Digital images and digital preservation

IPTC Photo Metadata Conference, Zagreb, 2016

Ed Pinsent, ULCC

About me

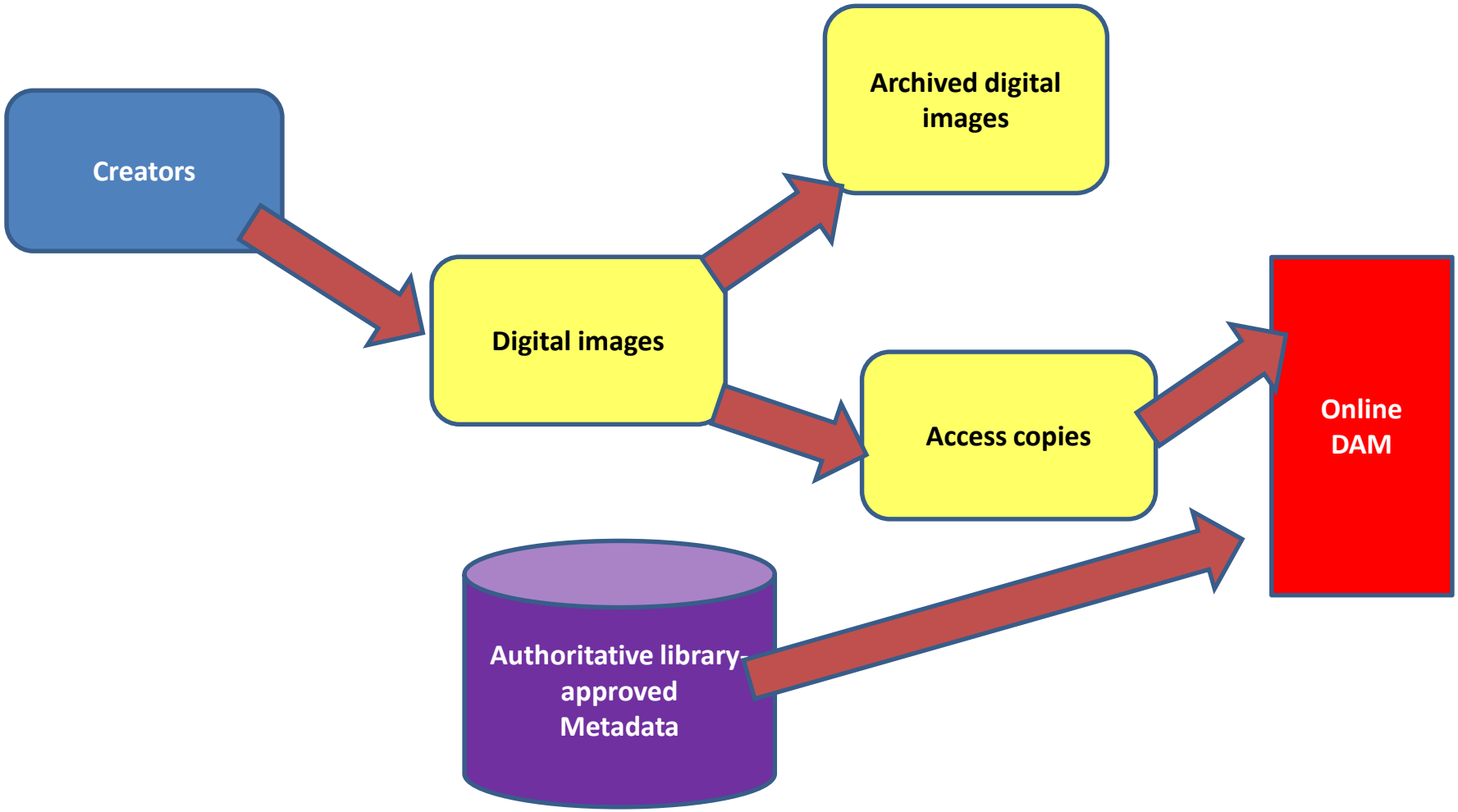
- Digital Archivist at ULCC since 2004
- Teaches digital preservation on the DPTP
- Background as archivist / records manager
- Experience in web-archiving, repository management, metadata projects, migration, digitisation, project management, etc.
- See more at digital archives blog http://dart.blogs.ulcc.ac.uk/

