

Digital Repositories and Manuscript Infrastructures

Maria Konstantinidou and Elpida Perdiki

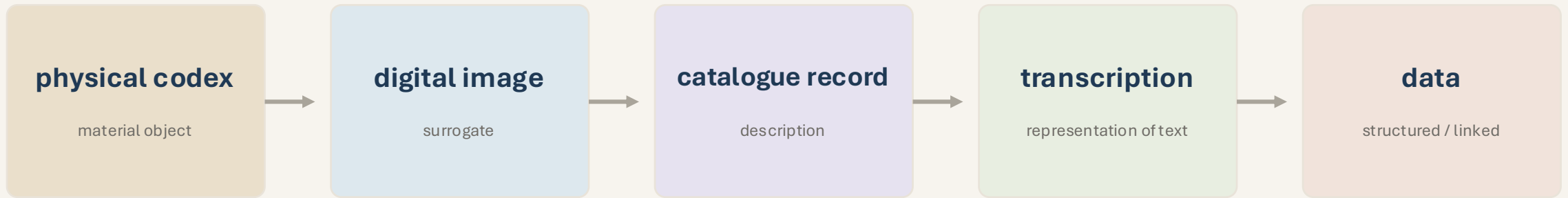


codexandcode.gr



Erasmus+

1. What is a “digital manuscript”?



Which of these is “the manuscript”?

2. From access to infrastructure

DIGITAL REPOSITORY

stores · describes · presents

digital objects

RESEARCH INFRASTRUCTURE

connects · exchanges · reuses

objects + metadata + identifiers + tools + people

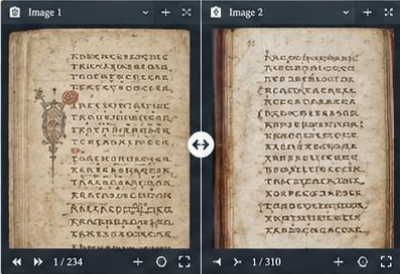
ACCESS ≠ INFRASTRUCTURE

2. From access to infrastructure

1

IIIF: Compare Manuscript Images

INTEROPERABLE VIEWING



View and compare manuscripts from different collections in a shared (IIIF) environment.

2

Manuscript Identification

PERSISTENT IDS & LINKED RECORDS



Sinaiticus, Greek New Testament

Shelfmark Sin. gr. 437

Persistent ID <https://doi.org/10.1234/ms.437>

Library Monastery of St Catherine (Sinai)

Related records

[ViaF 12345678](#)

[Wikidata Q123456](#)

[CERL cnp012345](#)

[Europeana 2048123](#)

Connected across collections, databases and authority files.

3

Metadata & Cataloguing

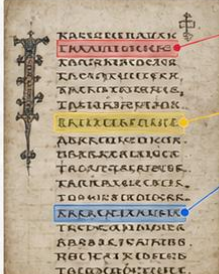
RICH, STRUCTURED DATA

Metadata	Cataloguing	Technical	Identifiers
Title	New Testament (selection)		
Shelfmark	Sin. gr. 437		
Repository	Monastery of St Catherine (Sinai)		
Date	4th century		
Material	Parchment		
Language	Greek		
Script	Uncial		
Contents	Gospels (with lacunae)		
Dimensions	24 x 16 cm		
Notes	Important early witness; palimpsest traces.		

4

Annotation

COLLABORATIVE READING



Textual variant

Alternative reading: λόγος

Marginal note

Later hand (12th c.?)

Liturgical use

Lectionary mark

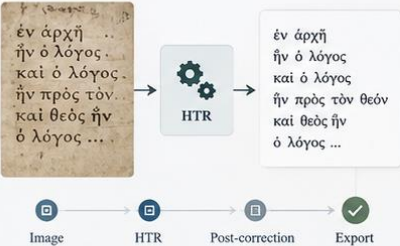
Add a note...

Highlight, annotate and discuss manuscript content together.

5

Text Extraction / HTR

FROM IMAGE TO MACHINE-READABLE TEXT



Transform manuscript images into searchable, reusable text.

