INTRODUCTION

DIGITAL PRESERVATION

1990S: PERSONAL COMPUTING & THE INTERNET

DIGITAL GOES MAINSTREAM, HUMANS GO INSANE

1990S: PERSONAL COMPUTING & THE INTERNET

DIGITAL GOES MAINSTREAM, HUMANS GO INSANE

1990S: PERSONAL COMPUTING & THE INTERNET

DIGITAL GOES MAINSTREAM, HUMANS GO INSANE

Lynch, Clifford (1993) Accessibility and Integrity of Networked Information Collections. Office of Technology Assessment, Congress of the United States, July 5.

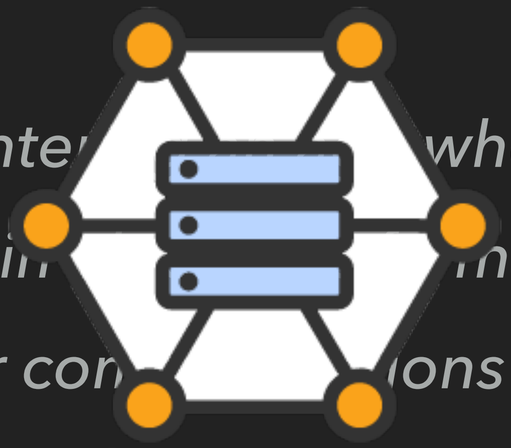
1990S: PERSONAL COMPUTING & THE INTERNET

DIGITAL GOES MAINSTREAM, HUMANS GO INSANE

We are entering an age when a great deal of information is available in electronic formats and can be obtained through computer communications networks. Sometimes, such collections of ***network-accessible electronic information*** *are referred to as* ***“digital libraries.”***

1990S: PERSONAL COMPUTING & THE INTERNET

DIGITAL GOES MAINSTREAM, HUMANS GO INSANE



We are ente[illegible] when a great deal of information is available in [illegible]mats and can be obtained through computer con[illegible]ions networks. Sometimes, such collections of ***network*****-accessible electronic information** *are referred to as* ***"digital libraries."***

1990S: PERSONAL COMPUTING & THE INTERNET

DIGITAL GOES MAINSTREAM, HUMANS GO INSANE

*... I view this terminology as **somewhat misleading**; indeed, one of the issues explored here is the developing roles of such electronic information collections and their relationships to institutions such as libraries.*

1990S: PERSONAL COMPUTING & THE INTERNET

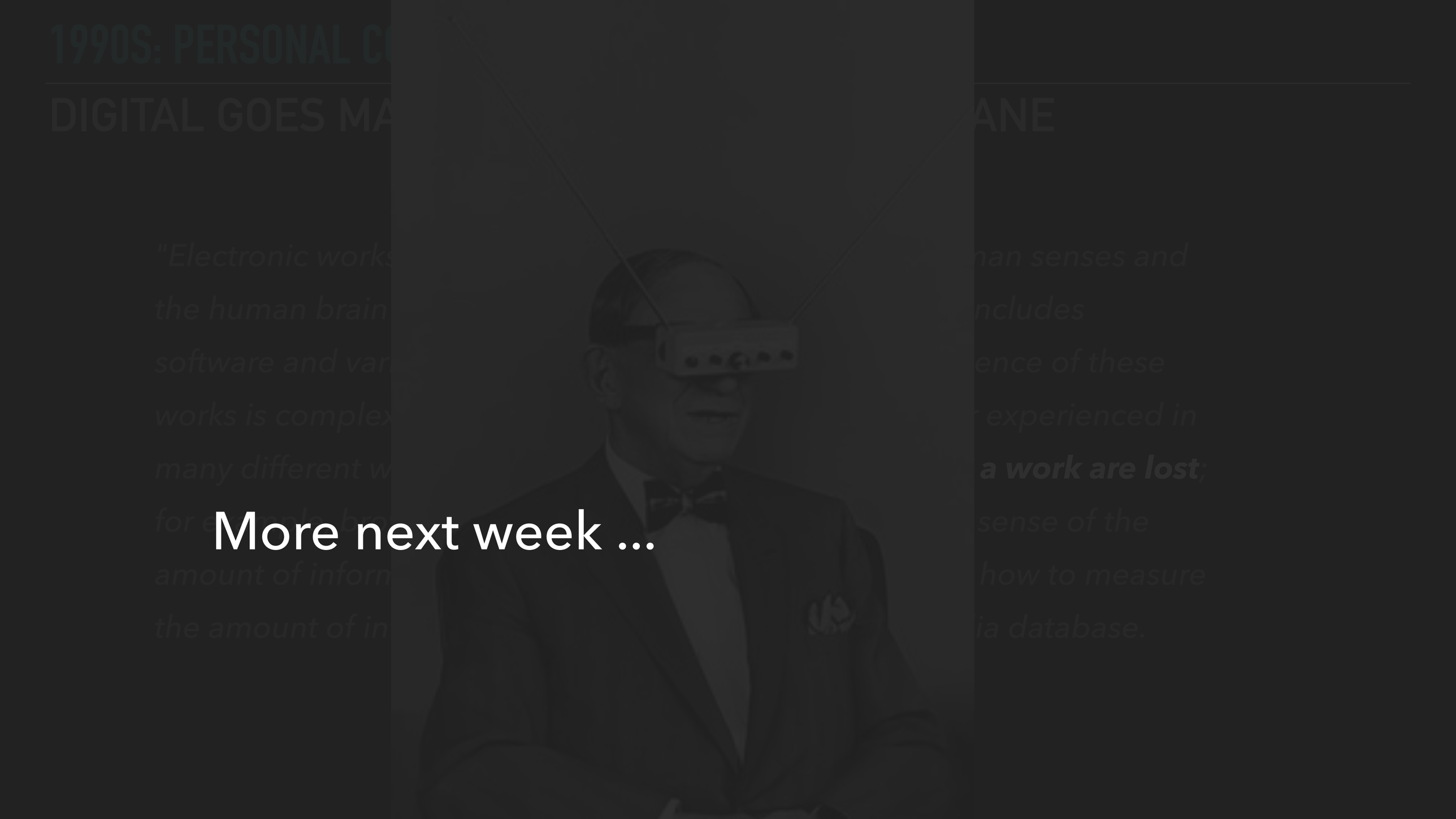
DIGITAL GOES MAINSTREAM, HUMANS GO INSANE

***"Electronic works"**... can only be apprehended by the human senses and the human brain interacting with a computer system that includes software and various input and output devices. The experience of these works is complex and interactive; a work can be viewed or experienced in many different ways. Further, **other intuitive measures of a work are lost**; for example, browsing a printed work gives the browser a sense of the amount of information that the work contains.*

1990S: PERSONAL C

DIGITAL GOES MA ANE

"Electronic works *uman senses*
and the human br that includes
software and varic rience of
these works is cor wed or
experienced in m ***ive measures***
of a work are lost *gives the*
browser a sense o rk contains.



