# **Digital Preservation**

## **DMFUG 2017**



### The need, the goal, a tutorial

- In 2000, the University of California, Berkeley estimated that 93% of the world's yearly intellectual output is produced in digital form
- To maintain the ability to display, retrieve, and use digital material in the face of rapidly changing technological and organizational infrastructures
- MIT Libraries tutorial is here http://www.dpworkshop.org/index.html

### Symptoms

### If the media is still online but data can't be accessed, probably the reasons are:

- 1. Corrupted file or media
- 2 Un-identified file format
- 3. Missing rendering application

**Compressed (zipped) Folders Error** 

OK



### **Digital Preservation Strategies**

### Refreshing

The transfer of data between two types of the same storage medium so there are no bitrot changes or alteration of data

#### Replication

Exact duplicates allow archives and libraries to manage, store, and provide access to identical copies of data across multiple systems and/or environments

### Emulation

Combines software and hardware to reproduce the performance of another computer of a different design, allowing programs or media designed for a particular environment to operate in a different, usually newer environment. Requires the creation of programs that translate code and instructions from one computing environment so it can be properly executed in another.

#### Migration

To copy or convert data, from one technology to another

### • Fixity

File fixity is the property of a digital file (aka bit stream) being fixed, or unchanged. File fixity checking is the process of validating that a file has not changed or been altered from a previous state. This effort is often enabled by the creation, validation, and management of checksums.

#### Open Archival Information System (OAIS) - Functional Reference Model (ISO 14721)



#### MANAGEMENT

OAIS has become the standard model for digital preservation systems.

5

OAIS-compliance has been a stated fundamental design requirement for major digital preservation and repository development efforts at the <u>National Archives and Records</u> <u>Administration, Library of Congress, British Library, Bibliothèque nationale de France, National Library of the Netherlands</u>, the <u>Digital Curation Centre</u> in the UK, <u>OCLC</u> (the Online Computer Library Center), the <u>JSTOR</u> (Journal Storage) scholarly journal archive, as well as several university library systems. The OAIS has been the basis of numerous prominent digital preservation initiatives and standards including the <u>Preservation Metadata: Implementation</u> <u>Strategies</u> working group and the <u>Trustworthy Repositories Audit & Certification</u> (TRAC) document from OCLC.

https://en.wikipedia.org/wiki/Open\_Archival\_Information\_System

### Active preservation principles

- 1) Ensuring bit integrity over the long term (decades/centuries)
- 2) Ensuring content health
  - Format viability
  - Complete metadata
  - Provenance
- 3) Open Standards compliance

### Format Viability – part I

• Media becomes unreadable when the devices to read it are no longer available





7

©2017 SGI

### Format Viability – part II

- File formats become obsolete for a number of reasons:
  - Software upgrades fail to support legacy files
  - The format itself is superseded by another or evolves in complexity
  - "take up" is low or industry fails to create compatible software
  - The format fails, stagnates, or is no longer compatible with the current environment
  - Software supporting the format fails in the marketplace or is bought by a competitor and withdrawn

- Access requires two things to happen:
  - Positive ID of format e.g. pdf v1.7
  - SW application availability e.g. Acrobat reader



### Metadata

#### Technical metadata

- ✓ Size
- ✓ Resolution
- ✓ Compression
- ✓ etc

#### Descriptive metadata

- ✓ Author
- ✓ Title
- ✓ publisher
- ✓ etc

### Provenance Metadata

- ✓ Date, time, occasion
- ✓ Why created
- ✓ IP ownership
- ✓ etc



### Migration

- Conversion from legacy format to;
  - ✓ an Open standard,
  - ✓ application neutral,
  - ✓ self describing format,
    - ✤ XML (eg. ODF)
  - ✓ or flat text



#### ISO 32000-1:2008

### ISO 8859-1 ≠ UTF-8

### ISO/IEC 26300:2006



### Information Package – Book example



11

©2017 SGI

### Information Package – Image example



©2017 SGI

### Active Preservation – Ingest Processing

Manages preservation planning process from risk to action

13

- Allows evaluation and comparison of alternatives
- Manual or automatic



©2017 SGI

Conditional or free text search. Web based services use indexes extracted from DPS metadata



**Dissemination Information Package** 

SGI Hierarchical storage lends itself well to DPS applications

- T10-PI RAID for working storage and dissemination data
- scalable capacity tier preserves integrity using hardware and DMF software features

http://digitalpowrr.niu.ed	u <b>/tool-grid,</b> Azure	Amazon S3	Archivematica	Chronopolis	Dropbox	DSPS (Digital Preservation Software Platform)	Glacier	MIXED (Migration to Intermediate XML for Electronic Data)	Portico	Preservica	RackSpace	Roda	Rosetta	Xena
Сору		х	x	х	х	х	х			х		х	х	
Fixity check		х	x	х		х	х		х	х		х	х	
Virus Scan			x			х				х		х	х	
File Dedupe				х	х	х			х	х			х	
Auto Unique ID		х	x	х		х	х		х	х		х	х	
Auto Metadata Creation		х	x	х		х	х		х	х		х	х	х
Auto Metadata Harvest			x	х					x	х			х	х
Manual Metadata			x	х					х	х		х	х	
Rights Management			x	х					х	х		х	х	
Package Metadata			x	х		х			x	х		х	х	х
Auto SIP creation			x	х					x	х		х	х	
Public Interface	x				х				х	х	х	x	х	
Auto DIP Creation			x						x	х		x	x	
Auto AIP Creation			x	х		х			x	х		x	х	
Redundancy		х		х		х	х		x	х		x	х	
Reliable, Long-Term Bit Preservation	x	x		x	x	x	x		x	x	x	x		
Geographically Dispersed Data Storage Model	x	x		x	x		x		x	x	x	x		
Exit Strategy		х		х	х	х	х			х	х	x	х	
Migration			х	х		х		х	x	х	х	х	х	
Monitoring		х		х		х	х		х	х	х	х	х	
Auto Recovery		x		х			х			х	х	х	х	
Open Source			х			х					х	х		х
Clear Documentation		х	x	х	х	х	х		x	х	х	x	х	х
Cost	Varies	Varies	Free	\$1,500 – \$2,200	Varies	Free	Varies	Free	Various	Varies	Varies	Free	Varies	Free



sg

