ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χρήση μεθόδων και εργαλείων Τεχνητής Νοημοσύνης στην Πολιτιστική Πληροφορία

Αντιστοίχιση ελεύθερων αρχειακών περιγραφών στο ISAD(G): μια συγκριτική αξιολόγηση Μεγάλων Γλωσσικών Μοντέλων

Ευφροσύνη Αποστόλου

Επιβλέπων: Επίκ. Καθ. Ματθαίος Δαμίγος
ΠΜΣ «Διαχείριση Ψηφιακής Πληροφορίας – Υπηρεσίες Πληροφόρησης»
Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας και Μουσειολογίας — Ιόνιο Πανεπιστήμιο
Κέρκυρα, Μάιος 2026

ΕΠΙΣΚΟΠΗΣΗ

Δομή της παρουσίασης

Οι ανεπεξέργαστες συσσωρεύσεις (backlogs)

Η αρχειακή περιγραφή είναι χρονοβόρα και εξειδικευμένη εργασία

Όγκος συλλογών ↑, πόροι σταθεροί → υλικό που μένει χρόνια χωρίς περιγραφή

Το αποτέλεσμα: «ανεπεξέργαστες συσσωρεύσεις» (backlogs) — υλικό απρόσιτο στους ερευνητές

Greene & Meissner (2005), «More Product, Less Process», λιγότερη επεξεργασία, ευρύτερη πρόσβαση

Τα born-digital αρχεία εντείνουν το πρόβλημα

Η Τεχνητή Νοημοσύνη σε ραγδαία εξέλιξη

Τέσσερις γενιές μεθόδων — κάθε μία λύνει τα όρια της προηγούμενης

Έμπειρα Συστήματα (Rule-based) [AA18]

Κανόνες διάταξης & γλωσσικά μοτίβα

+Διαφανές, σταθερό
- Εύθραστο σε αλλαγές μορφής

Μηχανική Μάθηση (ML) [GHK+12]

μάθηση από δεδομένα
CRF, SVM με χαρακτηριστικά σελίδας

+ Προσαρμόσιμο
-Χρειάζεται annotated data

Deep Learning [NR23]

νευρωνικά δίκτυα
CNN + bi-LSTM, ~94% ακρίβεια

+Σύνθετα μοτίβα
- Μεγάλα datasets

2017 — Transformer [VSP+17]

σταθμός για τη γλώσσα
Αρχιτεκτονική Νευρωνικού Δικτύου με μηχανισμό shelf-attention

2022+ — LLMs [Bag24]

Κατανόηση & γένεση γλώσσας

+ Ευέλικτο, χωρίς training
- Hallucinations

Νέες δυνατότητες → νέα ερωτήματα: η δυνατότητα δεν συνεπάγεται αξιοπιστία, ιδίως σε ευαίσθητους τομείς όπως η πολιτιστική κληρονομιά

Μπορεί η ΤΝ να βοηθήσει;

NLP / Μηχανική μάθηση: αναγνώριση οντοτήτων (πχ agents), ταξινόμηση, εξαγωγή μεταδεδομένων

– Όμως απαιτούν annotated datasets — σπάνια στον αρχειακό χώρο

LLMs: εκτέλεση εργασίας χωρίς ειδική εκπαίδευση ανά task (zero-/few-shot)

– Η συμπεριφορά καθορίζεται από τις οδηγίες, όχι από εξειδικευμένη εκπαίδευση

Η δυνατότητα ≠ αξιοπιστία

ΤΟ ΠΡΟΒΛΗΜΑ & ΤΟ ΚΙΝΗΤΡΟ

Τα τρία ερευνητικά ερωτήματα

i Μπορούν τα LLMs να μετατρέπουν ελεύθερο κείμενο σε δομημένα στοιχεία συμβατά με το ISAD(G);

ii Συμμορφώνονται με αυστηρούς κανόνες και διατηρούν την αρχειακή ιεραρχία;

iii Πώς διαφοροποιείται η απόδοση ανάλογα με το μέγεθος του μοντέλου;

Μεγάλα Γλωσσικά Μοντέλα (LLMs)

Πώς λειτουργούν με απλά λόγια

1. Pre-training

Εκπαίδευση σε τεράστιες ποσότητες κειμένου. Το μοντέλο μαθαίνει να προβλέπει την επόμενη λέξη.

2. Instruction tuning

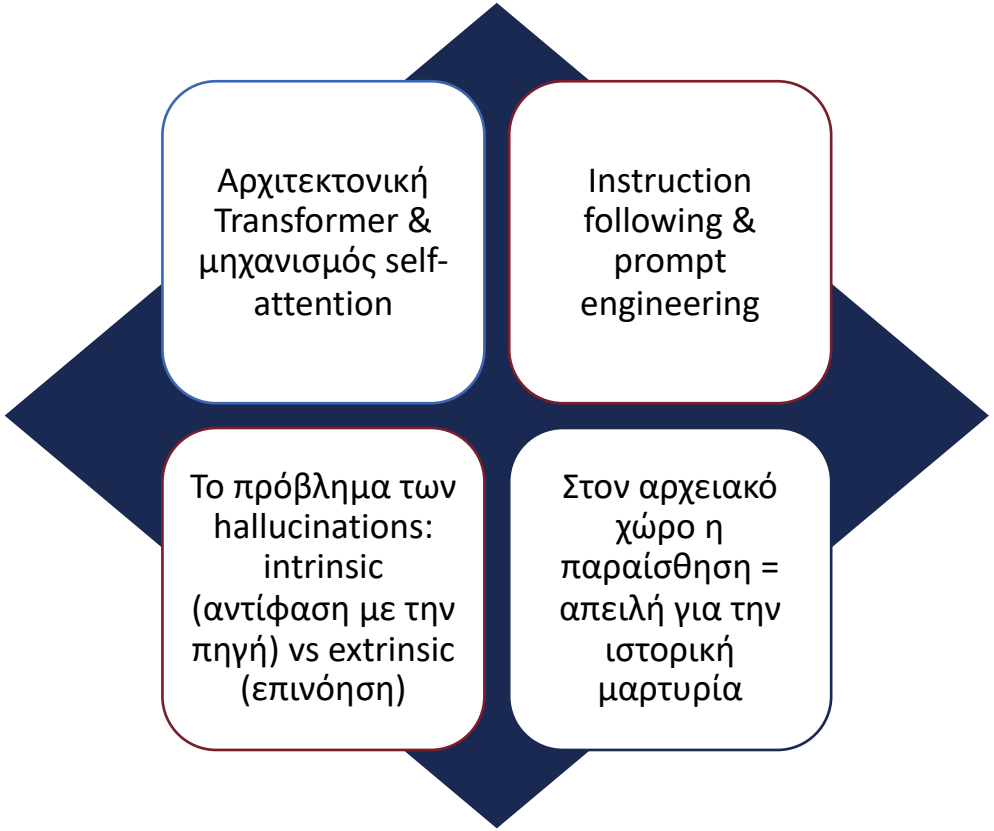
Δεύτερη φάση: μαθαίνει να ακολουθεί οδηγίες σε φυσική γλώσσα — όχι απλώς να συνεχίζει κείμενο.

