



ΨΗΦΙΑΚΟ ΤΟΠΙΟ

Οδηγός Προτύπων και Σχημάτων Μεταδεδομένων

Παραδοτέο Π3 / Ενότητα Εργασίας 2.3

Εμβληματική Δράση: «Το αναδυόμενο τοπίο ψηφιακής εργασίας στις Ανθρωπιστικές Επιστήμες στο πλαίσιο των ευρωπαϊκών υποδομών DARIAH και CLARIN» (Αριθμός Έργου:7982)

Τελική έκδοση



Έγγραφο:	7982_Π3_Οδηγός προτύπων
Τίτλος παραδοτέου:	Π3: Οδηγός Προτύπων και Σχημάτων Μεταδεδομένων / D3: Standards and Metadata Guide
Ενότητα εργασίας:	ΕΕ2: Ψηφιακές ροές εργασιών στις Ανθρωπιστικές και Κοινωνικές Επιστήμες / WP2: Digital work processes in the Humanities and Social Sciences
Υπεύθυνος φορέας:	Ακαδημία Αθηνών
Συμμετέχων φορέας:	Ε.Κ. Αθηνά
Συντάκτες:	Ε. Γουλή (ΑΑ), Κ. Καλαφατά (ΑΑ), Η. Σουγιουτζόγλου (ΑΑ), Γ. Τζεδόπουλος (ΑΑ)
Με τη συμβολή των:	Λ. Βουκούτη (ΕΚ ΑΘΗΝΑ), Κ. Γκίρτζου (ΕΚ ΑΘΗΝΑ), Δ. Γκούμα (ΕΚ ΑΘΗΝΑ), Μ. Ηλβανίδου (ΕΚ ΑΘΗΝΑ), Π. Κωνσταντόπουλου (ΕΚ ΑΘΗΝΑ), Σοφία Μπουγά (ΑΑ), Β. Ντρίτσου (ΕΚ ΑΘΗΝΑ) Μ. Σπηλιωτοπούλου (ΑΑ)
Έκδοση εγγράφου:	Τελική
Ημερομηνία έκδοσης:	31 Μαρτίου 2024

ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα.....	2
ΑΚΡΩΝΥΜΙΑ	4
ΠΕΡΙΛΗΨΗ	6
1. Εισαγωγή.....	7
2. ΜΕΘΟΔΟΛΟΓΙΑ	8
3. ΠΡΟΤΥΠΑ ΚΑΙ ΣΧΗΜΑΤΑ ΜΕΤΑΔΕΔΟΜΕΝΩΝ	9
3.1 Ο ΜΕΤΑ-ΘΗΣΑΥΡΟΣ BACKBONE (BBT)	9
3.2 Εννοιολογικό Μοντέλο Αναφοράς CIDOC	11
3.3 Component MetaData Infrastructure (CMDI)	15
3.4 CRMarchaeo Excavation Model.....	16
3.5 CRMgeo	17
3.6 Data Documentation Initiative (DDI) και CESSDA Metadata Management (CMM)	18
3.7 Dublin Core.....	21
3.8 Encoded Archival Description (EAD).....	23
3.9 Europeana Data Model (EDM)	25
3.10 General International Standard Archival Description ISAD(G)	27
3.11 ISLE Metadata Initiative (IMDI)	27
3.12 MACHINE-Readable Cataloging (MARC)	28
3.13 META-SHARE Metadata Schema	29
3.14 Midas Heritage	30
3.15 Metadata Object Description Schema (MODS)	31
3.16 Open Archives Initiative / Object Reuse and Exchange (OAI-ORE).....	32
3.17 Open Language Archives Community [OLAC] standards.....	33
3.18 Qualitative Data Exchange Schema (QuDEX).....	33
3.19 Statistical Data and Metadata eXchange (SDMX).....	34
3.20 Text Encoding Initiative (TEI).....	35
4. ΣΥΝΑΦΕΙΑ ΜΕ ΣΤΑΔΙΑ ΚΑΙ ΤΥΠΟΥΣ ΔΡΑΣΤΗΡΙΟΤΗΤΩΝ	39
5. ΕΡΕΥΝΗΤΙΚΑ ΔΕΔΟΜΕΝΑ, ΜΕΤΑΔΕΔΟΜΕΝΑ ΚΑΙ ΑΡΧΕΣ FAIR	42
6. ΣΥΜΠΕΡΑΣΜΑΤΑ	45
ΒΙΒΛΙΟΓΡΑΦΙΑ	47
ΔΙΚΤΥΟΓΡΑΦΙΑ.....	48

ΠΑΡΑΡΤΗΜΑ:Συνάφεια σχημάτων μεταδεδομένων	52
---	----

ΑΚΡΩΝΥΜΙΑ

ΑΚΕ: Ανθρωπιστικές και Κοινωνικές Επιστήμες

API: Application Programming Interface

DC: Dublin Core

ΔΙΠΑΠ: Διεθνές Πρότυπο Αρχειακής Περιγραφής

CESSDA: Consortium of European Social Data Archives

CIDOC CRM: International Committee for Documentation / Conceptual Reference Model

CLARIN: Common Language Resources and Technology Infrastructure

CMM: CESSDA Metadata Management

CMDI: Component MetaData Infrastructure

CODATA: Committee on Data

DCC: Digital Curation Centre

DC: Dublin Core

EAD: Encoded Archival Description

EDM: Europeana Data Model

EOSC: European Open Science Cloud

ESE: Europeana Semantic Elements

FAIR: Findable, Accessible, Interoperable, Reusable [data]

FISH: Forum on Information Standards in Heritage

FOAF: 'Friend Of A Friend' ontology

FRBRoo: Functional Requirements for Bibliographic Records - Object-Oriented

GoFAIR: Global Open FAIR

IMDI: ISLE Metadata Initiative

ISAD(G): General International Standard Archival Description

ITE: Ίδρυμα Τεχνολογίας και Έρευνας

MARC: Machine Readable Cataloguing

MODS: Metadata Object Description Schema

LIDO: Lightweight Information Describing Objects

OAI-ORE: Open Archives Initiative / Object Reuse and Exchange

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

OLAC: Open Language Archives Community

OODBMS: object-oriented database management system

QuDEx: Qualitative Data Exchange Schema

RDA: Research Data Alliance

RDBMS: Relational Database Management System

RDF: Resource Description Framework

SDMX: Statistical Data and Metadata eXchange

SDTL: Structured Data Transformation Language

SKOS: Simple Knowledge Organisation System

SSHOC: Social Sciences & Humanities Open Cloud

TEI: Text Encoding Initiative

XKOS: Extended Knowledge Organization System

XML: Extensible Markup Language

ΨΑΕ: Ψηφιακές Ανθρωπιστικές Επιστήμες

ΠΕΡΙΛΗΨΗ

Τα μεταδεδομένα αποτελούν θεμελιώδες στοιχείο του ψηφιακού οικοσυστήματος έρευνας, συμβάλλοντας στη σημασιολογική και τεχνική διαλειτουργικότητα των παραγώγων της. Σκοπός της παρούσας μελέτης είναι να συγκεντρώσει πρακτικές πληροφορίες σχετικά με την εφαρμογή σχημάτων και προτύπων μεταδεδομένων, και να αναδείξει τη συμβολή τους στη δημιουργία ενός ενιαίου ερευνητικού οικοσυστήματος.

Προς την κατεύθυνση αυτή, τεκμηριώνονται πρότυπα και σχήματα μεταδεδομένων που χρησιμοποιούνται στο ευρύ πεδίο των ανθρωπιστικών και κοινωνικών επιστημών, λαμβάνοντας υπόψη τις τρέχουσες ψηφιακές πρακτικές της ερευνητικής κοινότητας που προσδιορίστηκαν στο ΠΕ1 του έργου, καθώς και την κωδικοποίηση των συναφών δραστηριοτήτων που πραγματοποιήθηκε στο ΠΕ2.

Η μελέτη παρουσιάζει 20 σχήματα μεταδεδομένων, τα οποία εξετάζονται ως προς τη συνάφειά τους με επιμέρους επιστημονικούς κλάδους, το συσχετισμό τους με συγκεκριμένους τύπους δεδομένων και τη σύνδεσή τους με επιμέρους δραστηριότητες του οντολογικού πλαισίου που εισηγείται το παραδοτέο Π2 «Μοντέλα διεργασιών και ροών εργασίας». Συμπληρωματικά, αναδεικνύει τη συμβολή των προτύπων στη διασφάλιση της ανοικτότητας των ψηφιακών πόρων, καθώς και στην εφαρμογή των αρχών “FAIR”.

1. ΕΙΣΑΓΩΓΗ

Η παρούσα μελέτη αποτελεί το τρίτο παραδοτέο (Π3) του έργου «Το αναδυόμενο τοπίο ψηφιακής εργασίας στις Ανθρωπιστικές Επιστήμες στο πλαίσιο των ευρωπαϊκών υποδομών DARIAH και CLARIN». Στοχεύει να εντοπίσει, να αναλύσει και να κατηγοριοποιήσει πρότυπα και σχήματα μεταδεδομένων που υποστηρίζουν την ψηφιακή εργασία των ερευνητών των ανθρωπιστικών και κοινωνικών επιστημών (ΑΚΕ). Για το σκοπό αυτό λαμβάνει υπόψη της τις ανάγκες και προσδοκίες των ερευνητών, όπως αυτές αποτυπώθηκαν στο πλαίσιο του πρώτου πακέτου εργασίας (ΠΕ1) του έργου μέσω της συμπλήρωσης ερωτηματολογίου και των συναντήσεων των ομάδων εστίασης που διοργανώθηκαν. Η συστηματική ανάλυση των αποτελεσμάτων που προέκυψαν παρουσιάζεται στο πρώτο παραδοτέο του έργου (Π1) με τίτλο “Χρήση ψηφιακών μεθόδων στις Ανθρωπιστικές και Κοινωνικές Επιστήμες”.¹

Η μελέτη λαμβάνει επίσης υπόψη της το δεύτερο παραδοτέο του έργου (Π2) με τίτλο “Μοντέλα διεργασιών και ροών εργασίας”.² Άλλωστε η εκπόνηση των δύο αυτών παραδοτέων (Π2 και Π3) έγινε με συνεχή επικοινωνία μεταξύ των δύο ομάδων εργασίας, τα μέλη των οποίων παρατίθενται στα στοιχεία ταυτότητας του εγγράφου.

Σκοπός της μελέτης είναι να παρουσιάσει και να προτείνει τη χρήση καθιερωμένων σχημάτων μεταδεδομένων για τη διευκόλυνση και τον εμπλουτισμό της έρευνας στις ανθρωπιστικές και κοινωνικές επιστήμες. Γι’ αυτόν τον λόγο, η ανάλυση που ακολουθεί έχει περισσότερο πρακτικό παρά θεωρητικό χαρακτήρα.

Κατόπιν μιας σύντομης αναφοράς στη μεθοδολογία και τη συνάφεια της παρούσας μελέτης με την οντολογική αναπαράσταση των ερευνητικών πρακτικών του Π2, στο κύριο κεφάλαιο της μελέτης παρουσιάζονται και αναλύονται πρότυπα και σχήματα μεταδεδομένων για την έρευνα που εκπονείται στο πλαίσιο των ανθρωπιστικών και κοινωνικών επιστημών. Συμπληρωματικά, προσδιορίζεται η συνάφειά τους με τις επιμέρους περιοχές αυτού του ευρέως φάσματος επιστημών. Επιπλέον, τα σχήματα μεταδεδομένων εξετάζονται ως προς τους διαφορετικούς τύπους δεδομένων (οι οποίοι μπορεί να τέμνουν οριζόντια τις εκάστοτε υποκατηγορίες των ανθρωπιστικών και κοινωνικών επιστημών), αλλά και ως προς τις κοινότητες των χρηστών και τις εφαρμογές που τα αξιοποιούν. Στόχος είναι να δειχθεί η εφαρμοσιμότητα και χρησιμότητα των σχημάτων και προτύπων μεταδεδομένων μέσα από μια συνδυαστική και συγκριτική προσέγγιση.

Στο επόμενο κεφάλαιο εξετάζεται η συνάφεια των σχημάτων μεταδεδομένων με συγκεκριμένα στάδια και ροές εργασιών που είναι κοινές για τις επιστημονικές περιοχές που συναπαρτίζουν τις ανθρωπιστικές και κοινωνικές επιστήμες. Η ανάλυση βρίσκεται σε συμφωνία με το παραδοτέο Π2 του έργου, όπου αναπαρίστανται οντολογικά οι ψηφιακές εργασίες που επιτελούνται από τους ερευνητές.

Ακολουθεί ένα κεφάλαιο που αφορά στην ευθυγράμμιση με τις “αρχές FAIR” (FAIR Principles), ένα σύνολο κατευθυντήριων οδηγιών για τη διαχείριση και επιμέλεια των μεταδεδομένων με στόχο την ευρεσιμότητα, την προσβασιμότητα, τη διαλειτουργικότητα και την επανάχρηση των δεδομένων

¹ Ψηφιακό Τοπίο: Χρήση ψηφιακών μεθόδων στις ανθρωπιστικές και κοινωνικές επιστήμες / Ανάλυση τάσεων, Ιούλιος 2023 <https://zenodo.org/records/10797453> (20/3/2024).

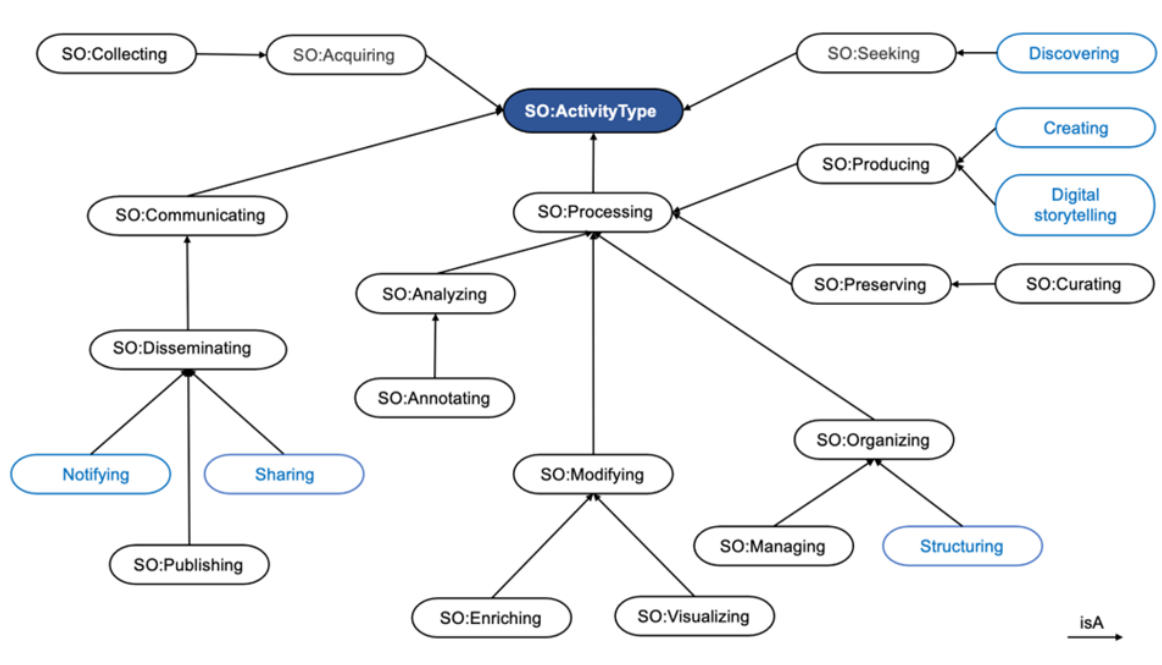
² Υπό δημοσίευση

της έρευνας. Η μελέτη κλείνει συνοψίζοντας σε ένα τελευταίο κεφάλαιο τα αποτελέσματα και συμπεράσματα της ανάλυσης.

2. ΜΕΘΟΔΟΛΟΓΙΑ

Η υιοθέτηση/ανάπτυξη προτύπων και σχημάτων μεταδεδομένων βρίσκεται στον πυρήνα των ψηφιακών πρακτικών εργασίας των επιστημονικών κοινοτήτων. Η πληθώρα των προτύπων και σχημάτων μεταδεδομένων καταδεικνύει, αφενός, την αναγκαιότητα του σημασιολογικού εμπλουτισμού των πηγών της έρευνας και, αφετέρου, την πολυπλοκότητα του εγχειρήματος επιλογής ενός κατάλληλου σχήματος/προτύπου, συμβατού με το είδος των τεκμηρίων ή συλλογών και τις κατά περίπτωση πληροφοριακές ανάγκες.

Η παρούσα μελέτη αξιοποιεί το οντολογικό πλαίσιο που αναπτύσσεται στο Π2, προκειμένου να αναδείξει τη συνάφεια των σχημάτων και προτύπων μεταδεδομένων με τύπους δραστηριοτήτων που ταυτοποιήθηκαν στο Π1 και κωδικοποιήθηκαν στο Π2. Η σημασιολογική αναπαράσταση των δραστηριοτήτων βασίζεται στην οντολογία αναφοράς Scholarly Ontology, η οποία επιτρέπει τον συσχετισμό τεσσάρων δομικών στοιχείων της επιστημονικής εργασίας (*Δραστηριότητα, Διαδικασία, Πόρος και Δράστης*). Κατόπιν επεξεργασίας των στοιχείων που προέκυψαν από τις ομάδες εστίασης και τη διαδικτυακή έρευνα, η οντολογία Scholarly Ontology αναπτύχθηκε περαιτέρω, με την προσθήκη ιεραρχικά συσχετιζόμενων εξειδικεύσεων των υφιστάμενων τύπων δραστηριοτήτων, ως εξής:



Εικ. 1: Αναπαράσταση δραστηριοτήτων βάσει των Τύπων Δραστηριοτήτων της SO και εξειδικεύσεων αυτών.
Πηγή: “Π2-Μοντέλα διεργασιών και ροών εργασίας”

Αφετηρία της διαδικασίας επιλογής των προτεινόμενων προτύπων και σχημάτων μεταδεδομένων αποτέλεσε το υλικό που συγκεντρώθηκε από τη διαδικτυακή έρευνα και τις ομάδες εστίασης που συγκροτήθηκαν στα αρχικά στάδια του έργου. Η ομάδα εργασίας ανέτρεξε στο εν λόγω υλικό αναζητώντας συγκεκριμένες αναφορές σε σχήματα και πρότυπα που οι συμμετέχοντες και συμμετέχουσες δήλωσαν ότι χρησιμοποιούν, καθώς και τις καταγεγραμμένες ερευνητικές δραστηριότητες με τις οποίες αυτά συνδέθηκαν. Συμπληρωματικά, πραγματοποιήθηκε έρευνα με

σκοπό να εντοπιστούν και να καταγραφούν σχήματα και πρότυπα που δεν συμπεριλαμβάνονται στα δεδομένα που προέκυψαν από το ΠΕ1. Μέσα από τη συνδυαστική αυτή προσέγγιση προέκυψε ένα μητρώο το οποίο περιλαμβάνει 20 επιλογές σχημάτων που αφορούν, όπως ήδη επισημάνθηκε, τόσο εξειδικευμένα πρότυπα, όσο και πρότυπα που έχουν ευρύτερο πεδίο εφαρμογής.

Προκειμένου να τεκμηριωθούν οι εν λόγω επιλογές και να στοιχειοθετηθεί, συνδυαστικά, ένα μοντέλο περιγραφής που περιλαμβάνει διαφορετικά μεν, αλλά συμπληρωματικά δομικά στοιχεία, ορίστηκε μια κοινή δομή ώστε να παρέχεται κατανοητά και με οργανωμένο τρόπο η σχετική πληροφορία. Τέλος, όπου κρίθηκε σκόπιμο, παρατέθηκαν συγκεκριμένοι τύποι δραστηριοτήτων, στους οποίους έχει εφαρμογή το κάθε πρότυπο ή σχήμα μεταδεδομένων.

Εκ των πραγμάτων οι περιγραφές των σχημάτων δεν περιλαμβάνουν τον ίδιο όγκο πληροφορίας και βαθμό ανάλυσης, καθώς σε κάποιες περιπτώσεις το πεδίο εφαρμογής και το εύρος των δραστηριοτήτων τις οποίες υποστηρίζουν είναι περιορισμένα. Ωστόσο έγινε προσπάθεια να υπάρξει μια επαρκής τεκμηρίωση -με πρακτικό κυρίως προσανατολισμό- ώστε να υπάρχει η δυνατότητα αξιοποίησης των περιεχομένων της μελέτης από την ερευνητική κοινότητα.

Το κοινό μοντέλο τεκμηρίωσης περιγράφει και τη συνάφεια του κάθε σχήματος με τις επιμέρους επιστημονικές περιοχές. Έμφαση δόθηκε επίσης στην ποικιλομορφία που παρουσιάζουν οι μορφότυποι των δεδομένων και πηγών που χρησιμοποιούνται στα πεδία αυτά, που καλύπτουν από κειμενικές πηγές και φυσικά αντικείμενα πολιτιστικής κληρονομιάς, έως πρωτογενώς ψηφιακές πηγές -όπως πολυμέσα και βάσεις δεδομένων. Τα σχήματα συνδέονται με διαφορετικούς τύπους δεδομένων (που μπορεί να χρησιμοποιούνται σε ένα ή περισσότερα επιστημονικά πεδία) και, αντίστοιχα, με τις κοινότητες ερευνητών που δυνητικά τα αξιοποιούν στην έρευνά τους. Συμπληρωματικά, παρατίθενται, όπως ήδη αναφέρθηκε, παραδείγματα επιτυχούς εφαρμογής, που αναδεικνύουν όχι μόνο την συμβολή των σχημάτων και προτύπων στον εμπλουτισμό της έρευνας, αλλά και πρακτικές ορθής αξιοποίησής τους.

3. ΠΡΟΤΥΠΑ ΚΑΙ ΣΧΗΜΑΤΑ ΜΕΤΑΔΕΔΟΜΕΝΩΝ

Τα σχήματα και πρότυπα μεταδεδομένων που τεκμηριώθηκαν παρατίθενται ακολούθως, με αλφαβητική σειρά:

3.1 Ο ΜΕΤΑ-ΘΗΣΑΥΡΟΣ BACKBONE (BBT)

Οι θησαυροί αποτελούν μορφή ελεγχόμενων λεξιλογίων που συγκροτούν μια δομημένη συλλογή των εννοιών που χρησιμοποιούν οι επιστημονικές κοινότητες για την τεκμηρίωση και την ευρετηρίαση πόρων με καθιερωμένους (κοινά αποδεκτούς και με ενιαίο τρόπο συντεταγμένους) όρους.

Η αξιοποίηση λεξιλογίων, θησαυρών και καταλόγων καθιερωμένων όρων ως ελεγχόμενων λεξιλογίων για συγκεκριμένα πεδία μεταδεδομένων αποτελεί κοινή επιστημονική πρακτική: χρησιμοποιούνται για την οργάνωση και διαχείριση των μεταδεδομένων ενός συστήματος και στοχεύουν στην αντιμετώπιση της σημασιολογικής ανομοιογένειας των δεδομένων και τη συγκρότηση κατανοητής και ενιαίας διαχείρισης της πληροφορίας. Συμβάλλουν, δε, στην υποστήριξη της τεκμηρίωσης ή της καταλογογράφησης, στον εμπλουτισμό των μεταδεδομένων των

τεκμηρίων, τη βελτίωση της σημασιολογικής διαλειτουργικότητας και της ποιότητας τεκμηρίωσης πρωτογενών συλλογών. Ως εκ τούτου, οι καθιερωμένοι όροι και οι θησαυροί αποτελούν σημαντικό τμήμα της ροής εργασιών για τη συγκρότηση ενός συστήματος ενιαίας διαχείρισης της πληροφορίας σε οργανισμούς πληροφόρησης (βιβλιοθήκες, αρχεία, μουσεία).

Για την υποστήριξη της σημασιολογικής διασύνδεσης επί μέρους θησαυρών αναπτύχθηκε ένας πολύγλωσσος και διεπιστημονικός μεταθησαυρός γενικών καθολικών όρων και ιεραρχιών (Back Bone Thesaurus, BBT),³ ο οποίος είναι το αποτέλεσμα της εργασίας των συμμετεχόντων στην ομάδα εργασίας Thesaurus Maintenance Working Group του DARIAH-EU, μιας διεθνούς ομάδας επιστημόνων της οποίας προΐστανται η Ακαδημία Αθηνών και το Ίδρυμα Τεχνολογίας και Έρευνας (ITE).⁴

Ο μεταθησαυρός απαντά στις ανάγκες των ερευνητών για διεπιστημονικότητα και διαλειτουργικότητα, ενώ σε επίπεδο οργανισμών θεραπεύει τον κατακερματισμό της πληροφορίας και διασφαλίζει οικονομία στη διαδικασία οργάνωσης των όρων σε ένα κοινό πλαίσιο και ένα καταναμεμημένο, συνεργατικό περιβάλλον.

