Digital Preservation & Archiving

"It's more than just backups!"

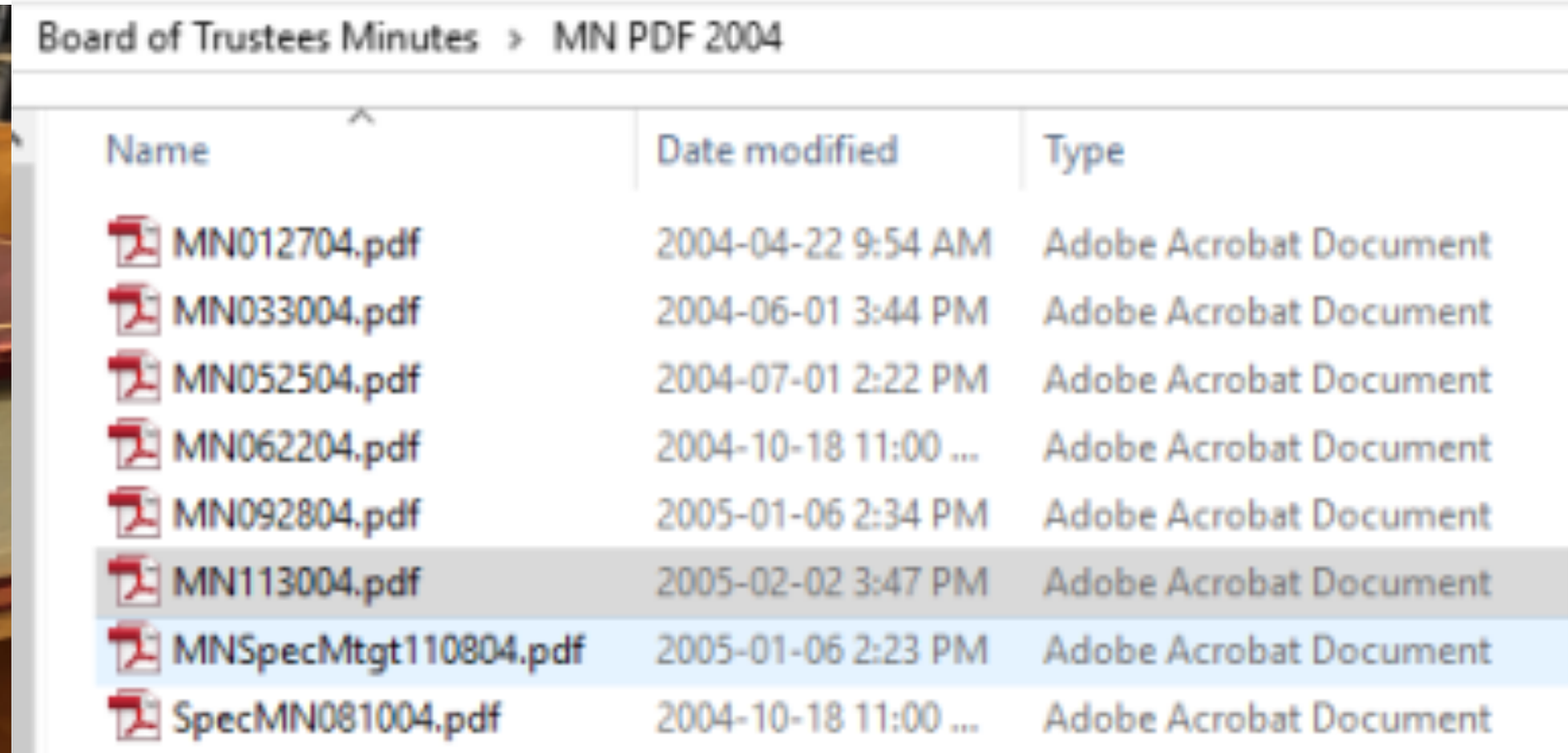
With Thanks and Acknowledgement to
Eira Tansey,
Records Manager/Digital Archivist, University of Cincinnati

What is a Digital Archive(s)?