3. RLHF

Άνθρωποι αξιολογητές προτιμούν συγκεκριμένες απαντήσεις. Το μοντέλο γίνεται πιο χρήσιμο και ασφαλές.

Στο πείραμά μου χρησιμοποίησα μόνο instruction-tuned μοντέλα *(Instruct / it)* ώστε να μπορούν να ακολουθήσουν τους έξι κανόνες του prompt μου.

LLMs εν συντομία



LLMs και τα όριά τους

Τα LLMs μετατοπίζουν τη «νοημοσύνη» στο ίδιο το μοντέλο

Federated intelligence (Groppe κ.ά.) — συνδυασμός πολλαπλών μοντέλων που το ένα ελέγχει το άλλο

Αποδοτικά σε δομημένες εργασίες, ασθενή σε σύνθετη σημασιολογία

Σύσταση της βιβλιογραφίας: human-in-the-loop, όχι πλήρης αυτονομία

Εφαρμογές ΤΝ στην πολιτιστική κληρονομιά

Τι δείχνει η έρευνα — και μια κρίσιμη αντίφαση

Ψηφιακές ανθρωπιστικές

3D ψηφιοποίηση, ανάγνωση επιγραφών, έξυπνα εικονικά μουσεία

Pavlidis [Pav22]

Πρόσβαση & ερμηνεία

Συστήματα AI+HCI που προσαρμόζουν περιεχόμενο στον χρήστη

Ardissono κ.ά. [ARM22]

Ερευνητική ατζέντα ΕΕ

Προτεραιότητες για ΤΝ στην ψηφιακή κληρονομιά στην Ευρώπη

Münster κ.ά. [MML+24]

Συμπεριληπτικότητα

LLMs εντοπίζουν προκαταλήψεις & ελλείψεις για περιθωριοποιημένες ομάδες

Osti & Roke [OR24]

Η αντίφαση: στα άρθρα «επίδειξης» η ΤΝ φαίνεται σχεδόν μαγική — στις μελέτες της καθημερινότητας των φορέων εμφανίζεται εύθραυστη, γεμάτη σφάλματα.

Υβριδικά σενάρια & ποιότητα

Η τάση δεν είναι αντικατάσταση — είναι επέκταση των υπαρχουσών pipelines

AI4DiTraRe

[JJG+25]

Πολλαπλοί agents — LLMs + κανόνες + επιλεκτική ανθρώπινη παρέμβαση σε κλιματικά repositories

ArcGPT

[ZHP+23]

LLM εκπαιδευμένο σε αρχειακά δεδομένα, με benchmark AMBLE — υπερέχει γενικών LLMs

Federated Intelligence

[GMW+25]

Συνδυασμός πολλαπλών LLMs που αλληλοελέγχονται για παραγωγή αρχειακών περιγραφών

Κοινό μοτίβο όλων των υβριδικών συστημάτων

LLM παράγει → schema validator ελέγχει → άνθρωπος επικυρώνει

Η ποιότητα δεν είναι μόνο ορθότητα — είναι και πληρότητα, σαφήνεια, συνέπεια [LRT21]. Και η ερμηνευσιμότητα είναι κρίσιμη: ένα αποτέλεσμα που δεν εξηγείται, δεν εγκρίνεται [TAP25].

Το ερευνητικό κενό

Τι υπάρχει στη βιβλιογραφία

Πολλές μελέτες για **βιβλιογραφικά μεταδεδομένα** *(GROBID, CERMINE κ.λπ.)*

- Αυτόματη εξαγωγή από επιστημονικά άρθρα [NR23, GHK+12]
- Generic LLM pipelines για μεταδεδομένα [Bag24, AAG25]
- Domain-specific αρχειακά μοντέλα όπως ArcGPT [ZHP+23]

Τι λείπει

Συστηματική αξιολόγηση LLMs ειδικά για **ISAD(G)**

- Σύγκριση πολλαπλών μοντέλων σε ίδιες συνθήκες
- Αξιολόγηση ιεραρχικής αρχειακής λογικής
- Επαγγελματική αξιολόγηση από αρχειονόμο, όχι μετρικές μηχανής
- Έρευνα σε ελληνικά αρχειακά δεδομένα

Σε αυτό το κενό τοποθετείται η παρούσα εργασία

ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ

Πλαίσια & Πρότυπα μεταδεδομένων

Από απλά σχήματα πεδίων προς πλήρη οντολογικά μοντέλα

Πρότυπο	Τύπος	Πεδίο εφαρμογής	Σχέση με LLMs
Dublin Core	15 στοιχεία πεδίων	Ψηφιακές βιβλιοθήκες	Εύκολη — επίπεδη δομή
ISAD(G)	Ιεραρχικό, 26 πεδία	Αρχεία	**Δύσκολη — απαιτεί ιεραρχική λογική**
EAD	XML finding aid	Αρχεία (κωδικοποίηση)	Μέτρια — αυστηρό schema
CIDOC-CRM	Οντολογία (RDF/OWL)	Μουσεία, κληρονομιά	Δύσκολη — υψηλή πολυπλοκότητα
RiC-CM / RiC-O	Εννοιολογικό + οντολογία	Αρχεία (νέα γενιά)	Υπό διερεύνηση — 2023/2025

Το κρίσιμο σημείο: τα περισσότερα συστήματα ΤΝ σχεδιάζονται για επίπεδα σχήματα (Dublin Core).
Το ISAD(G) είναι ιεραρχικό — και αυτή η μετάβαση από «ένα έγγραφο» σε «δομή επιπέδων» *παραμένει ανοιχτό ζητούμενο.*

ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ

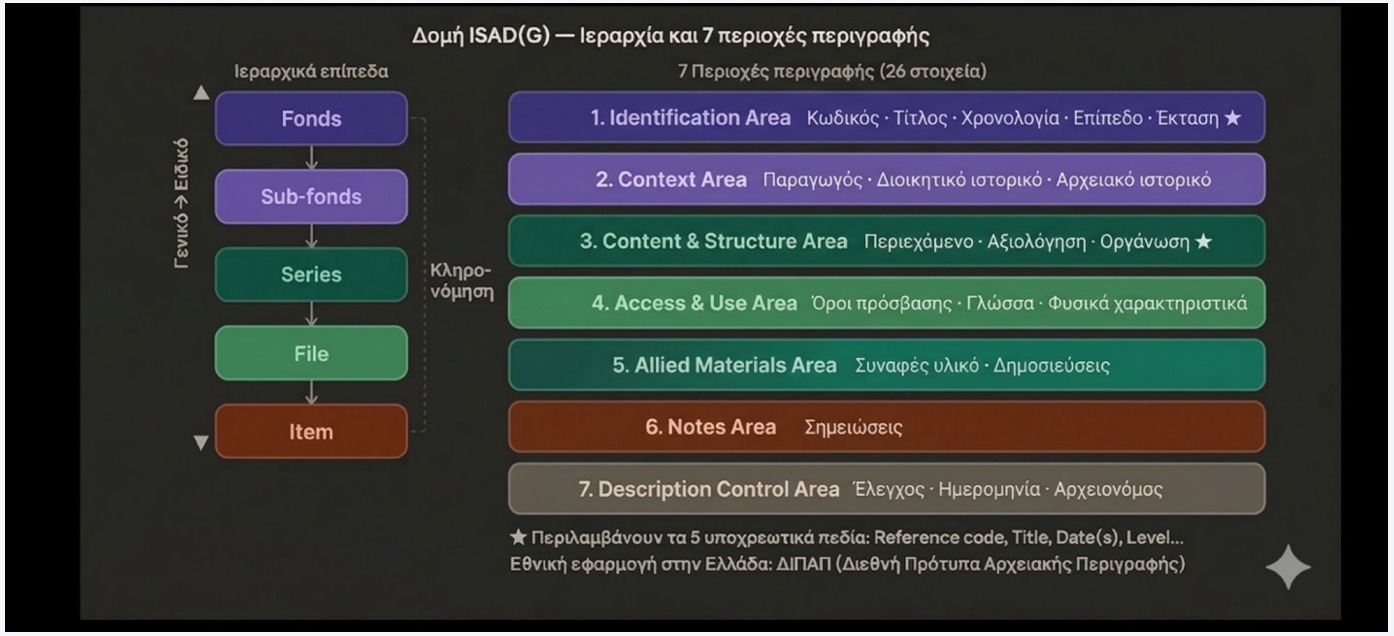
Το πρότυπο ISAD(G)

Σεβασμός προέλευσης & διατήρηση αρχικής τάξης

Ιεραρχική περιγραφή: fonds → series → file → item

26 στοιχεία περιγραφής σε 7 περιοχές

Αρχή της κληρονομικότητας: η πληροφορία δηλώνεται μία φορά, στο ανώτερο επίπεδο



Σχήμα: Ιεραρχική δομή και 7 περιοχές περιγραφής του ISAD(G)

ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ

Οι επτά περιοχές περιγραφής του ISAD(G)

Περιοχή	Βασικά στοιχεία περιγραφής
1. Αναγνώριση	κωδικός, τίτλος, χρονολογία, επίπεδο, έκταση
2. Πλαίσιο	παραγωγός, διοικητικό/βιογραφικό & αρχειακό ιστορικό
3. Περιεχόμενο & Δομή	παρουσίαση περιεχομένου, αξιολόγηση, σύστημα ταξινόμησης
4. Πρόσβαση & Χρήση	όροι πρόσβασης & αναπαραγωγής, γλώσσα, φυσικά χαρακτηριστικά
5. Συναφές Υλικό	πρωτότυπα, αντίγραφα, σχετικές ενότητες, δημοσιεύσεις
6. Σημειώσεις	ειδικές παρατηρήσεις εκτός των άλλων περιοχών
7. Έλεγχος Περιγραφής	ποιος, με ποιους κανόνες και πότε συνέταξε την εγγραφή

ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ

Τα σημαντικότερα εργαλεία & πρότυπα

Πρότυπα	ISAD(G)
Πλατφόρμες πρόσβασης	Hugging Face Chat · OpenRouter
Τεχνικές prompting	zero-shot · few-shot · chain-of-thought · RLHF

Τα τρία ερευνητικά ερωτήματα

i Μπορούν τα LLMs να μετατρέπουν ελεύθερο κείμενο σε δομημένα στοιχεία συμβατά με το ISAD(G);

ii Συμμορφώνονται με αυστηρούς κανόνες και διατηρούν την αρχειακή ιεραρχία;

iii Πώς διαφοροποιείται η απόδοση ανάλογα με το μέγεθος του μοντέλου;

ΤΟ ΠΙΛΟΤΙΚΟ ΠΕΙΡΑΜΑ

Σχεδιασμός σε τέσσερα στάδια

1

Δεδομένα εισόδου

Αρχειακή περιγραφή ειδικού → plain text

2

Prompt 6 κανόνων

ROLE, SOURCE, FILL, AMBIGUITY, UNITS, COPY

3

11 μοντέλα

Από 3B έως 235B παραμέτρους

4

Αξιολόγηση

Από αρχειονόμο, σε σχέση με gold standard

Πηγή: Αρχείο Επτανήσου Πολιτείας (1800–1807), ΓΑΚ-ΑΝΚ • **Παράμετροι:** temperature = 0 (ντετερμινιστική παραγωγή) και ίδιο prompt για όλα τα μοντέλα

