Driving digital transformation in Cultural Heritage Institutions

Session 3

Good practices and experiences for creation, access and re-use

2025-12-03

Prof. Dr. Frederik Temmermans

Vrije Universiteit Brussel - imec

Digitization

- Capture of a **physical** object into **digital data**
  - A physical, continuous object → reduced to samples
  - Convert reality into discrete measurements through sensors

- Conversion from **continuous** to **discrete**
  - Limited resolution, dynamic range, colour depth, …
  - Some information is not captured at all (reflectance, material properties, hyperspectral bandwidths, hidden surfaces, …)
  - Interpolation artefacts (demosaicing, upsampling, compression)
  - Missing information may be reconstructed or inferred

Images: closertovaneyck.org

Digitization

- Properties inherent to **equipment and setup**
  - Lens distortions, chromatic aberration
  - Lighting conditions, shadows, glare, reflections
  - Sensor noise, temperature, stability
  - Operator decisions (exposure, aperture, angles)

- Impact of the **processing pipeline**
  - Transcoding & compression
  - Composition (stitching, registration, photogrammetry)
  - Image processing (tone mapping, denoising, sharpening, …)
  - AI-based enhancement, reconstruction, segmentation
  - Every step alters the representation, most are irreversible

Images: closertovaneyck.org

Digitization

- **Every digitization is a unique representation**
  It can *never* capture all properties of the physical counterpart.
- **Choices shape outcome**
  Equipment, setup, file formats, and processing workflows all have a direct impact on the final result.
- **Physical artefacts are dynamic objects**
  A digitization records the object at a specific moment in time.
- **Interpretation depends on understanding those choices**
  Without insight into the pipeline, a digital asset cannot be fully trusted or correctly interpreted.
- **Therefore: documentation and transparency are essential**
  We need to preserve meaning, context, and authenticity, not just pixels.

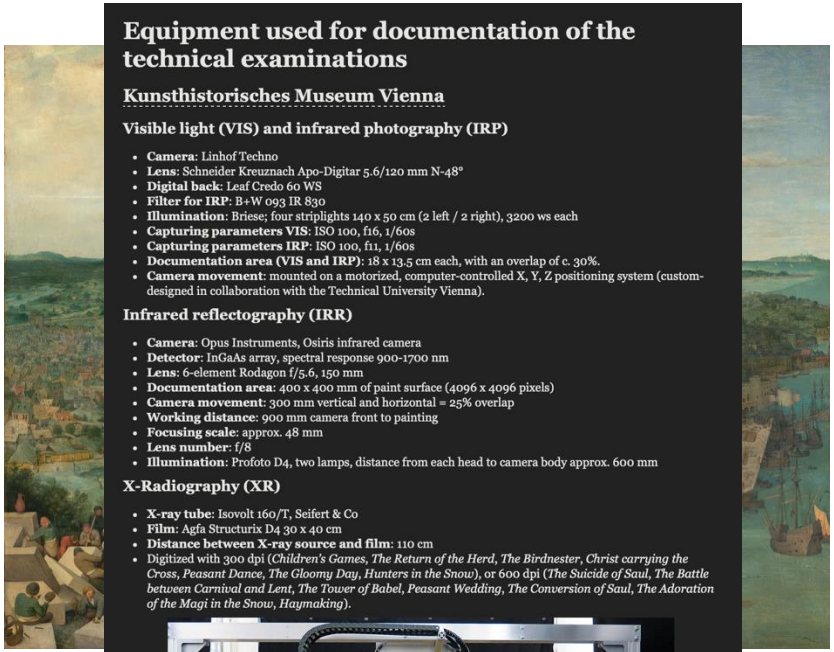
Images: closertovaneyck.org

Bruegel: Tower of Babel

InsideBruegel.net

Google Art Project

Bruegel: Tower of Babel



Equipment used for documentation of the technical examinations

Kunsthistorisches Museum Vienna

Visible light (VIS) and infrared photography (IRP)

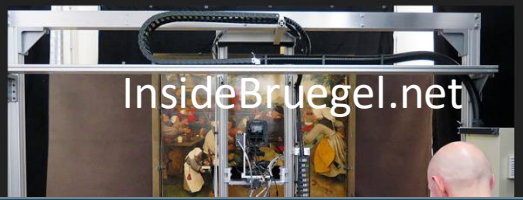
- **Camera**: Linhof Techno
- **Lens**: Schneider Kreuznach Apo-Digitar 5.6/120 mm N-48°
- **Digital back**: Leaf Credo 60 WS
- **Filter for IRP**: B+W 093 IR 830
- **Illumination**: Briese; four striplights 140 x 50 cm (2 left / 2 right), 3200 ws each
- **Capturing parameters VIS**: ISO 100, f16, 1/60s
- **Capturing parameters IRP**: ISO 100, f11, 1/60s
- **Documentation area (VIS and IRP)**: 18 x 13.5 cm each, with an overlap of c. 30%.
- **Camera movement**: mounted on a motorized, computer-controlled X, Y, Z positioning system (custom-designed in collaboration with the Technical University Vienna).