More next week ...

MICRO LEVEL

MICRO LEVEL

MACRO LEVEL

MICRO LEVEL

DIGITAL OBJECT

MICRO LEVEL

DIGITAL OBJECT

MICRO LEVEL

FINAL CUT 7 PRO PROJECT FILE

OBSOLETE DIGITAL OBJECT

FINAL CUT 7 PRO PROJECT FILE

OBSOLETE DIGITAL OBJECT

- FCP7 software *support dropped* by Apple in 2009
- FCP7 code is *proprietary*
- FCP7 project files *not compatible* with FCPX
- FCP7 *does not function* in current macOS

WHAT WILL BE PRESERVED?

WHAT WILL BE PRESERVED?

CONTENT

WHAT WILL BE PRESERVED?

CONTENT

- The contents of the file itself
  - instructions for edit itself (EDL)
  - color correction info
  - graphics, motion, fades, subtitles, etc.

WHAT WILL BE PRESERVED?

CONTENT

HOW WILL IT BE PRESERVED?

PHYSICAL PRESERVATION

HOW WILL IT BE PRESERVED?

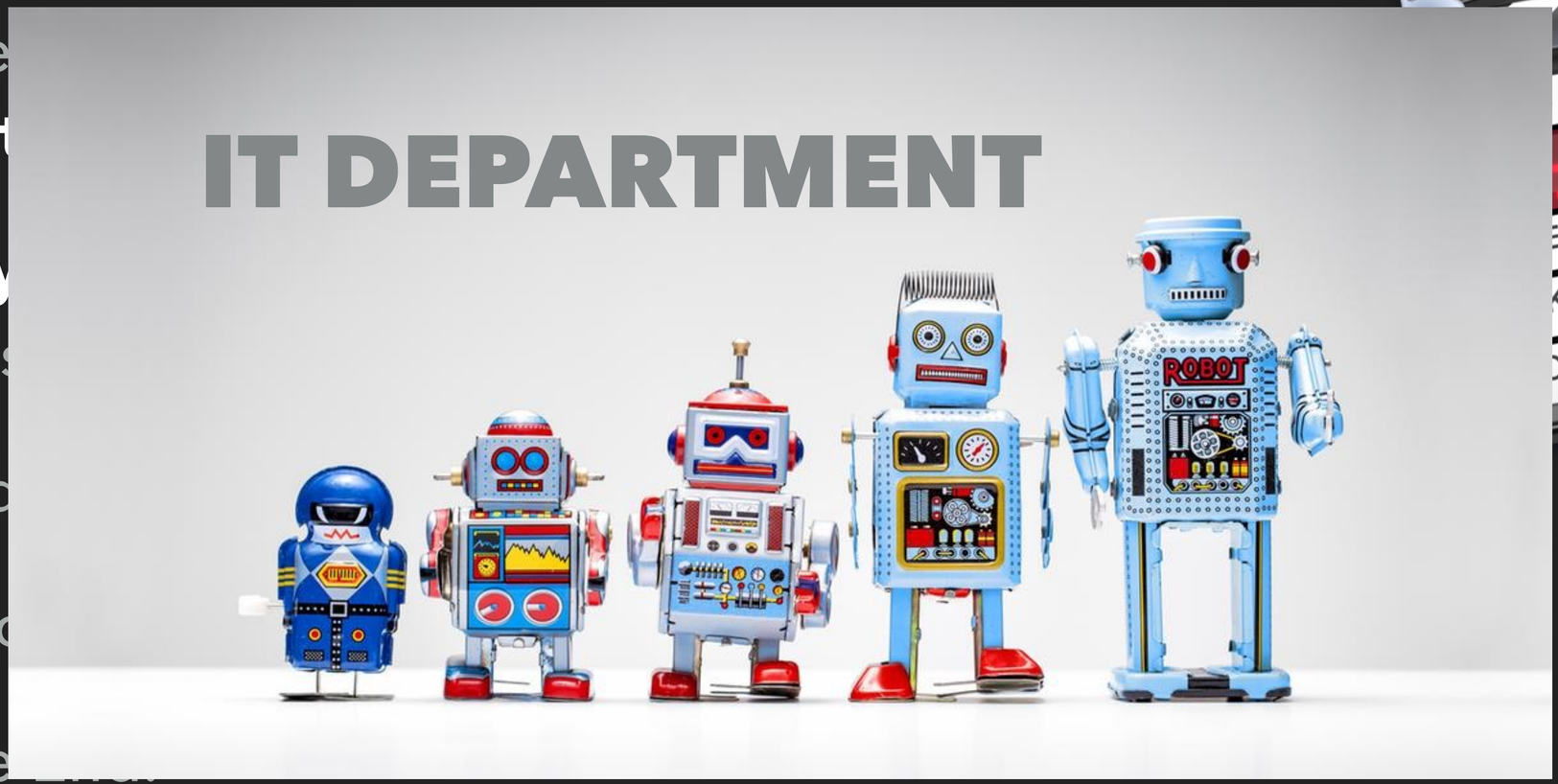
PHYSICAL PRESERVATION

- Every file is made of 0s and 1s (bits)
  "Bitstream"
- **Physical or "Bit" Preservation**
  Preserve bits in sequential order
- Redundant backups
- Geographical separation
- The End.

HOW WILL IT BE PRESERVED?