ΤΟ ΠΙΛΟΤΙΚΟ ΠΕΙΡΑΜΑ

Δεδομένα: το Αρχείο της Επτανήσου Πολιτείας

Αρχείο 473 κιβωτίων, ταξινομημένο θεματικά και χρονολογικά

Μελέτη περίπτωσης: φάκελος Β.2 / Φ.8 — παλαιό κατάστιχο με 3 τεκμήρια

– Συνθήκη & Σύνταγμα Κωνσταντινουπόλεως, Επιστολή Επτανησιακής Γερουσίας

5 εγγραφές αναφοράς (gold standard) από εξειδικευμένη αρχειονόμο

Πέντε ιεραρχικά επίπεδα: fonds → φάκελος → 3 τεκμήρια

ΤΟ ΠΙΛΟΤΙΚΟ ΠΕΙΡΑΜΑ

Ο σχεδιασμός του prompt: έξι κανόνες

ROLE	Το μοντέλο λειτουργεί ως αρχειακός καταλογογράφος
SOURCE	Χρήση μόνο του δοσμένου κειμένου — καμία εξωτερική γνώση
FILL	«N/A» για κάθε πεδίο που δεν αναφέρεται ρητά
AMBIGUITY	Οι αντιφάσεις αντιγράφονται ως έχουν
UNITS	Μία μονάδα ISAD(G) ανά τεκμηριωμένη αρχειακή μονάδα
COPY	Οι τιμές μεταφέρονται κατά λέξη, χωρίς παράφραση

ΤΟ ΠΙΛΟΤΙΚΟ ΠΕΙΡΑΜΑ

Τα τέσσερα στάδια της αξιολόγησης

1	**Προετοιμασία**	μετατροπή της περιγραφής σε απλό ελεύθερο κείμενο
2	**Παραγωγή**	ίδιο prompt στα 11 μοντέλα, μέσω API
3	**Σύγκριση**	με τις 5 εγγραφές αναφοράς (gold standard)
4	**Αξιολόγηση**	χειροκίνητα, από ομάδα 3 ειδικών αρχειονόμων

Γιατί όχι semantic similarity; Ο κανόνας COPY απαιτεί κατά λέξη απόδοση — η σημασιολογική ομοιότητα δεν προσφέρει επιπλέον πληροφορία.

Κριτήρια αξιολόγησης

Expected Units Generated — πόσες από τις 5 αναμενόμενες μονάδες παρήχθησαν

Hallucinated Units — πόσες επινοημένες μονάδες προστέθηκαν

Correct Level Assignment — σωστό ιεραρχικό επίπεδο ανά μονάδα

Επιπλέον: πληρότητα 26 πεδίων, σύμβαση «id στο Αρχείο», διαφοροποίηση παραγωγού

Συνολική εικόνα

Μεγάλη διακύμανση μεταξύ των 11 μοντέλων

Από μοντέλα με 0/5 σωστές μονάδες έως πλήρη δομική κάλυψη

Καλύτερο δομικά: google-gemma-3-12b-it — 5/5 μονάδες, 0 παραισθήσεις

Ο αριθμός εγγραφών από μόνος του είναι παραπλανητικός δείκτης

ΑΠΟΤΕΛΕΣΜΑΤΑ & ΕΥΡΗΜΑΤΑ

Ο πίνακας αξιολόγησης

Μοντέλο	Αναμεν. μονάδες	Μονάδες παραίσθησης	Ορθό επίπεδο
meta-llama-Llama-3.1-8B	1/5	0	1/1
openai-gpt-oss-20b	5/5	2	4/5
deepseek-ai-DeepSeek-V3.2	5/5	1	4/5
meta-llama-Llama-3.2-3B	1/5	0	1/1
meta-llama-Llama-3.3-70B	5/5	1	4/5
Qwen-Qwen3-235B-A22B-Instruct	4/5	1	4/4
arcee-ai-trinity-large	2/5	0	1/2
arcee-ai-trinity-mini	1/5	1	1/1
deepseek-r1-0528	0/5	0	—
google-gemma-3-4b-it	0/5	1	—
google-gemma-3-12b-it	**5/5**	**0**	**4/5**

Αναμεν. μονάδες: σωστά αναπαραγόμενες μονάδες (σύνολο 5) · Μονάδες παραίσθησης: μη τεκμηριωμένες μονάδες · Ορθό επίπεδο: σωστή ιεραρχική ανάθεση

Παράδειγμα: από το ελεύθερο κείμενο στο ISAD(G)

ΕΙΣΟΔΟΣ — ελεύθερο κείμενο

«Στην περιφερειακή Υπηρεσία των ΓΑΚ — Αρχεία Νομού Κέρκυρας, εντός του Αρχείου Ιονίου Γερουσίας, διασώζεται το Αρχείο της Επτανήσου Πολιτείας, το οποίο απαρτίζεται από 473 κιβώτια με φακέλους και τεκμήρια. Όλο το Αρχείο είναι ταξινομημένο θεματικά και μέσα στο κάθε θέμα χρονολογικά.

Η Επτάνησος Πολιτεία υπήρξε το πρώτο ημιανεξάρτητο κρατίδιο στον ελλαδικό χώρο… Ιδρύθηκε το 1800 … καταργήθηκε στις 07.07.1807 με τη Συνθήκη του Τιλσίτ…»

(αδόμητη αφήγηση, χωρίς πεδία)

ΕΞΟΔΟΣ — gemma-3-12b-it

1. Ταυτότητα:
   Κωδικός: GR GSA ANK
   Τίτλος: Αρχείο της Επτανήσου Πολιτείας
   Χρονολογία: 1800-1807
   Επίπεδο: Fonds · Έκταση: 473 κιβώτια

2. Πλαίσιο:
   Παραγωγός: οι εκπρόσωποι θεσμών κεντρικής Διοίκησης…
   Διοικητική ιστορία: Η Επτάνησος Πολιτεία υπήρξε το πρώτο ημιανεξάρτητο κρατικό μόρφωμα…

3. Περιεχόμενο & Δομή:
   Σύστημα ταξινόμησης: θεματική και εντός θέματος χρονολογική

(δομημένη εγγραφή στα 7 πεδία ISAD(G))

ΑΠΟΤΕΛΕΣΜΑΤΑ & ΕΥΡΗΜΑΤΑ

Η κρίσιμη αστοχία: η πληροφορία υπήρχε, αλλά χάθηκε

ΣΤΗΝ ΕΙΣΟΔΟ ΥΠΗΡΧΕ

Το ελεύθερο κείμενο ανέφερε ρητά:

«…Συντάκτης των τριών αυτών τεκμηρίων είναι το Νομοθετικό Σώμα της Επτανήσου Πολιτείας με συνταγματικές αρμοδιότητες.»

→ Ο παραγωγός των τεκμηρίων HTAN δηλωμένος στην πηγή.