Ο BBT παρέχει ένα σχήμα υπερκείμενων θεμάτων (facets) και ιεραρχιών για την ανάπτυξη θησαυρών, υπό το οποίο μπορούν να υπαχθούν όροι τοπικών θησαυρών που αναπτύσσονται από άλλους φορείς ή ερευνητές με διαλειτουργικό και βιώσιμο τρόπο, όπως και δομημένα λεξιλόγια και θησαυροί που έχουν αναπτυχθεί σε διαφορετικά γνωστικά πεδία των ανθρωπιστικών και κοινωνικών επιστημών. Διευκολύνει κατ' αυτόν τον τρόπο την κατηγοριοποίηση της επιστημονικής ορολογίας, χωρίς να υποχρεώνει τους ερευνητές να τροποποιήσουν τις μεθοδολογικές παραδοχές του γνωστικού τους αντικειμένου. Βασικά πλεονεκτήματα του συγκεκριμένου μεθοδολογικού σχήματος είναι η δυνατότητα επέκτασης σε νέες περιοχές γνώσης, η ενίσχυση της διεπιστημονικής προσέγγισης στην εξόρυξη πηγών και δεδομένων, η συμβατότητα και η διαλειτουργικότητα των υπαρχόντων θησαυρών και οντολογιών, η βελτιστοποίηση της συντήρησης και της διαχείρισης της ορολογίας.

Βασικές μεθοδολογικές αρχές που διέπουν την ανάπτυξή του είναι:

- Η ταξινόμηση σύμφωνα με facets
- Η κατηγορηματική σημασιολογία (σχέση IsA για τη δόμηση μιας ιεραρχίας)
- Η μέθοδος από κάτω προς τα πάνω (bottom-up)
- Ο μικρός αριθμός εννοιών (καθολικές έννοιες, ιεραρχίες)
- Ο ορισμός των υπερκείμενων εννοιών ανεξαρτήτως του περιβάλλοντος χρήσης τους
- Η αρχή των intensional properties (των αναγκαίων και επαρκών συνθηκών για την υπαγωγή μιας έννοιας σε μία κατηγορία)⁵

³ <https://backbonethesaurus.eu/> (20/3/2024).

⁴ Goulis, H. "The BBT meta-thesaurus model: building interoperable thesauri for humanities researchers". In Almeida, B., Costa, R., & Medeiros, F. (2021), *Controlled vocabularies and knowledge organisation for the digital humanities: Proceedings*, pp. 4-7. <https://doi.org/10.34619/pgtp-upne>

⁵ Τεκμηρίωση του Θησαυρού: https://www.backbonethesaurus.eu/sites/default/files/DARIAH_BBT%20v%202.pdf (20/3/2024).

Στην τρέχουσα έκδοσή του περιλαμβάνει 10 θέματα και 30 ιεραρχίες. Τα θέματα, οι όροι κορυφής και οι ορισμοί τους είναι διαθέσιμα σε 5 γλώσσες (ελληνικά, αγγλικά, γερμανικά, γαλλικά, πορτογαλικά). Η διαδικτυακή πλοήγηση στο περιβάλλον της ομοσπονδίας τοπικών θησαυρών γίνεται στην πλατφόρμα που έχει αναπτύξει το κέντρο Ψηφιακών Ανθρωπιστικών Σπουδών και Πολιτιστικής Κληρονομιάς της Αυστριακής Ακαδημίας Επιστημών (ACDH-CH) για την υποστήριξη υπηρεσιών της ευρωπαϊκής υποδομής DARIAH EU.

The screenshot shows the Backbone Thesaurus interface. On the left, there is a tree view of the hierarchy under 'activities'. The main panel displays the details for the term '000013 functions'. It includes a 'PREFERRED TERM' section, a 'BROADER CONCEPT' section pointing to '000001 activities', a 'NOTE' section with a detailed definition of functions, a 'SCOPE NOTE' section, a 'BELONGS TO GROUP' section pointing to 'Activities', a 'CREATOR' section pointing to 'BBT maintenance WG', a 'NOTATION' section pointing to '000013', and an 'IN OTHER LANGUAGES' section listing the term in French ('fonctions'), German ('Funktionen'), Greek ('Λειτουργίες'), and Portuguese ('funções').

Εικ. 2. Το περιβάλλον εργασίας του μεταθησαυρού BBT. Πηγή: <https://www.backbonethesaurus.eu/>

3.2 Εννοιολογικό Μοντέλο Αναφοράς CIDOC

Το CIDOC CRM⁶ είναι μία τυπική οντολογία (εκφρασμένη στη γλώσσα παράστασης γνώσης SIS-TELOS) κατάλληλη για την τεκμηρίωση και ενοποίηση πολιτισμικών δεδομένων. Τα στιγμιότυπα μπορούν να εκφραστούν σε πολλές μορφές (π.χ. RDBMS, ooDBMS, XML, RDF/S).

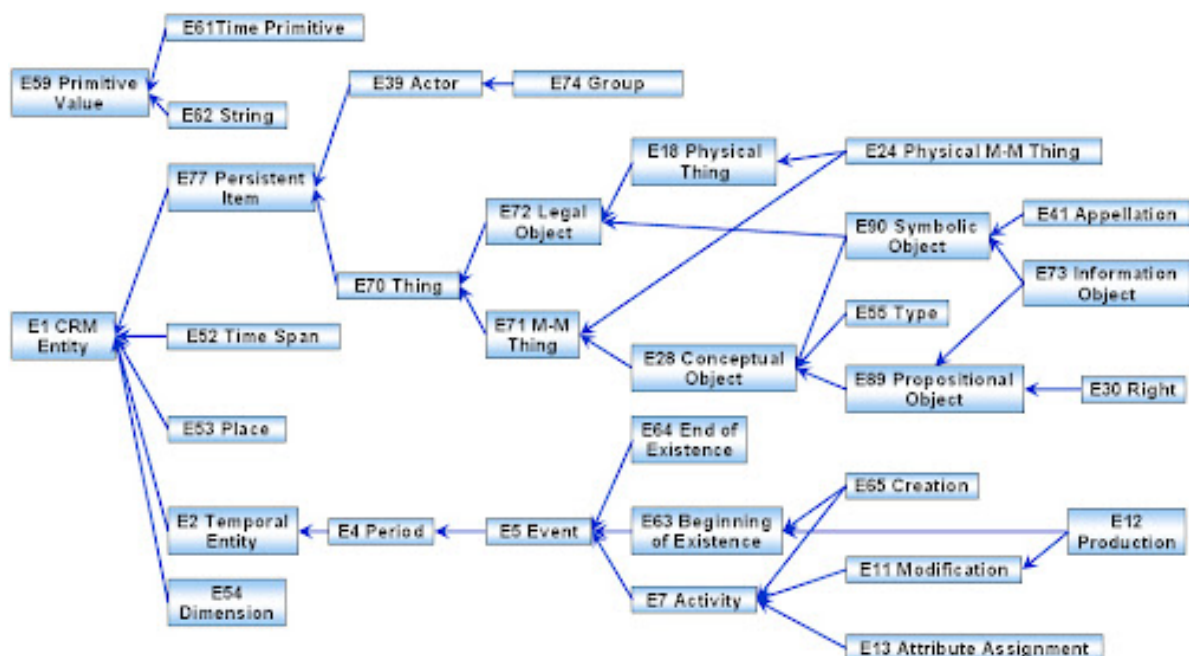
Το CIDOC CRM αποτελεί οντολογικό μοντέλο περιγραφής μεταδεδομένων και ανάπτυξης σημασιολογικών οντοτήτων, παράδειγμα δομικής μοντελοποίησης από την εννοιολογική σκοπιά με ευρύ πεδίο εφαρμογής. Σκοπό έχει να διευκολύνει τη συγχώνευση, διαμεσολάβηση, και ανταλλαγή ετερογενών πληροφοριών πολιτισμικής κληρονομιάς. Επιτρέπει την ερμηνεία των ψηφιακών και φυσικών αντικειμένων που περιγράφουν τα μουσεία και τα αρχεία, τυποποιώντας τις έννοιες που βοηθούν την ολοκλήρωση των δεδομένων και την εξεύρεση πόρων. Είναι εστιασμένο στη σημασιολογία των δομών δεδομένων, παρέχοντας μια επεκτάσιμη οντολογία για έννοιες και πληροφορίες στην τεκμηρίωση της πολιτιστικής κληρονομιάς και των μουσείων.

Το μοντέλο αναπτύχθηκε σε συνεργασία με το Διεθνές Συμβούλιο Μουσείων (International Council of Museums). Η έκδοση 3.4.9 της οντολογίας αναφοράς του CIDOC CRM έγινε διεθνές πρότυπο με κωδικό ISO 21127 το 2006. Έκτοτε δημοσιοποιήθηκαν δύο νέες εκδόσεις του προτύπου, η πρώτη

⁶CIDOC Conceptual Reference Model <https://www.cidoc-crm.org/> (20/3/2024).

τον Δεκέμβριο του 2014 και η τελευταία τον Φεβρουάριο του 2024 (version 7.1.3). Στην τελευταία του έκδοση, το CIDOC CRM περιλαμβάνει 81 κλάσεις και 160 ιδιότητες.⁷

Το CIDOC - CRM λειτουργεί ως κοινό σχήμα περιγραφής μεταδεδομένων, ενώ χάρη στην πληθώρα δυνατοτήτων και επεκτάσεων που διαθέτει στον πηγαίο του κώδικα (συγχώνευση, ενσωμάτωση, επεκτάσεις), μπορεί να δρα συνδυαστικά με άλλα συμβατά οντολογικά μοντέλα. Απαντά στις ανάγκες που επισημάνθηκαν κατά τη συνάντηση της ομάδας εστίασης των ιστορικών και αρχαιολόγων ("Ανάγκες συμβατότητας και διαλειτουργικότητας των δεδομένων και χρήση σημασιολογικών οντολογιών όπως εκείνη του CIDOC") και αποτυπώθηκαν στο Π1 του έργου.⁸



Εικ. 3: Ιεραρχία “τάξεων” του CIDOC CRM. Πηγή: CIDOC CRM Home page (<https://www.cidoc-crm.org/>)

Συνοπτικά, τα βασικά χαρακτηριστικά του μοντέλου είναι τα ακόλουθα:

Βασικές λειτουργικότητες

Το μοντέλο αναπτύχθηκε σύμφωνα με τις ακόλουθες αρχές:

- Πολλαπλή κληρονομικότητα (multiple inheritance) για την αποφυγή επανωρισμού ιδιοτήτων
- Πολλαπλή ταξινόμηση (multiple classification) για να μπορεί να συνδυάζει (παριστάνει) συνδυασμούς κλάσεων που δεν είναι πάντα έγκυροι ώστε να έχει οριστεί ξεχωριστή κλάση για αυτούς

⁷ Πληροφορίες για την πρόσφατη έκδοση του μοντέλου παρέχονται στους συνδέσμους <https://cidoc-crm.org/Version/version-7.1.3> και https://cidoc-crm.org/html/cidoc_crm_v7.1.3.html (24/3/2024).

⁸ Ψηφιακό Τοπίο: Χρήση ψηφιακών μεθόδων στις ανθρωπιστικές και κοινωνικές επιστήμες / Ανάλυση τάσεων, Ιούλιος 2023, <https://zenodo.org/records/10797453>, σελ. 75.

- Πολλαπλή κληρονομικότητα στις συσχετίσεις για να μπορεί να παραστήσει διαφορετικούς βαθμούς αφαίρεσης των ιδιοτήτων.

Οι λειτουργικότητες που υποστηρίζει το μοντέλο επιτρέπουν την εφαρμογή του ως

- Εννοιολογικό οδηγό για τη δημιουργία σχημάτων και μορφοτύπων
- Γλώσσα για την ανάλυση των περιεχομένων υπαρχουσών πηγών με στόχο την ενοποίηση τους (δημιουργώντας μια νέα πηγή, ή έναν μεσολαβητή)
- Μορφότυπος μεταφοράς δεδομένων (για ενοποίηση, migration)

Επίσης, το CIDOC - CRM επιτρέπει τον εντοπισμό στοιχείων (δεδομένων) που έχουν την ίδια σημασία.

Ο τρόπος χρήσης της κάθε έννοιας ή συσχέτισης περιγράφεται με κείμενο (scope note). Κάθε κλάση/συσχέτιση έχει ως ταυτότητα ένα γράμμα (E για Entities, P για Properties) ακολουθούμενο από έναν αριθμό. Αυτό αποτελεί την ταυτότητα της κλάσης ή της συσχέτισης. Ακολουθεί ένα σύντομο ενδεικτικό όνομα (π.χ. E4_Period, E5_Period, PX).

Ταξινόμηση συσχετίσεων

- Αναγνώριση (identification) στοιχείων του πραγματικού κόσμου μέσω ονομάτων (του πραγματικού κόσμου)
- Ταξινόμηση (classification) στοιχείων του πραγματικού κόσμου
- Συγκρότηση/Αποσύνθεση (part-decomposition) και άλλων δομικών ιδιοτήτων των Φυσικών και Εννοιολογικών Αντικειμένων, Περιόδων, Δραστών, Τοποθεσιών και Χρόνων
- Συμμετοχή (participation) σε χρονικές (έγχρονες) οντότητες
- Θέση (location) περιόδων στο χωροχρόνο και θέσης αντικειμένων στον χώρο
- Επίδραση (influence) αντικειμένων σε δραστηριότητες και προϊόντα και το αντίστροφο
- Αναφορά (reference) αντικειμένων πληροφορίας σε αντικείμενα του πραγματικού κόσμου

Επεκτάσεις του CIDOC CRM

Το βασικό πρότυπο συμπληρώνεται από μια σειρά επεκτάσεων οι οποίες έχουν σχεδιαστεί για να υποστηρίζουν διαφορετικούς τύπους εξειδικευμένων ερευνητικών αναγκών (π.χ. βιβλιογραφική τεκμηρίωση ή γεωπληροφορική). Οι επεκτάσεις που παρατίθενται ακολούθως έχουν αναπτυχθεί από την Ομάδα Ειδικού Ενδιαφέροντος του CIDOC CRM σε συνεργασία με επιμέρους ερευνητικές κοινότητες.

FRBRoo: στοχεύει στη σύλληψη και αναπαράσταση της υποκείμενης σημασιολογίας των βιβλιογραφικών πληροφοριών, στη διευκόλυνση της ενσωμάτωσης και τη διαλειτουργικότητά τους.⁹

PRESSoo: στοχεύει στη σύλληψη και αναπαράσταση της υποκείμενης σημασιολογίας των βιβλιογραφικών πληροφοριών σχετικά με πόρους όπως τα περιοδικά και οι εφημερίδες.¹⁰

CRMinf: χρησιμοποιείται ως σχήμα για την ενσωμάτωση μεταδεδομένων σχετικά με την επιχειρηματολογία και την εξαγωγή συμπερασμάτων σε αναλυτικές και εμπειρικές επιστήμες όπως, η γεωλογία, η γεωγραφία, η αρχαιολογία, η πολιτιστική κληρονομιά, η συντήρηση, η έρευνα περιβαλλόντων πληροφορικής, η βιοποικιλότητα και οι βιβλιοθήκες ερευνητικών δεδομένων.¹¹

CRMarchaeo: χρησιμοποιείται για την τεκμηρίωση της αρχαιολογικής ανασκαφής, επεκτείνοντας τις ιδιότητες του CRMsci (τεκμηριώνεται αναλυτικά σε ακόλουθη ενότητα).¹²

CRMsci: χρησιμοποιείται για την ενσωμάτωση μεταδεδομένων σχετικά με την επιστημονική παρατήρηση, τις μετρήσεις και τα επεξεργασμένα δεδομένα στις αναλυτικές και εμπειρικές επιστήμες.¹³

CRMgeo: χωροχρονικό μοντέλο για τη διασύνδεση των χωροχρονικών ιδιοτήτων των έγχρονων οντοτήτων και των και γεωπληροφοριών με χρήση των εννοιολογήσεων, των ορισμών, των προτύπων κωδικοποίησης και των τοπολογικών σχέσεων που ορίζονται από το Open Geospatial Consortium (τεκμηριώνεται αναλυτικά σε ακόλουθη ενότητα).¹⁴

CRMdig: χρησιμοποιείται για την τεκμηρίωση των σταδίων και των μεθόδων παραγωγής (προέλευσης) προϊόντων ψηφιοποίησης και συνθετικών ψηφιακών αναπαραστάσεων όπως 2D, 3D ή ακόμη και μοντέλων animation που δημιουργούνται από διάφορες τεχνολογίες.¹⁵

CRMba: μοντέλο το οποίο υποστηρίζει την καταγραφή των μαρτυριών και των ασυνεχειών της ύλης στα αρχαία κτίρια, για τον εντοπισμό της εξέλιξης της δομής τους ανά τους αιώνες και την καταγραφή των σχέσεων μεταξύ καθενός από τα δομικά στοιχεία μεταξύ τους, καθώς και με το κτίριο στο σύνολό του.¹⁶

CRMtex: χρησιμοποιείται για την τεκμηρίωση της μελέτης και την έκδοση αρχαίων χειρόγραφων κειμένων περιγράφοντάς τα με κατάλληλα οντολογικά εργαλεία σε μια διεπιστημονική οπτική. Παρέχει κλάσεις και ιδιότητες που ανταποκρίνονται περισσότερο στις συγκεκριμένες ανάγκες των ειδικοτήτων αναφοράς, συμπεριλαμβανομένης της παπυρολογίας, της παλαιογραφίας, της κωδικολογίας και της επιγραφικής.¹⁷

⁹ <https://www.cidoc-crm.org/frbroo/home-0> (24/3/2024).

¹⁰ <https://www.cidoc-crm.org/pressoo/home-6> (24/3/2024).

¹¹ <https://www.cidoc-crm.org/crminf/home-4> (20/3/2024).

¹² <https://www.cidoc-crm.org/crmarchaeo/home-3> (20/3/2024).

¹³ <https://www.cidoc-crm.org/crmsci/home-1> (20/3/2024).

¹⁴ <https://www.cidoc-crm.org/crmgeo/home-5> . Σχετικά με τα πρότυπα του Open Geospatial Consortium βλ. <https://www.ogc.org/standards/> (20/3/2024).

¹⁵ <https://www.cidoc-crm.org/crmdig/home-2> (20/3/2024).

¹⁶ <https://www.cidoc-crm.org/crmba/home-7> (20/3/2024).

¹⁷ <https://www.cidoc-crm.org/crmtex/home-8> (20/3/2024).

CRMsoc: μοντέλο για την ενοποίηση δεδομένων σχετικών με κοινωνικά φαινόμενα και έννοιες που ενδιαφέρουν τις ανθρωπιστικές και κοινωνικές επιστήμες.¹⁸

3.3 Component MetaData Infrastructure (CMDI)

Το Component MetaData Infrastructure αναπτύχθηκε από γλωσσολόγους και ερευνητές στο πλαίσιο του CLARIN (Common Language Resources and Technology Infrastructure)¹⁹, αποσκοπώντας στη διασύνδεση πόρων που χρησιμοποιούνται στην έρευνα και, ειδικότερα, στην εφαρμογή ενός ενιαίου πλαισίου για τη διαχείριση μεταδεδομένων που αφορούν τις γλωσσικές πηγές και τη γλωσσική τεχνολογία. Από το 2008, το CMDI εξελίσσεται και έχει πλέον ενσωματωθεί ως ένα κεντρικό δομικό στοιχείο της υποδομής CLARIN.

Το CMDI βασίζεται σε μια προσαρμόσιμη δομή, η οποία επιτρέπει την απόδοση μεταδεδομένων σε διάφορα επίπεδα λεπτομέρειας, καθώς και τη διαλειτουργικότητά τους με ένα εύρος τεχνολογιών και ψηφιακών εφαρμογών. Ως εκ τούτου, ανταποκρίνεται στο σύνθετο τοπίο και τις κλιμακούμενες ανάγκες των κοινοτήτων που δραστηριοποιούνται στα επιμέρους πεδία της γλωσσολογίας, οι οποίες καθιστούν την ευρεία εφαρμογή ενός κοινού προτύπου (όπως για παράδειγμα το IMDI) σχεδόν αδύνατη. Το CMDI έχει αναπτυχθεί ως μια υποδομή που επιτρέπει τη δημιουργία και τη συσχέτιση επιμέρους δομικών στοιχείων πληροφορίας για την περιγραφή γλωσσικών πόρων.

Το CMDI παρέχει ένα ενιαίο οντολογικό πλαίσιο που επιτρέπει στους ερευνητές να δημιουργήσουν και να κατηγοριοποιήσουν επιμέρους στοιχεία μεταδεδομένων (metadata components) τα οποία μπορούν να επαναχρησιμοποιηθούν -είτε μεμονωμένα, είτε συνδυαστικά. Συνεπώς, το CMDI συνιστά ένα “μέτα-σχήμα”, που υποστηρίζει προσαρμοσμένες περιγραφές (profiles), την ενσωμάτωση στοιχείων από άλλα σχήματα (όπως τα IMDI και OLAC που τεκμηριώνονται στη συνέχεια), καθώς και τον ad hoc σχεδιασμό πρόσθετων σχημάτων βάσει συγκεκριμένων προδιαγραφών.

Τα στοιχεία μεταδεδομένων καταχωρούνται σε κοινούς καταλόγους (registries) με αναζητήσιμο και προσαρμόσιμο περιεχόμενο, στοιχεία (fields) των οποίων οι ερευνητές συνδυάζουν για να παράγουν τα περιγραφικά μεταδεδομένα μιας διαθέσιμης πηγής.²⁰ Κατά τη δημιουργία (ή επιλογή) μιας προσαρμοσμένης περιγραφής, παρέχεται η δυνατότητα δημιουργίας αρχείων XML, τα οποία παρέχουν πληροφορίες σχετικά με τα στοιχεία που αυτή περιέχει. Τα στοιχεία αυτά είναι: α) η κεφαλίδα (header), με πληροφορίες σχετικά με τη διαδικασία δημιουργίας και τον παραγωγό του εν λόγω δομικού στοιχείου, β) οι πόροι (resources), με συνδέσμους προς τις πηγές που περιγράφονται ή/και άλλες συναφείς περιγραφές και γ) τα στοιχεία (components), με αναφορά στα δομικά στοιχεία της εν λόγω προσαρμοσμένης περιγραφής.

Σε όρους σημασιολογικής και τεχνικής διαλειτουργικότητας, το CMDI παρέχει τη δυνατότητα ενσωμάτωσης στοιχείων του IMDI και της παραγωγής αρχείων XML, υποστηρίζοντας, παράλληλα, τη δημιουργία, επιμέλεια και επικύρωσή τους με την εφαρμογή εργαλείων που παρέχει η υποδομή CLARIN. Πέραν της ανάλυσης φυσικής γλώσσας και της μηχανικής μετάφρασης, το CMDI εξυπηρετεί

¹⁸ <https://cidoc-crm.org/crmsoc/home-10> (20/3/2024).

¹⁹ CMDI: <https://www.clarin.eu/content/cmdi-component-metadata-infrastructure>
CLARIN: <https://www.clarin.eu/> (20/3/2024).

²⁰ CMDI Component Registries: <https://catalog.clarin.eu/ds/ComponentRegistry/#/> (20/3/2024).

τη διαχείριση μεταδεδομένων πολιτιστικών πόρων -στοιχείο που το καθιστά συναφές και με άλλα επιστημονικά πεδία όπως η ιστορία και η ανθρωπολογία.

3.4 CRMarchaeo Excavation Model

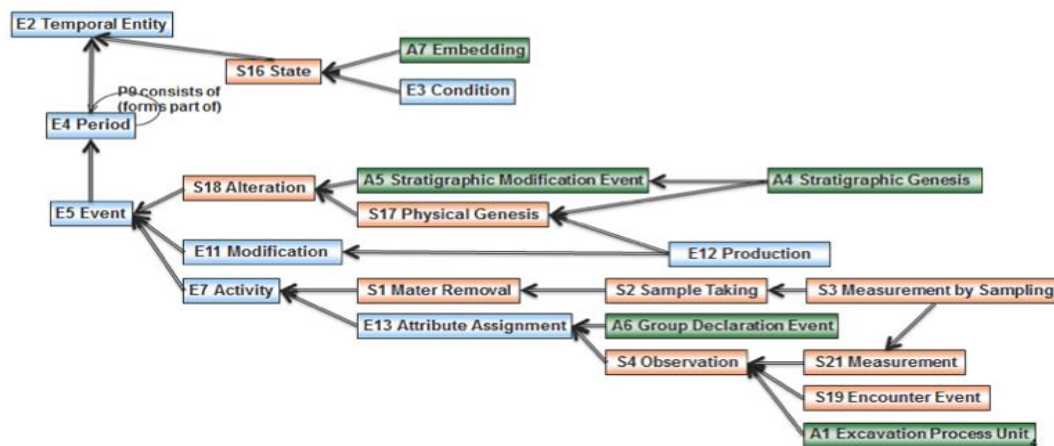
Το CRMarchaeo²¹ είναι μια επέκταση του οντολογικού προτύπου CIDOC CRM που προσανατολίζεται στην ψηφιακή αποτύπωση της διαδικασίας της αρχαιολογικής ανασκαφής με όλες τις σχετικές οντότητες και δραστηριότητες. Για την ανάπτυξή του ελήφθησαν υπόψη πρότυπα και μοντέλα που χρησιμοποιούνται από οργανισμούς πολιτιστικής κληρονομιάς και πραγματοποιήθηκε σε βάθος ανάλυση των μεταδεδομένων που προκύπτουν από την αρχαιολογική τεκμηρίωση. Το πρότυπο αξιοποιεί το οντολογικό πρότυπο CRMsci, από το οποίο κληρονομεί τις έννοιες σχετικά με τη γεωλογία και τη στρωματογραφία.

