018
March 21-23, 2018

The FAIR Guiding Principles

The goal of ‘long-term’ care is **re-use**

The FAIR Guiding Principles

Analysis revealed **a wide range of file formats**
represented across data sets reviewed

The FAIR Guiding Principles

Two-thirds of the formats identified were complex, proprietary or binary formats that **introduce version-specific software dependencies**

Challenge 1

No single organization can collect all the software (and hardware) needed to facilitate access to digital data

Implication:

Coordinated collection development and a mechanism to support sharing and reuse of software collections across organizations

Challenge 2

Emulation approaches have not scaled and there are dependencies on third-party emulators

Implication:

Leverage existing ecosystem of digital preservation networks/service providers

Create a formal business model to support sustained development of emulators

Build an emulation network that facilitates distributed preservation, sharing and reuse of software and configured software environments

Challenge 3

Copyright culture and DRM associated with software distributed on installation media

Implication:

Leverage existing legal tools

Build consensus

Challenge 4

Changing distribution models: we can't copy what we don't have in hand.

Implication:

Aligning shared interests and representing the needs of cultural memory and research organizations

Coherent articulation of broad research and education reuse needs in licensing discussions with industry representatives

Software Preservation Network

Saving Software Together.

Preserving software through community engagement,
infrastructure support, and knowledge generation.

Our Mission

Collective Impact Approach to Software Preservation

Moving the needle for a broader set of stakeholders.

Our Mission

Software Preservation Network and Affiliated Projects

Framing the work.

EaaSY

Develop a scalable emulation infrastructure.

Addresses:

Challenge 1: No single organization can collect everything they need

Challenge 2: Emulation approaches have not scaled.

Details:

- **Distributed management:** a network of emulation nodes
- **Sharing:** enable sharing of software images and configured software environments among nodes in the network
- **Discovery:** software and configured in the network will be discoverable via a shared index and the integration of Wikidata
- **Access:** APIs for network sharing, virtual reading rooms, reproducibility for computationally-dependent research

ARL Best Practices Code

Outline clear guidelines for the application of fair use to the preservation of software.

Addresses:

Challenge 1: No single organization can collect everything they need

Challenge 3: Copyright culture and DRM associated with software distributed on installation media

Details:

- Extensive interviews
- Focus Groups
- Building consensus among stakeholders regarding best practices for software preservation

FCoP

Cohort of six different types of organization and six different software preservation use cases.

Addresses:

Challenge 3: Copyright culture and DRM associated with software distributed on installation media

Challenge 4: Changing distribution models: we can't copy what we don't have in hand.

Details:

Creating documentation about how to integration emulation and software preservation activities into existing digital curation workflows.

Sign Up to Learn More About:

SPN
FCoP Project Newsletter
EaaSY Project Newsletter
SPN Technological Infrastructure Working Group