PHYSICAL PRESERVATION

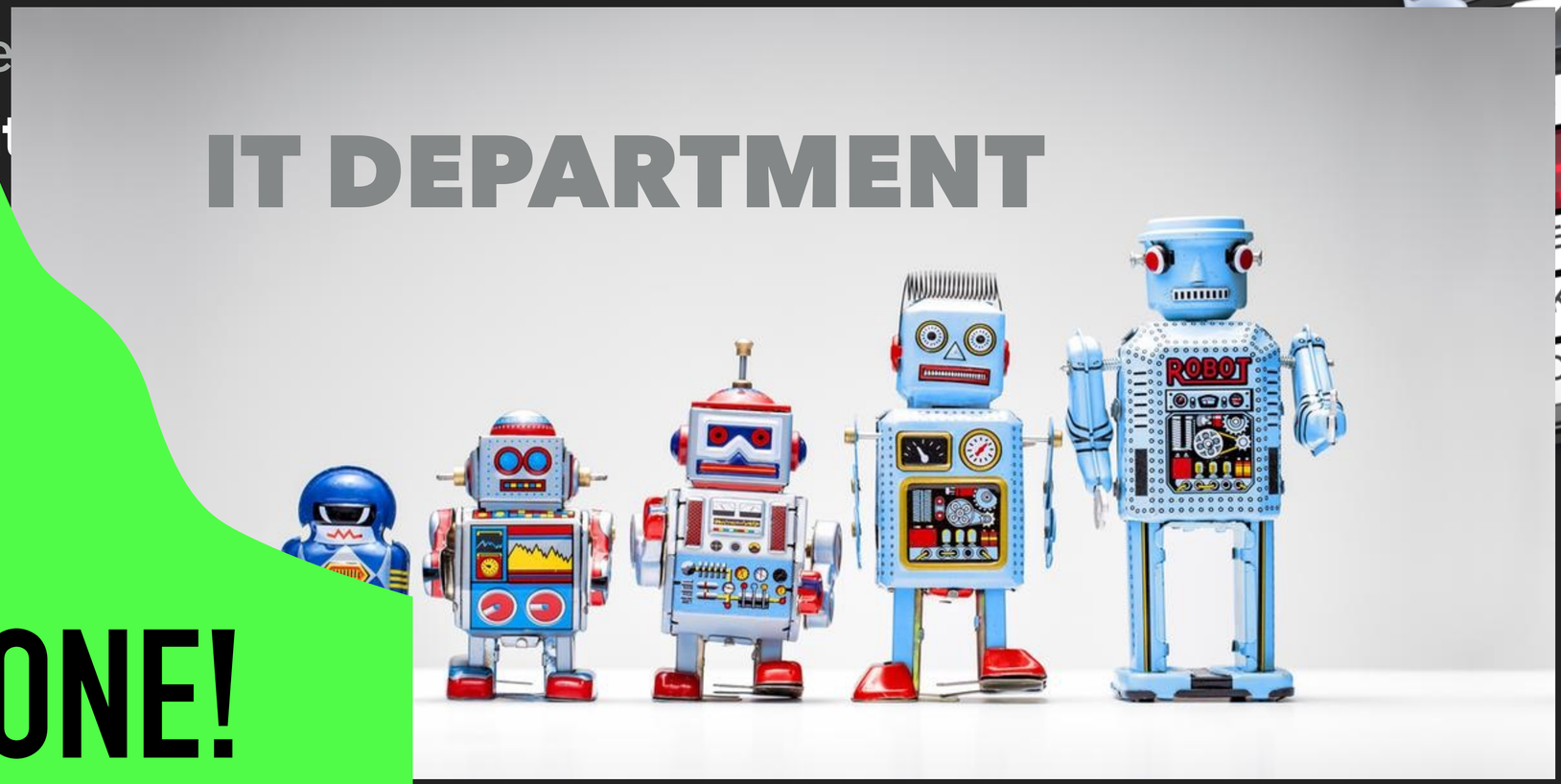
- Eve
  "Bit
- **Phy**
  Pres
- Rec
- Geo
- The



HOW WILL IT BE PRESERVED?

PHYSICAL PRESERVATION

- Eve
  "Bit
- **I**
  P



DONE!

CAN HUMANS USE IT?

RENDERABILITY

CAN HUMANS USE IT?

RENDERABILITY

- **Renderable**
  Digital object must be viewable and playable.

- **Logical Preservation**
  "...ensure [files] are still *understandable* and *readable*, regardless of the evolving technologies"

  - Digitising Contemporary Art

HOW DO WE RENDER FILES?

HOW DO WE RENDER FILES?

COMPUTER HARDWARE & SOFTWARE

- Digital objects are dependent on hardware and software

HOW DO WE RENDER FILES?

COMPUTER HARDWARE & SOFTWARE

- Digital objects are dependent on hardware and software
- Q: What dependencies are required to run a Final Cut Pro 7 project files?

HOW DO WE RENDER FILES?

COMPUTER HARDWARE & SOFTWARE

- Digital objects are dependent on hardware and software
- Q: What dependencies are required to run a Final Cut Pro 7 project files?
  - FCP7 software program
  - Operating system: Optimized for Mac OSX 10.6 (2009)
  - Hardware that runs Mac OSX 10.6

HOW DO WE RENDER FILES?

COMPUTER HARDWARE & SOFTWARE

- Digital objects are dependent on hardware and software
- Q: What dependencies are required to run a Final Cut Pro 7 project files?

Document this info as preservation metadata

  - FCP7 software program
    - Operating system: Optimized for Mac OSX 10.6 (2009)
    - Hardware that runs Mac OSX 10.6

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation
  - Backup : LTO, redundancy, geo separation

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation
  - Backup : LTO, redundancy, geo separation
- Logical preservation

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation
  - Backup : LTO, redundancy, geo separation
- Logical preservation
  - Software Environment: Mac OSX 10.6, software applications and dependencies

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation
  - Backup : LTO, redundancy, geo separation
- Logical preservation
  - Software Environment: Mac OSX 10.6, software applications and dependencies
  - Hardware Environment: Mac Pro (2006)

MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation
  - Backup : LTO, redundancy, geo separation
- Logical preservation
  - Software Environment: Mac OSX 10.6, software applications and dependencies
  - Hardware Environment: Mac Pro (2006)
  - Emulation/Migration