Το CRMarchaeo επιζητεί να υποστηρίξει την ενσωμάτωση της τεκμηρίωσης με σκοπό την τυποποίηση της γνώσης που αντλείται από την αρχαιολογική παρατήρηση. Απώτερος στόχος του είναι να διευκολύνει τη σημασιολογική κωδικοποίηση, τη διαλειτουργικότητα και την πρόσβαση στην αρχαιολογική τεκμηρίωση.

Επίκεντρο της ανάπτυξης του CRMarchaeo είναι το στρωματογραφικό συγκείμενο της αρχαιολογικής διαδικασίας. Το μοντέλο περιλαμβάνει οντότητες και ιδιότητες για την ενδελεχή περιγραφή της εκάστοτε στρωματογραφικής κατάστασης, των φυσικών φαινομένων ή των ανθρώπινων παρεμβάσεων που οδήγησαν στη δημιουργία τους, των υφιστάμενων στρωμάτων και επιφανειών, καθώς και των ανθρώπινων λειψάνων ή αντικειμένων που ανευρέθηκαν. Τούτο βοηθά στον ακριβή προσδιορισμό των χρονικών ακολουθιών και συμβάλλει στην ερμηνεία των ευρημάτων και στην ανασύσταση της ανθρώπινης παρουσίας και δράσης στον συγκεκριμένο χώρο.

Το CRMarchaeo αρθρώνεται σε 10 “τάξεις” (classes), οι οποίες παρέχουν το σημασιολογικό υπόβαθρο της ανασκαφικής διαδικασίας και προκύπτουν από τις ιεραρχίες της οντολογίας του CIDOC CRM, και σε 32 “ιδιότητες” (properties), οι οποίες περιγράφουν σχέσεις μεταξύ των τάξεων. Η εννοιολόγηση αυτή αποσκοπεί στο να διαμορφώσει μια κοινή, αντικειμενική γλώσσα μεταξύ διαφορετικών ανασκαφικών μεθόδων και παραδόσεων, η οποία διευκολύνει τη διαλειτουργικότητα και την επανάχρηση των δεδομένων. Οι συνδυασμοί τάξεων και ιδιοτήτων μπορούν να αναπαρασταθούν σε σημασιολογικούς γράφους, επιτρέποντας έτσι την περαιτέρω επεξεργασία από μηχανικούς της πληροφορίας.

²¹ CRMarchaeo Excavation Model, <https://www.cidoc-crm.org/crmarchaeo/> (20/3/2024).



Εικ 4. Παράδειγμα όρων και συσχετίσεων του μοντέλου. Πηγή: [Definition of the CRMarchaeo](#)

Το CRMarchaeo αναπτύσσεται σε συνεργασία με θεσμούς πολιτιστικής κληρονομιάς, προγράμματα και υποδομές, όπως το ARIADNE.²² Η πλέον πρόσφατη ενημέρωση του προτύπου (CRMarchaeo 2.0) έγινε τον Σεπτέμβριο του 2023.²³ Παράλληλα, οι ομάδες εργασίας συνεχίζουν την επεξεργασία του προτύπου, όπως φαίνεται στην αποτύπωση και δημοσιοποίηση των διαδικασιών για τη λύση συγκεκριμένων θεμάτων που ανακύπτουν.²⁴ Μέχρι στιγμής, πάντως, δεν είναι διαθέσιμες πηγές πληροφορίας και εργαλεία για την εκπαίδευση στη χρήση του προτύπου.

Δεδομένης της απόλυτης συμβατότητας του CRMarchaeo με την οντολογία CIDOC CRM, η οποία έχει σημαντική διάχυση στο πλαίσιο της σημασιολογικής διαχείρισης της πολιτιστικής κληρονομιάς, το πρότυπο μπορεί να σταθεί χρήσιμο στην αρχαιολογική ερευνητική κοινότητα. Άλλωστε, όπως προέκυψε από την έρευνα μέσω ερωτηματολογίου που πραγματοποιήθηκε στο πρώτο πακέτο εργασίας του Ψηφιακού Τοπίου και αποτυπώθηκε στο αντίστοιχο παραδοτέο του έργου, η αρχαιολογία είναι, μετά τις ψηφιακές ανθρωπιστικές επιστήμες, το επιστημονικό πεδίο που παρουσιάζει τη μεγαλύτερη χρήση ψηφιακών μεθόδων και εργαλείων.²⁵ Αυτή η εξοικείωση έγινε εμφανής και στη συνάντηση της αντίστοιχης ομάδας εστίασης.²⁶

3.5 CRMgeo

Το CRMgeo υλοποιήθηκε ως επέκταση του CIDOC CRM, με σκοπό τη σημασιολογική διασύνδεση γεωχωρικών στοιχείων και πολιτισμικών φαινομένων, αξιοποιώντας τις λεπτομερείς μεθοδολογίες και τα εξειδικευμένα εννοιολογικά μοντέλα που εφαρμόζονται στο πεδίο των γεωχωρικών τεχνολογιών.²⁷ Προς την κατεύθυνση αυτή, το CRMgeo εξειδικεύει και επεκτείνει τα στοιχεία και τις συσχετίσεις του CIDOC CRM, ώστε να είναι δυνατή η χαρτογράφηση των δεδομένων που παράγονται από γεωπληροφοριακά συστήματα στο οντολογικό πλαίσιο που ορίζει το CIDOC CRM. Συγκεκριμένα, το CRMgeo εφαρμόζει το αναπαραστατικό σχήμα, τους ορισμούς και τις σχέσεις που

²² ARIADNE Research Infrastructure, <https://www.ariadne-research-infrastructure.eu/> (20/3/2024).

²³ CRMarchaeo 2.0, <https://www.cidoc-crm.org/crmarchaeo/ModelVersion/crmarchaeo-version-2.0> (20/3/2024).

²⁴ CRMarchaeo. Issue Processing, https://www.cidoc-crm.org/crmarchaeo/issue_processing_fm (20/3/2024).

²⁵ Ψηφιακό Τοπίο: Χρήση ψηφιακών μεθόδων στις ανθρωπιστικές και κοινωνικές επιστήμες / Ανάλυση τάσεων, Ιούλιος 2023, σελ. 28.

²⁶ Ο.π. , σελ. 75.

²⁷ Ιστότοπος: <https://www.cidoc-crm.org/crmgeo/home-5> (20/3/2024).

ορίζονται στο πρότυπο GeoSPARQL²⁸ στο αναπαραστατικό μοντέλο του CIDOC CRM, επιτρέποντας την ενσωμάτωση πληροφοριών που διατίθενται σε μορφή GIS (Geographic Information Systems).

Το CRMgeo αποτελεί ένα λεπτομερές σχήμα, το οποίο υποστηρίζει τη σύνδεση γεγονότων, φαινομένων και υλικών αντικειμένων με αντικειμενικά προσδιορίσιμα χωροχρονικά στοιχεία: αναπτύσσεται ως σχεσιακό μοντέλο μεταξύ 10 κύριων “τάξεων” (classes) που περιγράφουν οντότητες, περιόδους και τόπους και 57 “ιδιοτήτων” (properties) που αναπαριστούν σημασιολογικά τους συσχετισμούς μεταξύ 2 ή περισσότερων τάξεων. Χρησιμοποιεί και επεκτείνει, δε, τις τιμές του CIDOC CRM (ISO21127) -διατηρώντας τον προσανατολισμό του εν λόγω προτύπου προς την πολιτιστική κληρονομιά, καθώς και τα στοιχεία διαλειτουργικότητάς του (π.χ. δεδομένα εξαγωγήμα σε μορφότυπους όπως RDFS, OWL).²⁹

Λόγω της συνάφειας με ένα ευρύ πεδίο επιστημών, το CRMgeo χρησιμοποιείται από φορείς πολιτιστικής διαχείρισης που ενδιαφέρονται να πλαισιώσουν τις περιγραφές των -φυσικών και ψηφιακών- αντικειμένων που διαχειρίζονται με στοιχεία γεωχωρικού προσδιορισμού, καθώς και από ερευνητές των ΑΚΕ που επιθυμούν να αποδώσουν με λεπτομέρεια το χωροχρονικό πλαίσιο των αντικειμένων μελέτης τους (όπως αρχαιολογικοί χώροι και ιστορικά γεγονότα).

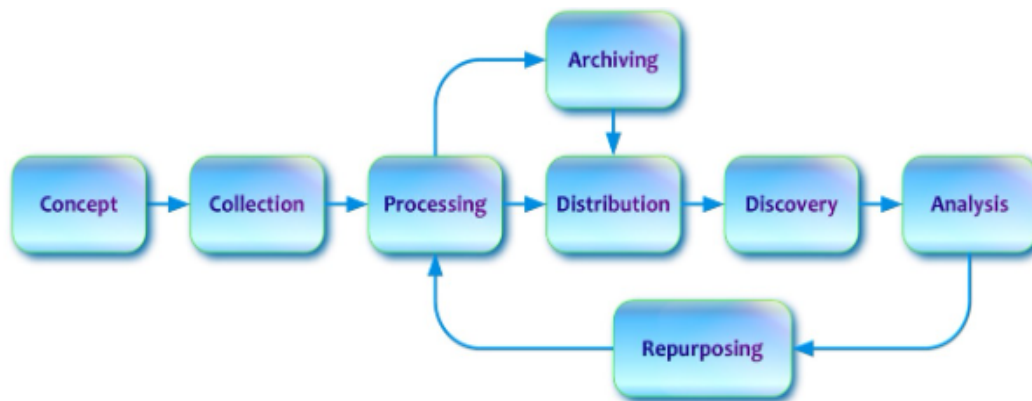
3.6 Data Documentation Initiative (DDI) και CESSDA Metadata Management (CMM)

Το Data Documentation Initiative (DDI)³⁰ είναι ένα ελεύθερο διεθνές πρότυπο για την περιγραφή των δεδομένων που παράγονται από έρευνες πεδίου (surveys) και άλλες μεθόδους παρατήρησης στις κοινωνικές και οικονομικές επιστήμες, καθώς και στις επιστήμες συμπεριφοράς και υγείας. Το DDI, που αναπτύσσεται με τη συνεργασία και συνεπιμέλεια των μελών της διεθνούς συνεργασίας Data Documentation Initiative Alliance (DDI Alliance), τεκμηριώνει και διαχειρίζεται διάφορα στάδια του κύκλου ζωής των ερευνητικών δεδομένων, όπως η εννοιολογική αναπαράσταση, η συλλογή, η επεξεργασία, η διανομή, η ανάκτηση και η αρχειοθέτηση. Η τεκμηρίωση δεδομένων με το DDI διευκολύνει την κατανόηση, την ερμηνεία και τη χρήση των δεδομένων από ανθρώπους, συστήματα λογισμικού και δίκτυα υπολογιστών. Στο πλαίσιο αυτό το DDI ευθυγραμμίζεται με τις αρχές FAIR, καθώς προωθεί τη διαλειτουργικότητα και τον διαμοιρασμό των δεδομένων.

²⁸ Ιστότοπος: <https://www.ogc.org/standard/geosparql/> (20/3/2024).

²⁹ Τεκμηρίωση: https://www.cidoc-crm.org/crmgeo/sites/default/files/CRMgeo1_2.pdf (20/3/2024).

³⁰ Data Documentation Initiative (DDI), <https://ddialliance.org/> (20/3/2024).



Εικ 5. Τα στάδια του κύκλου ζωής των δεδομένων, κατά το πρότυπο DDI. Πηγή: [DDI Lifecycle Documentation](#)

Το DDI αποτελείται από μια σειρά “προϊόντων”, τα οποία έχουν σχεδιαστεί ως ειδικότερα πρότυπα που καλύπτουν συγκεκριμένα πεδία εργασιών: την εννοιολόγηση, την ερευνητική μεθοδολογία, την επεξεργασία των δεδομένων, την ποσοτική και ποιοτική ανάλυση, και τη διαχείριση της πρόσβασης και κοινοποίησης δεδομένων.

Τα “προϊόντα” αυτά είναι:

- το DDI Codebook, ένα πρότυπο περιγραφής και τεκμηρίωσης των συνόλων δεδομένων,
- το DDI Lifecycle, ένα πρότυπο διαχείρισης των δεδομένων στο ανάπτυγμα του κύκλου ζωής τους,
- μια σειρά από Controlled Vocabularies, ελεγχόμενα λεξιλόγια που χρησιμοποιούνται στις κοινωνικές επιστήμες για τον προσδιορισμό ερευνητικών δεδομένων,
- το XKOS (Extended Knowledge Organization System), το οποίο επιτρέπει τη διαχείριση στατιστικών ταξινομήσεων και εννοιών, και
- το SDTL (Structured Data Transformation Language), μια γλώσσα αναπαράστασης που χρησιμοποιείται για τον μετασχηματισμό δεδομένων.³¹

Το DDI αφορά κυρίως τους κλάδους των κοινωνικών επιστημών οι οποίοι επεξεργάζονται, αναλύουν και αρχειοθετούν δεδομένα που προκύπτουν από έρευνες πεδίου και συναφείς δράσεις (εκλογές, δημοψηφίσματα, ερωτηματολόγια, δημοσκοπήσεις, επιτόπιες έρευνες, στατιστικά στοιχεία οργανισμών κλπ.). Όπως φαίνεται από τον κατάλογο των μελών της DDI Alliance, που προέρχονται από τις ΗΠΑ, πολλές ευρωπαϊκές χώρες και την Ωκεανία, το DDI καλύπτει ένα ευρύ επιστημονικό ανάπτυγμα, από πανεπιστημιακά τμήματα, ινστιτούτα, αποθετήρια και δημόσιους φορείς κοινωνικής έρευνας μέχρι συγκεκριμένες ερευνητικές εξειδικεύσεις (δημογραφία, απασχόληση, δημόσια υγεία, κοινή γνώμη κλπ.). Ιδιαίτερη σημασία έχει για τη στατιστική ανάλυση των δεδομένων και για την εξασφάλιση της διαλειτουργικότητας των μεταδεδομένων. Ως εκ τούτου, απαντά στις ανάγκες και προσδοκίες που επισημάνθηκαν κατά τη συνάντηση της ομάδας εστίασης

³¹ Πληροφορίες για τα εν λόγω πρότυπα παρέχονται στους ακόλουθους συνδέσμους:

DDI Codebook: <https://ddialliance.org/Specification/DDI-Codebook/2.5/>

DDI Lifecycle: <https://ddialliance.org/Specification/DDI-Lifecycle/>

XKOS: <https://ddialliance.org/Specification/RDF/XKOS>

SDTL: <https://ddialliance.org/products/sdtl/1.0> (20/3/2024).

των κοινωνικών και πολιτικών επιστημών στο πλαίσιο του πρώτου πακέτου εργασίας του Ψηφιακού Τοπίου και αποτυπώθηκαν στο αντίστοιχο παραδοτέο του έργου, όπου σημειώνονται τα εξής: “Υπάρχει, επίσης, μεγάλο ενδιαφέρον για τη διαχείριση κοινωνικών δεδομένων, με απλοποίηση του κύκλου ζωής τους, την αξιοποίηση τους και την επικέντρωση στην ανάλυση με αυτοματοποίηση των υπόλοιπων διαδικασιών. [...]”.³²

Τέλος, το DDI περιλαμβάνει ένα σύνολο εκπαιδευτικών εφαρμογών που εξειδικεύονται σε συγκεκριμένες κατηγορίες ενδιαφερόμενων χρηστών (βιβλιοθηκονόμους, αρχειονόμους, διαχειριστές δεδομένων, ερευνητές, διαχειριστές μεταδεδομένων) και στα συγκεκριμένα πρότυπα που έχει αναπτύξει. Παράλληλα, διοργανώνει θεματικά συνέδρια και εργαστήρια, όπως και εκπαιδευτικά σεμινάρια.

Πολύ κοντά στο πρότυπο DDI, στο οποίο άλλωστε βασίζεται, βρίσκεται το σχήμα μεταδεδομένων (ή “κορμός μεταδεδομένων”, όπως έχει αποδοθεί στα ελληνικά)³³ CMM (CESSDA Metadata Management), το οποίο έχει αναπτυχθεί από το Consortium of European Social Data Archives (CESSDA).³⁴ Περιλαμβάνει σαφείς και εξαντλητικούς ορισμούς μεταδεδομένων και επιχειρεί να καλύψει ένα εύρος μεθοδολογικών απαιτήσεων, όπως η επαναληψιμότητα. Το μοντέλο, που αφορά τα ποσοτικά δεδομένα των κοινωνικών επιστημών, στηρίζεται στο πρότυπο μεταδεδομένων DDI Lifecycle 3.2. Στόχος του είναι να παράσχει υποστήριξη τόσο στην αρχειακή και βιβλιοθηκονομική διαχείριση των δεδομένων (αρχεία, αποθετήρια, βιβλιοθήκες) όσο και στην έρευνα σε πανεπιστήμια και ερευνητικά ινστιτούτα.

Το δημοσιευμένο πλήρες σχήμα του μοντέλου, το οποίο περιγράφεται σε αρχείο xls,³⁵ περιλαμβάνει κατηγορίες μεταδεδομένων σχετικά με μελέτες, πρόσωπα, ιδρύματα, σύνολα δεδομένων, μέσα, ερωτηματολόγια, έννοιες, έγγραφα, δημοσιεύσεις, ομάδες μελετών και περιγραφές εγγράφων.

Ιδιαίτερα σημαντικό είναι ότι το μοντέλο συνοδεύεται από ένα σύνολο εκπαιδευτικών εφαρμογών, όπως ο οδηγός χρήσης, οι παρουσιάσεις και οι πλατφόρμες διαδικτυακής εκπαίδευσης, που είναι διαβαθμισμένες ανάλογα με το επίπεδο των εκπαιδευόμενων.³⁶

Τα παραπάνω καθιστούν το πρότυπο DDI και το σχήμα CMM εξαιρετικά προσιτά εργαλεία για την επεξεργασία των μεταδεδομένων των ποσοτικών ερευνών των κοινωνικών επιστημών. Ένα πρόσθετο στοιχείο που ενισχύει την ελκυστικότητα και εφαρμοσιμότητά τους είναι ότι μέλος της κοινοπραξίας CESSDA είναι το So.Da.Net, η ελληνική εθνική ερευνητική υποδομή για τις κοινωνικές επιστήμες, η οποία συγκεντρώνει και τεκμηριώνει δεδομένα και μεταδεδομένα ερευνών των κοινωνικών επιστημών στην Ελλάδα.³⁷ Το So.Da.Net παρέχει συμβουλευτικές υπηρεσίες για τη διαχείριση των δεδομένων, οι οποίες στηρίζονται στο CESSDA Data Management Expert Guide, και συνιστά τη χρήση του DDI στην επεξεργασία των μεταδεδομένων.³⁸ Το So.Da.Net και οι φορείς που συναπαρτίζουν την υποδομή μπορούν να συνδράμουν ουσιαστικά σε ό,τι αφορά την τεκμηρίωση

³² Ψηφιακό Τοπίο, σελ. 71.

³³ Social Data Network in Action, <https://sodanet.gr/storage/files/SodanetInAction.pdf>, (20/3/2024).

³⁴ CESSDA Metadata Model, <https://www.cessda.eu/Training-Resources/Resource-crd-2592> (20/3/2024).

³⁵ CESSDA Metadata Office (MDO), <https://zenodo.org/records/4751455> (20/3/2024).

³⁶ Training Resources, <https://www.cessda.eu/Training-Resources?keywords=CMM> (20/3/2024).

³⁷ Social Data Network, <https://sodanet.gr/>, (20/3/2024).

³⁸ Συμβουλευτικές υπηρεσίες για τη δημιουργία Σχεδίου Διαχείρισης Δεδομένων, <https://sodanet.gr/data-services/dmp>, (20/3/2024).

των δεδομένων και τη διαχείριση των μεταδεδομένων με όρους που διασυνδέουν τις ψηφιακές πρακτικές των κοινωνικών επιστημόνων με τις διεθνείς απαιτήσεις διαλειτουργικότητας.

3.7 Dublin Core

Το Σύνολο Στοιχείων Μεταδεδομένων Dublin Core³⁹ (Dublin Core™ Metadata Element Set - DCMES) ή Dublin Core (DC) αποτελεί ένα διεθνές πρότυπο μεταδεδομένων, το οποίο χρησιμοποιείται για την περιγραφή πόρων κάθε τύπου και μορφής, με σκοπό τον εντοπισμό και την ανάκτησή τους σε διαδικτυακό περιβάλλον. Περιλαμβάνει 15 βασικά στοιχεία (elements) μέσω των οποίων μπορούν να περιγραφούν ψηφιακά αντικείμενα, όπως βίντεο, ήχος, εικόνες, κείμενα ή ιστοσελίδες, αλλά και φυσικά, όπως βιβλία ή έργα τέχνης. Το όνομα του προκύπτει από την πόλη Dublin του Οχάιο (ΗΠΑ), όπου το 1995 πραγματοποιήθηκε η συνάντηση εργασίας στην οποία ορίστηκε ο πυρήνας (core) των βασικών στοιχείων που το αποτελούν. Στόχος του προτύπου ήταν εξ αρχής η απλότητα στη δομή και στη σημασιολογία των μεταδεδομένων. Για τον λόγο αυτό, τα στοιχεία του ορίζονται κατάλληλα ώστε να έχουν μια κοινά κατανοητή (και διαθεματική) σημασιολογία, η οποία προκύπτει από τη σύγκλιση κοινών εννοιών σε γενικά στοιχεία.

Ο πυρήνας των 15 στοιχείων του προτύπου αποτελεί μέρος ενός ευρύτερου συνόλου λεξιλογίων και τεχνικών προδιαγραφών μεταδεδομένων που αναπτύσσονται και συντηρούνται από τον οργανισμό DCMI (Dublin Core Metadata Initiative),⁴⁰ ενώ νεότερες εκδόσεις προβλέπουν τη χρήση εξειδικευτών (qualifiers), για την εξειδίκευση των στοιχείων αυτών όπου χρειάζεται.⁴¹ Το Dublin Core έχει υιοθετηθεί από το ISO και η Τεχνική Επιτροπή του TC 46 έχει εκπονήσει τα πρότυπα ISO 15836-1:2017⁴² και 15836-2:2019,⁴³ ενώ η αντίστοιχη επιτροπή ΕΛΟΤ/ΤΕ 22 έχει εκπονήσει το ελληνικό πρότυπο ΕΛΟΤ ISO 15836-1:2019.⁴⁴

Βασική δομή

Το συγκεκριμένο πρότυπο περιλάμβανε αρχικά δύο επίπεδα εφαρμογής: το απλό (Simple) Dublin Core (DC) και το εξειδικευμένο (Qualified) Dublin Core (DC). Το απλό DC χρησιμοποιούσε τον βασικό πυρήνα των 15 στοιχείων για την περιγραφή των τεκμηρίων, ενώ η εκτεταμένη του εκδοχή χρησιμοποιούσε κάποια επιπλέον στοιχεία και ταυτόχρονα έδινε τη δυνατότητα εισαγωγής εξειδικευτών (qualifiers), παρέχοντας τη δυνατότητα για πιο λεπτομερείς περιγραφές και ακριβέστερη αναζήτηση ψηφιακών πόρων. Ωστόσο από το 2012 τα δύο σχήματα έχουν ενσωματωθεί στο DCMI Metadata Terms ως ένα σύνολο όρων εκφρασμένων σε λεξιλόγιο RDF.

Η αρχική έκδοση του DC (DCMES Version 1.1)⁴⁵ - DC Simple: Η αρχική έκδοση του DC περιλάμβανε 15 στοιχεία για την περιγραφή πόρων. Κάθε στοιχείο έχει μια περιγραφή («ετικέτα») για την

³⁹ Dublin Core, <https://www.dublincore.org/> (20/3/2024).

⁴⁰ DCMI Dublin Core, <https://www.dublincore.org/> (20/3/2024).

⁴¹ Λεξικό [Επιστήμης](https://www.dublincore.org/terms/) της Πληροφόρησης, <https://liblex.gr/term.php?termid=8> (20/3/2024)

⁴² ISO 15836-1:2017, <https://www.iso.org/standard/71339.html> (20/3/2024).

⁴³ ISO 15836-2:2019 <https://www.iso.org/standard/71341.html> (20/3/2024).