Infrared reflectography (IRR)

- **Camera**: Opus Instruments, Osiris infrared camera
- **Detector**: InGaAs array, spectral response 900-1700 nm
- **Lens**: 6-element Rodagon f/5.6, 150 mm
- **Documentation area**: 400 x 400 mm of paint surface (4096 x 4096 pixels)
- **Camera movement**: 300 mm vertical and horizontal = 25% overlap
- **Working distance**: 900 mm camera front to painting
- **Focusing scale**: approx. 48 mm
- **Lens number**: f/8
- **Illumination**: Profoto D4, two lamps, distance from each head to camera body approx. 600 mm

X-Radiography (XR)

- **X-ray tube**: Isovolt 160/T, Seifert & Co
- **Film**: Agfa Structurix D4 30 x 40 cm
- **Distance between X-ray source and film**: 110 cm
- Digitized with 300 dpi (*Children's Games, The Return of the Herd, The Birdnester, Christ carrying the Cross, Peasant Dance, The Gloomy Day, Hunters in the Snow*), or 600 dpi (*The Suicide of Saul, The Battle between Carnival and Lent, The Tower of Babel, Peasant Wedding, The Conversion of Saul, The Adoration of the Magi in the Snow, Haymaking*).



InsideBruegel.net

Google Art Project

Case studies and good practices

- **File formats**
  Making the right choices by defining requirements.

- **The role of AI in digitization**
  How AI is reshaping entire digitization pipelines.

- **Documenting provenance**
  Why transparency enables authenticity and responsible re-use.

Image: closertovaneyck.org

File formats: requirements

- **Lossless** or **lossy**?
- **File size** (storage, bandwidth, …)
- **Perceptual quality**
- **Encoding** and **decoding** efficiency
- File formats **features** and **capabilities**
  - Transparency, colour spaces, HDR, resolution scalability, …
- **Interoperability**
  - Web browser support, reader and writer software, …
- **Licensing**

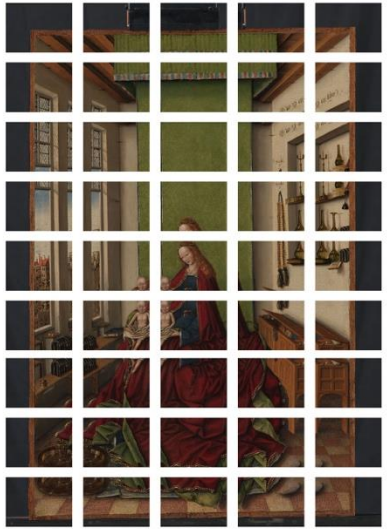
→ **There is no holy grail of file formats!**

File formats: strengths and weaknesses

File format	Strengths	Weaknesses
JPEG 1	Compatibility Decent compression efficieny for photographic content	No lossless, no transparancy, limited bit depth, … Compression artefacts for graphic content
PNG	Transparancy Great performance for graphical content Compatibility	Not ideal for photographic content
TIFF	Widely supported Flexibilty (multiple channels, colour spaces, …)	4GB file size limit (partially resolved by bigTIFF) Mostly uncompressed
JPEG 2000	Great compression efficiency Still + motion Wavelet based (resolution scalability)	Higher complexity compared to JPEG 1 More limited support compared to JPEG 1
Camera RAW	RAW sensor information High bit depth Editing flexibility	Lack of standardization Inconsistent rendering Limited compression
JPEG XL	Lossless JPEG 1 transcoding State of the art compression Wide set of features	Compatibility and adoption still in early phase