6

Aggregation & Networks

CONNECTING PEOPLE, PLACES, TEXTS

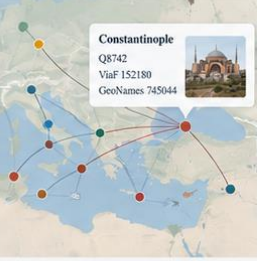


Aggregate data from multiple sources to reveal new relationships.

7

Map & Linked Data

PLACES, PEOPLE, ENTITIES



Constantinople

Q8742

ViaF 152180

GeoNames 745044

Related entities

Constantinople

People (12)

Manuscripts (34)

Texts (18)

Places (7)

Explore places, people and works through maps and linked data.

8

Export, Reuse & Open Data

SHARE AND BUILD FURTHER

Export

TEI, CSV, JSON

API

Programmatic access

Linked Data

RDF, LOD

Open Data

CC0 / CC BY

Reuse in:

Research & Teaching

Digital Editions

New Datasets

Scholarly Collaboration

Open data. More research. Greater impact.

CLASSICS IN A CONNECTED WORLD

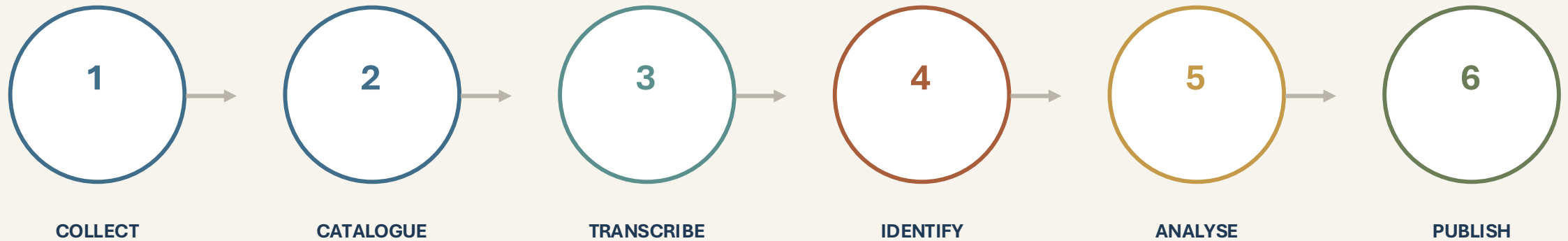
MANUSCRIPTS BRIDGE PAST AND FUTURE

ACCESS ≠ INFRASTRUCTURE

Digital Repositories and Manuscript Infrastructures

4

3. What do scholars actually do?



The outputs of one stage are the inputs of the next.

Donald J. Waters, “The Emerging Digital Infrastructure for Research in the Humanities,” *International Journal on Digital Libraries* 24 (2023), 87–102.

4. The first digital revolution: access

BEFORE

- travel & permissions
- microfilm / limited photography
- serial consultation
- difficult comparison

AFTER

- remote high-resolution access
- zoom & colour inspection
- simultaneous comparison
- mass availability



ACCESS IS NOT INTEROPERABILITY

ACCESS IS NOT COMPUTABILITY

ACCESS IS NOT SUSTAINABILITY

5. The digital surrogate is not neutral

What is preserved?

- sequence
- visual detail
- colour (approximately)
- annotations
- some codicological structure

What is transformed or lost?

- scale & thickness
- tactility / material feel
- lighting & colour choices
- cropping / omissions
- interface-defined sequence
or not...

A digital surrogate is a scholarly representation.

Bridget Whearty, *Digital Codicology: Medieval Books and Modern Labor* (Stanford, 2022).

6. Throwback: we can only suspect what might work

“A one-size-fits-all solution for digital manuscript studies remains difficult to envision.”