⁴⁴ ΕΛΟΤ ISO 15836-1:2019, http://193.218.125.20/SEP_ELOT-ISO_15836-1.pdf (20/3/2024).

⁴⁵ Dublin Core™ Metadata Element Set, Version 1.1
<https://www.dublincore.org/specifications/dublin-core/dces/> (20/3/2024).

αναγνώρισή του από άνθρωπο και ένα URI, ως ένα καθολικά μοναδικό αναγνωριστικό για το συγκεκριμένο στοιχείο για χρήση στη μηχανική επεξεργασία.

Τα 15 βασικά στοιχεία μεταδεδομένων του απλού DC είναι προαιρετικά και χωρίζονται σε 3 κατηγορίες, που αφορούν την πληροφορία που περιγράφεται από αυτά:

- Το περιεχόμενο, το οποίο περιλαμβάνει τα στοιχεία που περιγράφουν το περιεχόμενο του πόρου: τον τίτλο του (Title), την περιγραφή του (Description), τη γλώσσα του (Language), το θέμα του (Subject), τη σχέση του με άλλο περιεχόμενο (Relation), τον χρόνο και τον τόπο που αφορά (Coverage) κτλ.
- Την πνευματική ιδιοκτησία, που περιλαμβάνει τα στοιχεία τα οποία σχετίζονται με: τη δημιουργία (Creator), τη διάθεση και την πνευματική ιδιοκτησία του πόρου (Rights) κτλ.
- Το στιγμιότυπο, το οποίο περιλαμβάνει τα στοιχεία που περιγράφουν: τον τύπο (Type), το μορφότυπο (Format), την ημερομηνία (Date), που σχετίζεται με ένα συμβάν στον κύκλο ζωής του πόρου, το Αναγνωριστικό (Identifier) του πόρου.

Qualified DC: Το εξειδικευτικό (Qualified) DC προκύπτει από την επέκταση του απλού DC. Το εν λόγω σχήμα περιλαμβάνει κάποια επιπλέον στοιχεία, όπως επίσης τους εξειδικευτές (qualifiers) των 15 βασικών ετικετών του προτύπου. Οι εξειδικευτές αυτοί χρησιμοποιούνται για τη «βελτίωση» της σημασιολογικής ακρίβειας των στοιχείων όπου αυτή χρειάζεται. Για παράδειγμα, στην ετικέτα Description (περιγραφή) προστίθενται οι εξειδικευτές Abstract (περίληψη) και TableOfContents (λίστα περιεχομένων). Με τον τρόπο αυτό επιτυγχάνεται εξειδίκευση οποιουδήποτε στοιχείου (element refinement) σε ένα ή περισσότερα νέα στοιχεία. Επίσης περιλαμβάνει ένα σύνολο συνιστώμενων σχημάτων κωδικοποίησης, σχεδιασμένων για να βοηθήσουν στην απόδοση τιμής σε ένα στοιχείο. Τα σχήματα αυτά περιλαμβάνουν ελεγχόμενα λεξιλόγια και συντακτικούς κανόνες. Έτσι μια τιμή είναι ένας όρος από ελεγχόμενο λεξιλόγιο (π.χ. Library of Congress Subject Headings).

DCMI Metadata Terms:⁴⁶ Περιλαμβάνει το τρέχον σύνολο όρων που απαρτίζουν το λεξιλόγιο DC. Περιλαμβάνει τα 15 στοιχεία πυρηνικά στοιχεία του προτύπου, όπως και τους εξειδικευτές τους. Κάθε όρος έχει ένα μοναδικό URI στον ονοματοχώρο,⁴⁷ και όλοι οι όροι ορίζονται ως RDF ιδιότητες.

Πλεονεκτήματα και μειονεκτήματα στη χρήση του

Το DC αποτελεί τη lingua franca στον χώρο των μεταδεδομένων. Είναι ίσως το πιο ευρέως χρησιμοποιούμενο διεθνώς σχήμα μεταδεδομένων και αυτό οφείλεται κατ' αρχάς στην απλότητα της δομής και της σημασιολογίας του. Ακόμα και ένας μη ειδικός μπορεί να δημιουργήσει περιγραφικές εγγραφές για τα τεκμήρια που τον ενδιαφέρουν, χωρίς να χρειάζεται να επενδύσει ιδιαίτερο χρόνο για την εξοικείωση με το πρότυπο.

Επίσης η απλότητά του το καθιστά κοινό παρονομαστή των προτύπων για επικοινωνία και διαλειτουργικότητα, καθώς μέσω αυτού συνδέονται κύρια σημεία των περιγραφών που εκφράζονται με τα πρότυπα αυτά.⁴⁸ Το DC είναι σχεδιασμένο για εύκολη και προφανή αντιστοίχιση σε αυτό άλλων σχημάτων μεταδεδομένων, και η συμβολή του στη διαλειτουργικότητα είναι προφανής.

⁴⁶ DCMI Metadata Terms, <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> (20/3/2024).

⁴⁷ Βλ. σχετικά: <http://purl.org/dc/terms> (20/3/2024).

⁴⁸ Dublin Core, https://users.ionio.gr/~toraki/metadata_met/metadata_2007_dc.pdf (20/3/2024).

Επίσης, καθώς το DC είναι η σημασιολογική τομή των σχημάτων δεδομένων που χρησιμοποιούνται ευρέως, μπορεί να χρησιμοποιηθεί ως βάση στοιχείων σε προφίλ εφαρμογών για ευκολότερη (και προσχεδιασμένη) διαλειτουργικότητα.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description>
    <dc:title>Homage to Catalonia,</dc:title>
    <dc:creator>Orwell, George, 1903-1950.</dc:creator>
    <dc:type>text</dc:type>
    <dc:publisher>London, Secker and Warburg</dc:publisher>
    <dc:date>[1938]</dc:date>
    <dc:language>eng</dc:language>
  </rdf:Description>
</rdf:RDF>
```

Εικ. 6. Παράδειγμα εφαρμογής του προτύπου Dublin Core σε RDF-XML. Πηγή: [Wikimedia Commons](#)

Το πρότυπο είναι επεκτάσιμο για να καλύψει επιπρόσθετες ανάγκες περιγραφής και ανακάλυψης πόρων. Ωστόσο δεν προορίζεται για την περιγραφή σύνθετων πόρων, ούτε μπορεί να υποκαταστήσει πλουσιότερα περιγραφικά πρότυπα. Υπάρχουν, π.χ., είδη πόρων, όπως βιβλία ή περιοδικά, για τους οποίους κρίνεται καταλληλότερη η χρήση άλλων περιγραφικών προτύπων όπως το Marc, παρ' όλη την πολυπλοκότητά του σε σχέση με το DC.

Αναφορά στο Παραδοτέο ΠΕ1

Το DC είναι και στην Ελλάδα το πιο ευρέως διαδεδομένο πρότυπο και στην Ελλάδα για την περιγραφή ψηφιακού και μη περιεχομένου, όπως κατέδειξε η έρευνα που διεξήχθη για την αποτύπωση των τάσεων στη χρήση ψηφιακών μεθόδων, στο πλαίσιο του πρώτου παραδοτέου του έργου.⁴⁹ Η έρευνα έδειξε ότι το DC είναι πρώτο στη σειρά τόσο στις ανθρωπιστικές επιστήμες όσο και στις κοινωνικές επιστήμες ανάμεσα στα πρότυπα μεταδεδομένων που χρησιμοποιούνται περισσότερο. Όπως διαπιστώθηκε και στην εν λόγω έρευνα, η προτίμηση αυτή οφείλεται σε μεγάλο βαθμό στην απλότητα του προτύπου, το οποίο με τα ευάριθμα πεδία μεταδεδομένων που διαθέτει στη βασική του μορφή μπορεί να καλύψει περιγραφικά ποικίλες κατηγορίες ψηφιακών και μη αντικειμένων.

3.8 Encoded Archival Description (EAD)

Το EAD είναι ένα διεθνές πρότυπο μεταδεδομένων για ιεραρχικές περιγραφές αρχειακών καταγραφών με χρήση XML.⁵⁰ Έχει αναπτυχθεί από την Ένωση Αμερικανών Αρχειονόμων και βασίζεται στο Διεθνές Πρότυπο Αρχειακής Περιγραφής ΔΙΠΑΠ (Γ) του Διεθνούς Συμβουλίου Αρχείων. Το ΔΙΠΑΠ έχει αναδειχθεί ως το σημαντικότερο πρότυπο αρχειακής περιγραφής και έχει χρησιμοποιηθεί ευρέως σε ολόκληρο τον κόσμο. Με τη συνδρομή της Ελληνικής Αρχειακής

⁴⁹ Ψηφιακό Τοπίο, σ. 38

⁵⁰ Encoded Archival Description, <https://www.loc.gov/ead/> (20/3/2024).

Εταιρείας,⁵¹ το ΔΙΠΑΠ αποτέλεσε τη βάση όπου στηρίχθηκε η περιγραφή όλων των μεγάλων δημόσιων και ιδιωτικών αρχειακών συνόλων στην Ελλάδα.

Το EAD διατίθεται σήμερα στην τρίτη έκδοσή του (EAD3), η οποία ολοκληρώθηκε το 2023.⁵² Στην επεξεργασία της τρίτης αυτής έκδοσης δόθηκε ιδιαίτερη σημασία στα εξής:

- Στην επίτευξη μεγαλύτερης εννοιολογικής και σημασιολογικής συνέπειας
- Στους μηχανισμούς με τους οποίους οι κωδικοποιημένες με EAD πληροφορίες μπορούν απρόσκοπτα να ανταλλάξουν, να ενσωματώσουν ή να διασυνδεθούν με δεδομένα που διατηρούνται σύμφωνα με άλλα πρωτόκολλα
- Στη βελτίωση της αναπαράστασης περιγραφικών πληροφοριών σε διεθνή και ιδιαίτερα σε πολύγλωσσα περιβάλλοντα, και
- Στις επιπτώσεις της νέας έκδοσης ως προς τους χρήστες

Το EAD, όπως και το ΔΙΠΑΠ, είναι γνωστό στην ελληνική αρχειακή κοινότητα και έχει με τη σειρά του αξιοποιηθεί στο πλαίσιο ταξινόμησης ή/και ψηφιοποίησης μεγάλων αρχειακών συλλογών (π.χ. Γενικά Αρχεία του Κράτους). Τα ιδιαίτερα πλεονεκτήματά του είναι η πολυεπίπεδη περιγραφή των αρχειακών τεκμηρίων και συνόλων, η ανταπόκριση στις ανάγκες μιας διεθνούς κοινότητας χρηστών που έχει αποδεχθεί την αναζήτηση κοινών προτύπων, και το γεγονός πως έχει σχεδιαστεί από την ίδια την κοινότητα των αρχειονόμων. Βασικές διαδικασίες του EAD είναι η λεπτομερής τεκμηρίωση της αρχειακής περιγραφής (δημιουργός της περιγραφής, ταυτοποίηση, βιβλιογραφία, διάθεση, κωδικοποίηση περιγραφής κλπ.), από τη μια, και η περιγραφή του αρχειακού συνόλου (περιεχόμενο, δομή, παραγωγός, χρονολογίες παραγωγής, όροι πρόσβασης κλπ.), από την άλλη.⁵³

Όπως δείχθηκε από την έρευνα στο πλαίσιο του πρώτου παραδοτέου του έργου, οι αρχειονόμοι και βιβλιοθηκονόμοι είναι οι πλέον συστηματικοί χρήστες ψηφιακών μεθόδων και εφαρμογών στο στάδιο της διατήρησης των δεδομένων.⁵⁴ Μάλιστα το EAD είναι ένα από τα πιο διαδεδομένα πρότυπα μεταδεδομένων.⁵⁵ Εκείνο στο οποίο επικεντρώνονται οι συμμετέχοντες στην έρευνα είναι η ανάγκη εκπαίδευσης των επιστημόνων της πληροφορίας στις ιδιαίτερες ανάγκες της εργασίας τους και η ανάγκη ενημέρωσης στα συνεχώς εξελισσόμενα ψηφιακά εργαλεία που την υποστηρίζουν, όπως το EAD.⁵⁶

⁵¹ Ελληνική Αρχειακή Εταιρεία - Διεθνές Πρότυπο Αρχειακής Περιγραφής, <https://www.eae.org.gr/index.php/protypa/isad-dipap> (20/3/2024).

⁵² *Encoded Archival Description Tag Library Version EAD3 1.1.2, Edition 2023*, <https://www.loc.gov/ead/EAD3taglib/EAD3-TL-eng.pdf> (20/3/2024).

⁵³ Γιαννακόπουλος, Γ., Β. Μπουντούρη, *Εισαγωγή στην Αρχειονομία*, Αθήνα: ΣΕΑΒ, 2015, σελ. 80-83,

⁵⁴ Ψηφιακό Τοπίο, σελ. 28-29.

⁵⁵ Ο.π., σελ. 38-39.

⁵⁶ Ο.π., σελ. 64.

3.9 Europeana Data Model (EDM)

Το Europeana Data Model (EDM)⁵⁷ είναι ένα RDF μοντέλο που προτείνεται από την Europeana.⁵⁸ Παρέχει ψηφιακή πρόσβαση σε υλικό της ευρωπαϊκής πολιτιστικής κληρονομιάς για την αναπαράσταση μεταδεδομένων πολιτιστικού περιεχομένου. Η Europeana προβάλλει υλικό από περίπου 3.600 διαφορετικούς φορείς προέλευσης (γκαλερί, βιβλιοθήκες, αρχεία και μουσεία ανά την Ευρώπη) παρέχοντας πρόσβαση σε εκατομμύρια τεκμήρια (έργα τέχνης, βιβλία, μουσική και βίντεο για την τέχνη, τις εφημερίδες, την αρχαιολογία, τη μόδα, την επιστήμη, τον αθλητισμό κ.ά.). Οι φορείς προέλευσης δεν συνεργάζονται ξεχωριστά με την Europeana. Ένα δίκτυο συσσωρευτών συλλέγει τα δεδομένα, τα ελέγχει και τα εμπλουτίζει με πληροφορίες όπως η γεωτοποθεσία, ή τα συνδέει με άλλα υλικά ή σύνολα δεδομένων μέσω σχετικών ατόμων, τόπων ή θεμάτων.⁵⁹

Σκοπός είναι το EDM να αποτελέσει ένα μέσο ολοκλήρωσης για τη συλλογή, τη σύνδεση και τον εμπλουτισμό των μεταδεδομένων πολιτιστικής κληρονομιάς που παρέχονται από τους παρόχους περιεχομένου της Europeana. Περιλαμβάνει οποιοδήποτε στοιχείο (π.χ. κατηγορία ή ιδιότητα) που βρίσκεται στην περιγραφή ενός παρόχου περιεχομένου.⁶⁰ Αναπτύχθηκε το 2010, αλλά χρησιμοποιείται επίσημα από τον Ιανουάριο του 2012. Αντικατέστησε το πρότυπο Europeana Semantic Elements (ESE) που χρησιμοποιήθηκε επίσημα για τη δημιουργία μεταδεδομένων στην Europeana, το οποίο βασίστηκε στη λογική του ελάχιστου κοινού παρονομαστή για την επίτευξη της σημασιολογικής διαλειτουργικότητας μεταξύ της ανομοιογενούς πληροφορίας που προήλθε από μουσεία, αρχεία, βιβλιοθήκες και πολιτιστικούς φορείς. Αντίθετα με τον προκάτοχό του, το EDM βασίζεται στις τεχνολογίες του σημασιολογικού ιστού (Web 2.0) και των διασυνδεδεμένων δεδομένων και υιοθετεί ένα ανοικτό δυναμικό πλαίσιο πολυεπίπεδης αναπαράστασης ικανό να αξιοποιήσει τον πλούτο και την ποικιλότητα των διαφορετικών προτύπων τεκμηρίωσης που χρησιμοποιούνται στους επιμέρους τομείς από τους πολιτιστικούς φορείς και τους φορείς μνήμης (ενδεικτικά, το LIDO⁶¹ για τη μουσειακή τεκμηρίωση, το EAD για την αρχειακή και το MARC αναφορικά με τις ψηφιακές βιβλιοθήκες) και διαφορετικές οντολογίες (DC, SKOS, FOAF).⁶²

⁵⁷ <https://pro.europeana.eu/page/edm-documentation> (20/3/2024).

⁵⁸ <https://www.europeana.eu/el> (20/3/2024).

⁵⁹ <https://pro.europeana.eu/share-your-data/metadata> (20/3/2024).

⁶⁰ Τεκμηρίωση του μοντέλου

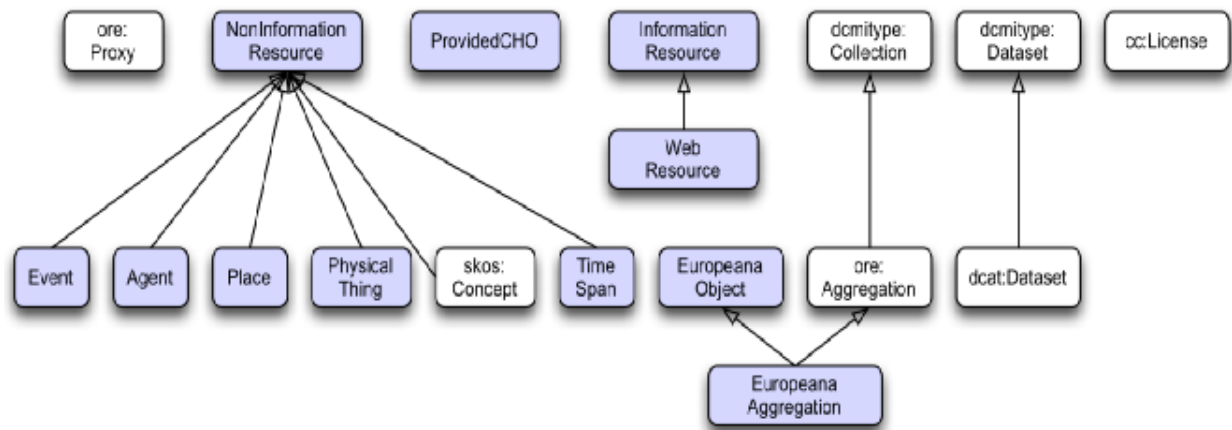
:https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.8_102017.pdf

Βλ. επίσης: <https://europeana.atlassian.net/wiki/spaces/EF/pages/987791389/EDM+-+Mapping+guidelines>

⁶¹ Lightweight Information Describing Objects (LIDO) <https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/about-lido/what-is-lido/> (20/3/2024).

⁶² SKOS: <https://www.w3.org/2004/02/skos/> (20/3/2024).

FOAF: <http://www.foaf-project.org/> (20/3/2024).



Εικ. 7: Ιεραρχία “τάξεων” του EDM. Πηγή: [Definition of the Europeana Data Model v5.2.8](#)

Η ομάδα εργασίας για τα μεταδεδομένα της Europeana, σε συνεργασία με τους εταίρους του έργου, παρέχει τεκμηριωτικό υλικό σχετικά με το EDM για τους παρόχους περιεχομένου. Οι βασικές εκδόσεις είναι:

- Το Europeana Data Model Primer που περιέχει επεξήγηση της διαδικασίας δημιουργίας μοντέλου EDM, του υποβάθρου και του τρόπου με τον οποίο μπορούν να χρησιμοποιηθούν μαζί τάξεις και ιδιότητες για τη μοντελοποίηση δεδομένων.⁶³
- Το Europeana Data Model Definition που περιέχει εξήγηση της ιεραρχίας όλων των κλάσεων και ιδιοτήτων που θα μπορούσαν να χρησιμοποιηθούν στη δημιουργία μεταδεδομένων για την Europeana.⁶⁴
- Τα EDM objects templates φανερώνουν ποιες ιδιότητες ισχύουν σε κάθε κλάση και δηλώνουν τους τύπους δεδομένων και την υποχρεωτικότητα των τιμών (έργο σε εξέλιξη).⁶⁵
- Οι EDM Mapping Guidelines που περιέχουν οδηγίες για τον τρόπο με τον οποίο οι πάροχοι μπορούν να χαρτογραφήσουν τα δεδομένα τους σε EDM και να δώσουν έναν ορισμό ιδιοτήτων, πληροφορίες σχετικά με τους τύπους δεδομένων που μπορούν να χρησιμοποιηθούν ως τιμές.⁶⁶

Το EDM αφορά παρόχους περιεχομένου (βιβλιοθήκες, μουσεία, αρχεία), ειδικούς στον τομέα της πολιτιστικής κληρονομιάς που θέλουν να δώσουν το υλικό τους στην Europeana. Η σχετική διαδικτυακή πύλη έχει αναπτύξει εφαρμογές για επιμέρους δίκτυα και κοινότητες χρηστών όπως για επαγγελματίες του τουρισμού (Europeana Pro), εκπαιδευτικούς που αναπτύσσουν πόρους για μαθήματα (βλ. τον δικτυακό τόπο Teaching with Europeana blog,⁶⁷ πρωτοβουλία του European Schoolnet με σκοπό να ενθαρρύνει τους εκπαιδευτικούς να εντάξουν το υλικό της Europeana στην

⁶³ https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf (20/3/2024).

⁶⁴ https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.8_102017.pdf (20/3/2024).

⁶⁵ <https://github.com/europeana/corelib/wiki/EDMObjectTemplatesProviders> (20/3/2024).

⁶⁶ <https://europeana.atlassian.net/wiki/spaces/EF/pages/987791389/EDM+-+Mapping+guidelines> (20/3/2024).

⁶⁷ <https://teachwitheuropeana.eun.org/> (20/3/2024).

τάξη τους και να μοιραστούν την εμπειρία τους) και προγραμματιστές που χρησιμοποιούν το API (Application Programming Interface) ανοιχτού κώδικα για να φτιάξουν παιχνίδια.

3.10 General International Standard Archival Description ISAD(G)

Το ISAD(G) είναι ένα ευρέως χρησιμοποιούμενο πρότυπο περιγραφής αρχειακών συλλογών, το οποίο αναπτύχθηκε από το Διεθνές Συμβούλιο Αρχείων (International Council on Archives).⁶⁸ Το ISAD(G) περιλαμβάνει 26 στοιχεία τα οποία παρέχουν πληροφορίες για τα ακόλουθα χαρακτηριστικά συλλογών και αρχείων:

- Αναγνωριστικά στοιχεία όπως ονομασία, τίτλος, ημερομηνία και τόπος δημιουργίας της συλλογής.
- Περιγραφή του περιεχομένου της συλλογής, συμπεριλαμβανομένων θεμάτων και όρων πρόσβασης
- Δομή και κατανομή της συλλογής σε κατηγορίες ή υποκατηγορίες, για τη διευκόλυνση της οργάνωσης και της αναζήτησης.
- Λεπτομερής περιγραφή του περιεχομένου κάθε αρχείου ή ενότητας της συλλογής.
- Άλλες πληροφορίες που μπορεί να είναι χρήσιμες για την κατανόηση της συλλογής, όπως πληροφορίες σχετικά με τη δημιουργία, τη διατήρηση, ή τη χρήση των αρχείων.

Τα εν λόγω στοιχεία ταξινομούνται σε 7 ενότητες. Εξ αυτών, τα ακόλουθα στοιχεία θεωρούνται υποχρεωτικά κατά την περιγραφή συλλογών: Κωδικός αναφοράς (Reference code), Τίτλος (Title), Δημιουργός (Creator), Ημερομηνία (Date), Έκταση και μορφότυπος (Extent and medium) και επίπεδο περιγραφής (Level of description).

Το ISAD(G) είναι ένα σημαντικό εργαλείο για τους αρχειονόμους και τους ερευνητές που επιθυμούν να περιγράψουν αρχειακά υλικά, καθώς επιτρέπει τη δημιουργία συνεκτικών, εμπεριστατωμένων και ομοιογενών περιγραφών. Ως εκ τούτου, η αναθεωρημένη έκδοση του ISAD(G) έχει χαρτογραφηθεί ως προς μια σειρά άλλων ευρέως χρησιμοποιούμενων προτύπων (π.χ. Dublin Core) και χρησιμοποιείται από διεθνείς οργανισμούς όπως η UNESCO και η Παγκόσμια Τράπεζα.

3.11 ISLE Metadata Initiative (IMDI)

Ο σκοπός του IMDI (ISLE Metadata Initiative) που ανέπτυξε το Max Planck Institute for Psycholinguistics είναι να συγκροτήσει ένα σχήμα μεταδεδομένων για την περιγραφή πολυμεσικών και γλωσσικών πόρων, όπως ηχητικά και βιντεοσκοπημένα υλικά, μεταγραφές και σώματα

⁶⁸ ISAD: <https://www.ica.org/resource/isadg-general-international-standard-archival-description-second-edition/>
International Council on Archives: <https://www.ica.org/> (20/3/2024).

κειμένων.⁶⁹ Βασίζεται σε μια ιεραρχική περιγραφική δομή που διακρίνει κύρια (corpora) και επιμέρους δομικά στοιχεία (subcorpora): ένα corpus αποτελεί συλλογή γλωσσικών πόρων σε κειμενική ή πολυμεσική μορφή, οι οποίοι μπορεί να κατηγοριοποιηθούν σε μικρότερες συλλογές σύμφωνα με προκαθορισμένα κριτήρια – όπως κοινά γλωσσολογικά στοιχεία και θέματα. Η σχέση μεταξύ των δομικών στοιχείων αποδίδεται ως τμήμα της περιγραφής τόσο των ίδιων των στοιχείων, όσο και της “ανώτερης”, κατά περίπτωση, συλλογής – διευκολύνοντας με τον τρόπο αυτό την αναζήτηση και την περιήγηση στο περιεχόμενο εκτεταμένων συλλογών.

Συμπληρωματικά, το υλικό οργανώνεται και περιγράφεται σε συγκεκριμένες ενότητες (sessions), οι οποίες περιλαμβάνουν πηγές σχετικές με ένα συγκεκριμένο γλωσσολογικό ‘γεγονός’ (ενδεικτικά, οι ηχογραφήσεις, οι μεταγραφές, το φωτογραφικό υλικό και οι σημειώσεις που προκύπτουν από μια συνέντευξη πεδίου). Οι πληροφορίες σχετικά με τις διαθέσιμες ενότητες μιας συλλογής περιλαμβάνουν τον τίτλο και ένα μοναδικό αναγνωριστικό, το χωρικό και χρονικό στίγμα του γλωσσολογικού γεγονότος, τους συμμετέχοντες και τις θεματικές κ.ο.κ.⁷⁰

Το IMDI έχει υποστεί αρκετές εκδόσεις και βελτιώσεις και εξελίχθηκε σε ένα σύγχρονο και ευέλικτο σύνολο μεταδεδομένων που ανταποκρίνεται στις προκλήσεις της διατήρησης και του διαμοιρασμού ψηφιακών γλωσσικών πόρων. Το σχήμα υποστηρίζεται από ελεγχόμενα λεξιλόγια και μια σειρά εργαλείων (όπως τα Abril και IMDI Editor) που αναπτύσσει το Max Planck Institute for Psycholinguistics, ενώ είναι συμβατό με άλλα σχήματα μεταδεδομένων.⁷¹ Λόγω του σημασιολογικού εύρους που καλύπτει, το IMDI είναι συναφές και με επιστημονικά πεδία που βασίζονται σε έρευνα πεδίου και συλλογή μη κειμενικών πηγών. Το IMDI χρησιμοποιήθηκε για την τεκμηρίωση του γλωσσικού αρχείου του Ινστιτούτου, καθώς και σε ερευνητικά προγράμματα εθνογλωσσολογίας, όπως τα Spoken Dutch Corpus και Documentation of Endangered Languages.⁷²

3.12 MACHINE-READABLE CATALOGING (MARC)

Το MARC είναι ένα πρότυπο για τη συγκρότηση βιβλιοθηκονομικών μεταδεδομένων, σχεδιασμένο αρχικά ως τρόπος ανταλλαγής βιβλιογραφικών εγγραφών μεταξύ των καταλόγων βιβλιοθηκών.⁷³ Υπάρχουν διαφορετικές εκδόσεις του προτύπου, κυρίως με εθνική ή περιφερειακή εμβέλεια, από τις οποίες το MARC 21 είναι πιθανότατα αυτό που χρησιμοποιείται ευρέως.

Πρόκειται για ένα μορφότυπο δεδομένων που προέκυψε από την πρωτοβουλία της Βιβλιοθήκης του Κογκρέσου στις ΗΠΑ, η οποία ξεκίνησε πριν από σχεδόν σαράντα χρόνια. Παρέχει τον μηχανισμό με τον οποίο οι υπολογιστές ανταλλάσσουν, χρησιμοποιούν και ερμηνεύουν τις βιβλιογραφικές πληροφορίες. Τα στοιχεία δεδομένων του αποτελούν τη βάση των περισσότερων καταλόγων βιβλιοθηκών που χρησιμοποιούνται σήμερα.

⁶⁹ IMDI: <https://archive.mpi.nl/forums/t/imdi-metadata-information/2933>

Max Planck Institute for Psycholinguistics: <https://www.mpi.nl/> (20/3/2024).

⁷⁰ <https://archive.mpi.nl/forums/uploads/short-url/c07lIVbfaLy2LYCGFpToQnhvddS.pdf> (20/3/2024).

⁷¹ Abril: <https://www.mpi.nl/corpus/manuals/manual-arbil-imdi.pdf> (20/3/2024).

⁷² Language Archive at the Max Planck: <https://archive.mpi.nl/>
Spoken Dutch Corpus: <https://web.archive.org/web/20061003165337/http://lands.let.kun.nl/cgn/ehome.htm>
Documentation of Endangered Languages: <https://dobes.mpi.nl/> (20/3/2024)

⁷³ MARC Standards, <https://www.loc.gov/marc/> (20/3/2024).

Το MARC βασίζεται στη χρήση κωδικοποιημένων πληροφοριών που ομογενοποιούν τις βιβλιογραφικές εγγραφές και τις καθιστούν αναγνώσιμες από τους υπολογιστές, επιτρέποντας έτσι τη διαλειτουργικότητα και τον διαμοιρασμό. Τα πεδία πληροφορίας (fields) που καλύπτουν οι βιβλιογραφικές εγγραφές ορίζονται από κωδικούς πεδίου (tags) και δείκτες (indicators), ενώ τα ειδικότερα υποπεδία (subfields) από κωδικούς υποπεδίου (subfield codes).⁷⁴

Το πρότυπο MARC και τα παράγωγά του χρησιμοποιούνται ευρέως στην Ελλάδα από φορείς όπως η Εθνική Βιβλιοθήκη και οι πανεπιστημιακές βιβλιοθήκες. Έτσι διατίθενται οδηγοί χρήσης του MARC 21 στα ελληνικά,⁷⁵ αλλά και διοργανώνονται σεμινάρια για την εφαρμογή του προτύπου.⁷⁶