ΣΤΗΝ ΕΞΟΔΟ ΤΟΥ ΜΟΝΤΕΛΟΥ

Τεκμήριο 14a — Συνθήκη Κωνσταντινουπόλεως
Παραγωγός: N/A

Τεκμήριο 14b — Σύνταγμα Κωνσταντινουπόλεως
Παραγωγός: N/A

Τεκμήριο 14c — Επιστολή Γερουσίας
Παραγωγός: N/A

→ Και στα τρία τεκμήρια το πεδίο έμεινε κενό.

Το μοντέλο είχε την πληροφορία μπροστά του όμως δεν τη συνέδεσε ιεραρχικά με τα τεκμήρια.

Εύρημα 1: το μέγεθος δεν καθορίζει την ποιότητα

Καλύτερο μοντέλο: gemma-3-12b-it — μόλις 12B παράμετροι

Ξεπέρασε το Qwen3-235B (παρέλειψε επίπεδο, πρόσθεσε λάθος series)

Ξεπέρασε το Llama 3.3-70B (πρόσθεσε αδικαιολόγητο ενδιάμεσο επίπεδο)

Τα πολύ μικρά μοντέλα (3B–8B), σχεδόν καθολική αποτυχία

Καθοριστικό: η ποιότητα του instruction tuning, όχι η κλίμακα

Εύρημα 2: τα μοντέλα δεν σκέφτονται ιεραρχικά

Κανένα μοντέλο δεν χρησιμοποίησε τη σύμβαση «id στο Αρχείο»

Κανένα δεν διαφοροποίησε τον παραγωγό ανά τεκμήριο

Αντιμετωπίζουν κάθε εγγραφή ως ανεξάρτητο έγγραφο και όχι ως κόμβο ιεραρχίας

Βαθύτερη ασυμφωνία: λογική των LLM ↔ αρχές της αρχειακής επιστήμης

ΣΥΜΠΕΡΑΣΜΑΤΑ

Απαντήσεις στα τρία ερωτήματα

i	**Ναι, υπό προϋποθέσεις**	Η δομική συμμόρφωση είναι εφικτή· η ποιότητα πέφτει στα κατώτερα επίπεδα.
ii	**Όχι πλήρως**	Κανένα μοντέλο δεν τήρησε όλους τους κανόνες χωρίς παρεκκλίσεις.
iii	**Το μέγεθος δεν αρκεί**	Το instruction tuning βαρύνει περισσότερο από την κλίμακα.

Περιορισμοί: μία μελέτη περίπτωσης · απουσία συστηματικού συνόλου παραδειγμάτων εισόδου-εξόδου

ΣΥΜΠΕΡΑΣΜΑΤΑ

Μελλοντική έρευνα

Υβριδικά συστήματα — πολλαπλά μοντέλα + validation + ανθρώπινη εποπτεία

Reasoning LLMs (o1/o3, DeepSeek R1, extended thinking) για ιεραρχική λογική

Domain-specific μοντέλα εκπαιδευμένα σε ελληνικά αρχειακά corpora

Ενσωμάτωση ιεραρχικών δομών — knowledge graphs, Ric-Cm / Ric-O / EAD

Επέκταση σε πολύγλωσσα και born-digital σύνολα · συστηματικά παραδείγματα input-output, καθώς και πειράματα σε περισσότερα αρχεία

«Τα LLMs δεν αντικαθιστούν τον αρχειονόμο· τον υποστηρίζουν.»

Η ανθρώπινη επικύρωση παραμένει αναντικατάστατη —
για λόγους δεοντολογίας, λογοδοσίας και ερμηνευσιμότητας.

Ευχαριστώ θερμά τον επιβλέποντα καθηγητή και την εξεταστική επιτροπή.

Ερωτήσεις;

Παράρτημα Β
Prompt used for generating fonds – and sub-unit-level ISAD(G)

The prompt below is designed to generate ISAD(G)-compliant archival descriptions based exclusively on archivist-provided source text. It instructs the LLM not to use any outside knowledge and not to guess or ll in missing information. If an ISAD(G) eld is not clearly stated in the text, the model must write N/A (not stated in source). The model must also create separate ISAD(G) blocks for the fonds and for each sub-unit that is explicitly described in the input text.

ROLE: Archival description cataloguer.

SOURCE RULE: Use ONLY the text in SOURCE DESCRIPTION. No outside knowledge. No guessing.

FILL RULE: For any ISAD(G) eld not explicitly stated, write: N/A (not stated in source).

AMBIGUITY RULE: If the source conicts/unclear, copy the values exactly as written. Do not explain or choose.

UNITS RULE: Create one ISAD(G) block for: (1) Fonds, and (2) every subunit explicitly described (series/sub-series/le/item/part). Do NOT invent units or hierarchy. Only use hierarchy if explicitly indicated (e.g., Series, numbering, contains, part of).

COPY RULE: Field values must be copied verbatim from the source (you may remove bullet markers only).

OUTPUT: No intro, no summary, no commentary. One block per unit, in source order. For EACH unit output exactly these elds in this order: 1. Identity: - Reference code, - Title, - Date(s), - Level of description, - Extent and medium. 2. Context: - Name of creator(s), - Administrative/Biographical history, - Archival history, - Immediate source of acquisition or transfer. 3. Content and structure: - Scope and content, - Appraisal, destruction and scheduling information, - Accruals, - System of arrangement. 4. Access and use: - Conditions governing access, - Conditions governing reproduction, - Language/scripts of material, - Physical characteristics and technical requirements, - Finding aids. 5. Allied materials: - Existence and location of originals, - Existence and

location of copies, - Related units of description, - Publication note. 6. Notes: - Note(s). 7. Description control: - Archivist's note, - Rules or conventions, - Date(s) of description

SOURCE DESCRIPTION: Archivist's Source Description (Input Text)