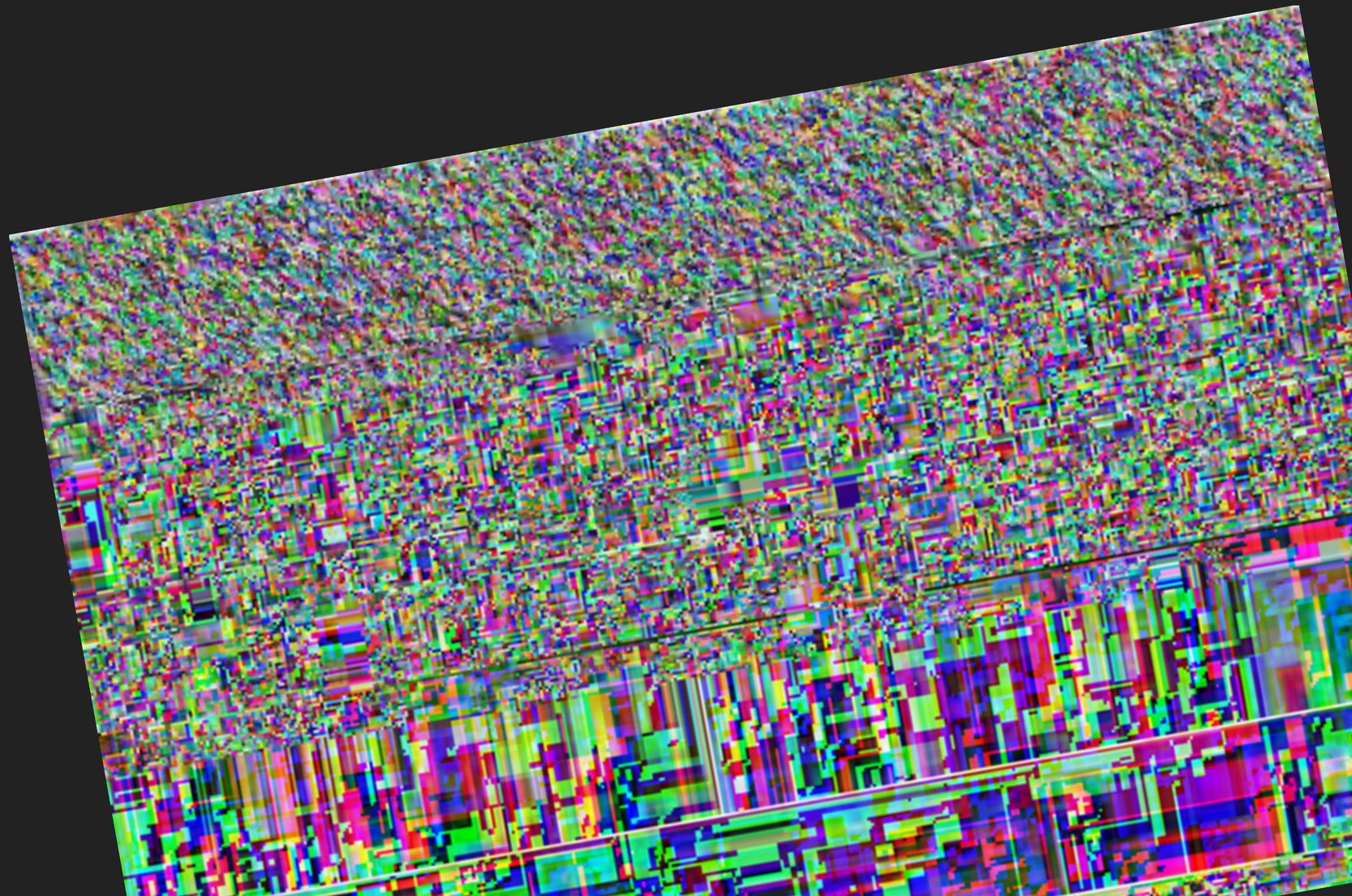
MICRO-LEVEL: DIGITAL OBJECT

REVIEW: BASIC DIGITAL PRESERVATION

- Physical, bit-level preservation
  - Backup : LTO, redundancy, geo separation
- Logical preservation
  - Software Environment: Mac OSX 10.6, software applications and dependencies
  - Hardware Environment: Mac Pro (2006)
  - Emulation/Migration
  - File Formats: Research compatibility

MICRO-LEVEL: DIGITAL OBJECT

DATA INTEGRITY & DEGRADATION



MICRO-LEVEL: DIGITAL OBJECT

DATA INTEGRITY & DEGRADATION

FLIPPED BITS

BIT ROT

CORRUPTION

MICRO-LEVEL: DIGITAL OBJECT

DATA INTEGRITY & DEGRADATION

FLIPPED BITS

BIT R

COR

MICRO-LEVEL: DIGITAL OBJECT

DATA VERIFICATION

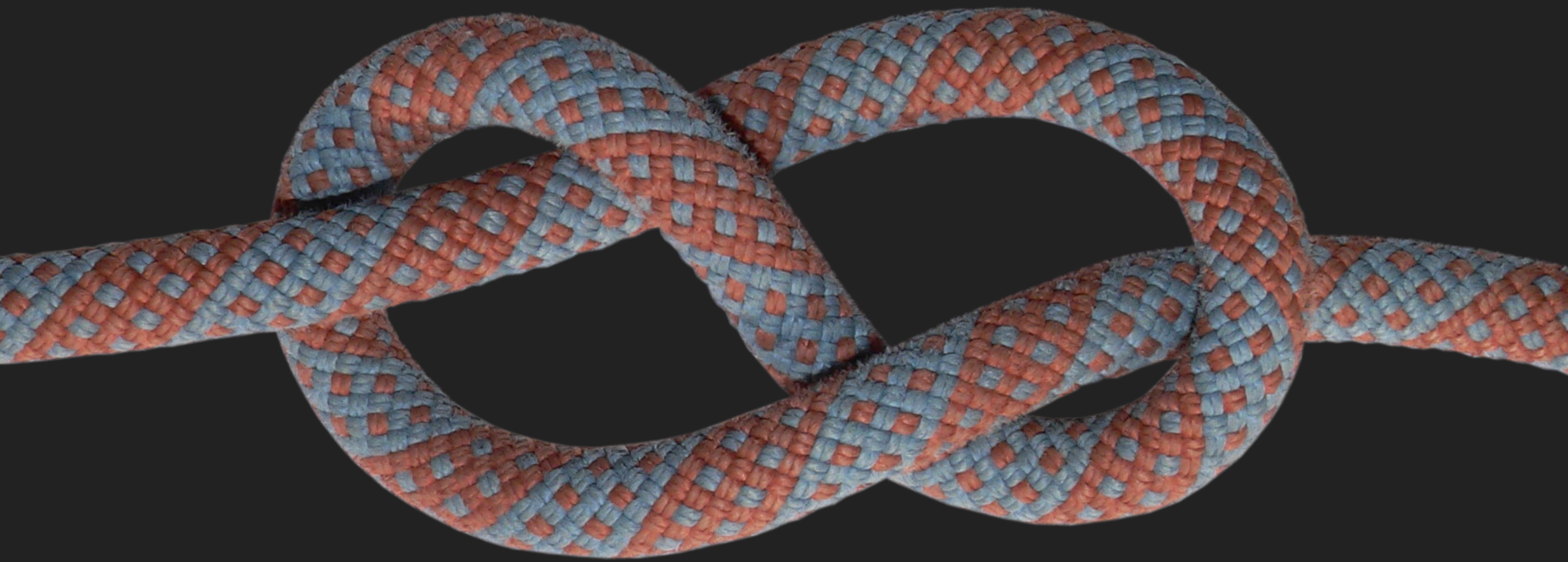
MICRO-LEVEL: DIGITAL OBJECT

DATA VERIFICATION

- Run automated or manual fixity checks
- Name/size checks for low-tech archives
- File format validation (JHOVE, DROID)
- Checksums or "hashes" in digital repository
  - MD5, SHA-1
- Create/monitor hashes throughout lifecycle