```
LDR 00626cam a22002053 4500
005 20120905083316.0
008 890802s1899 xx 000 0 eng d
0167 $a002622267$2Uk
0291 $aUNITY$b123402042
0291 $aUKRNI$b10162817
0291 $aUKRNI$b10162818
1001 $aNesbit, E.$q(Edith),$d1858-1924.
24514 $aThe Story of the Treasure Seekers. Being the adventures of the Bastable children in search of a fortune
260 $aT.F. Unwin:$bLondon,$c1899.
300 $axii. 296 p. ;$c8°.
```

Εικ. 8. Παράδειγμα καταχώρισης MARC, σε μορφή αναγνώσιμη από ανθρώπους. Πηγή: [Wikibooks](https://www.wikibooks.org/wiki/MARC_21)

Όπως και οι αρχειονόμοι, οι βιβλιοθηκονόμοι είναι από τις πιο εξοικειωμένες με ψηφιακές μορφές εργασίας κοινότητες στο πλαίσιο των ανθρωπιστικών σπουδών, πράγμα που επίσης αποτυπώθηκε στο πρώτο παραδοτέο του έργου.⁷⁷ Σε εκείνους, άλλωστε, οφείλεται ότι το MARC είναι έκτο στη σειρά στις ανθρωπιστικές επιστήμες και τέταρτο στις κοινωνικές επιστήμες ανάμεσα στα πρότυπα μεταδεδομένων που χρησιμοποιούνται περισσότερο.⁷⁸ Εκείνο, στο οποίο επιμένουν, είναι, ήδη πριν από την ανταπόκριση στις συνεχείς επικαιροποιήσεις των ψηφιακών εφαρμογών, η ανάγκη να γίνει κανόνας η συστηματοποίηση του ψηφιακού αποθέματος και η χάραξη πολιτικής για την ψηφιακή διατήρηση,⁷⁹ ζητήματα που άπτονται της γενίκευσης στη χρήση προτύπων όπως το MARC.

3.13 META-SHARE Metadata Schema

Η υποδομή META-SHARE παρέχει υπηρεσίες τεκμηρίωσης και απόθεσης γλωσσολογικών πόρων.⁸⁰ Το σχήμα META-SHARE⁸¹ αναπτύχθηκε στο πλαίσιο της υποδομής, με σκοπό να επενδύσει με μεταδεδομένα τις πηγές που φιλοξενούνται από αυτή, καθώς και τις εργασίες που οι ερευνητές εκτελούν μέσω της διεπαφής της. Τα δομικά στοιχεία του σχήματος αναφέρονται στο περιεχόμενο των γλωσσικών πόρων, στους μεταξύ τους συσχετισμούς, καθώς και στις ψηφιακές εφαρμογές που χρησιμοποιούνται για τη διαχείριση και την επεξεργασία των εν λόγω πόρων. Συνεπώς, η δομή και

⁷⁴ Understanding MARC, <https://www.loc.gov/marc/umb/> (20/3/2024): Καταλογογράφηση και καθιερωμένοι όροι MARC 21, https://www.lib.auth.gr/sites/default/files/docs_files/Koha_Cataloguing_MARC21.pdf (20/3/2024).

⁷⁵ Ελληνική μετάφραση MARC 21, <https://www.lib.aegean.gr/el/elliniki-metafrasi-marc-21> (20/3/2024).

⁷⁶ Διαδικτυακές συναντήσεις για την εφαρμογή των κανόνων του νέου RDA σε MARC, <https://www.nlg.gr/news/diadiktyakes-synantiseis-gia-tin-efarmogi-ton-kanonon-toy-neoy-rda-se-marc/> (20/3/2024).

⁷⁷ Ψηφιακό Τοπίο, σελ. 28-29.

⁷⁸ Ο.π., σελ. 38.

⁷⁹ Ο.π., σελ. 64.

⁸⁰ Ιστότοπος: <http://www.meta-share.org/> (20/3/2024).

⁸¹ Ιστότοπος: <http://www.meta-net.eu/meta-share/metadata-schema> (20/3/2024).

το περιεχόμενο του META-SHARE βασίζονται σε υπάρχοντα μοντέλα που χρησιμοποιούνται σε συναφή επιστημονικά πεδία, αντανακλώντας παράλληλα τις λειτουργικές απαιτήσεις των εκτεταμένων ψηφιακών υποδομών αναφορικά με την τεκμηρίωση, την ευρεσιμότητα και τη σημασιολογική συμβατότητα του διαθέσιμου υλικού.

Η βασική μονάδα αναφοράς του META-SHARE είναι ο “πόρος” (resource), ο οποίος αντιστοιχεί σε ομαδοποιήσεις πηγών (όπως, για παράδειγμα, κειμενικές πηγές ή αρχεία πολυμέσων), λεξικολογικών καταχωρήσεων, εργαλείων, κ.ο.κ. Η ταυτοποίηση των πόρων πραγματοποιείται μέσω της συνδυαστικής χρήσης περιγραφικών στοιχείων (elements), στη βάση δύο εννοιολογικών αξόνων που αποδίδουν τον τύπο του πόρου και το φέρον υπόστρωμά του (medium). Παράλληλα, αποτυπώνονται οι συσχετισμοί (relations) μεταξύ των τεκμηριωμένων πόρων.⁸²

Η συνδυαστική εφαρμογή των περιγραφικών στοιχείων και των συσχετισμών σε επίπεδο πόρων καθιστούν δυνατή την απόδοση των σταδίων του κύκλου ζωής των πόρων, καθώς και τη διασύνδεσή τους με οντότητες όπως, ενδεικτικά, πρόσωπα ή οργανισμοί που συμμετέχουν στην ανάπτυξή τους και άδειες χρήσης.

3.14 Midas Heritage

Το MIDAS Heritage είναι ένα πρότυπο μεταδεδομένων σχετικά με ιστορικούς τόπους και τοποθεσίες (historic sites).⁸³ Το πρότυπο, που έχει αναπτυχθεί στο Ηνωμένο Βασίλειο, επικεντρώνεται στην καταγραφή της πληροφορίας που υποστηρίζει την αποτελεσματική ανταλλαγή και τη μακροπρόθεσμη διατήρηση της γνώσης σχετικά με κτίρια, αρχαιολογικούς χώρους, ιστορικά ναυάγια, πάρκα και κήπους, πεδία μαχών, περιοχές πολιτιστικού ενδιαφέροντος αλλά και αντικείμενα. Ως εκ τούτου ανοίγεται στο ευρύτερο πεδίο της πολιτιστικής κληρονομιάς, το οποίο έχει ιδιαίτερη σημασία για χώρες όπως η Ελλάδα.

Το Midas Heritage προτείνει ένα ελάχιστο επίπεδο πληροφοριών που απαιτούνται για την καταγραφή των πόρων πολιτιστικής κληρονομιάς και καλύπτει τις διαδικασίες κατανόησης, προστασίας και διαχείρισης αυτών των πόρων με τρόπο που είναι συμβατός με αναλυτικότερα πρότυπα μεταδεδομένων, όπως π.χ. το CIDOC. Παρέχει επίσης κατευθυντήριες γραμμές για την υποστήριξη της αποτελεσματικής ανταλλαγής γνώσεων, της ανάκτησης και της μακροπρόθεσμης διατήρησης των δεδομένων. Χρησιμοποιείται από κυβερνητικούς οργανισμούς, τοπικές αρχές, εταιρείες διαχείρισης και συλλόγους πολιτιστικής κληρονομιάς, όπως και την ερευνητική κοινότητα.

Το πρότυπο οργανώνεται σε μια δέσμη θεμάτων (themes) που καλύπτουν την καταγραφή της πληροφορίας σχετικά με την πολιτιστική κληρονομιά (πόροι, δραστηριότητες, πηγές πληροφορίας, χωρική πληροφορία, χρονική πληροφορία, δράστες/ρόλοι). Τα θέματα αυτά αναλύονται στη συνέχεια σε ομάδες πληροφορίας (information groups) που καλύπτουν ειδικότερες θεματικές

⁸² Βλ. σχετικά: Gavrilidou, M., Labropoulou, Π. et al. The META-SHARE Metadata Schema for the Description of Language Resources. Proceedings of the Eighth International Conference on Language Resources and Evaluation, 1090-197. http://www.lrec-conf.org/proceedings/lrec2012/pdf/998_Paper.pdf (19/3/2024).

Τεκμηρίωση του σχήματος: <http://www.meta-net.eu/meta-share/META-SHARE%20%20documentationUserManual.pdf>

⁸³ Midas Heritage, <https://historicengland.org.uk/images-books/publications/midas-heritage/> (19/3/2024).

ενότητες ως προς την περιγραφή. Με αυτόν τον τρόπο χτίζεται ένα ιεραρχικό σύμπαν μεταδεδομένων που στοιχούν σε περιοχές πληροφορίας και στις σχέσεις μεταξύ τους.⁸⁴

Το Midas έχει αναπτυχθεί από την κοινοπραξία Forum on Information Standards in Heritage (FISH),⁸⁵ στην οποία συμμετέχουν πολλοί φορείς και οργανισμοί πολιτιστικής κληρονομιάς του Ηνωμένου Βασιλείου, όπως το Βρετανικό Μουσείο, το National Trust, το Εθνικό Μουσείο της Ουαλίας κλπ. Στόχος του FISH είναι να δώσει πρόσβαση σε πηγές πληροφορίας που υποστηρίζουν καλές πρακτικές στην πολιτιστική τεκμηρίωση, διατήρηση και διαχείριση με έμφαση στις αρχές FAIR.

Το Midas θα μπορούσε να σταθεί χρήσιμο σε δράσεις και έργα γενικής καταγραφής της υλικής πολιτιστικής κληρονομιάς (κυρίως του δομημένου περιβάλλοντος), καθώς παρέχει μια συνεκτική και εύχρηστη δέσμη πληροφοριών που είναι συμβατή με άλλα, λεπτομερέστερα και αναλυτικότερα σχήματα. Αυτή η πιο σφαιρική εικόνα θα μπορούσε να συμβάλει στην υποστήριξη στόχων όπως η ψηφιακή αφήγηση, η ανάδειξη σχέσεων και η ερμηνεία του πολιτισμικού πλούτου, ζητήματα που έχουν τεθεί από τους συμμετέχοντες στην έρευνα που οργανώθηκε στο πλαίσιο του πρώτου παραδοτέου του έργου.⁸⁶

3.15 Metadata Object Description Schema (MODS)

Το MODS είναι πρότυπο περιγραφικών μεταδεδομένων για βιβλιοθηκονομικές εγγραφές που εφαρμόζεται σε XML.⁸⁷ Αναπτύσσεται ως ένα υποσύνολο των στοιχείων του MARC, χρησιμοποιώντας λεκτικές αντί για αριθμητικές ετικέτες, σε μια προσπάθεια να αντιμετωπιστεί η πολυπλοκότητα του σχήματος MARC. Εισήχθη το 2002 από το Τμήμα Ανάπτυξης Δικτύων της Βιβλιοθήκης του Κογκρέσου και το Γραφείο Προτύπων MARC με τη συνδρομή ομάδας ενδιαφερόμενων ειδικών.

Η πιο πρόσφατη έκδοση του MODS είναι η 3.8. Προορίζεται πρωτίστως για τη δημιουργία πρωτότυπων εγγραφών, αν και μπορεί να αξιοποιηθεί για την τροποποίηση εγγραφών MODS που έχουν προκύψει από εγγραφές MARC 21 ή για χρήση στην ανάπτυξη λεπτομερών προδιαγραφών μετατροπής.⁸⁸

Το MODS χρησιμοποιεί στους κωδικούς πεδίου γενικότερες κατηγορίες από αυτές των εγγραφών MARC. Επιτρέπει μεν την αναπαράσταση δεδομένων που υπάρχουν ήδη σε συστήματα βασισμένα στο MARC, αλλά ορισμένα από αυτά τα δεδομένα παραλείπονται. Ωστόσο, το σχήμα δεν στοχεύει στη δυνατότητα εναλλαγής με το MARC 21. Μια αρχική εγγραφή MARC 21 που έχει μετατραπεί σε MODS μπορεί να μη μετατρέπεται αντιστρόφως σε MARC 21 στο σύνολό της χωρίς κάποιες απώλειες. Μερικές φορές τα δεδομένα δεν χάνονται, αλλά τοποθετούνται σε άλλα πεδία. Άλλοτε πάλι, όταν το συγκεκριμένο στοιχείο στο MARC δεν έχει αντίστοιχο στοιχείο στο MODS, τα στοιχεία

⁸⁴ Midas Heritage. The UK Historic Environment Data Standard, https://historicengland.org.uk/images-books/publications/midas-heritage/midas-heritage-2012-v1_1/ (20/3/2024).

⁸⁵ Forum on Information Standards in Heritage (FISH): Midas Heritage, <https://www.heritage-standards.org.uk/midas-heritage/> (20/3/2024).

⁸⁶ Ψηφιακό Τοπίο, σελ. 73.

⁸⁷ Metadata Object Description Schema, <https://www.loc.gov/standards/mods/> (20/3/2024).

⁸⁸ MODS 3.8 User Guidelines, <https://www.loc.gov/standards/mods/userguide/> (20/3/2024).

χάνονται κατά τη μετατροπή της εγγραφής στο MODS. Τα δύο πρότυπα δεν είναι λοιπόν απόλυτα συμβατά μεταξύ τους.

Αρκετοί φορείς, κυρίως στις ΗΠΑ, χρησιμοποιούν το πρότυπο MODS 3.8. Είναι ιδιαίτερα ενδιαφέρον ότι οι περισσότεροι, είτε είναι ψηφιακές βιβλιοθήκες είτε χρησιμοποιούν το πρότυπο στην οργάνωση ειδικά ψηφιακών βιβλιοθηκών.⁸⁹ Τούτο μπορεί να έχει ενδιαφέρον για την ανάπτυξη αντίστοιχων υπηρεσιών, οι οποίες άλλωστε, όπως φαίνεται από την έρευνα όπου στηρίχθηκε το πρώτο παραδοτέο του έργου, παρουσιάζουν ζήτηση από τους ερευνητές των ανθρωπιστικών και κοινωνικών επιστημών.⁹⁰

3.16 Open Archives Initiative / Object Reuse and Exchange (OAI-ORE)

Το Object Reuse and Exchange (ORE) αναπτύχθηκε από το Open Archives Initiative (OAI), μια διεθνή πρωτοβουλία για την προώθηση της ανοικτής πρόσβασης σε επιστημονικά έργα μέσω του πρωτοκόλλου OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).⁹¹

Η ανάπτυξη του OAI-ORE βασίζεται στη συνεργασία της OAI με ερευνητές και εμπειρογνώμονες από διάφορους τομείς - συμπεριλαμβανομένων των βιβλιοθηκονόμων, των επιστημόνων της πληροφορίας και των μηχανικών λογισμικού. Το ORE καλύπτει την ανάγκη εφαρμογής ενός κοινού πρωτοκόλλου που επιτρέπει τον διαμοιρασμό και την επανάχρηση σύνθετων ψηφιακών πηγών στο διαδίκτυο (όπως συλλογές αρχείων, επιστημονικές δημοσιεύσεις, εικόνες και πολυμέσα. Το OAI-ORE δημοσιεύθηκε το 2008, υπό τη μορφή προτύπου που περιγράφει τις αρχές και τις πρακτικές για τη δημιουργία και τη διαχείριση πολύπλοκων δομών δεδομένων στο διαδίκτυο. Το πεδίο χρήσης του έχει διευρυνθεί και πλέον εφαρμόζεται σε πλατφόρμες και αποθετήρια, υποστηρίζοντας τις διαδικασίες διαχείρισης σύνθετων ψηφιακών πηγών και δομών δεδομένων.

Το πρότυπο περιγραφής συγκροτείται από 4 δομικά στοιχεία:

- **Aggregation:** αναφέρεται σε ένα σύνολο ψηφιακών αντικειμένων που σχηματίζουν μια συλλογή με συγκεκριμένο σκοπό ή θέμα. Τα Aggregations υποστηρίζουν σύνθετες δομικές διατάξεις (για παράδειγμα ιεραρχικές ή/και επίπεδες δομές) και, κατ' επέκταση, τη διαχείριση σύνθετων δομών δεδομένων αναφορικά με ψηφιακές πηγές όπως δημοσιεύσεις, πολυμέσα και πρωτογενή δεδομένα.
- **Aggregated Resource:** οι μεμονωμένες πηγές που συναποτελούν το aggregation
- **Resource map (ReM):** το περιγραφικό μοντέλο του OAI-ORE βασίζεται στον εν λόγω πληροφοριακό "χάρτη" που διατίθεται σε μορφή XML ή RDF και παρέχει τις απαραίτητες σημασιολογικές πληροφορίες. Το συγκριτικό πλεονέκτημα της εν λόγω οντολογικής μονάδας είναι ότι επιτρέπει την τεκμηρίωση διαφορετικών μορφοτύπων ή/και τεκμηρίων που

⁸⁹ MODS Implementation Registry, <https://www.loc.gov/standards/mods/registry-search.php> (20/3/2024).

⁹⁰ Ψηφιακό Τοπίο, σελ. 32, 47, 50.

⁹¹ Object Reuse and Exchange: <https://www.openarchives.org/ore/>

Open Archives Initiative: <https://www.openarchives.org/>

Open Archives Initiative Protocol for Metadata Harvesting: <https://www.openarchives.org/pmh/> (20/3/2024).

βρίσκονται διάσπαρτα σε διαφορετικές ψηφιακές υποδομές, σε επίπεδο μεμονωμένου “αντικειμένου” (Aggregated Resource).

- Proxy: ένα εικονικό αντίγραφο των πηγών που περιλαμβάνονται στο Aggregation, το οποίο επιτρέπει την πρόσβαση και την αλληλεπίδραση με αντικείμενα και δομές δεδομένων που ανήκουν σε διαφορετικά περιβάλλοντα, λειτουργώντας ως σύνδεσμος ανάμεσα σε Aggregations ή Resource Maps

Το πρότυπο είναι εξαιρετικά χρήσιμο σε δραστηριότητες που σχετίζονται με την τεκμηρίωση, την οργάνωση και τη διάχυση ψηφιακών πηγών σε ποικιλία μορφοτύπων, καθώς παρέχει τον σημασιολογικό εμπλουτισμό των μεταδεδομένων και την ανταλλαγή συναφών πληροφοριών μεταξύ συστημάτων.

3.17 Open Language Archives Community [OLAC] standards

Τα OLAC standards (Open Language Archives Community) αποτελούν ένα ενιαίο σημασιολογικό πλαίσιο για την περιγραφή και την αναζήτηση γλωσσολογικών πόρων σε υποδομές και αποθετήρια.⁹² Το πρότυπο που περιλαμβάνουν χρησιμοποιεί τα 15 δομικά στοιχεία του Dublin Core, τα οποία αναπτύχθηκαν περαιτέρω, παράλληλα με τις ανάγκες της ερευνητικής κοινότητας που το OLAC εξυπηρετεί.⁹³ Πλέον, το OLAC περιλαμβάνει 23 στοιχεία και υποστηρίζει πιο εξειδικευμένες σημασιολογικές περιγραφές που εξυπηρετούν τη συμβατότητα και την ανταλλαγή πόρων μεταξύ των ψηφιακών υποδομών. Το OLAC παρέχει τη δυνατότητα προσαρμογής των στοιχείων του, σύμφωνα με τις ακόλουθες παραμέτρους:

- Προσαρμόσιμη περιγραφή των δεδομένων και δυνατότητα προσθήκης ή τροποποίησης πεδίων
- Εφαρμογή προτυποποιημένων όρων ή περιγραφών για την κωδικοποίηση τιμών σε ένα πεδίο μεταδεδομένων. Καθίσταται, έτσι, δυνατή η αναγνώριση του τύπου των δεδομένων που περιέχονται και η ομοιογενής περιγραφή των πόρων
- Δυνατότητα προσδιορισμού της γλώσσας στην οποία διατίθενται οι ψηφιακές πηγές και τα μεταδεδομένα.

Οι εν λόγω ιδιότητες του OLAC επιτρέπουν την παραμετροποιημένη αναζήτηση περιεχομένου, καθώς και την αυτοματοποιημένη ανάκτηση και επεξεργασία γλωσσικών πόρων.

3.18 Qualitative Data Exchange Schema (QuDEX)

Το QuDEX (Qualitative Data Exchange Schema) είναι ένα σχήμα ανταλλαγής ποιοτικών δεδομένων που επιτρέπει στους χρήστες να αναζητούν, να εντοπίζουν, να ανακτούν και να παραθέτουν

⁹² Βλ. σχετικά: <http://www.language-archives.org/documents.html#Standards> (20/3/2024).

⁹³ Βλ. σχετικά: <http://www.language-archives.org/OLAC/metadata.html> (20/3/2024).

σύνθετες συλλογές δεδομένων.⁹⁴ Το σχήμα έχει αναπτυχθεί από το Αρχείο Δεδομένων του Ηνωμένου Βασιλείου (UK Data Archive)⁹⁵ ως συμπληρωματικό εργαλείο του DDI και παρέχει τις εξής δυνατότητες:

- Τη συνεπή σήμανση και δόμηση δεδομένων,
- Τη συγκρότηση πλούσιων περιγραφικών μεταδεδομένων (π.χ. χαρακτηριστικά συνέντευξης, περιβάλλον συνέντευξης),
- Τη δημιουργία λογικών συνδέσμων μεταξύ αντικειμένων (μεταξύ κειμένου, ήχου, εικόνας και άλλων ερευνητικών παραγώγων),
- Τη διατήρηση των παραπομπών σε επισημειώσεις δεδομένων, και
- Την ανάπτυξη κοινών στοιχείων μεταδεδομένων που επιτρέπουν τη δημιουργία λειτουργικών καταλόγων μεταξύ διαφορετικών παρόχων και διεθνών φορέων.

Το QuDEx αξιοποιείται στο QualiBank της Υπηρεσίας Δεδομένων του Ηνωμένου Βασιλείου (UK Data Service), ένα διαδικτυακό εργαλείο για την αναζήτηση, τη φυλλομέτρηση και την παράθεση ποιοτικών δεδομένων.⁹⁶ Η χρήση του QuDEx καθιστά δυνατή την αναζήτηση όχι μόνο μέσω περιλήψεων, τίτλων και λέξεων-κλειδίων, αλλά μέσω των ίδιων των δεδομένων (όροι και φράσεις που περιλαμβάνονται στα δεδομένα και προέρχονται είτε από τους συμμετέχοντες είτε από τους ερευνητές). Η βάση δεδομένων του QualiBank περιέχει ένα ευρύ φάσμα ποιοτικών δεδομένων, συμπεριλαμβανομένων συνεντεύξεων, μελετών, απαντήσεων σε ερωτήσεις ανοικτού τύπου και αναφορών. Η γνωριμία με το εργαλείο αυτό, που είναι προσβάσιμο διαδικτυακά, μπορεί να παράσχει βοήθεια στους ερευνητές των κοινωνικών επιστημών σχετικά με τις δυνατότητες που προσφέρει το QuDEx, ιδιαίτερα ως προς τη διαλειτουργικότητα, τον διαμοιρασμό και την επανάχρηση των δεδομένων, αλλά και, πράγμα που επισημάνθηκε στο πρώτο παραδοτέο του έργου, ως προς την “παρουσίαση [των δεδομένων και των αποτελεσμάτων της έρευνας] με τρόπο εύληπτο και απλό στο ευρύτερο κοινό”.⁹⁷ Το σχήμα, πάντως, μπορεί να χρησιμοποιηθεί και ευρύτερα από τους επιστήμονες των ανθρωπιστικών σπουδών, στον βαθμό που είναι δυνατόν να εφαρμοστεί σε αντίστοιχα σύνολα δεδομένων (π.χ. κειμενικά και οπτικά ιστορικά τεκμήρια).