Digital preservation

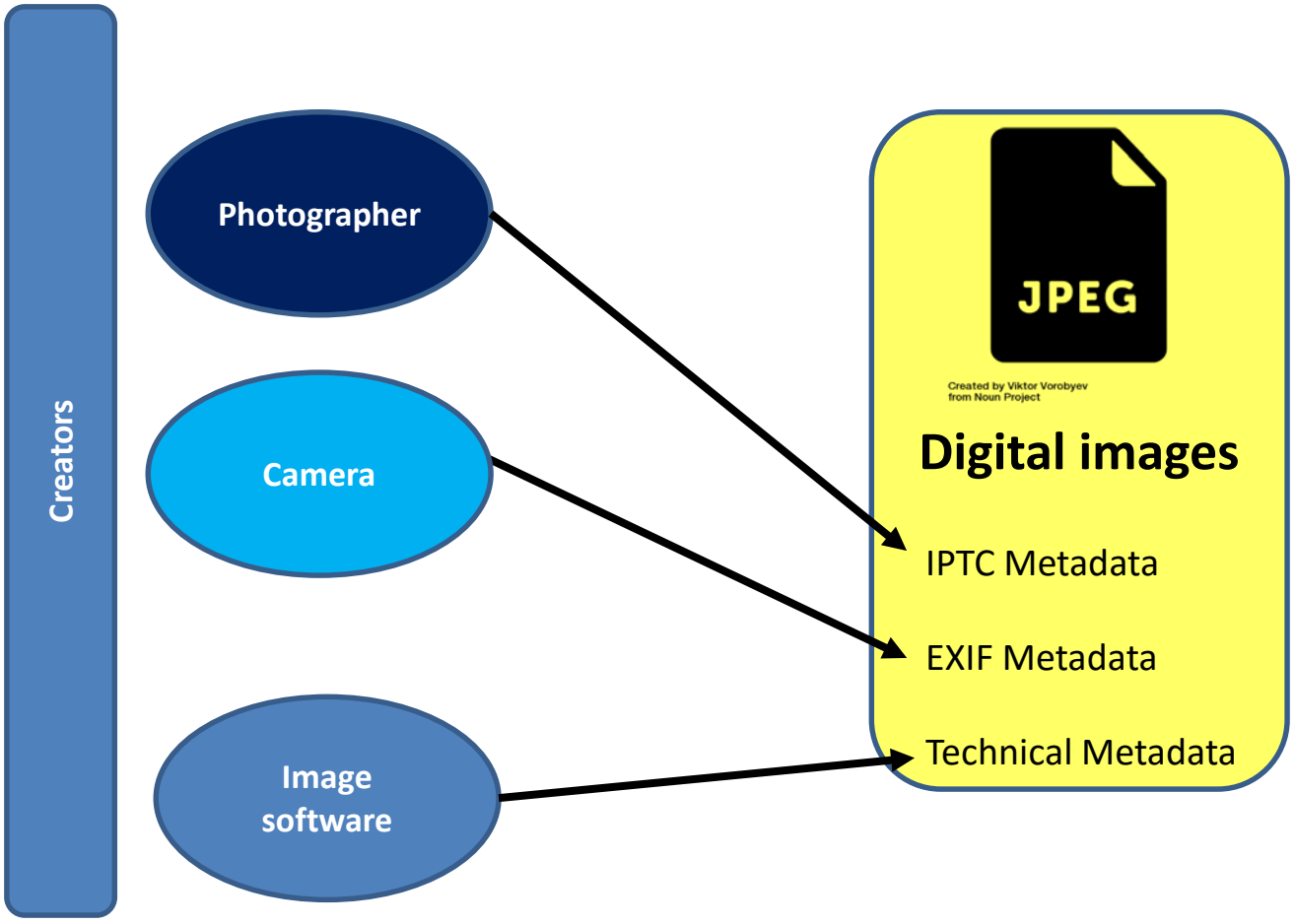
- Digital preservation means ensuring **sustained, continued access** to your content, over along time; not just backing-up
- For image collections, this might mean:
  1. Preservation of **digital image files**
  2. Preservation (and continued management) of **metadata**

This presentation will define these targets of preservation, and propose some interventions