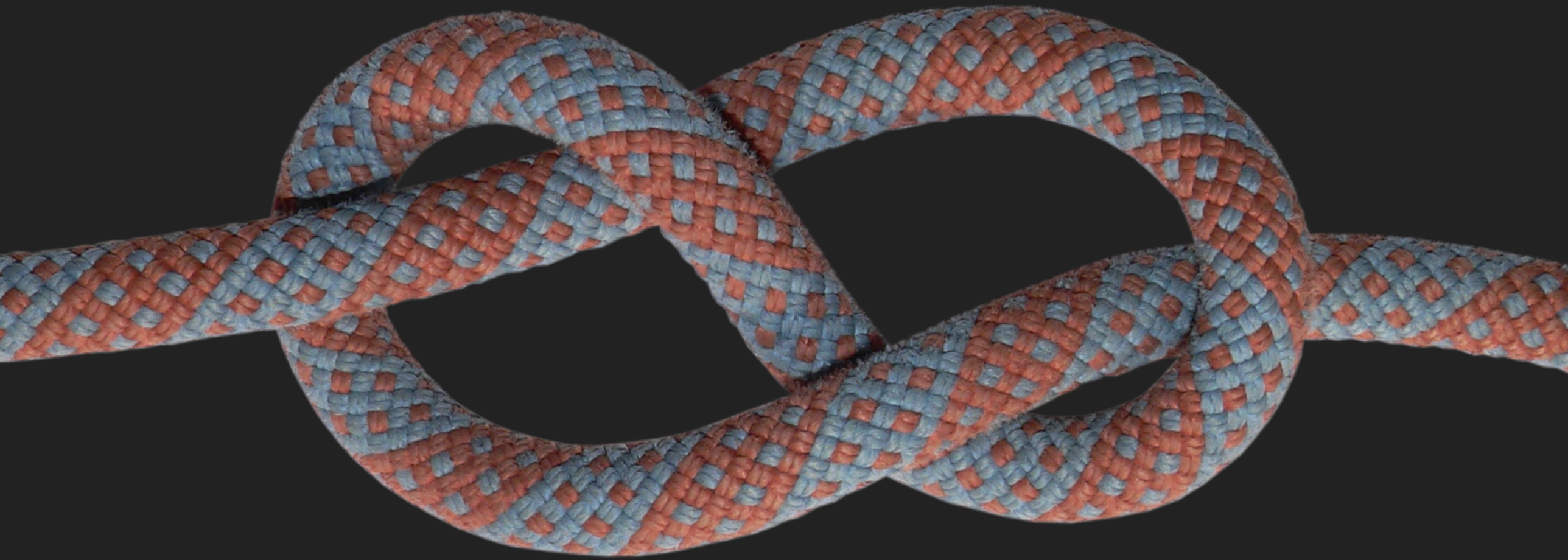
METADATA & DESCRIPTION

METADATA & DESCRIPTION



METADATA & DESCRIPTION

WHY DO WE NEED IT?



METADATA & DESCRIPTION

WHY DO WE NEED IT?

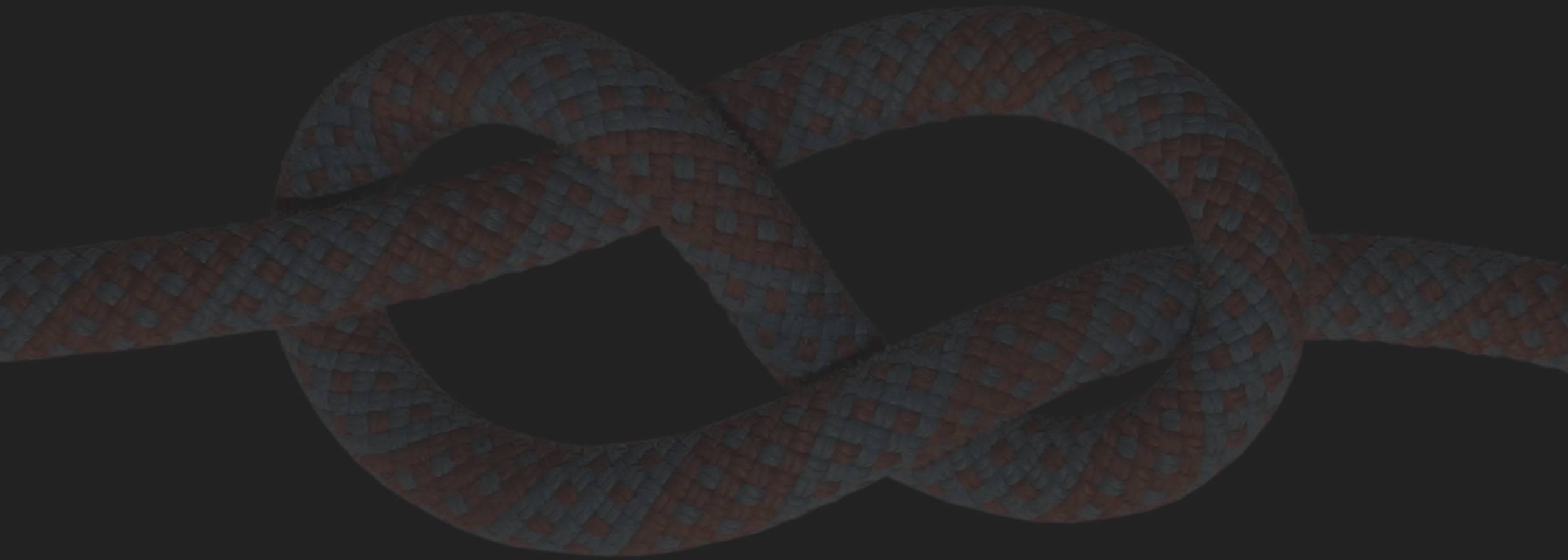
- **Discovery**
  - Digital objects have no physical form
  - If it's not in the catalog, it doesn't exist
- **Identification**
- **Structure**
- **Administration**

METADATA & DESCRIPTION

- Q: What metadata might be embedded in this file at the time of accession or ingest?