3.19 Statistical Data and Metadata eXchange (SDMX)

Το SDMX (Statistical Data and Metadata eXchange)⁹⁸ είναι ένα διεθνές πρότυπο που περιγράφει στατιστικά δεδομένα και μεταδεδομένα και βελτιώνει την ανταλλαγή και τον διαμοιρασμό τους μεταξύ στατιστικών και επιστημονικών φορέων.

⁹⁴ QuDEx, <https://www.data-archive.ac.uk/managing-data/standards-and-procedures/metadata-standards/qudex/> (20/3/2024).

⁹⁵ UK Data Archive, <https://www.data-archive.ac.uk/> (20/3/2024).

⁹⁶ Searching QualiBank, <https://ukdataservice.ac.uk/help/searching-data/searching-qualibank/> (20/3/2024).

⁹⁷ Ψηφιακό Τοπίο, σελ. 72.

⁹⁸ The official site for the SDMX community, <https://sdmx.org/> (12/3/2024).

Το SDMX υποστηρίζεται από οκτώ διεθνείς οργανισμούς, μεταξύ των οποίων η Ευρωπαϊκή Κεντρική Τράπεζα (ΕΚΤ), η Eurostat (Στατιστική Υπηρεσία της Ευρωπαϊκής Ένωσης), η Διεθνής Οργάνωση Εργασίας (ILO), το Διεθνές Νομισματικό Ταμείο (IMF), ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (OECD), η Στατιστική Διεύθυνση των Ηνωμένων Εθνών (UNSD) και η Παγκόσμια Τράπεζα.

Το πρότυπο SDMX εφαρμόζεται σε τρία γενικά πεδία στατιστικής:

- Δημογραφία και κοινωνία (Demographic and social statistics)
- Οικονομία (Economic statistics)
- Περιβάλλον και πολλαπλοί τομείς (Environment and multi-domain statistics).

Τα τρία αυτά πεδία διακλαδώνονται θεματικά σε μια μεγάλη σειρά ειδικότερων τομέων, οι οποίοι καλύπτουν μεγάλο φάσμα της κοινωνικής και οικονομικής δραστηριότητας. Στο τρίτο πεδίο, για παράδειγμα, εντάσσονται ειδικότεροι τομείς όπως οι συνθήκες διαβίωσης, η φτώχεια, το κοινωνικό φύλο, η βιώσιμη ανάπτυξη και η επιχειρηματικότητα.⁹⁹ Είναι σαφές πως το πρότυπο υπηρετεί την επεξεργασία και τον διαμοιρασμό στατιστικών μεταδεδομένων τα οποία σχετίζονται με τις πολιτικές και ερευνητικές πρωτοβουλίες των διεθνών οργανισμών που το υποστηρίζουν.

Το πρότυπο συνοδεύεται από ένα σετ οδηγιών (guidelines) που συνιστούν καλές πρακτικές για τη δημιουργία διαλειτουργικών συνόλων δεδομένων και μεταδεδομένων, όπως και από έναν ευρύ κατάλογο εργαλείων λογισμικού που έχουν αναπτυχθεί από οργανισμούς που συμμετέχουν στην πρωτοβουλία SDMX. Ο χρήστης μπορεί να περιηγηθεί στις εφαρμογές αυτές ανάλογα με το ιδιαίτερο ενδιαφέρον του. Παρέχεται επιπλέον ένας επίσης ευρύς κατάλογος εκπαιδευτικών εφαρμογών που προσφέρονται από χορηγούς και μέλη της ευρύτερης κοινότητας που λαμβάνει μέρος στην επεξεργασία του SDMX. Τέλος, η τεχνική ομάδα εργασίας (Technical Working Group) διατηρεί ένα φόρουμ για την επικοινωνία και τη συμμετοχή των μελών της κοινότητας SDMX στη διαχείριση και ανάπτυξη των ειδικότερων τεχνικών προτύπων.

Οι στόχοι του SDMX ανταποκρίνονται στα ζητούμενα των κοινωνικών επιστημόνων που, όπως αποτυπώθηκε στο πρώτο παραδοτέο του έργου, χρησιμοποιούν λογισμικά και ελεύθερα εργαλεία στατιστικής ανάλυσης.¹⁰⁰ Στο πλαίσιο αυτό, μπορεί να συμβάλει στην ανάπτυξη μεταδεδομένων που επιτρέπουν τη διαλειτουργικότητα και τον διαμοιρασμό των ευρημάτων.

3.20 Text Encoding Initiative (TEI)

Το TEI (Text Encoding Initiative) είναι ένα διεθνές και ανοιχτό πρότυπο για την αναπαράσταση κειμένων σε ηλεκτρονική μορφή, ευρέως διαδεδομένο στον χώρο των ψηφιακών ανθρωπιστικών επιστημών. Ειδικότερα το TEI παρέχει ένα σύνολο ετικετών (tag sets) και οδηγιών (guidelines), οι οποίες εκφράζονται σε XML, για την κωδικοποίηση κειμένων (text encoding), δηλαδή για την αναπαράσταση πληροφοριών σχετικά με τα κείμενα, οι οποίες ενσωματώνονται σε αυτά και μπορούν να διαβαστούν από τις μηχανές. Στόχος του προτύπου είναι να δημιουργήσει ένα

⁹⁹ SDMX list of statistical subject-matter domains, https://sdmx.org/wp-content/uploads/03_sdmx_cog_annex_3_smd_2009.pdf (7/3/2024).

¹⁰⁰ Ψηφιακό Τοπίο, σελ. 36, 71.

περιβάλλον για την κωδικοποίηση κειμένων κάθε είδους γραπτού λόγου (πεζός λόγος, ποίηση, δράμα), οποιασδήποτε φυσικής γλώσσας και χρονικής περιόδου.

Από το 1994, οι Οδηγίες του TEI (Guidelines)¹⁰¹ χρησιμοποιούνται από βιβλιοθήκες, μουσεία, εκδοτικούς οίκους, ακαδημαϊκά ιδρύματα, προγράμματα¹⁰² και μεμονωμένους ερευνητές σε όλο τον κόσμο για την επισημείωση κειμενικών πηγών με στόχο την έρευνα, τον διαμοιρασμό και τη διδασκαλία. Στην Ελλάδα χαρακτηριστικό παράδειγμα χρήσης του είναι η Ανεμόσκαλα,¹⁰³ το ψηφιακό corpus νεοελληνικής ποίησης του 19ου και 20ού αιώνα.

Το πρότυπο αναπτύσσεται και υποστηρίζεται από το TEI Consortium,¹⁰⁴ ένα δίκτυο ερευνητών που δραστηριοποιείται κυρίως στον χώρο των ανθρωπιστικών επιστημών.

Βασική δομή

Καθώς οι Οδηγίες του TEI φιλοδοξούν να παράσχουν ένα πλαίσιο για την κωδικοποίηση (θεωρητικά) οποιουδήποτε είδους κειμένου, το σύνολο ετικετών που περιλαμβάνεται στην τελευταία έκδοση (P5, Version 4.7.0)¹⁰⁵ του προτύπου ανέρχεται στα σχεδόν 500 XML στοιχεία (<sentence>, <word>, <person> κτλ.) επιτρέποντας μια εξαιρετικά λεπτομερή κωδικοποίηση οποιουδήποτε κειμένου. Ωστόσο, τα στοιχεία του TEI οργανώνονται σε ένα σύνολο 21 ενοτήτων (modules),¹⁰⁶ οι οποίες περιλαμβάνουν τους ορισμούς των στοιχείων που ανήκουν σε αυτές, έτσι ώστε, ανάλογα με το είδος κειμένου που έχει κανείς να κωδικοποιήσει, να επιλέγει τα στοιχεία της αντίστοιχης ενότητας, όπως, π.χ., την ενότητα *verse* για την κωδικοποίηση κειμένων ποίησης, ή να συνδυάζει στοιχεία από διαφορετικές ενότητες σε περίπτωση που το υλικό του το απαιτεί (π.χ. στην κωδικοποίηση κειμένων μιας ανθολογίας, με έργα σε πεζό αλλά και έμμετρο λόγο). Κάθε στοιχείο ανήκει, επίσης, σε μια ή περισσότερες κλάσεις, ενώ ανάλογα με τις κλάσεις που ανήκει, το περιεχόμενο και τα γνωρίσματά του μπορεί να είναι διαφορετικά. Αυτός ο δισδιάστατος χώρος ενοτήτων – κλάσεων των στοιχείων καθιστά το πρότυπο ιδιαίτερα ευέλικτο.¹⁰⁷

¹⁰¹ <https://tei-c.org/guidelines/> (20/3/2024).

¹⁰² <https://tei-c.org/Activities/Projects/> (20/3/2024).

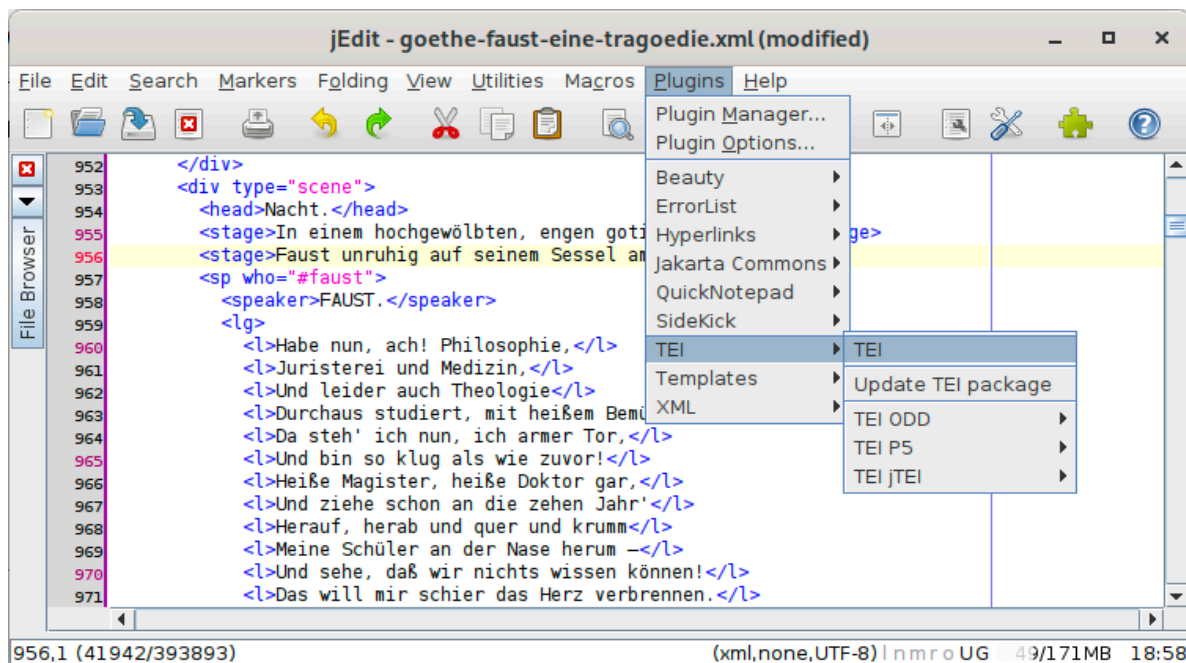
¹⁰³ <https://www.greek-language.gr/Resources/literature/tools/concordance/> (20/3/2024).

¹⁰⁴ <https://tei-c.org/> (20/3/2024).

¹⁰⁵ <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> (20/3/2024).

¹⁰⁶ <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>, σ.56 (20/3/2024).

¹⁰⁷ <https://eclass.aueb.gr/modules/document/file.php/INF391/lectures/tei%20intro.pdf> (20/3/2024).



Εικ. 9. Εφαρμογή του TEI στον Φάουστ του Γκαίτε. Πηγή: [DARIAH-EU](https://dariah.eu)

Στην πράξη, ωστόσο, οι περισσότεροι χρήστες του TEI χρησιμοποιούν συνήθως ένα πολύ μικρότερο υποσύνολο των ετικετών που διαθέτει το πρότυπο. Για τον λόγο αυτόν το TEI Consortium συστήνει μια απλοποιημένη, παραμετροποιημένη έκδοσή του TEI, το TEI Lite.¹⁰⁸ Το TEI Lite περιλαμβάνεται σε ένα αρχείο και έχει ήδη αξιοποιηθεί από μεγάλο αριθμό φορέων και έργων.

Αξίζει να αναφερθεί ότι με βάση τις αρχές του TEI και τις κατάλληλες παραμετροποιήσεις έχουν αναπτυχθεί από την κοινότητα του TEI υποκατηγορίες σχημάτων με βάση τις αρχές του, αλλά με πιο ειδική στόχευση, όπως:

- Το EpiDoc (Epigraphic Documents in TEI XML),¹⁰⁹ το οποίο παρέχει οδηγίες και εργαλεία για την κωδικοποίηση εκδόσεων αρχαίων κειμένων. Χρησιμοποιεί ένα υποσύνολο του προτύπου TEI για την αναπαράσταση κειμένων σε ψηφιακή μορφή και αναπτύχθηκε αρχικά για τη δημοσίευση ψηφιακών εκδόσεων αρχαίων επιγραφικών κειμένων (π.χ. Επιγραφές της Αφροδισιάδας, Πινακίδες της Βιντολάντα). Το πεδίο του έχει διευρυνθεί και περιλαμβάνει τη δημοσίευση παπύρων και χειρογράφων (π.χ. Papyri.info). Η τελευταία του έκδοση του προτύπου και των οδηγιών του είναι η 9.6¹¹⁰ (Μάρτιος 2024).
- TEI με μουσική σημειογραφία: Παραμετροποίηση¹¹¹ που επιτρέπει τη χρήση στοιχείων από το MEI (Music Encoding Initiative) για την κωδικοποίηση μουσικών κειμένων σε μηχαναγνώσιμη δομή.

Πλεονεκτήματα και μειονεκτήματα στη χρήση του

¹⁰⁸ <https://tei-c.org/guidelines/customization/lite/> (20/3/2024).

¹⁰⁹ <https://epidoc.stoa.org/> (20/3/2024).

¹¹⁰ <https://sourceforge.net/p/epidoc/wiki/LatestRelease/> (20/3/2024).

¹¹¹ <https://github.com/music-encoding/music-encoding/tree/develop/source/modules>

Το TEI αποτελεί από τα πιο δημοφιλή πρότυπα στις ανθρωπιστικές σπουδές, κυρίως στον χώρο των φιλολογικών/γλωσσολογικών μελετών, και αυτό οφείλεται στο ότι αποτελεί μια πολύ αναλυτική γλώσσα κωδικοποίησης για την υποστήριξη σήμανσης περίπλοκων κειμένων. Αυτό το καθιστά ένα πολύτιμο εργαλείο κωδικοποίησης των κειμένων εις βάθος, καθώς ο πλούτος της πληροφορίας που μπορεί να αποτυπωθεί είναι εντυπωσιακός.

Επίσης είναι ένα πρότυπο που επιδέχεται παραμετροποίηση και επέκταση, με σκοπό να ανταποκριθεί στις ιδιαίτερες ανάγκες περιγραφής των εκάστοτε τεκμηρίων. Το TEI Consortium - πέρα από την παραμετροποιημένη, συντομευμένη, εκδοχή του προτύπου που έχει αναπτύξει, το TEI Lite- διαθέτει εργαλεία λογισμικού για να απλοποιήσει το έργο της παραμετροποίησης από τους ίδιους τους χρήστες. Το εργαλείο παραμετροποίησης για την έκδοση P5 είναι το Roma.¹¹²

Ωστόσο, η δυνατότητα λεπτομερούς περιγραφής που παρέχεται από το πρότυπο στην εκτεταμένη του μορφή έχει και κάποια μειονεκτήματα: απαιτείται ειδική εκπαίδευση και χρόνος για την εκμάθησή του, όπως και εξοικείωση με τη γλώσσα XML. Επιπλέον, η λεπτομερής κωδικοποίηση/σήμανση δυσχεραίνει τη διαλειτουργικότητα και τις αντιστοιχίσεις / γενικεύσεις της σηματοδότησης που αποδίδεται στην πληροφορία.

Από την άλλη, για την κωδικοποίηση απλών κειμένων, χωρίς ιδιαίτερες απαιτήσεις, αρκεί η ενασχόληση κάποιων ημερών για την εξοικείωση με το TEI. Εκπαιδευτικό υλικό παρέχεται και στην ιστοσελίδα του οργανισμού, όπου ανά διαστήματα ανακοινώνονται σεμινάρια και εργαστήρια εκμάθησης του προτύπου.¹¹³

Τέλος ένα από πλεονεκτήματα του TEI είναι ότι διαθέτει μια μεγάλη κοινότητα χρηστών παγκοσμίως που λειτουργεί υποστηρικτικά και συμβουλευτικά σε προβλήματα και νέους χρήστες μέσω τακτικών συναντήσεων, ομάδων εργασίας, που εστιάζονται σε συγκεκριμένα προβλήματα, και ενός forum συζήτησης¹¹⁴ που διατηρεί στην ιστοσελίδα του οργανισμού.

Αναφορά στο παραδοτέο ΠΕ1

Το TEI είναι από τα πιο διαδεδομένα πρότυπα και στην Ελλάδα, όπως κατέδειξε η έρευνα που διεξήχθη για την αποτύπωση των τάσεων στη χρήση ψηφιακών μεθόδων, στο πλαίσιο του πρώτου παραδοτέου του έργου.¹¹⁵ Σύμφωνα με την έρευνα, το TEI είναι τρίτο στη σειρά στις ανθρωπιστικές επιστήμες ανάμεσα στα πρότυπα μεταδεδομένων που χρησιμοποιούνται περισσότερο. (βλ. σ. 38). Εκείνο, ωστόσο που επισήμαναν οι συμμετέχοντες στην έρευνα, ειδικότερα στη συνάντηση με την ομάδα εστίασης φιλολόγων/γλωσσολόγων, είναι ότι υπάρχει ανάγκη εκπαίδευσης στο πρότυπο και στις δυνατότητες που παρέχει, μέσω της διοργάνωσης σεμιναρίων και εργαστηρίων πρακτικής άσκησης.

¹¹² <https://roma2.tei-c.org/> (20/3/2024).

¹¹³ <https://tei-c.org/support/learn/> (20/3/2024).

¹¹⁴ <https://tei-c.org/support/#tei-list>

¹¹⁵ Ψηφιακό Τοπίο, σ. 38

4. ΣΥΝΑΦΕΙΑ ΜΕ ΣΤΑΔΙΑ ΚΑΙ ΤΥΠΟΥΣ ΔΡΑΣΤΗΡΙΟΤΗΤΩΝ

Ένας από τους βασικούς στόχους της παρούσας εργασίας είναι να προσδιοριστεί η συνάφεια των προαναφερθέντων σχημάτων και προτύπων μεταδεδομένων με τις ερευνητικές δραστηριότητες στον χώρο των ψηφιακών ανθρωπιστικών επιστημών, όπως αυτές ταυτοποιήθηκαν κατά την ανάλυση των στοιχείων που προέκυψαν από το Ερωτηματολόγιο και τις Ομάδες Εστίασης που παρουσιάζονται στο παραδοτέο ΠΕ1. Οι δραστηριότητες αυτές αναπαραστάθηκαν κατόπιν οντολογικά στο Παραδοτέο ΠΕ2 και είναι οι εξής: Αναζήτηση (Seeking), Πρόσκτηση (Acquiring), Επεξεργασία (Processing) και Επικοινωνία (Communicating). Οι δραστηριότητες αυτές λειτουργούν ως υπερκείμενοι όροι κάτω από τους οποίους τοποθετούνται ιεραρχικά εξειδικεύσεις τους, όπως παρουσιάζεται στην ενότητα 1 της παρούσας μελέτης. Με βάση αυτό το οντολογικό πλαίσιο, θα προσδιοριστεί η συνάφεια των προτύπων μεταδεδομένων με συγκεκριμένους τύπους ψηφιακών εργασιών/δραστηριοτήτων.

Θα πρέπει να σημειωθεί ότι η εφαρμογή προτύπων αφορά αποκλειστικά τη δραστηριότητα της Επεξεργασίας (Processing), καθώς και τις υποκατηγορίες δραστηριοτήτων που ταξινομούνται υπό αυτήν, κυρίως την Οργάνωση (Organising) και την Ανάλυση (Analysing). Στο στάδιο της επεξεργασίας και διαχείρισης των ψηφιακών πόρων η εφαρμογή καθιερωμένων προτύπων και σχημάτων μεταδεδομένων είναι κομβικής σημασίας, καθώς η δομημένη και συστηματική περιγραφή των δεδομένων εξυπηρετεί την οργάνωση, την ευρεσιμότητα και τη διαλειτουργικότητά τους με άλλα σύνολα δεδομένων. Από αυτό συνεπάγεται ότι η εφαρμογή προτύπων στο στάδιο της Επεξεργασίας έχει αντίκτυπο σε όλες σχεδόν τις δραστηριότητες που αποτυπώνονται στο οντολογικό πλαίσιο αναπαράστασης των ερευνητικών δραστηριοτήτων. Ενδεικτικά, η χρήση προτύπων διευκολύνει την Αναζήτηση (Seeking) και τον Εντοπισμό (Discovering) της πληροφορίας, βοηθά την εξόρυξη δεδομένων (data mining) από μεγάλα σώματα κειμένων, ενώ παράλληλα προάγει, όπως προαναφέρθηκε, τη διαλειτουργικότητα επιτρέποντας τον απρόσκοπτο διαμοιρασμό της πληροφορίας (Disseminating).

Ως προς τον διαμοιρασμό, ο οποίος εντάσσεται στον τύπο δραστηριότητας “Επικοινωνία”, θα γίνει ιδιαίτερη αναφορά στο τέλος του υποκεφαλαίου, αφού πρώτα επιχειρηθεί ένας συνοπτικός συσχετισμός των προτύπων και σχημάτων μεταδεδομένων με τις πέντε εξειδικεύσεις του τύπου δραστηριότητας “Επεξεργασία” (Processing), όπως προκύπτουν από την ανάλυση στο παραδοτέο 2.1 του έργου: Οργάνωση (Organizing), Ανάλυση (Analyzing), Τροποποίηση (Modifying), Παραγωγή (Producing), Διατήρηση (Preserving). Για τον συσχετισμό με καθεμία από τις πέντε αυτές κατηγορίες λαμβάνονται υπόψη επίσης οι περαιτέρω εξειδικεύσεις τους, οι οποίες δηλώνονται με βέλη.

Οργάνωση → Διαχείριση (Managing), Δόμηση (Structuring)

Η ετερογένεια των δεδομένων αποτελεί βασική πρόκληση για την αποτελεσματική οργάνωσή τους. Σκόπιμη είναι η ένταξη της τελευταίας σε έναν συνολικό σχεδιασμό για την κατάστρωση ψηφιακής πολιτικής εκ μέρους του φορέα, θεσμού ή και ερευνητικής ομάδας. Η διαθεσιμότητα των σχημάτων και προτύπων μεταδεδομένων που αφορούν ποικίλους πόρους και επιστημονικές περιοχές παρέχει δυνατότητες επιλογής και δικτύωσης για τη επίτευξη της ψηφιακής οργάνωσης, δόμησης και διαχείρισης σε ευθυγράμμιση με αναγνωρισμένες καλές πρακτικές, ιδιαίτερα στην κατευθυνση της διαλειτουργικότητας, η οποία με τη σειρά της διευκολύνει πρακτικές ανταλλαγής, διαμοιρασμού και επανάχρησης. Η δόμηση περιλαμβάνει εργασίες που σχετίζονται με τον εννοιολογικό σχεδιασμό όπως η συλλογή και ανάλυση των απαιτήσεων από τους χρήστες, ο σχεδιασμός (ορισμός των τύπων

οντοτήτων και τύπων συσχετίσεων), και η επιλογή του κατάλληλου προτύπου/σχήματος σε περίπτωση υιοθέτησης κάποιου από τα καθιερωμένα πρότυπα.

Ανάλυση → Επισημείωση (Annotating)

Η χρήση σχημάτων μεταδεδομένων παρέχει δυνατότητες για τη μηχανική εκτέλεση εργασιών ανάλυσης των δεδομένων. Τούτο εξυπηρετείται τόσο από την ομαδοποίηση και την ταξινόμηση των περιοχών πληροφορίας των δεδομένων, πράγμα που επιτρέπει την αναλυτική περιγραφή τους, όσο και από τις δυνατότητες που παρέχονται για ανάλυση του περιεχομένου. Τούτο γίνεται εμφανές στα σχήματα μεταδεδομένων για δεδομένα των κοινωνικών επιστημών (DDI, CMM, QuDEX) και, κυρίως, για κειμενικές πηγές.