- “A digital archives is a repository that stores one or more collections of digital information objects with the intention of providing **long-term access** to the information.”—archivemati.ca

Note: A digital information object contains *both* the object and associated metadata

Which set of records is at greater risk of damage?



What issues do archivists face to ensure long term access?

- “storage media instability and deterioration

- technology obsolescence and incompatibility (at the level of hardware, system software application software data and file formats, storage media readers and drivers)

- lack of metadata which results in failure to locate information, the inability to render and read the information, or the inability to attribute meaning or value to the information due to the lack of contextual information”—archivemati.ca

Bits are like the cells of our digital “stuff”

- Let’s take the word FOX…Here is how a computer might read the word “FOX”

F = 0100 0110

O = 0100 1111

X = 0101 1000

- If you have a text file that says “FOX” to your eyes, your computer reads it as 0100 0110 0100 1111 0101 1000

- But that long string of 0’s and 1’s could also represent music. Or a picture.

- So we have to give the computer instructions about how to *render* or display the bits on your screen. This is known as a file format.

How do archivists authenticate digital archives?

- A **checksum** is a unique algorithm you can generate for a file that represents the specific arrangement of 1's and 0's.

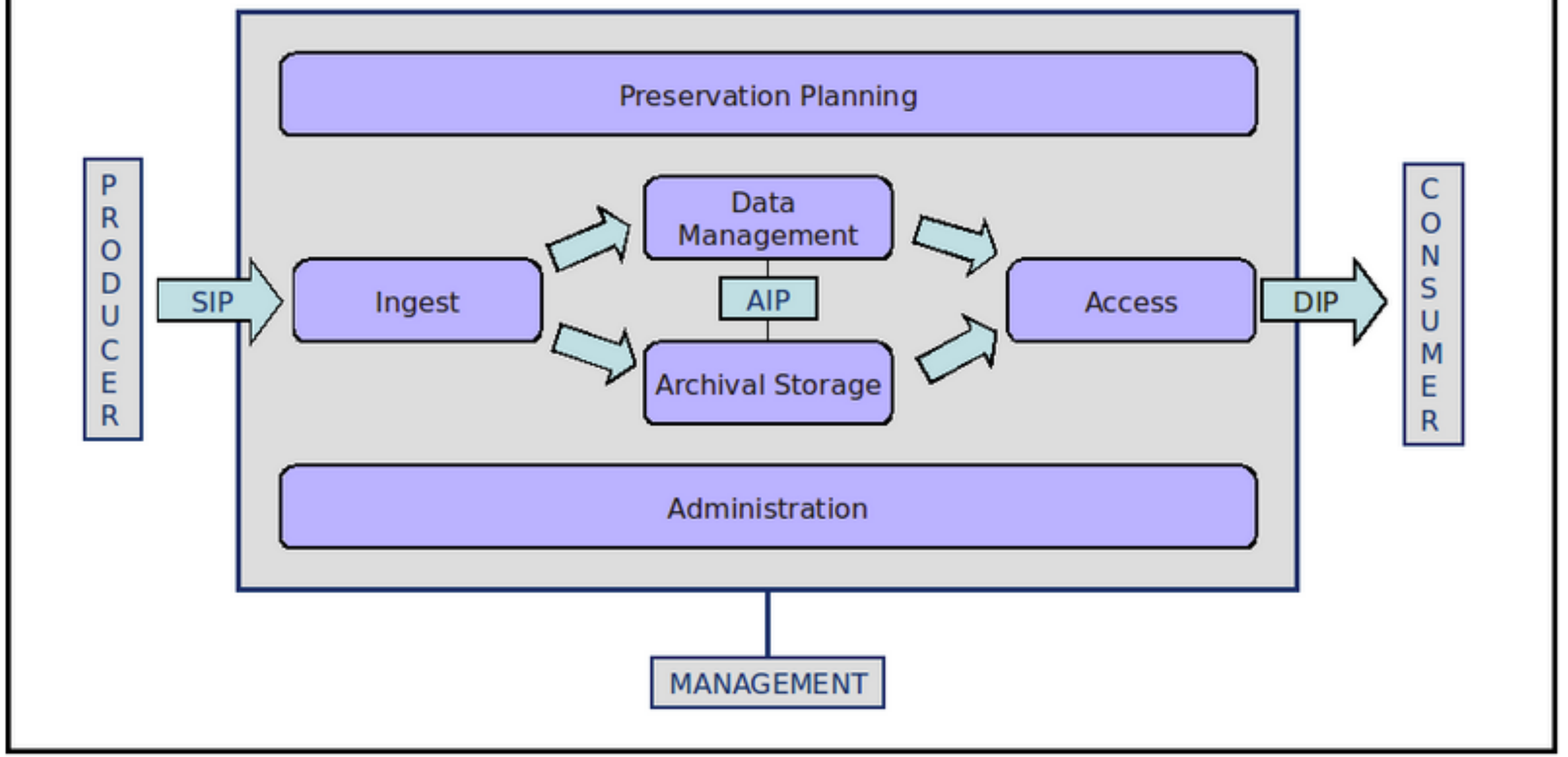
- Checksums are periodically checked on files for **fixity** to see if something changed