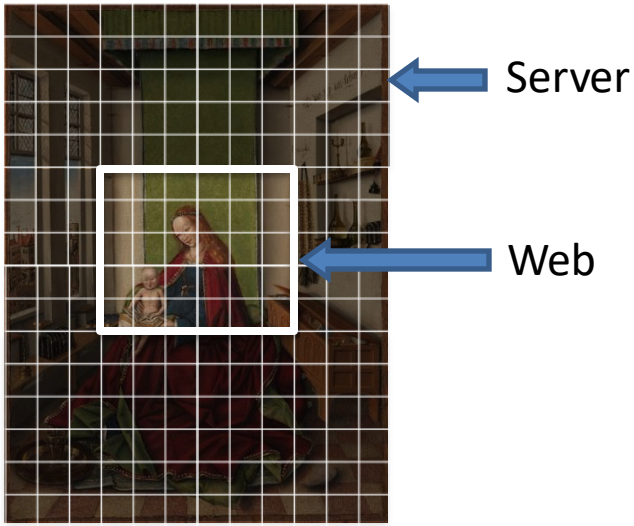
File formats: transcoding

Camera RAW → TIFF



TIFF → RAW → bigTIFF

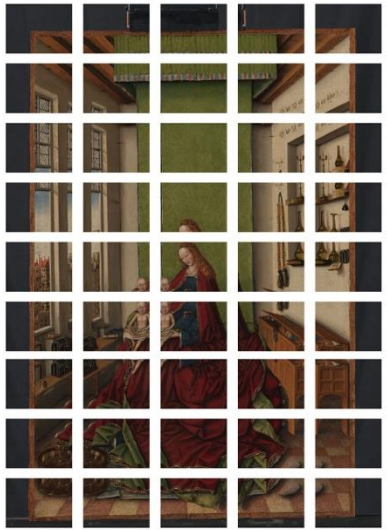
bigTIFF → JPEG 2000 → JPEG 1



File formats: what to archive?

Closest to capture / photographers' vision

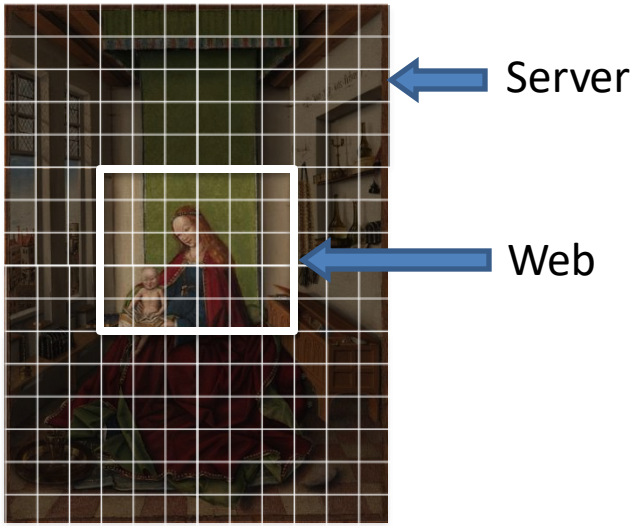
Camera RAW → TIFF



Intermmediate steps?

TIFF → RAW → bigTIFF

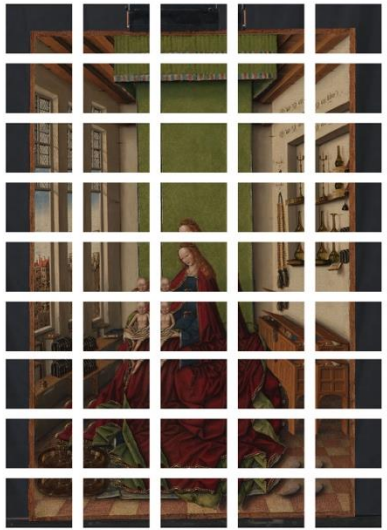
bigTIFF → JPEG 2000 → JPEG 1



File formats: what to archive?

Closest to capture / photographers' vision

Camera RAW → TIFF



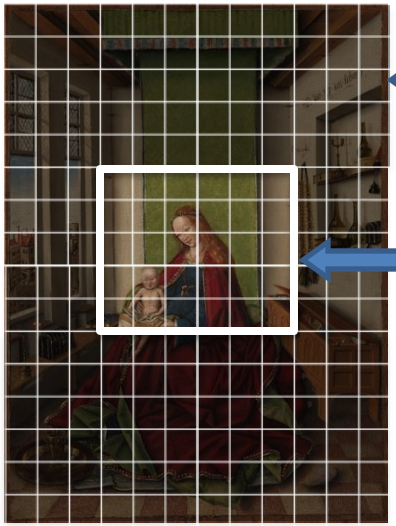
Intermmediate steps?