Η επισημείωση αφορά την αναπαράσταση πληροφοριών σχετικά με τα κείμενα, σε μηχαναγνώσιμη μορφή. Πολύ συχνά ο όρος χρησιμοποιείται για να δηλώσει την περιγραφή και σήμανση δομικών και σημασιολογικών στοιχείων, διευκολύνοντας έτσι τη μηχανική ανάλυση σε σώματα κειμένων. Σχετικά σχήματα μεταδεδομένων είναι τα CMDI, META-SHARE, OLAC, TEI. Ας σημειωθεί εδώ ότι η διαδικασία αυτή επιτρέπει εξαγωγές συμπερασμάτων που είναι πολύ χρήσιμα για τη γλωσσολογία και άλλες ανθρωπιστικές και κοινωνικές επιστήμες -όπως η αρχαιολογία (επιγραφική) και η φιλολογία- που ερευνούν την ανθρώπινη δράση και συμπεριφορά μέσα από γλωσσικές πηγές και αποθέματα.

Τροποποίηση → Εμπλουτισμός (Enriching), Οπτικοποίηση (Visualizing)

Κατά τον εμπλουτισμό των δεδομένων προστίθεται πληροφορία σε ένα αντικείμενο, σχετικά με την προέλευση, τη φύση, τη δομή ή τα στοιχεία του. Ο εμπλουτισμός της περιεχόμενης πληροφορίας γίνεται με παραγωγή μεταδεδομένων, μέσω της διασύνδεσης στοιχείων μεταδεδομένων με πηγές δεδομένων ή/και λεξιλόγια, θησαυρούς όρων και καταλόγους καθιερωμένων ονομάτων. Αυξάνει σημαντικά την ακρίβεια και την ποιότητα των μεταδεδομένων. Σχήματα που κατεχοχύν υποστηρίζουν τη διαδικασία του εμπλουτισμού είναι οι οντολογίες με εστίαση στη σημασιολογία των δομών δεδομένων όπως το CIDOC CRM (αρχικά για την τεκμηρίωση εννοιών και πληροφοριών για την πολιτιστική κληρονομιά αλλά πλέον με επεκτάσεις σε μία σειρά εξειδικευμένων επιστημονικών πεδίων) και ο μετα-θησαυρός Backbone με σαφή οντολογική προσέγγιση που αντλεί από το CIDOC αλλά συνομιλεί με τις έννοιες ανώτερου επιπέδου και όρους κορυφής των ιεραρχιών καθιερωμένων θησαυρών όπως ο AAT του Ινστιτούτου Getty. Υποστηρίζει τη σημασιολογική διασύνδεση επί μέρους θησαυρών και δομημένων λεξιλογίων.

Η οπτικοποίηση αφορά την οπτική αναπαράσταση δεδομένων και αποτελεσμάτων της ανάλυσης (γεωχωρική αναπαράσταση, χρονική ανάπτυξη, διαγράμματα, πίνακες κλπ.). Τα σχήματα μεταδεδομένων που την υποστηρίζουν ιδιαίτερα είναι, πέρα από τα πρότυπα που αφορούν γεωχωρικά δεδομένα, οι επεκτάσεις του CIDOC (CRMgeo, CRMarchaeo, CRMba, CRMdig), οι οποίες λαμβάνουν υπόψη τις ειδικότερες ανάγκες των ανθρωπιστικών επιστημών, όπως και τα σχήματα που επεξεργάζονται ποσοτικά δεδομένα των κοινωνικών επιστημών (DDI, CMM, SDMX).

Παραγωγή → Δημιουργία (Creating), Ψηφιακή αφήγηση (Digital Storytelling)

Η παραγωγή νέων ψηφιακών δεδομένων βασίζεται στην αξιοποίηση του υλικού που έχει συλλεγεί και οργανωθεί δομημένα για την ανάπτυξη επιπλέον περιεχομένου που προκύπτει από τις επιμέρους διασυνδέσεις. Σχετικά σχήματα μεταδεδομένων είναι εκείνα που αφορούν την οπτικοποίηση, αλλά και εκείνα που μπορούν να υποστηρίξουν αποτελεσματικά την ανάπτυξη ψηφιακής αφήγησης (EDM, Midas, OAI-ORE).

Διατήρηση → Επιμέλεια (Curating)

Η διατήρηση των δεδομένων υπηρετείται από όλα τα σχήματα και πρότυπα μεταδεδομένων. Ως εξειδίκευση της διατήρησης, η επιμέλεια είναι ιδιαίτερα σημαντικά στις περιπτώσεις συλλογών μεγάλης σημασίας για την έρευνα, την κοινωνική αναγνωρισιμότητα και την πολιτιστική κληρονομιά. Πρόκειται για συλλογές όπως εκείνες των βιβλιοθηκών, αρχείων και μουσείων, αλλά και για συσσωρευτές και αποθετήρια υλικού που έχουν παραχθεί είτε από ερευνητική εργασία ή από συμβολές επιμέρους φορέων κληρονομιάς και περιλαμβάνουν φυσικούς ή/και ψηφιακούς/ψηφιοποιημένους πόρους. Τα σχετικά σχήματα μεταδεδομένων είναι αυτά που επεξεργάζονται δεδομένα αρχείων, βιβλιοθηκών και μεγάλων αποθετηρίων (EAD, EDM, ISAD, MARC, MODS, OAI-ORE).

Πριν κλείσει αυτή η συνοπτική αντιστοίχιση σχημάτων μεταδεδομένων και τύπων δραστηριοτήτων, θα γίνει μια σύντομη αναφορά στον τύπο δραστηριότητας “Επικοινωνία” (Communicating) και στις εξειδικεύσεις του.

Επικοινωνία → Διάδοση (Disseminating) → Διαμοιρασμός (Sharing), Κοινοποίηση (Notifying)

Είτε πρόκειται για σχήματα μεταδεδομένων που αφορούν συγκεκριμένα είδη και κατηγορίες πόρων, είτε -και ενδεχομένως ακόμη περισσότερο- για σημασιολογικά πρότυπα που αναφέρονται στο ευρύτερο ανάπτυγμα της πολιτιστικής κληρονομιάς και του ερευνητικού αποθέματος (π.χ. BBT, CIDOC, EDM), η αποδοχή κοινών ή συμβατών μεταξύ τους κανόνων για τη δομημένη περιγραφή και διαχείριση του υλικού επιτρέπει τη διαλειτουργικότητα των δεδομένων και τον διαμοιρασμό τους ανάμεσα σε φορείς, οργανισμούς και ερευνητικές ομάδες. Τούτο με τη σειρά του διευκολύνει τη συσσώρευση υλικού προς έρευνα και επεκτείνει την ερμηνευτική ισχύ των συμπερασμάτων.

Η κοινοποίηση, ως διάθεση δεδομένων στο ερευνητικό κοινό, αποτελεί κι αυτή ένα στάδιο εργασιών που υποστηρίζεται από τα πρότυπα και σχήματα μεταδεδομένων χωρίς διάκριση τύπου δεδομένων και επιστημονικής περιοχής. Είτε πρόκειται για βιβλιογραφικά, γλωσσικά, αρχειακά, ποσοτικά και ποιοτικά δεδομένα, η δόμησή τους με μεταδεδομένα επιτρέπει στον ερευνητή/χρήστη, αλλά και στο ευρύτερο κοινό, να επωφεληθεί από τις διαδικασίες κοινοποίησης και τις υπηρεσίες αναζήτησης και περιήγησης που περιλαμβάνουν και να ανακαλύψει υλικό που απαντά στα ερωτήματα και τα ενδιαφέροντά του. Με αυτόν τον τρόπο ολοκληρώνεται και αναπαράγεται σωρευτικά και ποιοτικά ένας κύκλος εργασιών και πρακτικών που ανταποκρίνεται στα ζητούμενα των αρχών FAIR.

Η συνοπτική αυτή εξέταση της αντιστοίχισης προτύπων και σχημάτων μεταδεδομένων με τα στάδια ψηφιακών εργασιών στο πλαίσιο των ανθρωπιστικών και κοινωνικών επιστημών οδηγεί σε κάποιες γενικές παρατηρήσεις ως προς τη σχέση των παραπάνω με τύπους δεδομένων και επιμέρους επιστημονικές περιοχές:

- Υψηλότερος βαθμός εξειδίκευσης των σχημάτων και προτύπων μεταδεδομένων παρατηρείται σε επιστημονικά πεδία που είτε αναπτύσσουν κατά παράδοση δομημένες ταξινομίες και ειδολογικές ιεραρχίες δεδομένων (π.χ. γλωσσολογία, φιλολογία, αρχειονομία, βιβλιοθηκονομία, αρχαιολογία, μουσειολογία) είτε παράγουν τυποποιημένα ποσοτικά και ποιοτικά ερευνητικά δεδομένα με στοχευμένες δράσεις πεδίου (κοινωνικές επιστήμες).
- Επιστημονικές περιοχές όπως η ιστορία, η κοινωνική ανθρωπολογία και οι πολιτισμικές και θεατρικές σπουδές οι οποίες στηρίζονται σε μεγάλο βαθμό σε υλικό που δεν μπορεί εύκολα να μοντελοποιηθεί, μπορούν κάλλιστα να ωφεληθούν από την αξιοποίηση σχημάτων μεταδεδομένων π.χ. σε κειμενικούς και γλωσσικούς πόρους, καθώς άλλωστε η ερμηνευτική τους δραστηριότητα βασίζεται εν πολλοίς στην ανάλυση τέτοιων πηγών. Χαρακτηριστική είναι η περίπτωση των οδηγιών του προτύπου TEI (βλ. παραπάνω) για τα δραματικά και

παραστασιακά κείμενα.¹¹⁶ Εδώ ανοίγει ένα διεπιστημονικό ερευνητικό ζητούμενο: η ιστορική έρευνα π.χ. για τον Νεοελληνικό Διαφωτισμό θα μπορούσε να επωφεληθεί από την ψηφιακή επισημείωση και ανάλυση σωμάτων κειμένων της εποχής, ενώ οι γλωσσικές επιστήμες θα επέκτειναν το εύρος των ψηφιακών τους εργαλείων τους σε παλαιότερες μορφές της ελληνικής.

- Τα σημασιολογικά πρότυπα τείνουν να επεκτείνονται σε ολόκληρο το φάσμα των ανθρωπιστικών και κοινωνικών επιστημών μέσω της συνεργασίας στο πλαίσιο διεπιστημονικών προσεγγίσεων από ομάδες ερευνητών.
- Τα πρότυπα και σχήματα μεταδεδομένων τείνουν να εξετάζουν συνδυαστικά όλο και περισσότερες κατηγορίες υλικού, καλύπτοντας όλο και μεγαλύτερο εύρος της πολιτιστικής κληρονομιάς και ανοίγοντας τις υπηρεσίες που υποστηρίζουν σε όλο και ευρύτερο φάσμα χρηστών, από τους ερευνητές μέχρι την εκπαιδευτική κοινότητα, τις δημιουργικές βιομηχανίες και τον πολιτιστικό τουρισμό.

5. ΕΡΕΥΝΗΤΙΚΑ ΔΕΔΟΜΕΝΑ, ΜΕΤΑΔΕΔΟΜΕΝΑ ΚΑΙ ΑΡΧΕΣ FAIR

Η Ανοικτή Επιστήμη¹¹⁷ θεμελιώνεται στην έννοια της ελεύθερης προσβασιμότητας και σε πρακτικές που ενισχύουν τη συνεργασία, επιταχύνουν την επιστημονική έρευνα και διασφαλίζουν την ποιότητα των παραγώγων της. Οι αρχές της Ανοικτής Επιστήμης ανταποκρίνονται στο σύνολο των σταδίων της έρευνας στο ψηφιακό συγκείμενο και συνδέονται άρρηκτα με την εφαρμογή κατάλληλων τεχνολογικών εργαλείων και οντολογικών πλαισίων.

Το ακρωνύμιο FAIR (Findable, Accessible, Interoperable, Reusable)¹¹⁸ συνοψίζει τις ιδιότητες των δεδομένων και τις αρχές που διέπουν την παραγωγή, τη χρήση και τον διαμοιρασμό τους στο οικοσύστημα της Ανοικτής Επιστήμης. Ως προς την εφαρμογή σημασιολογικών προτύπων και σχημάτων μεταδεδομένων, οι εν λόγω ιδιότητες αποδίδονται ως ακολούθως:

- 1) **Findability** (ευρεσιμότητα): δεδομένα τα οποία είναι οργανωμένα και τεκμηριωμένα, με χρήση κοινών προτύπων και σχημάτων τα οποία επιτρέπουν τον εντοπισμό τους από ερευνητές ή/και ψηφιακές εφαρμογές
- 2) **Accessibility** (προσβασιμότητα): δεδομένα που πλαισιώνονται με σαφείς όρους πρόσβασης και αδειοδότησης, καθώς και με πληροφορίες για τα εργαλεία που απαιτούνται για την επεξεργασία τους

¹¹⁶ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>

¹¹⁷ Βλ. UNESCO Recommendation on Open Science <https://unesdoc.unesco.org/ark:/48223/pf0000379949>

¹¹⁸ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

- 3) **Interoperability** (διαλειτουργικότητα): δεδομένα που υποστηρίζονται από διαδεδομένα πρότυπα και σχήματα μεταδεδομένων, τα οποία διασφαλίζουν την κατανόηση του περιεχομένου από ερευνητές ή/και ψηφιακές εφαρμογές
- 4) **Reusability** (επανάχρηση): δεδομένα που τεκμηριώνονται ως προς τη δομή, το περιεχόμενο και τις συσχετίσεις μεταξύ των στοιχείων που τα συναποτελούν, καθώς και ως προς τους όρους περαιτέρω χρήσης και διαμοιρασμού τους

Οι αρχές FAIR αντικατοπτρίζουν την εξέλιξη της έρευνας σε μια πολυδιάστατη δραστηριότητα που ενσωματώνει σύνθετες μεθοδολογίες και προϋποθέτει την κατανόηση και υλοποίηση ψηφιακών ροών εργασίας για τον εντοπισμό, την επεξεργασία και την αναπαραγωγή ψηφιακών πηγών που πλέον διατίθενται σε ένα μεγάλο εύρος μορφότυπων. Ειδικά στις Ανθρωπιστικές και Κοινωνικές Επιστήμες, η ανάδυση των Ψηφιακών Ανθρωπιστικών Σπουδών (Digital Humanities) ως διακριτού επιστημονικού πεδίου αναδεικνύει τις προοπτικές -αλλά και τις προκλήσεις- του νέου παραδείγματος στο πλαίσιο του οποίου στοιχειοθετούνται οι αρχές FAIR: οι νέες ερευνητικές πρακτικές διευρύνουν μεν τη δυνατότητα παραγωγής, επεξεργασίας και διαμοιρασμού δεδομένων -απαιτώντας, ωστόσο, την εφαρμογή προτύπων και διαδικασιών που διασφαλίζουν την ανοικτότητα των ψηφιακών πόρων.

Η χρήση προτύπων και σχημάτων μεταδεδομένων ενισχύει την τεχνική και σημασιολογική διαλειτουργικότητα των ψηφιακών πηγών και επιτρέπει την αλληλεπίδραση ψηφιακών εφαρμογών μέσω ενιαίων πρωτοκόλλων και οντολογικών πλαισίων. Μια σειρά από διεθνείς πρωτοβουλίες εργάζονται προς την εφαρμογή δομημένων σημασιολογικά μεταδεδομένων στα παράγωγα της έρευνας και την ανάπτυξη υποδομών που υποστηρίζουν ένα ομοιογενές και ανοικτό περιβάλλον ψηφιακής έρευνας.

Ενδεικτικά, η Research Data Alliance (RDA) συγκροτήθηκε το 2013 ως ένα συνεργατικό φόρουμ με σκοπό την ανάπτυξη βέλτιστων πρακτικών στην επιστημονική έρευνα.¹¹⁹ Στο πλαίσιο των εργασιών της RDA, ερευνητές, εκπρόσωποι του επιχειρηματικού κλάδου και φορέων έρευνας συγκροτούν ομάδες εργασίας οι οποίες επεξεργάζονται προτάσεις σχετικά με τη διαχείριση των ερευνητικών δεδομένων και μεταδεδομένων, την προώθηση των αρχών FAIR και το σχεδιασμό εργαλείων για την εφαρμογή τους.¹²⁰ Η RDA δημοσιεύει τακτικά τεκμηριωτικό υλικό αναφορικά με τα ισχύοντα πρότυπα, καθώς και τη χρήση μεταδεδομένων από επιστημονικές (και μη) κοινότητες πρακτικής.¹²¹ Συμπληρωματικά, υποστηρίζει δραστηριότητες επιμόρφωσης και αναπτύσσει οδηγούς για συγκεκριμένα επιστημονικά πεδία, όπως οι Κοινωνικές Επιστήμες, οι Ψηφιακές Ανθρωπιστικές Επιστήμες και η Γλωσσολογία.¹²²

Η CODATA (Committee on Data) αποτελεί το τμήμα του Διεθνούς Επιστημονικού Συμβουλίου (International Science Council) που εστιάζει στην ανάπτυξη προτύπων, καθώς και την προώθηση της ανοικτής πρόσβασης στα παράγωγα της έρευνας.¹²³ Η Επιτροπή δραστηριοποιείται μέσω ομάδων εστίασης επί θεματικών ενοτήτων όπως η επεξεργασία δεδομένων, η Ανοικτή Επιστήμη και οι πολιτικές υποστήριξής της, καθώς και η ανάπτυξη δεξιοτήτων συναφών με τη διαχείριση

¹¹⁹ <https://rd-alliance.org> (20/4/2024).

¹²⁰ <https://www.rd-alliance.org/groups> (20/4/2024).

¹²¹ <https://www.rd-alliance.org/recommendations-outputs> (20/4/2024).

¹²² Οδηγός για τις ΚΕ: <https://www.rd-alliance.org/rda-disciplines/rda-and-social-sciences>

Οδηγός για τις ΨΑΕ: <https://www.rd-alliance.org/rda-disciplines/rda-and-digital-humanities>

Οδηγός για τη Γλωσσολογία: <https://www.rd-alliance.org/rda-disciplines/rda-and-linguistics> (20/3/2024)

¹²³ CODATA: <https://codata.org/>

International Science Council: <https://council.science> (20/3/2024).

δεδομένων· παράλληλα, δημοσιεύει το Data Science Journal, ένα επιστημονικό περιοδικό που ειδικεύεται σε θέματα δεδομένων.

Εστιάζοντας στις πρακτικές όψεις της διαχείρισης δεδομένων κατά τον κύκλο της έρευνας, το Digital Curation Centre (DCC) του Πανεπιστημίου του Εδιμβούργου αναπτύσσει τεκμηριωτικές πηγές και ψηφιακές εφαρμογές που απευθύνονται στα μέλη της ερευνητικής κοινότητας και σε επαγγελματίες που ειδικεύονται στη διαχείριση της πληροφορίας.¹²⁴ Οι οδηγοί¹²⁵ που δημοσιεύει το κέντρο πλαισιώνονται από συμβουλευτικές υπηρεσίες και εκπαιδευτικές δραστηριότητες, ελεύθερα προσβάσιμα εργαλεία και μελέτες περίπτωσης που επιτρέπουν στους ερευνητές να κατανοήσουν και να εφαρμόσουν τις αρχές FAIR, με έμφαση στην πλαισίωση των ψηφιακών πηγών με επαρκή και δομημένα μεταδεδομένα.

Η Ευρωπαϊκή Επιτροπή υποστηρίζει πρωτοβουλίες για τη δημιουργία ενός ανοικτού, διαλειτουργικού και βιώσιμου οικοσυστήματος δεδομένων για την επιστημονική έρευνα. Ενδεικτικά, το GO FAIR (Global Open FAIR), συγκροτεί επιμέρους Δίκτυα Εφαρμογής (Implementation Networks) με σκοπό την προαγωγή βέλτιστων πρακτικών διαχείρισης FAIR δεδομένων σε ένα εύρος επιστημονικών πεδίων -συμπεριλαμβανομένων των ΑΚΕ.¹²⁶ Παράλληλα, στο πλαίσιο της υλοποίησης του Ευρωπαϊκού Νέφους Ανοικτής Επιστήμης (European Open Science Cloud - EOSC)¹²⁷ η Ευρωπαϊκή Επιτροπή έχει χρηματοδοτήσει έργα όπως τα FAIRCORE4EOSC, FAIR-IMPACT και SSHOC (Social Sciences & Humanities Open Cloud), τα οποία ανέπτυξαν πόρους, τεχνικές υποδομές, πολιτικές και υπηρεσίες για την πρακτική εφαρμογή των αρχών FAIR στο ψηφιακό οικοσύστημα του EOSC.¹²⁸

Η παραγωγή ερευνητικών δεδομένων συμβατών με τις αρχές FAIR επικαλύπτει όλα τα στάδια του κύκλου της έρευνας στο ψηφιακό οικοσύστημα. Ως προς τα σχήματα και πρότυπα μεταδεδομένων, οι ορθές πρακτικές που παρουσιάζονται στο τεκμηριωτικό υλικό που δημοσιεύεται από τους φορείς που παρατέθηκαν συνοψίζονται ακολούθως:

Ευρέσιμα δεδομένα: απόδοση αναγνωριστικών και περιγραφικών μεταδεδομένων, απόθεση σε ελεύθερα προσβάσιμες ψηφιακές υποδομές και ευρετηρίαση σε μηχανές αναζήτησης περιεχομένου

Προσβάσιμα δεδομένα: εφαρμογή ανοικτών και τεκμηριωμένων πρωτοκόλλων ευρείας χρήσης (όπως HTTP, OAI-PMH), διάθεση των δεδομένων σε κοινούς μορφότυπους και απόθεσή τους σε ελεύθερα προσβάσιμες ψηφιακές υποδομές

¹²⁴ <https://www.dcc.ac.uk> (20/3/2024).

¹²⁵ Οδηγοί: <https://www.dcc.ac.uk/guidance>
Υπηρεσίες: <https://www.dcc.ac.uk/services>
Εργαλεία: <https://www.dcc.ac.uk/tools>

Μελέτες περίπτωσης: <https://www.dcc.ac.uk/guidance/case-studies> (20/3/2024).

¹²⁶ GoFAIR: <https://www.go-fair.org>

Implementation Networks: <https://www.go-fair.org/implementation-networks/> (20/3/2024).

¹²⁷ <https://eosc-portal.eu> Το Ευρωπαϊκό Νέφος Ανοικτής Επιστήμης αποτελεί ένα σύνολο ψηφιακών υποδομών, υπηρεσιών και πόρων που στοχεύουν στην υποστήριξη της επιστημονικής έρευνας σε όλους τους επιστημονικούς τομείς.

¹²⁸ FAIRCORE4EOSC: <https://faircore4eosc.eu>

FAIR-IMPACT: <https://fair-impact.eu>

SSHOC: <https://sshopencloud.eu/project> (20/3/2024).

Διαλειτουργικά δεδομένα: πλαισίωση των δεδομένων με προτυποποιημένα και μηχαναγνώσιμα σχήματα περιγραφής περιεχομένου (ενδεικτικά, τα σχήματα Dublin Core και TEI που τεκμηριώνονται στην προηγούμενη ενότητα) ή/ και ελεγχόμενα λεξιλόγια ευρείας χρήσης