Συγκριτικά αποτελέσματα του Google-gema-3-12b-it

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
3.1.3 Χρονολογία (ες)	3.1.4 Επίπεδο περιγραφής	3.1.5 Μέγεθος και υπόστρωμα της ενότητας	3.2.1 Όνομα παραγωγού (ών)	3.2.2 Διοικητική ιστορία / Βιογραφικό	3.2.3 Ιστορικό της ενότητας περιγραφής	3.2.4 Διαδικασία πρόσκτησης	3.3.1 Παρουσίαση περιεχομένου	3.3.2 Επιλογές, εκκαθαρίσεις και τελική διατύπωση	3.3.3 Προσθήκες υλικού	3.3.4 Σύστημα ταξινόμησης	3.4.1 Όροι πρόσβασης	3.4.2 Όροι αναπαραγωγής	3.4.3 Γλώσσα /γραφή των τεκμηρίων	3.4.4 Φυσικά χαρακτηριστικά και τεχνικές προϋποθέ	3.4.5 Εργαλεία έρευνας	3.5.1 Εντοπισμός πρωτοτύπων	3.5.2 Εντοπισμός αντιγράφων	3.5.3 Συμπληρωματικές πηγές / Σχετικές ενότητες	3.5.4 Δημοσιεύσεις / Βιβλιογραφία	3.6.1 Παρατηρήσεις	3.7.1 Παρατηρήσεις και όνομα του / της αρχειονό	3.7.2 Κανόνες ή πρότυπα περιγραφής	3.7.3 Χρονολογία (ες) περιγραφής	Θέματα												
1800 -1807	Αρχείο	473 κιβώτι	Επτάνησος	Η Επτάνησ	Το Αρχείο	Δημόσια Ει	Το Αρχείο	Δεν αναφέ	Δεν προβλ	Ταξινόμησ	Ανοικτή Πρ	Ελεύθερη	Ιταλική, ελ	Το αρχειακ	Γενικό (συν	Εντός του Α	Ψηφιακό Α	Αρχείο Βιβ	1.Ίδρυμα τι	Τα πρωτότ	Αλεξάνδρα	Διεθνές Πρ	Ιούνιος 202	Επτάνησος	Πολιτεία, Συνταγματικά Κείμενα.											
1800-1807	Fonds	473 κιβώτι	οι εκπρόσω	Η Επτάνησ	N/A	Δημόσια Ει	περιλαμβά	Επιλογές, εκκαθαρίσεις του Αρχείου δεν	N/A	θεματική κ	ανοικτή	ελεύθερη,	ιταλική, η ε	παρατηρεί	ένα γενικό	εντός του Α	στο ψηφια	Αρχείο Βιβ	N/A	N/A	Αλεξάνδρα	Διεθνές Πρ	Ιούνιος 2022													
21.3.1800 -	Φάκελος μ	Κατάστιχο-	Επτάνησος	Id στο Αρχε	Id στο Αρχε	Id στο Αρχε	Στο Κατάστ	Id στο Αρχε	Id στο Αρχε	Id στο Αρχε	Id στο Αρχε	Id στο Αρχε	Ιταλική	Όλα τα φύ	Id στο Αρχε	Εντός του κ	Id στο Αρχε	Id στο Αρχε	Id στο Αρχε	Id στο Αρχε	Αλεξάνδρα	Διεθνές Πρ	Ιούνιος 202	Επτάνησος	Πολιτεία, Βιβλίο-Κατάστιχο, Συνθήκη Κωνσταντινουπόλεως, Βυζαντινό Σύνταγμα, Επιστολή Γερουσίας προς Υψηλή Πύλη											
N/A	File	N/A	N/A	αναφέρον	N/A	N/A	περιέχει π	N/A	N/A	N/A	N/A	N/A	ιταλική, η ε	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A												
21.3.1800	Τεκμήριο	id στο Αρχε	Επτάνησος	id στο Αρχε	id στο Αρχε	id στο Αρχε	Στο τεκμήρ	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	Ιταλική	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	Αλεξάνδρα	Διεθνές Πρ	Ιούνιος 202	Επτάνησος	Πολιτεία, Συνθήκη Κωνσταντινουπόλεως											
N/A	Item	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	ιταλική, η ε	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A												
21.3.1800	Τεκμήριο	id στο Αρχε	Επτάνησος	id στο Αρχε	id στο Αρχε	id στο Αρχε	Στο τεκμήρ	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	Ιταλική	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	Αλεξάνδρα	Διεθνές Πρ	Ιούνιος 202	Επτάνησος	Πολιτεία, Σύνταγμα Βυζαντινό											
N/A	Item	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	ιταλική, η ε	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A												
25.10/06.1	Τεκμήριο	id στο Αρχε	Επτάνησος	id στο Αρχε	id στο Αρχε	id στο Αρχε	Στο τεκμήρ	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	Ιταλική	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	id στο Αρχε	Αλεξάνδρα	Διεθνές Πρ	Ιούνιος 202	Επτάνησος	Πολιτεία, Επιστολή Γερουσίας προς Υψηλή Πύλη											
N/A	Item	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	ιταλική, η ε	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A												

Βιβλιογραφία

[Καβ23]	**Καββαδία, Α.: Οι Θεσμοί της Κεντρικής Διοίκησης στα Συντάγματα των Ιονίων Νήσων κατά την Επτάνησο Πολιτεία (1800–1803) και η Παραγωγή Αρχείων. Διδακτορική διατριβή (2023). https://doi.org/10.12681/eadd/55516, http://hdl.handle.net/10442/hedi/55516 (κείμενο στα ελληνικά)**
[AA18]	J. Azimjonov, J. Alikhanov. Rule Based Metadata Extraction Framework from Academic Articles, arXiv preprint arXiv:1807.09009, 2018
[AAG25]	Z. Alyafeai, M. S. Al-Shaibani, B. Ghanem. MOLE: Metadata Extraction and Validation in Scientific Papers Using LLMs, In Findings of the Association for Computational Linguistics: EMNLP 2025, pages 12236–12264, 2025
[ARM22]	L. Ardissono, G. E. Raptis, N. Mauro. Special Issue on AI and HCI Methods and Techniques for Cultural Heritage Curation, Exploration and Fruition. Applied Sciences, 12 (19): 10118, 2022. DOI: 10.3390/app121910118.
[ATO20]	W. Z. Alma'aitah, A. Z. Talib, M. A. Osman. Opportunities and challenges in enhancing access to metadata of cultural heritage collections: a survey, Artificial Intelligence Review 53: 3621–3646, 2020, Springer Nature. https://doi.org/10.1007/s10462-019-09773-w
[Bag24]	M. Bagchi. A Generative AI-driven Metadata Modelling Approach, Accepted for publication in Library Trends, Special Issue "Generative AI and Libraries", Johns Hopkins University Press, 2024, https://doi.org/10.48550/arXiv.2501.04008
[BDD+23]	Bountouri, L., Damigos, M., Drakiou, M., Gergatsoulis, M. and Kalogeros, E. (2023). The Semantic Mapping of RiC-CM to CIDOC-CRM. In: Lecture Notes in Computer Science, σσ. 90–99. Springer Nature Singapore. International Conference on Asia-Pacific Digital Libraries (ICADL) 2023. https://doi.org/10.1007/978-981-99-8088-8_8
[Bod23]	J. Bodenhamer. The Reliability and Usability of ChatGPT for Library Metadata, Journal of Library Metadata, 23 (3–4): 123–137, 2023.
[CBJ+21]	G. Colavizza, T. Blanke, C. Jeurgens, J. Noordegraaf. Archives and AI: An Overview of Current Debates and Future Perspectives. Journal on Computing and Cultural Heritage, 15 (1): Article 4, 2021.
[CMK+13]	S. R. Choudhury, P. Mitra, A. Kirk, S. Szep, D. Pellegrino, S. Jones, C. L. Giles. Figure Metadata Extraction from Digital Documents, In Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR 2013), pages 135–139, 2013.