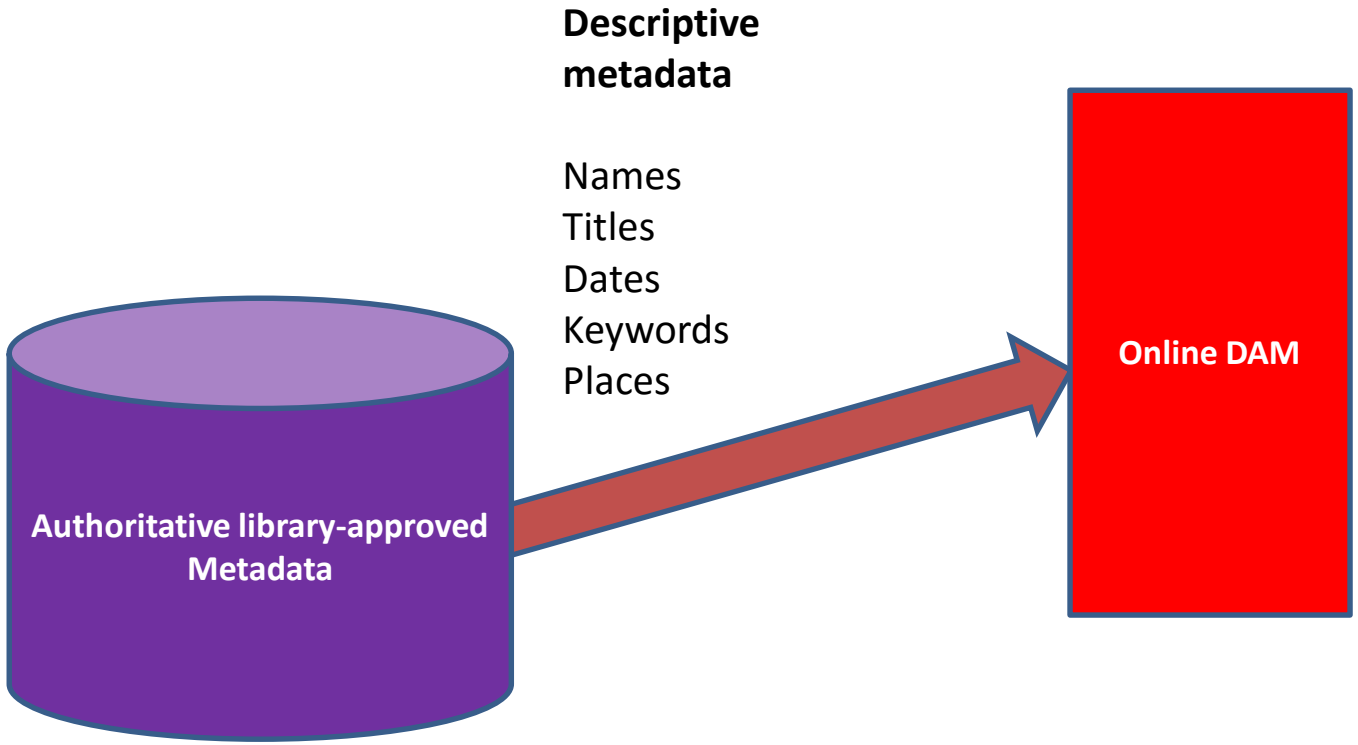
Production chain



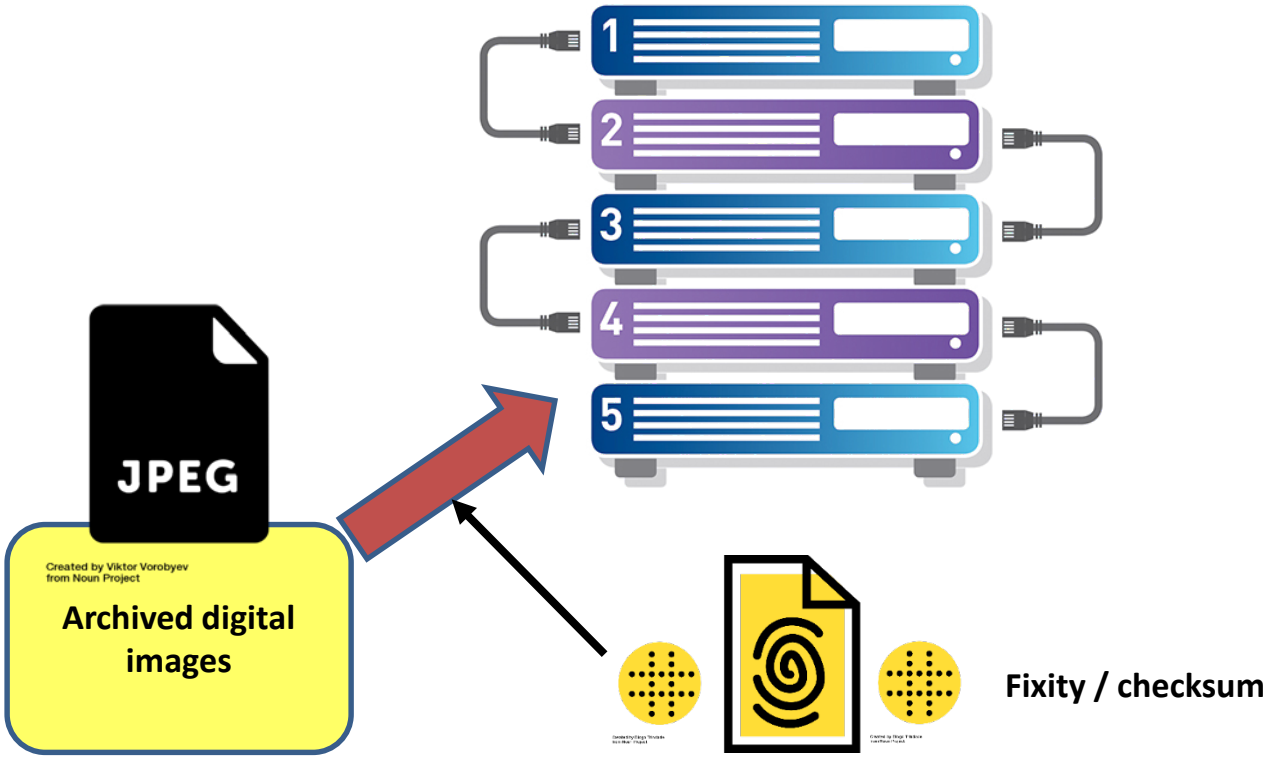
Metadata creation - embedded



Metadata creation – not embedded, descriptive