2018 problems

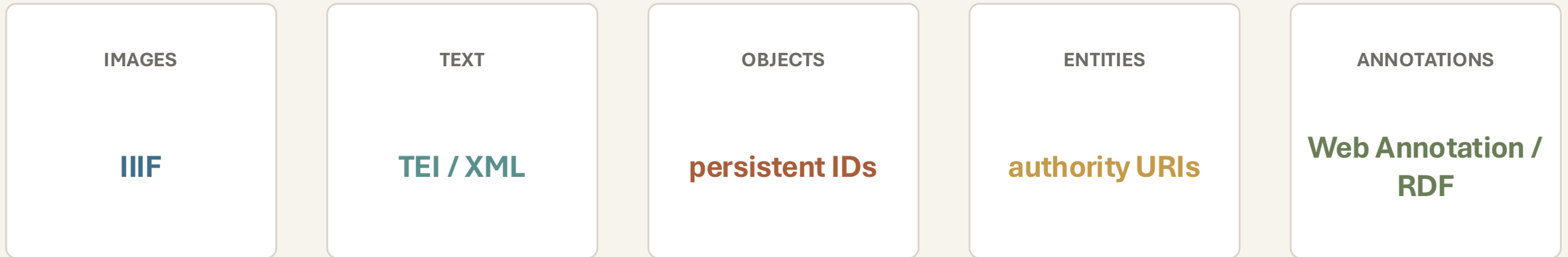
- fragmented tools
- developer dependence
- data trapped in platforms
- fragile project ecosystems

The prophetic principle

“use APIs that can outlast the tools”

Almas, B., Khazraee, E., Miller, M.T. & Westgard, J., “Manuscript Study in Digital Spaces: The State of the Field and New Ways Forward,” *DHQ* 12.2 (2018).

7. Interoperability as opposed to integration of tools in one giant platform



Independent systems can remain independent provided that their data can interact with other systems.

What is IIIF?

A common language that lets digital manuscript images move between repositories, viewers and research tools.

IIIF standardises

1. how images are delivered
2. how manuscripts/books are structured
3. how annotations are attached
4. how other systems can access and reuse them

IIIF is NOT viewer. Mirador is a viewer. IIIF is the standard that makes interoperability possible.

One institution stores it → many compatible tools can use it

8. Sustainability: beyond keeping the website alive

DATA

TOOLS

COMMUNITY

DOCUMENTATION

RESEARCH

“optimised for reuse”

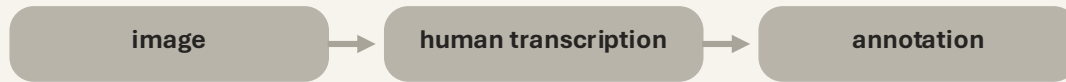
Sustainability is a process, not an end product.

A sustainable digital resource is one whose scholarly value can survive its current interface.

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

9. 2026: the second digital revolution

2018



The unknown variant in the 2018 workflow: machine learning

Digitised collections increasingly become computable corpora. New research questions: provenance, error, training data and reproducibility.