[CRM24]	CIDOC CRM Special Interest Group. Volume A: Definition of the CIDOC Conceptual Reference Model. Version 7.1.3. International Council of Museums (ICOM), February 2024. https://cidoc-crm.org/Version/version-7.1.3
[DCAT-AP24]	SEMIC (Semantic Interoperability Community), European Commission. DCAT Application Profile for data portals in Europe (DCAT-AP), Version 3.0.0. Publications Office of the European Union, 2024. https://semiceu.github.io/DCAT-AP/releases/3.0.0/
[DCMI98]	Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1. DCMI Recommendation, 1998. https://www.dublincore.org/specifications/dublin-core/dces/
[DDK+25]	Dimitriou, Y., Damigos, M., Kalogeros, E. and Boueti, C. (2025). Describing Corfu Criminal Court Archives Using RiC-CM. Moderna arhivistika, 8(1), 93–112. https://doi.org/10.54356/MA/2025/GMNE4635.
[GHK+12]	M. Granitzer, M. Hristakeva, R. Knight, K. Jack, R. Kern. A Comparison of Layout-Based Bibliographic Metadata Extraction Techniques, In Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS 2012), Craiova, Romania, pages 1–12, 2012, ACM, https://doi.org/10.1145/2254129.2254154
[Gil08]	A. J. Gilliland. Setting the Stage, In Introduction to Metadata 3.0, J. Paul Getty Trust, 2008, Available at: http://www.getty.edu/research/conducting_research/standards/intrometadata/
[GM05]	M. A. Greene, D. Meissner. More Product, Less Process: Revamping Traditional Archival Processing, The American Archivist, 68(2), 2005.
[GMW+25]	J. Groppe, A. Marquet, A. Walz, S. Groppe. Automated Archival Descriptions with Federated Intelligence of LLMs, arXiv preprint, 2025
[GWF24]	G. Griffin, E. Wennerström, A. Foka. AI and Swedish Heritage Organisations: challenges and opportunities, AI & Society, 39: 2359–2372, 2024. https://doi.org/10.1007/s00146-023-01689-y
[HNB21]	R. Hathurusinghe, I. Nejadgholi, M. Bolic. A Privacy-Preserving Approach to Extraction of Personal Information through Automatic Annotation and Federated Learning. In Proceedings of the Association for Computational Linguistics (ACL) Workshop / arXiv preprint arXiv:2105.09198, 2021.
[HS97]	S. Hochreiter, J. Schmidhuber. Long Short-Term Memory. Neural Computation, 9(8): 1735–1780, 1997.
[ICA00]	International Council on Archives (Ed.). ISAD(G): General International Standard Archival Description. 2nd edition, Ottawa, International Council on Archives, 2000. Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19–22 September 1999

[ICA00]	**International Council on Archives (Ed.). ISAD(G): General International Standard Archival Description. 2nd edition, Ottawa, International Council on Archives, 2000. Adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19–22 September 1999**
[ICA04]	International Council on Archives, Committee on Descriptive Standards. ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, Second Edition. Paris: ICA, 2004. https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families
[ICA07]	International Council on Archives, Committee on Best Practices and Standards. ISDF: International Standard for Describing Functions, First Edition. Dresden: ICA, 2007. https://www.ica.org/en/isdf-international-standard-describing-functions
[ICA23]	International Council on Archives Expert Group on Archival Description (EGAD): Records in Contexts Conceptual Model (RiC-CM), version 1.0 (2023), https://www.ica.org/app/uploads/2023/12/RiC-CM-1.0.pdf
[ICA25]	International Council on Archives Expert Group on Archival Description (EGAD): Records in Contexts Ontology (RiC-O), version 1.1 (2025), https://www.ica.org/standards/RiC/ontology/1.1
[ISO16]	International Organization for Standardization. ISO 15489-1:2016 Information and documentation — Records management — Part 1: Concepts and principles, International Standard, 2016.
[ISO17]	International Organization for Standardization. ISO 5127:2017 Information and documentation — Foundation and vocabulary. International Standard, 2017.
[JC22]	L. Jaillant, A. Caputo. Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data, AI & Society 37: 823–835, 2022.
[JJG+25]	A. Jacyszyn, S. Jiang, G. A. Gesese, S. Hertling, T. Kerzenmacher, P. Nowack, S. Barthlott, E. Posthumus, H. Sack. AI4DiTraRe: Towards LLM-Based Information Extraction for Standardising Climate Research Repositories, First AAAI Bridge on Artificial Intelligence for Scholarly Communication (AISc), 2025. DOI: 10.5281/zenodo.14872358
[LCD+20]	D. Lin, J. Crabtree, I. Dillo, R. R. Downs, R. Edmunds, D. Giaretta, M. DeGiusti, H. L'Hours, W. Hugo, R. Jenkyns, V. Khodiyar, M. E. Martone, M. Mokrane, V. Navale, J. Petters, B. Sierman, D. V. Sokolova, M. Stockhause, J. Westbrook. The TRUST Principles for Digital Repositories. Scientific Data, 7:144, 2020.

