

Metadata & Interoperability

Maria Konstantinidou and Elpida Perdiki



What metadata do

Metadata are structured information about objects.

Identify

Which object?
Which shelfmark?
Which ID?

Describe

What is it?
What does it contain?
How is it made?

Interpret

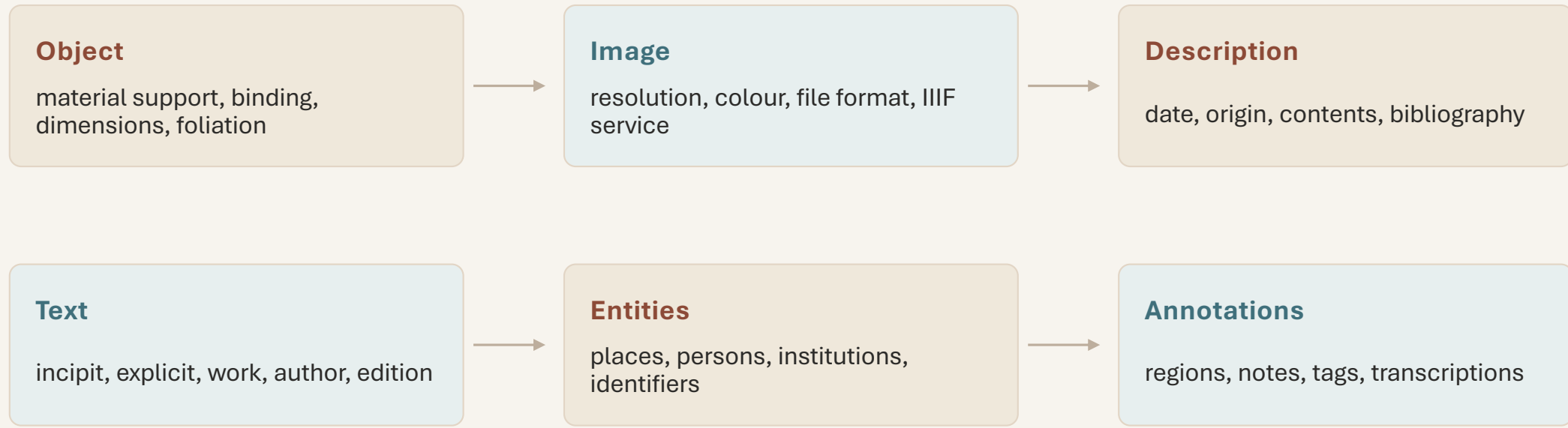
Date? Origin?
Hand? Attribution?
Confidence?

Enable use

Search
Manage
Compare
Reuse

Metadata decide what a source can become; interoperability decides whether it can travel beyond its platform.

One object, several metadata layers



The same manuscript is represented differently depending on which layer a system is designed to handle.

Kinds of metadata in manuscript work

Descriptive

title, shelfmark, date, origin, contents, bibliography

Technical / administrative

file format, resolution, encoding, rights, software/version

Structural

page order, foliation, quires, canvases, sequences

Codicological

support, dimensions, layout, binding, decoration

Palaeographical

script, hand(s), ductus, letterforms, dating criteria

Textual / semantic

authors, works, incipits, places, persons, annotations

Useful metadata are not “more fields”. They are fields chosen for a purpose: discovery, interpretation, interoperability or reuse.

Metadata are scholarship

Date	11th c.
Origin	Constantinople (?)
Contents	John Chrysostom
Scribe	unidentified
Provenance	?

FACT?

ASSERTION?

UNCERTAINTY?

Metadata encode attribution, interpretation and degrees of confidence.

A metadata schema is a model

A schema decides:

- 1 what counts as an entity
- 2 what gets a field
- 3 which relationships can be expressed
- 4 how uncertainty is recorded
- 5 what becomes searchable — and what disappears

**Not “everything about
the manuscript”**

but

a formalised account of what a
community chooses to make
computable.

Controlled description: useful, dangerous, necessary

Palaeographical description is often expert, comparative and deliberately elastic.
Metadata ask us to make some of it consistent enough to search.

Free description

“small, regular, slightly right-sloping
minuscule, with bouletée tendencies”

Controlled fields

script = minuscule
style = bouletée?
slope = right
certainty = medium

Risk

If the schema is too rigid, expert
judgement is flattened into bad data.

The aim is not to eliminate interpretation, but to encode it responsibly.

Identity is surprisingly difficult

One physical object may have...

old shelfmark
new shelfmark
catalogue number
project identifier
fragment or part
front/back of a papyrus
multiple texts/authors



Shared identity across systems

DIKTYON: Greek manuscript shelfmark identity
TRISMEGISTOS: papyri, texts, people, places, collections
PLEIADES: ancient places
TLG / work IDs: textual identity

Persistent identifiers let separate databases recognise that they are talking about the same object, place, person or text.

P. Oxy. XXVII 2468
(Oxford)



P. Oxy. LXXVI 5106
(British Museum)

[illegible]

From strings to entities

A string

Constantinople

John Chrysostom

Basel, Universitätsbibliothek, AN III 11



An entity

Constantinople → authority ID

John Chrysostom → person/work records

AN III 11 → stable manuscript identity

Metadata no longer only describe; they connect.

Interoperability: metadata must travel

Good metadata should travel

Technical interoperability

formats, APIs, IIIF manifests, stable URLs

Can another system read it?

Semantic interoperability

controlled terms, authority IDs, shared entities

Does it mean the same thing?

Scholarly interoperability

provenance, uncertainty, documentation, licences

Can it be trusted and reused?

Interoperability is preserving meaning among diverse systems.

ML exposed the metadata problem

Most manuscript infrastructures do not integrate ML. But they increasingly enable it — if their data can be extracted and trusted.

ML needs

- large, structured datasets
- bulk image/metadata access
- ground truth / labels
- provenance
- consistent identifiers

ML punishes

- messy dates
- missing values
- local terminology
- closed viewers
- unrecorded uncertainty

Good infrastructure gives

- IIIF / APIs
- exportable metadata
- TEI/XML, CSV, JSON
- Zenodo, GitHub, Dataverse, OSF

The problem moved from access to data readiness.

Sustainability. Again.

What makes a digital project sustainable?

Data that remain usable

Open, structured, well-documented data and standards that can outlive individual tools and platforms.

Communities that remain engaged

Users, maintainers and institutions that continue to understand, support and develop the resource.

Knowledge that remains transferable

Documented workflows, decisions, methods and tacit expertise that allow others to reproduce, adapt and reuse the work.

Sustainability means preserving the capacity for future reuse, not simply keeping a website online.

Edmond & Morselli 2020, “Sustainability of digital humanities projects as a publication and documentation challenge,” *Journal of Documentation* 76.5: 1019–1031.

Selected references

Karin Hansson, “Metadata as imaginary demands: Exploring metadata markets in digital heritage with speculative design,” *JASIST* 76.2 (2025), 413–427.

Almas, Khazraee, Miller & Westgard, “Manuscript Study in Digital Spaces” (2018)

Byoung-Goon An, “A metadata design for humanities research data based on semantic structure of research literature” (2025).

Koraljka Golub & Rick Szostak, “Information retrieval of humanities resources: subject searching from a user perspective” (2025).

Nappi et al., “Models and tools for the digital organisation of knowledge: accessible and adaptive narratives for cultural heritage,” *Heritage Science* 12 (2024).

Edmond & Morselli 2020, “Sustainability of digital humanities projects as a publication and documentation challenge,” *Journal of Documentation* 76.5: 1019–1031.