TIFF → RAW → bigTIFF

Master result

bigTIFF → JPEG 2000 → JPEG 1

Live version



Server

Web

File formats: what to archive?

- **Source** and **final result** are most important, but…
- What exactly is the source?
- Can all subsequent steps be **reproduced**?
- How does the **processing pipeline** look like?
  - Time consumption, expertise, software availability, …
- What are the requirements for the **final result**?
  - Efficiency, storage limitations, access, …
- **Transparency** and documentation of the **provenance** is paramount.
  - What equipment was used, what were the processing steps, …

AI in digitization: generative AI

Mona Lisa with a cat.
Generated with ChatGPT.

Selfportrait in the style of Van Gogh.
Generated with ChatGPT.

AI in digitization: image enhancement

Zoom shot of the moon rendered on a Samsung Galaxy S23.

AI enhance skin texture. Natural and soft options on the same person.

AI in digitization: demosaicing

Nikon D850 DSLR.

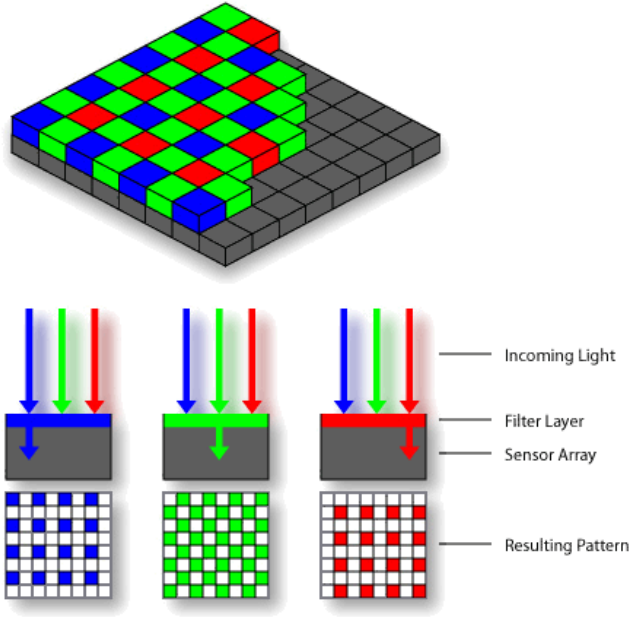
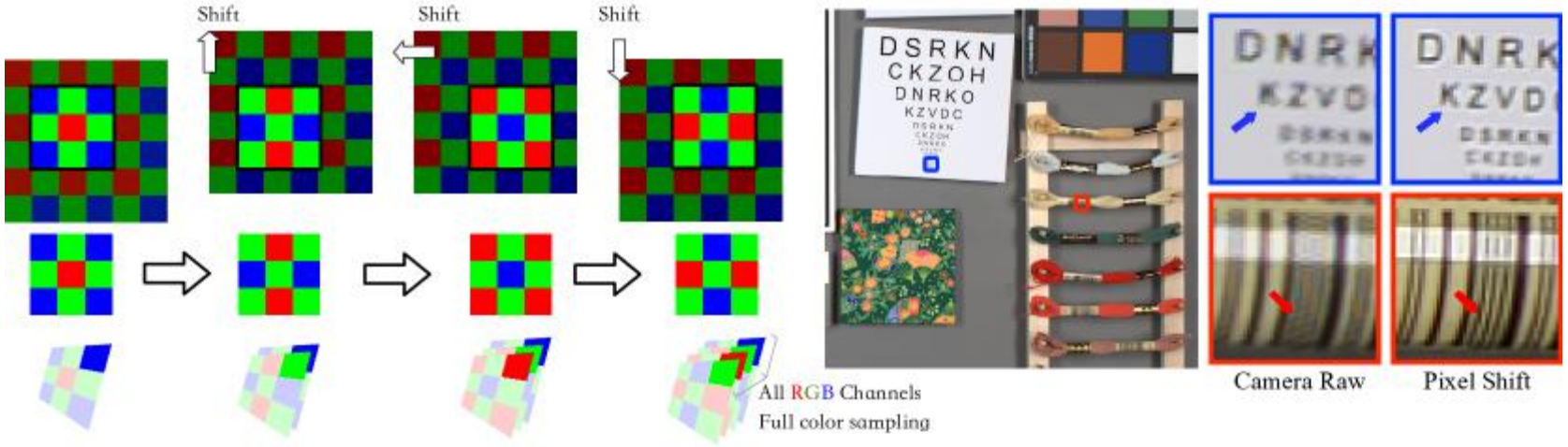


Illustration of how a Bayer filter works.