[LoC10]	**Library of Congress, Network Development and MARC Standards Office. METS: Metadata Encoding and Transmission Standard, Primer and Reference Manual, Version 1.6. Digital Library Federation / Library of Congress, 2010. https://www.loc.gov/standards/mets/**
[LoC18]	Library of Congress, Network Development and MARC Standards Office. Metadata Object Description Schema (MODS), Version 3.7. Washington, D.C.: Library of Congress, 2018. https://www.loc.gov/standards/mods/
[LoC23]	Library of Congress. EAD: Encoded Archival Description — Official Site. Network Development and MARC Standards Office, Library of Congress, 2023. https://www.loc.gov/ead/
[LoC99]	Library of Congress, Network Development and MARC Standards Office. MARC 21 Format for Bibliographic Data. Washington, D.C.: Library of Congress, 1999– (συνεχώς ενημερωμένο). https://www.loc.gov/marc/bibliographic/
[LRT21]	M. Lorenzini, M. Rospoche, S. Tonelli. Automatically evaluating the quality of textual descriptions in cultural heritage records, International Journal on Digital Libraries 22: 217–231, 2021, Springer, https://doi.org/10.1007/s00799-021-00302-1
[MAG+23]	W. Mardiati, S. Aisyah, N. Grataridarga, R. Wulandari. The Potential Use of Artificial Intelligence Technology in the Process of Collecting Metadata in Photo Archive Description Activities. In D. V. Ferezagia et al. (Eds.), Proceedings of the International Conference on Vocational Education Applied Science and Technology (ICVEAST 2023). Advances in Social Science, Education and Humanities Research 783, Atlantis Press, 2023. DOI: 10.2991/978-2-38476-132-6_78.
[MML+24]	S. Münster, F. Maiwald, I. di Lenardo, J. Henriksson, A. Isaac, M. M. Graf, C. Beck, J. Oomen. Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe. Heritage, 7(2): 794–816, 2024, MDPI. https://doi.org/10.3390/heritage7020038
[NR23]	P. R. Nayaka, R. Ranjan. An Efficient Framework for Algorithmic Metadata Extraction over Scholarly Documents Using Deep Neural Networks, SN Computer Science 4:341, 2023, Springer Nature Singapore, https://doi.org/10.1007/s42979-023-01776-3
[OE24]	C. Ozogul, E. Ergen. Metadata Extraction of RFIs Using Natural Language Processing and Machine Learning Algorithms, In Proceedings of the European Conference on Computing in Construction (ECCIC 2024), Chania, Greece, July 14–17, 2024.

[OR24]	**G. Osti, E. R. Roke. Collaborating for Change? Assessing Metadata Inclusivity in Digital Collections with Large Language Models (LLMs), In 2024 IEEE International Conference on Big Data (Big Data), 2024.**
[OWJ+22]	L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, R. Lowe. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems (NeurIPS), 35: 27730–27744, 2022.
[Pav22]	G. Pavlidis. AI trends in digital humanities research, Trends in Computer Science and Information Technology 7(2): 26–34, 2022. https://doi.org/10.17352/tcsit.000048
[PGF23]	A. Pacheco, C. Guardado Da Silva, M. C. Vieira De Freitas. A metadata model for authenticity in digital archival descriptions. Archival Science 23: 629–673, 2023. https://doi.org/10.1007/s10502-023-09422-w
[PRM15]	PREMIS Editorial Committee. PREMIS Data Dictionary for Preservation Metadata, Version 3.0. Library of Congress, June 2015. https://www.loc.gov/standards/premis/v3/
[RB24]	P. Raval, H. Bhaidasna. A Review of Extracting Metadata from Scholarly Articles using Natural Language Processing (NLP), In Proceedings of the 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS 2024), pages 1355–1359, 2024, IEEE, https://doi.org/10.1109/ICACRS62842.2024.10841656
[RSD23]	V. Rawte, A. Sheth, A. Das. A Survey of Hallucination in “Large” Foundation Models, arXiv preprint arXiv:2309.05922, 2023. Available at: https://arxiv.org/abs/2309.05922
[SCK+24]	E. Shepherd, J. Cowan, J. J. Kaye, J. McLeod. Practices and pain points in personal records, Archival Science, 24: Article 3, 2024.
[SGB+24]	P. Svärd, E. Guerrero, T. Balogun, N. Saurombe, L. Jacobs, P. Henttonen. Local regulations for the use of artificial intelligence in the management of public records – a literature review, Records Management Journal, 2024, Emerald Publishing
[Sil21]	M. B. da Silva, F. C. C. Silva, R. C. F. Silva. Metadados para la preservación digital de datos abiertos: un estudio de identificación. Biblios: Journal of Librarianship and Information Science, 81: 1–22, 2021. https://doi.org/10.5195/biblios.2021.793
[SKD+25]	S. Stamou, E. Kalogeros, M. Damigos, M. Gergatsoulis. Leveraging LLMs to Build Text Narratives from CIDOC CRM. In: Metadata and Semantic Research, Proceedings of the 19th International Conference (MTSR 2025), Communications in Computer and Information Science, Springer, Cham, 2025.

[TSKK19]	**Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Introduction to Data Mining (2nd ed.). Pearson Education Limited.**
[Und08]	W. Underwood. Automatic Metadata Extraction for Archival Description and Access. In Proceedings of the Society of American Archivists Research Forum, San Francisco, CA, 2008.
[W3C24]	Data Catalog Vocabulary (DCAT) — Version 3. W3C Recommendation, 22 August 2024. https://www.w3.org/TR/vocab-dcat-3/
[WBM+23]	M. Wu, H. Brandhorst, M.-C. Marinescu, J. More Lopez, M. Hlava, J. Busch. Automated metadata annotation: What is and is not possible with machine learning, Data Intelligence 5(1): 122–138, 2023, MIT Press. https://doi.org/10.1162/dint_a_00162
[YFA+25]	W. Yang, R. Fu, M. B. Amin, B. Kang. The Impact of Modern AI in Metadata Management, Human-Centric Intelligent Systems 5: 323–350, 2025, Springer, https://doi.org/10.1007/s44230-025-00106-5
[ZHP+23]	Zhang, S., Hou, J., Peng, S., Li, Z., Hu, Q., and Wang, P. ArcGPT: A Large Language Model Tailored for Real-world Archival Applications, arXiv:2307.14852, 2023. https://doi.org/10.48550/arXiv.2307.14852
[ZSP+24]	K. Zoutsou, M. Sfakakis, L. Papachristopoulos, C. Papatheodorou, "Automated Topic Exploration in a Cultural Heritage Corpus", Metadata and Semantic Research, Proceedings of the 17th International Conference (MTSR 2023), Milan, Italy, October 25 - 27, 2023, Communications in Computer and Information Science (CCIS), No. 2048: Springer, Cham., 2024, pp. 229 - 240.
[ZZL15]	X. Zhang, J. Zhao, Y. LeCun. Character-level Convolutional Networks for Text Classification. In Advances in Neural Information Processing Systems (NeurIPS), 28: 649–657, 2015.