METADATA & DESCRIPTION

AUTOMATED METADATA HARVESTING



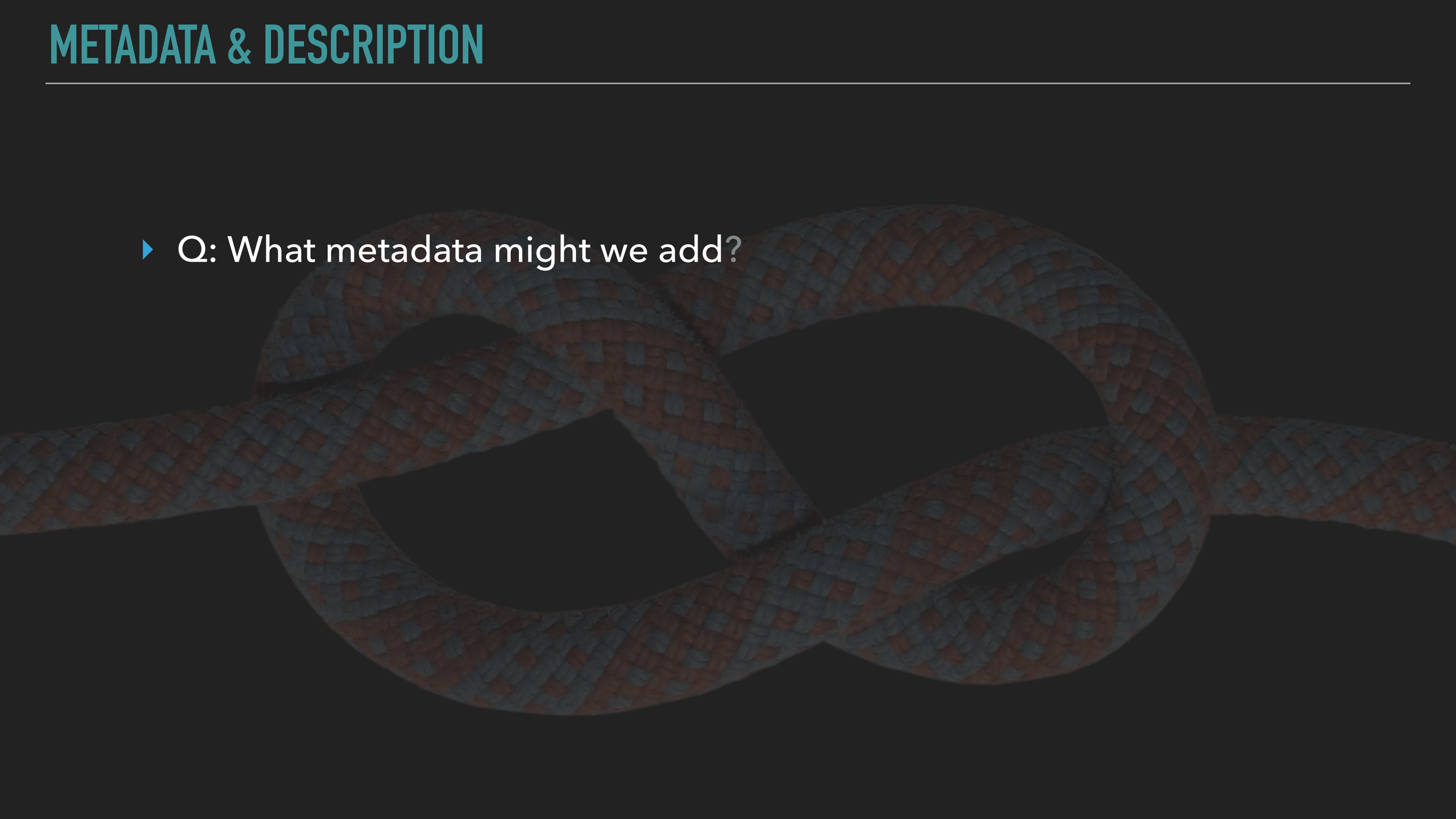
METADATA & DESCRIPTION

AUTOMATED METADATA HARVESTING

- Descriptive
  - File name, creator/author
- Administrative
  - Environment: Software/hardware used to create this file
  - Provenance: Creating application, date
- Structural
  - Technical specifications
  - Structure: Pages, chapter markers (ebook)