Intervention 1: preservation storage



Long-term storage

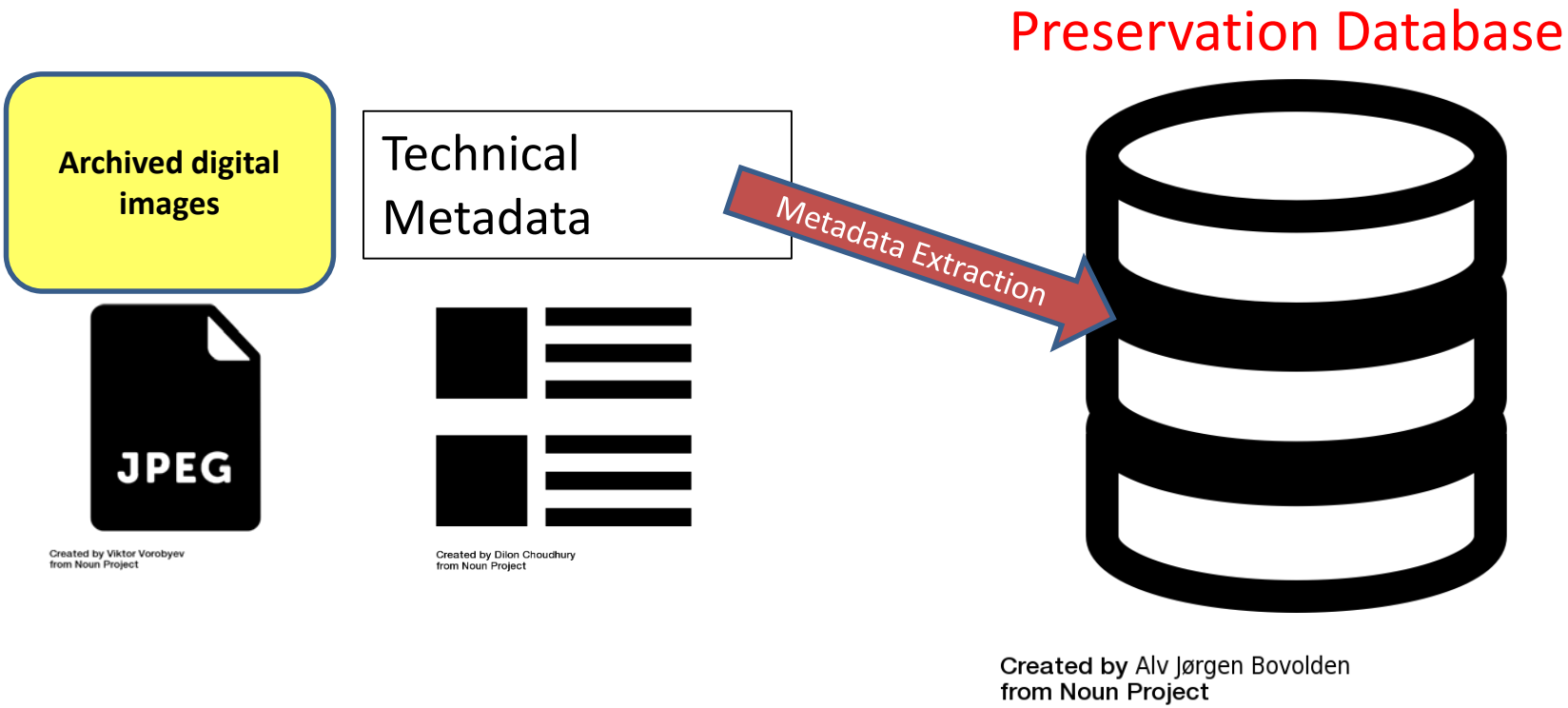
ulcc
University of London Computer Centre