2026



10. What we missed completely

2018: we were preparing for

access

metadata

annotation

interoperability

reuse

2026: ML requires

large datasets

image/metadata
access

ground truth

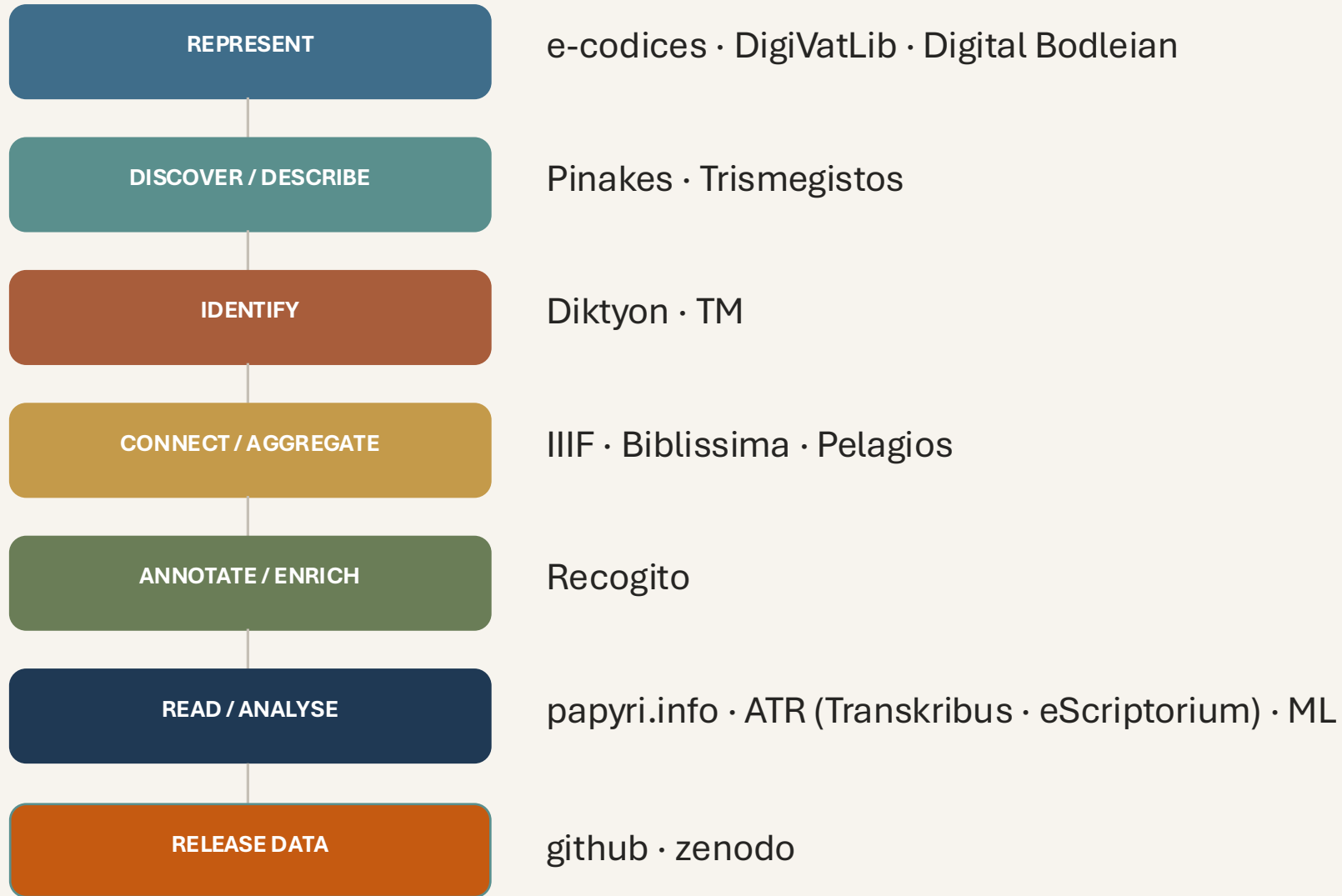
machine-readable
formats

APIs / IIIF / open
repositories

analysis

Problem shifts from access to data availability

10. A map of a possible manuscript ecosystem



11. Three tools we are focusing on today

e-codices

REPRESENT

How is the manuscript digitised,
structured and exposed?

Pinakes

IDENTIFY / DESCRIBE

What manuscript is this? What
does it contain? How is it related
to other records?

Recogito

ANNOTATE / ENRICH

What scholarly knowledge can
we add and make reusable?

Selected references

- Almas, B., Khazraee, E., Miller, M. T. & Westgard, J. (2018). “Manuscript Study in Digital Spaces: The State of the Field and New Ways Forward.” *Digital Humanities Quarterly* 12.2.
- Edmond, J. & Morselli, F. (2020). “Sustainability of digital humanities projects as a publication and documentation challenge.” *Journal of Documentation* 76.5: 1019–1031.
- Waters, D. J. (2023). “The Emerging Digital Infrastructure for Research in the Humanities.” *International Journal on Digital Libraries* 24: 87–102.
- Whearty, B. (2022). *Digital Codicology: Medieval Books and Modern Labor*. Stanford University Press.
- e-codices: TEI-P5 manuscript metadata and interoperable digital manuscript presentation.
- Pinakes / Diktyon: Greek manuscript discovery, identifiers and bibliographic aggregation.
- Recogito / Pelagios: annotation, entity linking and exportable structured data.