AI in digitization: demosaicing



Rethinking Learning-based Demosaicing, Denoising, and Super-Resolution Pipeline

Guocheng Qian, Yuanhao Wang, Jinjin Gu, Chao Dong, Wolfgang Heidrich, Bernard Ghanem, Jimmy S. Ren

Image and Video Processing (eess.IV); Computer Vision and Pattern Recognition (cs.CV)

AI in digitization: image stitching

Process of stitching multiple photos into a single giga-pixel image.
Image from closertovaneyck.org, stitching by Universum Digitalis BV.

AI in digitization: gaussian splatting

Granite gneiss ram of Amun.
Artifact scanned using 3DGS, kiriengine.app.

AI in digitization

- **AI is everywhere in modern imaging pipelines**, from capture to enhancement to (3D) reconstruction.
- **It brings major benefits**: efficiency, quality, automation, enrichment.
- **But it also introduces risks** for interpretation, authenticity, and reproducibility.
- **Therefore, transparency about AI use is essential.**
  - What AI tools were used?
  - What data was used or inferred?
  - Is RAW sensor data available?

Documenting provenance: metadata

- Metadata is **embedded** in images at the **moment of capture**.
  - Contains information such as **location**, **time**, **camera model** and **settings**, etc.
- However:
  - Often **not retained after transcoding.**
  - Not always **consistently updated** after editing or processing.
  - **Manipulation** of this metadata is **trivial** and often **unnoticeable.**
  - Not all information on equipment, context, circumstances is recorded, neither any information on processing pipelines.