- Ideally, we'd want preservation-standard storage
- This is not the same as "backing-up" a network, or storing content on HD / USB Drives
- Preservation-standard storage is:
  - Dedicated archival storage
  - Little if any network traffic
  - Fixity, validation, monitoring and reporting
  - Multiple independent geo-redundant copies
- A place to keep our "master copies" of digital images, probably hi-res images stored in lossless formats, so they have a big footprint

Fixity

- Crucial part of archival storage / long-term preservation
- Means of detecting change or corruption
- We must generate a checksum for each file
- Checksum = “fingerprint” of digital object; also called *hash* or *fixity*
- If checksum changes, this is an indicator that object has changed
- Regular validation of checksums is strongly recommended

Intervention 2: metadata management



Technical metadata

- Metadata about the digital image, very often embedded in the file header
- Helps us identify the format with high degree of certainty
- <identity format="Graphics Interchange Format" mimetype="image/gif”>

Technical metadata

- Helps us identify specific elements of the image encoding that are necessary for **rendition**
- <byteOrder>little endian</byteOrder>
- <compressionScheme>LZW</compressionScheme>
- <imageWidth>550</imageWidth>
- <imageHeight>428</imageHeight>
- <colorSpace>RGB Palette</colorSpace>
- <orientation>normal*</orientation>
- <bitsPerSample>8</bitsPerSample>

EXIF metadata

- **EXIF metadata**: very detailed record of camera (or scanner) information
- Embedded (automatically) in the file when the image is created
- May have some value / meaning as:
  - a history of the file's creation
  - a history of hardware use
  - a history of the production chain of the image

IPTC metadata

- **IPTC metadata**: rights metadata and descriptive metadata; used for expressing copyright, IPR and ownership of a digital image
- Tends to be authored / created by the photographer, agency, image owner
- In some image file formats, such as JPEG, TIFF, and PSD, the metadata is standard and supported
- Likely to have long-term value:
  - Protects the owner's rights
  - Protects the image from unauthorised copying
  - Adds meaning and context to the image
  - Translates into a business / commercial value

Intervention 2: metadata management

ulcc
University of London Computer Centre

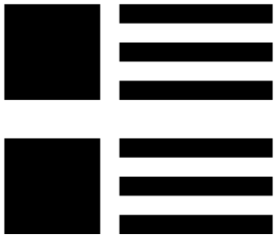
Preservation Database

Archived digital images

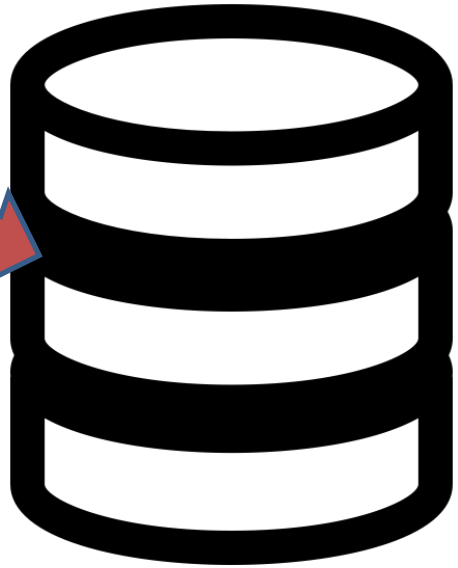
Technical metadata
+ EXIF metadata
+ IPTC metadata
+ UID