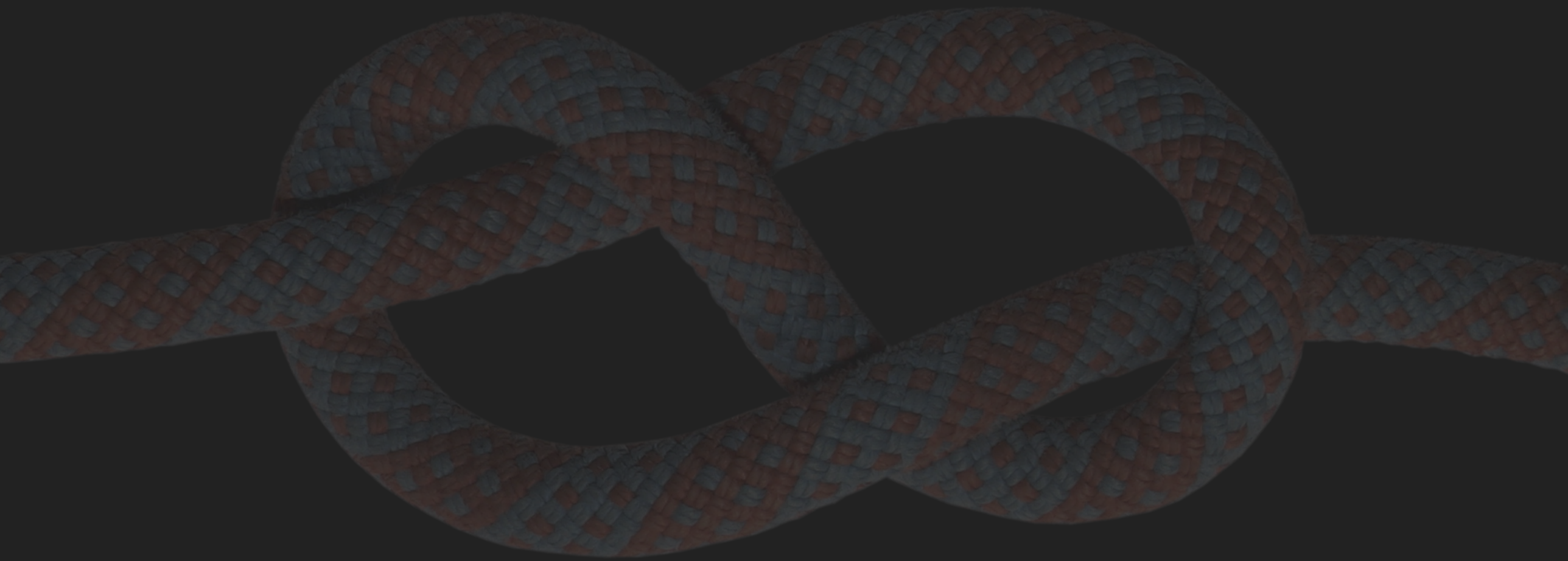
METADATA & DESCRIPTION

- Q: What metadata might we add?



METADATA & DESCRIPTION

MANUAL, HUMAN-POWERED CATALOGING



METADATA & DESCRIPTION

MANUAL, HUMAN-POWERED CATALOGING

- Descriptive
  - Title, subject, language, additional creators (credits)
- Administrative
  - Rights (copyright, license info, terms of use)
  - Environment: Additional hardware/software compatibility information
  - Provenance: prior use context
- Structural
  - References/relationships (original media)

MACRO LEVEL

DIGITAL REPOSITORY

MACRO LEVEL

DIGITAL REPOSITORY

MACRO LEVEL

MACRO-LEVEL: DIGITAL REPOSITORY

ACTIVITIES OF A DIGITAL REPOSITORY

MACRO-LEVEL: DIGITAL REPOSITORY

ACTIVITIES OF A DIGITAL REPOSITORY

- Appraisal & Selection
- Ingest & Accession Policy
- Storage
- Migration & Emulation
- System Sustainability
- Economic & Environmental Sustainability
- Risk Management

MACRO-LEVEL: DIGITAL REPOSITORY

APPRAISAL & SELECTION

MACRO-LEVEL: DIGITAL REPOSITORY

APPRAISAL & SELECTION

- Unique, obsolete or at-risk content
- Concept of "rare" may apply differently for digital objects
- Collection assessment evaluation

MACRO-LEVEL: DIGITAL REPOSITORY

INGEST & ACCESSION POLICY

MACRO-LEVEL: DIGITAL REPOSITORY

INGEST & ACCESSION POLICY

- Receive/prepare objects for ingest
  - Receive objects (drives, download, FTP, etc.)
  - Identify, validate, run fixity checks
  - Describe/catalog objects
  - Impose standards (naming & arrangement, etc.)
- Perform these tasks in accordance with Accession Policy

MACRO-LEVEL: DIGITAL REPOSITORY

STORAGE

MACRO-LEVEL: DIGITAL REPOSITORY

STORAGE

- Store locally
  - server
  - client
  - LTO tape
- Create redundant copies (duplicate or triplicate)
- Geographical separation ("cloud" or elsewhere)

MACRO-LEVEL: DIGITAL REPOSITORY

MIGRATION/EMULATION

MACRO-LEVEL: DIGITAL REPOSITORY

MIGRATION/EMULATION

- Migration: Change file format (transcode)
- Emulation: Create virtual computing environment

MACRO-LEVEL: DIGITAL REPOSITORY

MIGRATION/EMULATION

- Identify obsolete, at-risk (ex: FCP7 file)
- Produce renderable alternative (ex: EDL)
- Challenges of migration/emulation:
  - Expensive (requires engineering, storage)
  - Not always successful (information loss)
  - Works best for simple formats (text)
- Migration: Not great for AV or complex media (video, layered image formats: CAD, graphics)
- Emulation: More successful than migration for comp

MACRO-LEVEL: DIGITAL REPOSITORY

SYSTEMS SUSTAINABILITY

MACRO-LEVEL: DIGITAL REPOSITORY

SYSTEMS SUSTAINABILITY

Digital preservation is complex and expensive (see: Clifford Lynch)

- Systems Sustainability
  - Technologically-sound infrastructure
    - Non-proprietary // open source
    - Well-maintained // community investment (ex: Archivematica)
    - Interoperable //compatible
    - Modular // export-friendly
  - Use of information science standards (naming & arrangement)
  - Use of digipres standards (OAIS, TDR, METS, XML, PREMIS, TRAC)