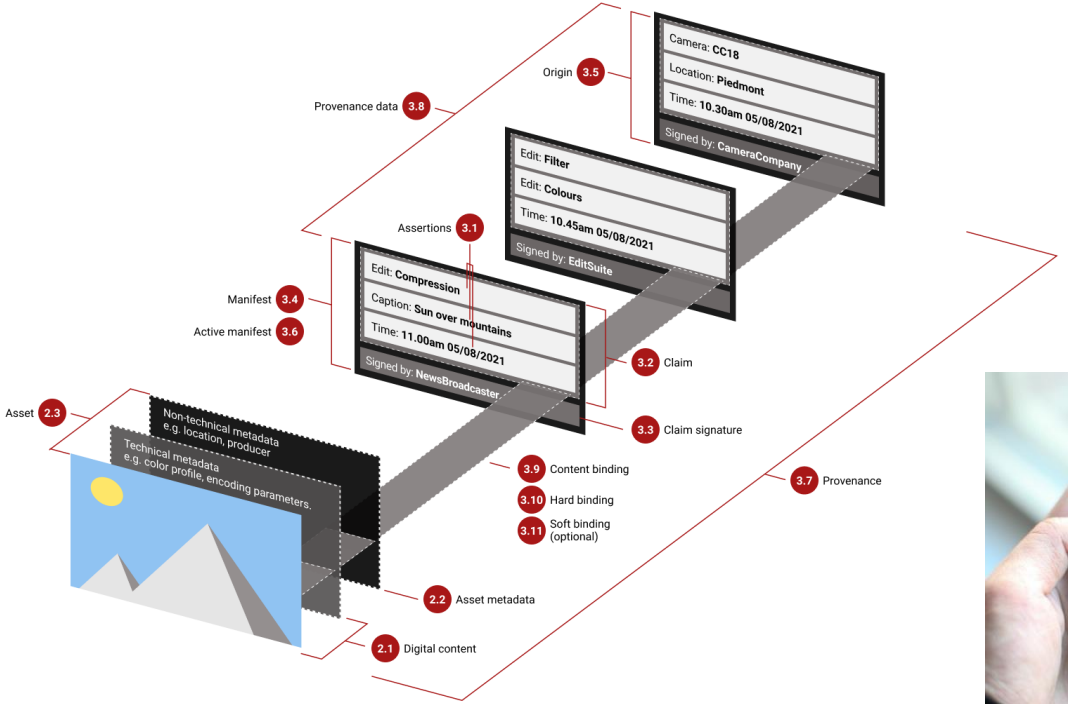
Documenting provenance: aims

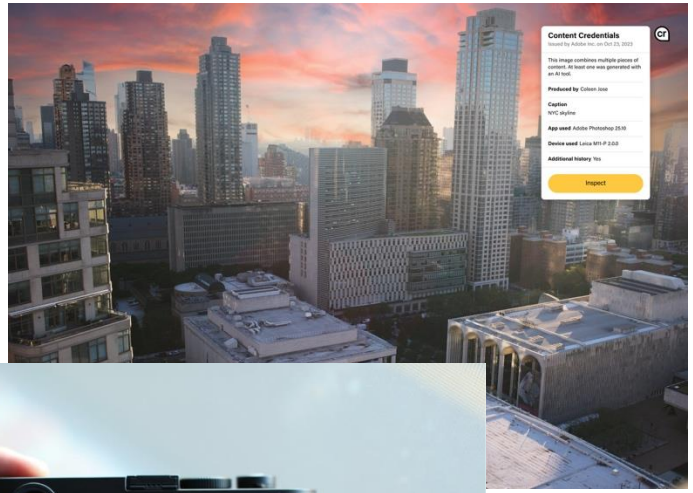
- Several improvements to traditional metadata are needed:
  - **Securing metadata**, allowing identification of modifications.
  - Documenting not only creation but the entire **provenance**, a **persistent chain of information** documenting the **creation** details, as well as **all changes** made to a digital file since it's creation.
  - Provide means to **interpret the metadata**.
  - **Interoperable** frameworks.
  - Allow **embedding** as well as **referencing**.

C2PA Coalition for content provenance and authenticity

JPEG Trust ISO/IEC 21627: JPEG Trust

Documenting provenance: C2PA





JPEG Trust Part 1: Core Foundation

Signalling provenance

C2PA

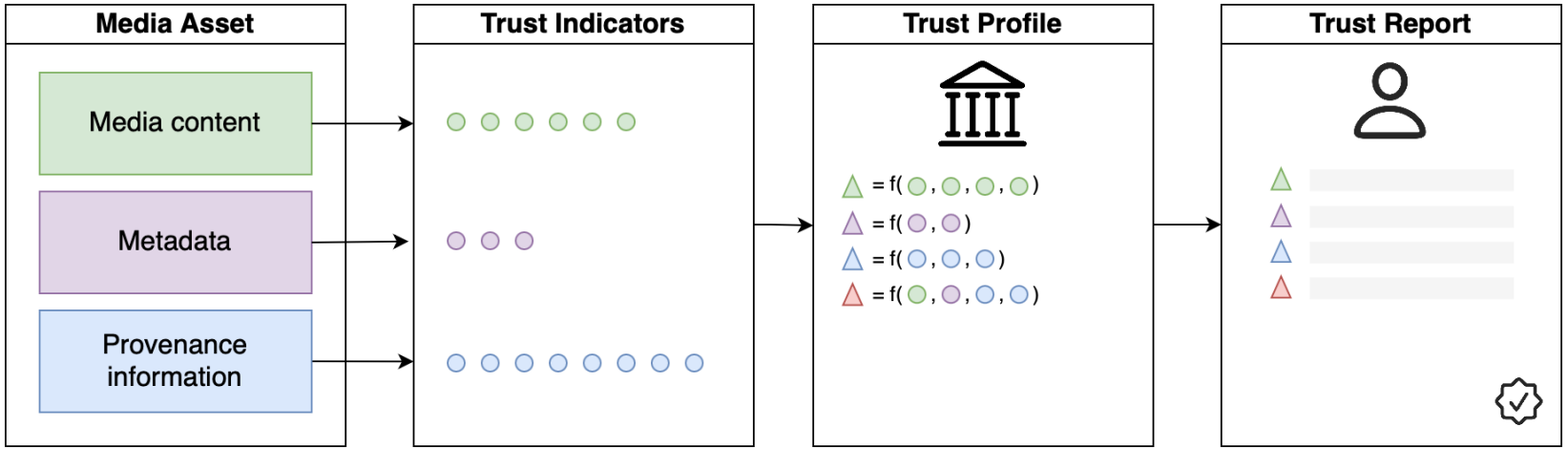
Extracting and evaluating Trust Indicators

Handling privacy and security concerns

NEW

Intellectual Property Rights

Extracting and evaluating trust indicators



Conclusions

- **A digitization is unique capture of a dynamic object**
  - It is always a representation, shaped by equipment, setup, and processing.
- **File formats involve trade-offs**
  - There is no holy grail, choose formats based on clear requirements and document the decisions.
- **AI is everywhere in the imaging pipeline**
  - It brings efficiency and quality, but also interpretive changes.
  - Understanding AI's role is essential for correct interpretation.
- **Provenance and transparency are essential**
  - To ensure trust, authenticity, and responsible re-use, we must document how a digital asset was created and how it was transformed.