Metadata Extraction

Created by Viktor Vorobyev from Noun Project

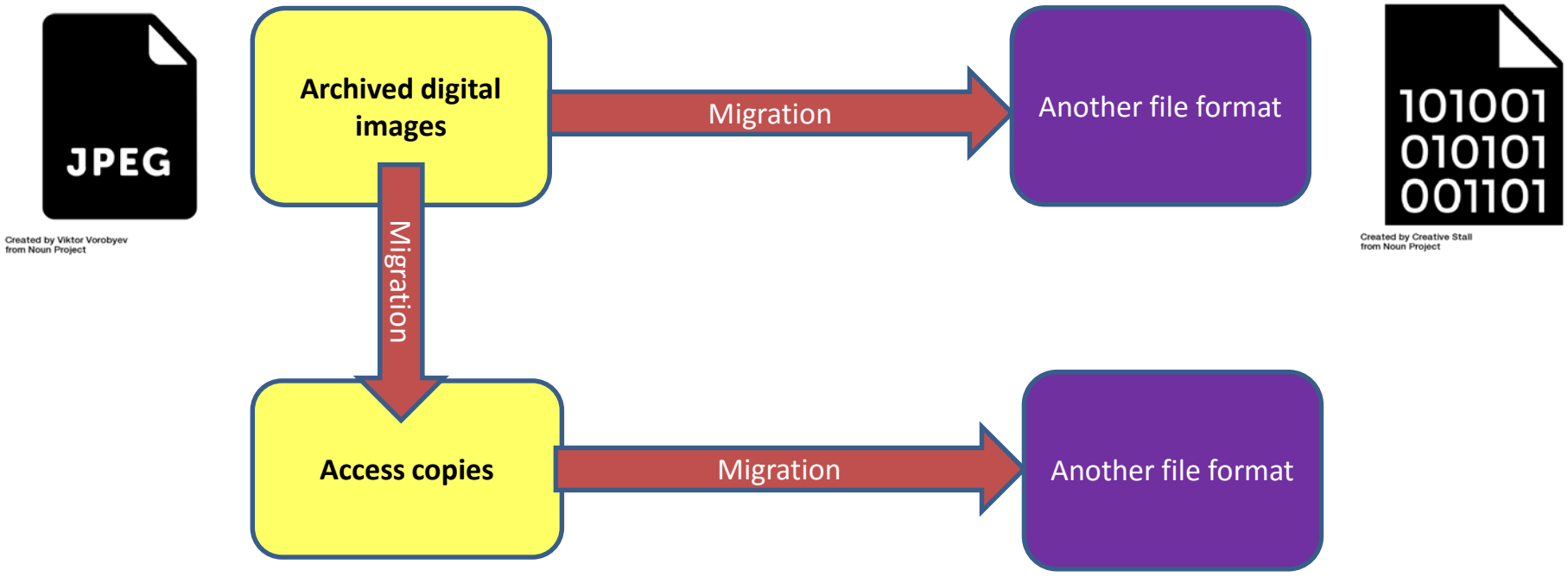


Created by Dilon Choudhury from Noun Project



Created by Alv Jørgen Bovolden from Noun Project

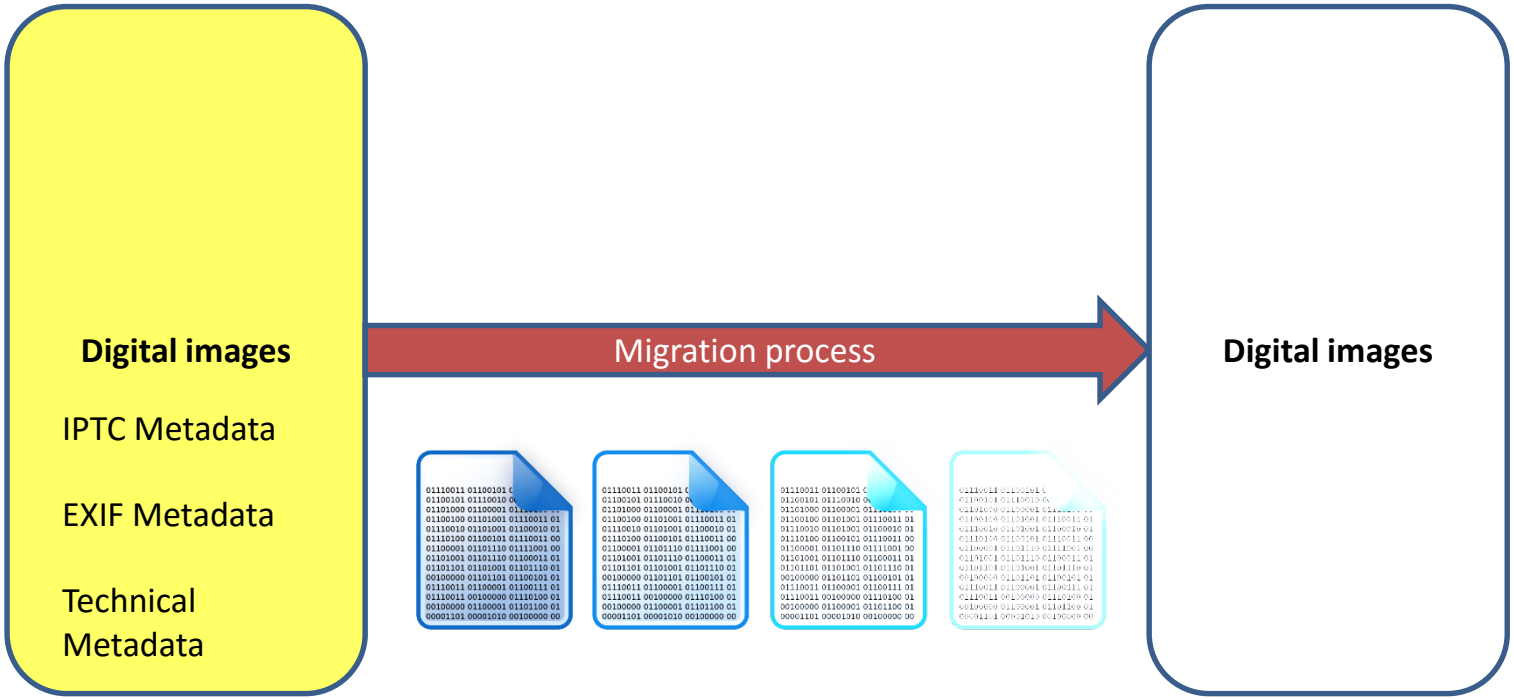
Intervention 3: migration



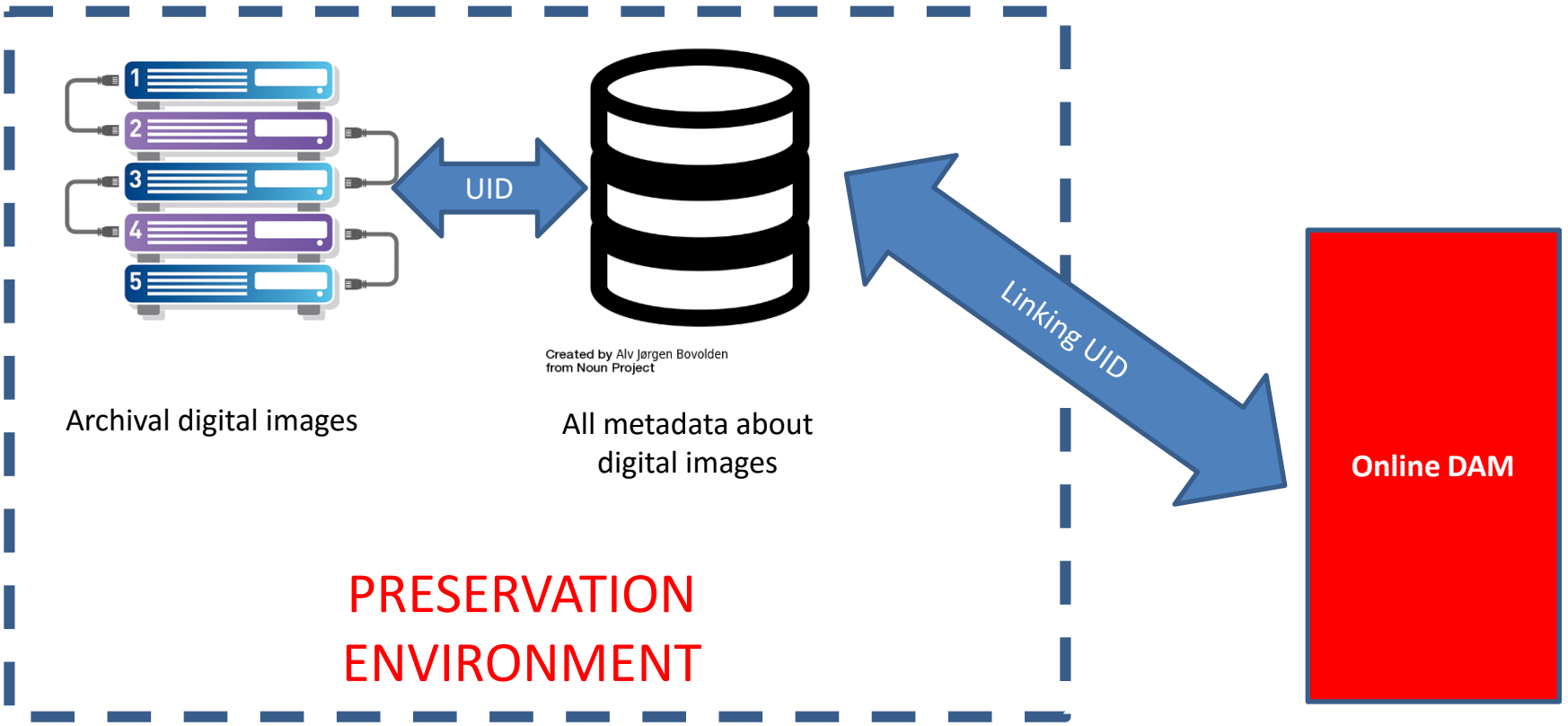
Migration

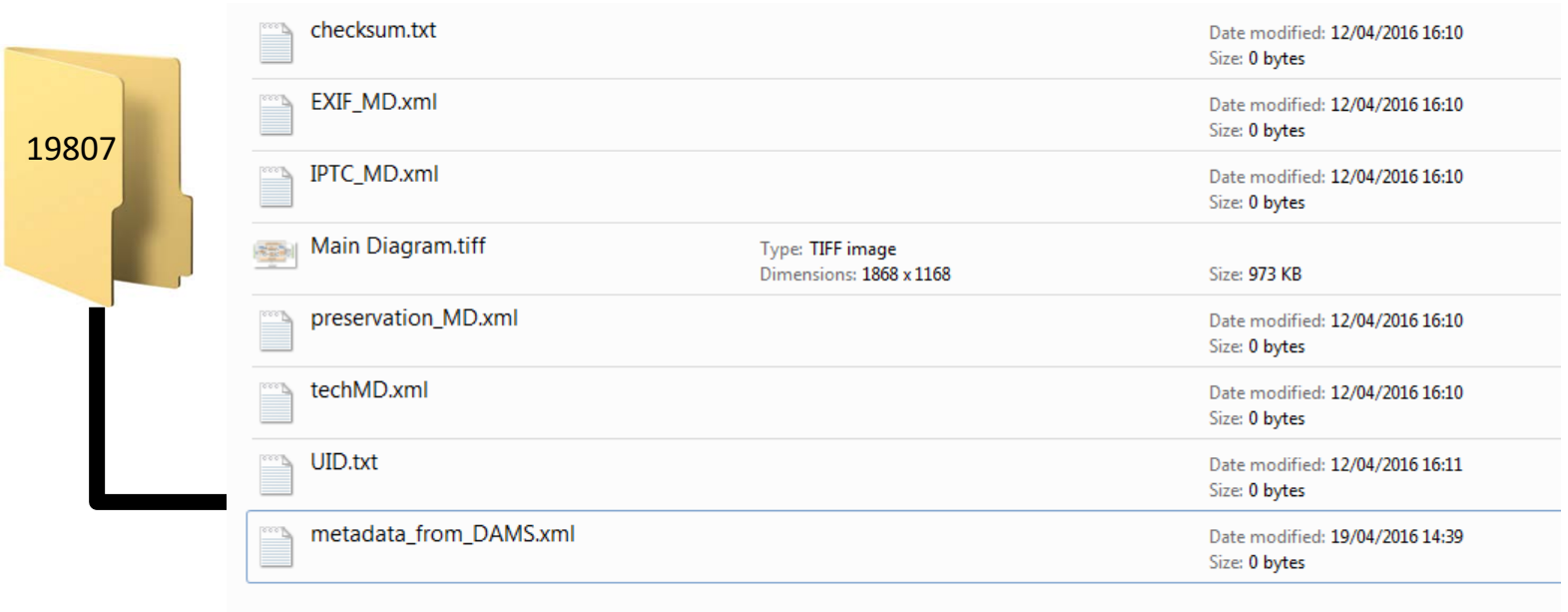
- Common and well-understood approach to digital preservation
- E.g. migrate TIFF to JPEG 2000, if JP2 gives you more confidence as a long-term format for preservation
- E.g. migrate TIFF to PNG, to create access copies for dissemination or sale
- You would need to support migration tools, and target formats
- However, migration also introduces risks of *loss*, especially metadata

Risks of loss



How this all joins up…





View of your preserved assets and metadata in preservation storage

Next steps…

- How do we do all this?
- What is the software / who are the service providers?
- Is it expensive?
- What are the steps towards doing preservation?
- Who is usually responsible for preservation?
- How can we as archivists/image library people get the process going?
- What do we need to say to IT people to achieve best preservation?
- Who else is doing it?
- Is it commercially viable to preserve images in commercial image libraries?
- How do we adjust our workflow now to encompass the need for preservation later?
- How does a DAM system relate to preservation?

Some links

- **AOR Toolkit**: info.ulcc.ac.uk/aortoolkit-iptc
- **Free OAIS Course**: info.ulcc.ac.uk/oais-iptc
- **Image metadata**: info.ulcc.ac.uk/metadata-iptc16