MACRO-LEVEL: DIGITAL REPOSITORY

ECONOMIC STABILITY

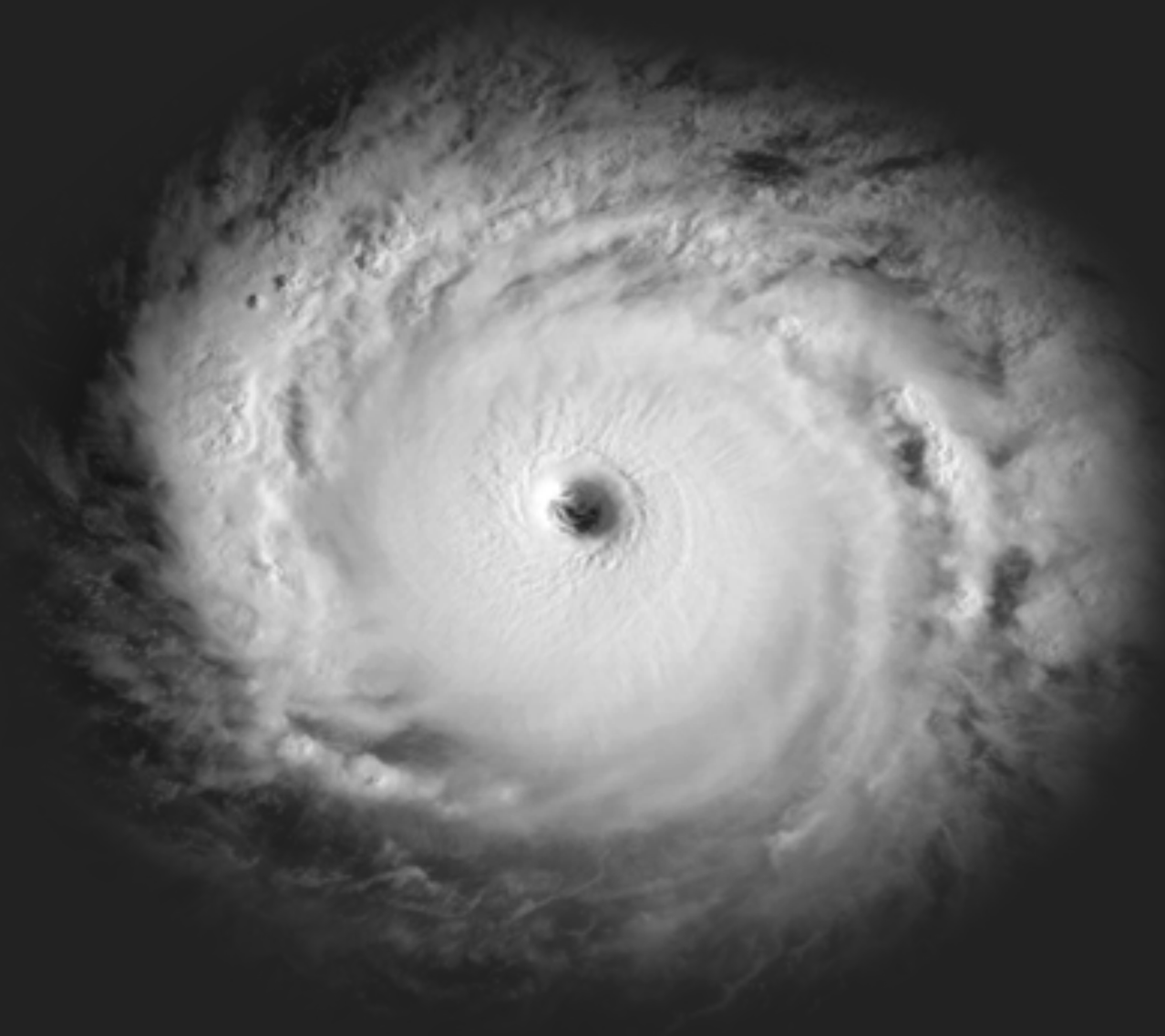
MACRO-LEVEL: DIGITAL REPOSITORY

ECONOMIC & ENVIRONMENTAL STABILITY

- Economic Stability
  - Archives must persist over long periods of time
  - Perfect bit-level preservation, description, standards adherence ≠ financial success
  - Viable access models for *Designated Community*
- Environmental Stability
  - Energy efficient
  - Built for longevity (LTO tape vs. hard drive)

MACRO-LEVEL: DIGITAL REPOSITORY

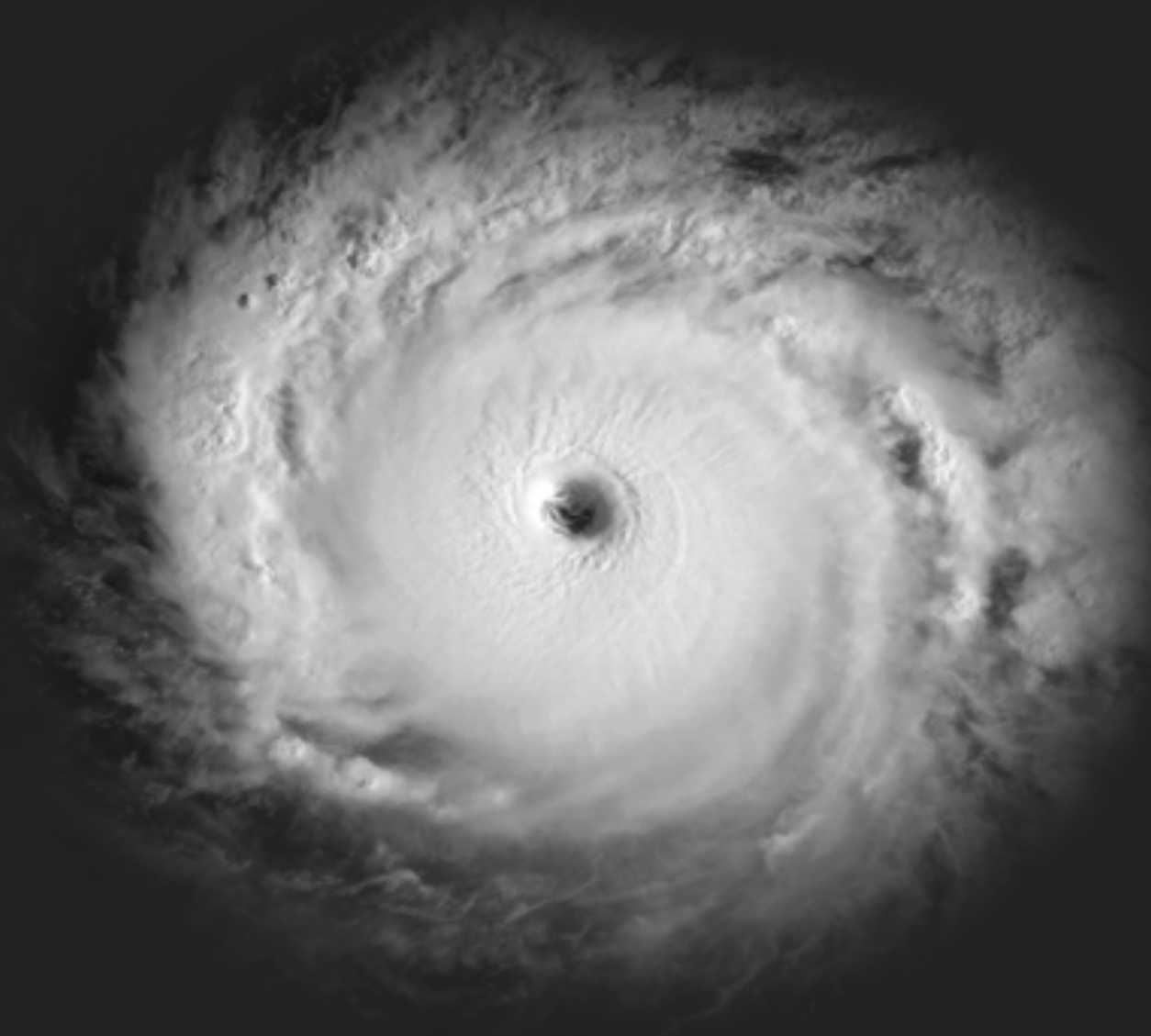
RISK MANAGEMENT



MACRO-LEVEL: DIGITAL REPOSITORY

RISK MANAGEMENT

- Financial risk
  - High cost of digital & physical infrastructure
  - Sustainable funding models
- Disaster planning



THE

END.