Imagine a text file that started with “FOX”	The bits look almost the same…	But the checksums are very different!
FOX	0100 0110 0100 1111 0101 1000	1cbdb8043b37c7c9ca83582f9218c8b4
FoX	0100 0110 0110 1111 0101 1000	66c29de3d89a74fdde085cb3d8025b64
FAX	0100 0110 0100 0001 0101 1000	8e467397ce3ea1d84b02719bd3abc595

Archivists manage digital preservation using the Open Archival Information System Reference Model (OAIS)

Note: The OAIS Model is an ISO Standard

Open Archival Information System (OAIS) reference model (ISO-STD 14721)



https://wiki.archivematica.org/File:OAIS.png

Decoding the OAIS Model

An **Information Package** contains:

- Content Information
- Preservation Description Information
- Packaging Information
- Descriptive Information

Information Packages (Cont.)

Content Information

- The data object and its representation information

Preservation Description Information

- Contains information necessary to preserve the content information (ex. checksum)

Information Packages (Cont.)

Packaging Information

- Holds the components of the information package together (Ex. METS, BagIT)

Description Information

- Metadata about the object

Types of Information Packages

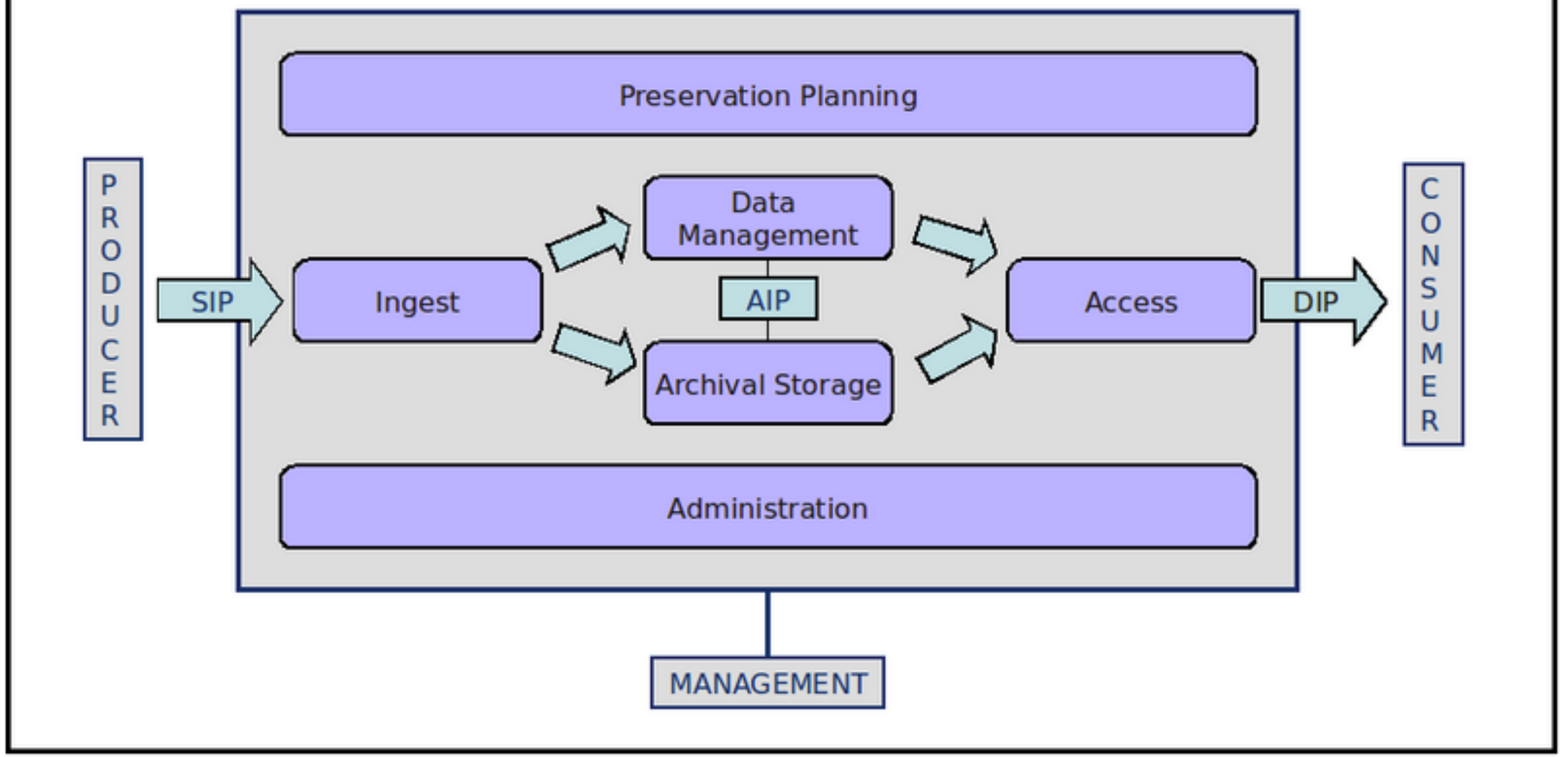
SIP = Submission information package;

information sent from producer to archive

AIP = Archival information package: information stored by the archive

DIP = Dissemination information package: information sent to the user when requested

Open Archival Information System (OAIS) reference model (ISO-STD 14721)



https://wiki.archivematica.org/File:OAIS.png

What about PREMIS?

- **PREMIS** (Preservation Metadata: Implementation Strategies) builds on the OAIS model by taking the general, conceptual elements of the model (information objects, packages, and metadata) and translates them into implementable semantic units
- PREMIS includes a data model, data dictionary, and XML schema
- PREMIS is generally used with METS (Metadata Encoding & Transmission Standard)

Why do digital preservation and archiving matter for DH projects?

Should all DH projects be archived?

At what stage should DH projects be archived if they are ongoing?

Best Practices for Preserving DH Projects

- Have multiple file backups (ideally stored in separate physical spaces)
- Use clear, consistent file naming strategies; create metadata that explains your project's content and any idiosyncrasies in your data (ex. abbreviations)
- Define what components of your project you want to preserve long term (if any); define an archival v. access copy (especially with dynamic, Web-based content)
- Use a trusted repository for long term preservation (ex. university IR, discipline-specific repository)