Επαναχρησιμοποιήσιμα δεδομένα: επαρκής σημασιολογική τεκμηρίωση των δεδομένων, ως προς το περιεχόμενο, τις δυνατότητες επεξεργασίας και τα εργαλεία που απαιτούνται, σαφής περιγραφή των όρων αδειοδότησης

Η κατανόηση της σημασίας των μεταδεδομένων και η ικανότητα παραγωγής και εφαρμογής τους συνιστούν πλέον βασικό στοιχείο των ψηφιακών ροών εργασίας. Μια σειρά ευρωπαϊκών ερευνητικών έργων, όπως τα FAIRsFAIR, PARTHENOS και FOSTER, έχουν αναπτύξει μια σειρά οδηγών, πόρων και εκπαιδευτικών σεναρίων για την καθοδήγηση και την υποστήριξη της ερευνητικής κοινότητας προς την επίτευξη της σημασιολογικής και τεχνολογικής διαλειτουργικότητας που προϋποθέτουν οι αρχές FAIR.¹²⁹

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα μεταδεδομένα αποτελούν ένα θεμελιώδες δομικό στοιχείο του ψηφιακού οικοσυστήματος έρευνας. Τα πρότυπα και σχήματα μεταδεδομένων συμβάλλουν καθοριστικά στη σημασιολογική και τεχνική διαλειτουργικότητα των πηγών και των παραγώγων της έρευνας παρέχοντας στους ερευνητές τη δυνατότητα να εντοπίζουν, να κατανοούν, και να επεξεργάζονται το ψηφιακό περιεχόμενο. Επιτρέπουν, επιπρόσθετα, σε συστήματα ή ψηφιακές εφαρμογές να αλληλεπιδρούν και να ανταλλάσσουν πληροφορίες μεταξύ τους, συμβάλλοντας στην αυτοματοποιημένη ανάλυση και τον συσχετισμό του περιεχομένου με άλλες πηγές.

Η συνδυαστική χρήση των προτύπων και σχημάτων μεταδεδομένων που καταγράφει η μελέτη επιτρέπει στην ερευνητική κοινότητα να δημιουργήσει ένα πλούσιο σημασιολογικό οικοσύστημα πηγών που είναι ελεύθερα προσβάσιμες και επαναχρησιμοποιήσιμες. Όπως προκύπτει, η δημιουργία κοινών συνόλων δεδομένων και η εφαρμογή ανοικτών προτύπων ενθαρρύνει την ανάπτυξη νέων ερευνητικών προσεγγίσεων, καθώς και καινοτόμων εφαρμογών και υπηρεσιών εκ μέρους των ψηφιακών υποδομών υποστήριξης της έρευνας.

Η ανάδυση ωστόσο της Ανοικτής Επιστήμης ως κύριο επιστημονικό παράδειγμα αναδεικνύει παράλληλα την ανάγκη εξοικείωσης και εκπαίδευσης της επιστημονικής κοινότητας στις ψηφιακές ροές διαχείρισης της πληροφορίας. Το συμπέρασμα που προκύπτει από την έρευνα είναι ότι τα μεταδεδομένα είναι ζωτικής σημασίας σε ένα οικοσύστημα όπου η ποικιλία και ο αριθμός των ψηφιακών πηγών και των πρωτογενών δεδομένων αυξάνεται εκθετικά.

Η ενσωμάτωση της χρήσης προτύπων και σχημάτων μεταδεδομένων ευρέως αποδεκτών από την κοινότητα σε όλα τα στάδια της έρευνας και της επιστημονικής επικοινωνίας, θεμελιώνεται επίσης στην υποστήριξη και την προώθηση πολιτικών που ενθαρρύνουν τη διαθεσιμότητα και την ανοικτή χρήση δεδομένων και πηγών. Στο πλαίσιο αυτό, οι ψηφιακές υποδομές καλούνται να αναπτύξουν πόρους και υπηρεσίες που εστιάζουν στην εκπαίδευση της επιστημονικής κοινότητας σχετικά με την εφαρμογή προτύπων και σχημάτων μεταδεδομένων -η οποία αποτελεί τον καμβά επί του οποίου θα αναπτυχθεί περαιτέρω η ψηφιακή ερευνητική δραστηριότητα, καθώς και ένα αναγκαίο

¹²⁹ FAIRsFAIR: <https://www.fairsfair.eu>

PARTHENOS: <http://www.parthenos-project.eu>

FOSTER: <https://www.fosteropenscience.eu> (20/3/2024).

βήμα προς την ενίσχυση του αποτυπώματος της Ανοικτής Επιστήμης στο πεδίο των ανθρωπιστικών και κοινωνικών επιστημών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Αγγελής, Σ., Πάζα, Ε., Παπαχριστόπουλος, Λ., Παπαθεοδώρου, Χ. *Text Encoding Initiative – TEI*. Μονάδα Ψηφιακής Επιμέλειας Ινστιτούτο Πληροφοριακών Συστημάτων Ερευνητικό Κέντρο «Αθήνα». https://eclass.aueb.gr/modules/document/file.php/INF391/lectures/tei%20_intro.pdf

Doerr, M., Daskalaki, M., Bekiari, Ch., Katsiadakis, H., Goulis, H., Terzis, Ch. (2015): *Thesaurus Maintenance Methodological Outline*. Thesaurus Maintenance Working Group, Greece. https://www.backbonethesaurus.eu/sites/default/files/workingpaperonthesaurusmaintenance29_05_2015.pdf

Doerr, M., Felicetti, A., Hermon, S., et al. (2015). *Definition of the CRMarchaeo An Extension of CIDOC CRM to support the archaeological excavation process*. https://www.cidoc-crm.org/crmarchaeo/sites/default/files/CRMarchaeo_v1.4.4.pdf

Δρίτσου, Β., Κωνσταντόπουλος, Π. *Ψηφιακό Τοπίο: Π2-Μοντέλα διεργασιών και ροών εργασίας*. Υπό δημοσίευση.

Gavrilidou, M., Labropoulou, P et al. (2012). *'The META-SHARE Metadata Schema for the Description of Language Resources'*. Proceedings of the Eighth International Conference on Language Resources and Evaluation, pp. 1090-197. http://www.lrec-conf.org/proceedings/lrec2012/pdf/998_Paper.pdf

Goulis, H. (2021). "The BBT meta-thesaurus model: building interoperable thesauri for humanities researchers". In Almeida, B., Costa, R., & Medeiros, F (ed). *Controlled vocabularies and knowledge organisation for the digital humanities: Proceedings*, pp. 4-7. <https://doi.org/10.34619/pgtp-upne>

Γιαννακόπουλος, Γ., Μπουντούρη, Β. (2015). *Εισαγωγή στην Αρχειονομία*. Αθήνα: ΣΕΑΒ

Ηλβανίδου, Μ., Δρίτσου, Β., Πουλή, Κ., Σουγιουλτζόγλου, Η. Τζεδόπουλος, Γ. (2023). *Ψηφιακό Τοπίο: Χρήση ψηφιακών μεθόδων στις ανθρωπιστικές και κοινωνικές επιστήμες / Ανάλυση τάσεων*. <https://zenodo.org/records/10797453>

Καπιδάκης, Σ., Λαζαρίνης, Φ., & Τοράκη, Κ. (2015). "Dublin Core" [Κεφάλαιο]. Στο Καπιδάκης, Σ., Λαζαρίνης, Φ., & Τοράκη, Κ. *Θέματα βιβλιοθηκονομίας και επιστήμης των πληροφοριών [Προπτυχιακό εγχειρίδιο]*. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. <https://hdl.handle.net/11419/1683>

Καρασίμος, Α. (2016). "XML και TEI: μια σύντομη εισαγωγή στην κωδικοποίηση κειμένων. ΔΥΑΣ" — Δίκτυο Υποδομών για την Έρευνα στις Ανθρωπιστικές Επιστήμες. <https://dyas-net.gr/wp-content/uploads/2021/10/Δυάς-λογοι-3-Εισαγωγή-στο-TEI.pdf>

Μπουντούρη, Λ. (2004). "Text Encoding Initiative: Επισκόπηση, προβλήματα και εφαρμογές". *Μετα-Βιβλιοθήκες, οι Βιβλιοθήκες μετά το Διαδίκτυο και τον Παγκόσμιο Ιστό: προς μια νέα λογική συγκρότησης, λειτουργίας και εργαλείων - 13ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών*. <https://lekythos.library.ucy.ac.cy/handle/10797/10819>

UNESCO. (2012). *UNESCO Recommendation on Open Science* <https://unesdoc.unesco.org/ark:/48223/pf0000379949>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). 'The FAIR Guiding Principles for scientific data management and stewardship'. Sci Data 3, 160018.
<https://doi.org/10.1038/sdata.2016.18>

ΔΙΚΤΥΟΓΡΑΦΙΑ

[Abril](#) (εφαρμογή διαχείρισης του προτύπου IMDI)

[Ανεμόσκαλα](#) (ψηφιακό corpus νεοελληνικής ποίησης)

[ARIADNE](#) (ψηφιακή υποδομή)

[BBT](#) (μετα-θησαυρός BackBone)

[Τεκμηρίωση](#)

[CESSDA](#) (Σύνδεσμος Ευρωπαϊκών Αρχείων Κοινωνικών Δεδομένων)

[Πρότυπο μεταδεδομένων CESSDA](#)

[Εκπαιδευτικοί πόροι](#)

[CLARIN](#) (ψηφιακή υποδομή)

[CMDI](#) (πρότυπο μεταδεδομένων)

[CODATA](#) (Επιτροπή για τα Δεδομένα του Διεθνούς Επιστημονικού Συμβουλίου)

[CIDOC](#) (πρότυπο μεταδεδομένων, το οποίο εξειδικεύεται στα ακόλουθα)

[CRMarchaeo](#)

[CRMdig](#)

[CRMba](#)

[CRMgeo](#)

[Τεκμηρίωση](#)

[CRMinf](#)

[CRMsci](#)

[CRMsoc](#)

[CRMtex](#)

[PRESSoo](#)

[DDI](#) (πρότυπο μεταδεδομένων)

[DDI Codebook](#)

[DDI Lifecycle](#)

[DDI Lifecycle Documentation](#)

[DCC](#) (κέντρο ψηφιακής τεκμηρίωσης του Πανεπιστημίου του Εδιμβούργου)

[DCC Guidance](#)

[DCC Services](#)

[DCC Tools](#)

[DCC Case Studies](#)

[Διεθνές Πρότυπο Αρχειακής Περιγραφής](#)

[Documentation of Endangered Languages](#) (Αρχείο απειλούμενων γλωσσών)

[Dublin Core](#) (πρότυπο μεταδεδομένων)

[DC](#) (Λεξικό Επιστήμης της Πληροφόρησης)

[Δομή και περιεχόμενα](#) (έκδοση 1.1)

[Ετικέτες](#)

[ISO Standard 15836-2:2019](#) (Dublin Core)

[ISO Standard 15836-1:2017](#) (Dublin Core metadata elements)

[Encoded Archival Description](#) (πρότυπο μεταδεδομένων)

[EpiDoc](#) (σχήμα μεταδεδομένων)

[Europeana](#)

[Europeana Pro](#)

[Europeana Data Model](#) - EDM (πρότυπο μεταδεδομένων)

[EDM Definition](#)

[EDM objects templates](#)

[EDM Mapping Guidelines](#)

[EDM Primer](#)

[EOSC](#) (Ευρωπαϊκό Νέφος Ανοικτής Επιστήμης)

[FAIRCORE4EOSC](#) (ευρωπαϊκό έργο για τα ανοικτά δεδομένα)

[FAIR-IMPACT](#) (ευρωπαϊκό έργο για τα ανοικτά δεδομένα)

[FAIRsFAIR](#) (ευρωπαϊκό έργο για τα ανοικτά δεδομένα)

[FOSTER](#) (ευρωπαϊκό έργο για την Ανοικτή Επιστήμη)

[FOAF](#) (σχήμα μεταδεδομένων)

[FRBRoo](#) (σχήμα μεταδεδομένων)

[GeoSPARQL](#) (πρότυπο μεταδεδομένων)

[GO FAIR](#) (διεθνής πρωτοβουλία για τα FAIR δεδομένα)

[GO FAIR Implementation Networks](#)

[International Council on Archives](#) (Διεθνές Συμβούλιο Αρχείων)

[IMDI](#) (πρότυπο μεταδεδομένων)

[IMDI vocabulary](#)

[ISAD](#) (πρότυπο μεταδεδομένων)

[International Science Council](#) (Διεθνές Επιστημονικό Συμβούλιο)

[LIDO](#) (πρότυπο μεταδεδομένων)

[MARC Standards](#)

[MARC 21](#)

[MARC 21](#) (ελληνικά)

[Understanding MARC](#)

[Max Planck Institute for Psycholinguistics](#)

[Language Archive at the Max Planck](#)

[Spoken Dutch Corpus](#)

[META-SHARE](#) (ψηφιακή υποδομή)

[META-SHARE](#) metadata (σχήμα μεταδεδομένων)

[Τεκμηρίωση](#)

[Midas Heritage](#) (πρότυπο μεταδεδομένων)

[Midas Heritage](#) (εφαρμογές)

[MODS](#) (Σχήμα μεταδεδομένων)

[MODS Implementation Registry](#)

[MODS 3.8 User Guidelines](#)

[OLAC](#) (πρότυπο μεταδεδομένων)

[Open Archives Initiative](#) (Πρωτοβουλία Ανοικτών Αρχείων)

[Open Archives Initiative Protocol for Metadata Harvesting](#)

[Object Reuse and Exchange](#) (πρότυπο μεταδεδομένων)

[Open Geospatial Consortium standards](#) (πρότυπα για γεωχωρικά δεδομένα)

[PARTHENOS](#) (ευρωπαϊκό έργο για την εφαρμογή καλών πρακτικών στις ΑΚΕ)

[QualiBank](#) (αποθετήριο δεδομένων)

[QuDEx](#) (σχήμα μεταδεδομένων)

[Research Data Alliance](#) (διεθνής οργάνωση για τα ερευνητικά δεδομένα)

[Research Data Alliance \(groups\)](#)

[Research Data Alliance \(recommendations\)](#)

[Research Data Alliance and the Social Sciences](#)

[Research Data Alliance and the Digital Humanities](#)

[Research Data Alliance and Linguistics](#)

[SSHOC](#) (ψηφιακή υποδομή για τις ΑΚΕ)

[SKOS](#) (πρότυπο μεταδεδομένων)

[SDTL](#) (γλώσσα προγραμματισμού)

[SDMX community](#) (διεθνής πρωτοβουλία για τα στατιστικά δεδομένα)

[SODANET](#) (ερευνητική υποδομή για τις κοινωνικές επιστήμες)

[TEI](#) (κοινοπραξία για την ανάπτυξη του προτύπου TEI)

[TEI](#) (πρότυπο μεταδεδομένων)

[Τεκμηρίωση](#)

[Εργαλεία εφαρμογής](#)

[Εκπαιδευτικοί πόροι](#)

[TEI για μουσικές πηγές](#)

Εφαρμογές TEI

[TEI lite](#)

[UK Data Archive](#) (βρετανικό αρχείο δεδομένων)

[Wikibooks](#)

[XKOS](#) (πρότυπο μεταδεδομένων)

ΠΑΡΑΡΤΗΜΑ: ΣΥΝΑΦΕΙΑ ΣΧΗΜΑΤΩΝ ΜΕΤΑΔΕΔΟΜΕΝΩΝ

Σχήμα	Επεκτάσεις/ εξειδικεύσεις	Περιγραφή	Επιστημονικά πεδία	Τύποι δραστηριοτήτων (ΠΕ2)
BBT		Μεταθησαυρός υπερκείμενων γενικών καθολικών όρων και ιεραρχιών που υποστηρίζει τη σημασιολογική διασύνδεση επί μέρους τοπικών θησαυρών και δομημένων λεξιλογίων με διαλειτουργικό και βιώσιμο τρόπο.	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Οργάνωση (Δόμηση, Διαχείριση) Τροποποίηση (Εμπλουτισμός) Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση)
CIDOC CRM		Εννοιολογικό μοντέλο για την ενσωμάτωση πληροφορίας σε ζητήματα πολιτιστικής κληρονομίας.	Ανθρωπιστικές Επιστήμες	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMarchaeo	Χρησιμοποιείται για την τεκμηρίωση και τη διαχείριση αρχαιολογικών δεδομένων και διαδικασιών ανασκαφής.	Ανθρωπιστικές Επιστήμες / Αρχαιολογία	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMba	Υποστηρίζει την καταγραφή των μαρτυριών και των ασυνεχειών της ύλης στα αρχαία κτίρια, για τον εντοπισμό της εξέλιξης της δομής τους.	Ανθρωπιστικές Επιστήμες / Αρχαιολογία	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMdig	Χρησιμοποιείται για την τεκμηρίωση των σταδίων και των μεθόδων παραγωγής προϊόντων ψηφιοποίησης και συνθετικών ψηφιακών αναπαραστάσεων.	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMgeo	Συνδέει εννοιολογικά γεγονότα, φαινόμενα και υλικά αντικείμενα με αντικειμενικά προσδιορισίμα χωροχρονικά στοιχεία.	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)

Σχήμα	Επεκτάσεις/ εξειδικεύσεις	Περιγραφή	Επιστημονικά πεδία	Τύποι δραστηριοτήτων (ΠΕ2)
	CRMinf	Χρησιμοποιείται ως σχήμα για την ενσωμάτωση μεταδεδομένων σχετικά με την επιχειρηματολογία και την εξαγωγή συμπερασμάτων σε αναλυτικές και εμπειρικές επιστήμες	Ανθρωπιστικές και Κοινωνικές Επιστήμες, (ιδιαίτερα Γεωγραφία, Αρχαιολογία)	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMtex	Χρησιμοποιείται για την τεκμηρίωση της μελέτης και την έκδοση αρχαίων χειρόγραφων κειμένων περιγράφοντάς τα με κατάλληλα οντολογικά εργαλεία σε μια διεπιστημονική οπτική.	Παπυρολογία, Παλαιογραφία, Κωδικολογία, Επιγραφική	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMsci	Επιτρέπει την ενσωμάτωση μεταδεδομένων σχετικά με την επιστημονική παρατήρηση, τις μετρήσεις και τα επεξεργασμένα δεδομένα στις αναλυτικές και εμπειρικές επιστήμες.	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
	CRMsoc	Μοντέλο για την ενοποίηση δεδομένων σχετικών με κοινωνικά φαινόμενα και κατασκευές (έννοιες;) που ενδιαφέρουν τις ανθρωπιστικές και κοινωνικές επιστήμες.	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Τροποποίηση (Εμπλουτισμός) Ανάλυση (Επισημείωση)
CMDI		Πλαίσιο για τη δημιουργία και χρήση μορφότυπων μεταδεδομένων για γλωσσικούς πόρους και εργαλεία.	Ανθρωπιστικές Επιστήμες, Γλωσσολογία	Αναζήτηση (Εντοπισμός) Οργάνωση (Διαχείριση, Δόμηση) Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση)
CMM		Υποστηρίζει τη διαχείριση ποσοτικών δεδομένων, κυρίως στις κοινωνικές επιστήμες. Περιέχει ορισμούς και περιγραφές τύπων μεταδεδομένων.	Κοινωνικές Επιστήμες	Ανάλυση (Επισημείωση) Τροποποίηση (Εμπλουτισμός, Οπτικοποίηση)

Σχήμα	Επεκτάσεις/ εξειδικεύσεις	Περιγραφή	Επιστημονικά πεδία	Τύποι δραστηριοτήτων (ΠΕ2)
DDI		Πρότυπο για την περιγραφή δεδομένων που παράγονται από έρευνα πεδίου. Τεκμηριώνει τα στάδια του κύκλου ζωής των ερευνητικών δεδομένων.	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Ανάλυση (Επισημείωση) Τροποποίηση (Εμπλουτισμός, Οπτικοποίηση)
Dublin Core		Πρότυπο γενικής χρήσης για την περιγραφή ψηφιακών αντικειμένων	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Ανάλυση (Επισημείωση) Τροποποίηση (Εμπλουτισμός, Οπτικοποίηση)
EAD		Πρότυπο περιγραφικών μεταδεδομένων βιβλιογραφίας.	Κυρίως Αρχειονομία και Επιστήμες της Πληροφορίας	Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση) Αναζήτηση (Εντοπισμός) Διατήρηση (Επιμέλεια)
EDM		Πρότυπο που χρησιμοποιείται από την ψηφιακή υποδομή Europeana για τον εμπλουτισμό και τη διασύνδεση μεταδεδομένων πολιτιστικής κληρονομιάς.		Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση) Αναζήτηση (Εντοπισμός) Διατήρηση (Επιμέλεια)
IMDI		Περιγράφει πολυμεσικούς και γλωσσικούς πόρους, όπως ηχητικά και βιντεοσκοπημένα υλικά, μεταγραφές και σώματα κειμένων.	Ανθρωπιστικές και Κοινωνικές Επιστήμες, (ιδιαίτερα Γλωσσολογία)	Οργάνωση (Διαχείριση, Δόμηση) Ανάλυση (Επισημείωση) Διατήρηση (Επιμέλεια)
ISAD		Ευρέως χρησιμοποιούμενο πρότυπο περιγραφής αρχειακών συλλογών	Ανθρωπιστικές Επιστήμες, Επιστήμες της Πληροφορίας	Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση) Αναζήτηση (Εντοπισμός) Διατήρηση (Επιμέλεια)

Σχήμα	Επεκτάσεις/ εξειδικεύσεις	Περιγραφή	Επιστημονικά πεδία	Τύποι δραστηριοτήτων (ΠΕ2)
MARC		Πρότυπο για την παρουσίαση βιβλιογραφικών μεταδεδομένων, σχεδιασμένο για να υποστηρίξει την ανταλλαγή βιβλιογραφικών εγγραφών μεταξύ των καταλόγων βιβλιοθηκών.	Επιστήμες της Πληροφορίας	Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση) Αναζήτηση (Εντοπισμός) Διατήρηση (Επιμέλεια)
META - SHARE		Πρότυπο που εξειδικεύεται στην περιγραφή του περιεχομένου γλωσσικών πόρων, των μεταξύ τους συσχετισμών και των εργαλείων επεξεργασίας τους.	Γλωσσολογία	Οργάνωση (Διαχείριση, Δόμηση) Ανάλυση (Επισημείωση) Διατήρηση (Επιμέλεια)
MIDAS		Σχήμα που χρησιμοποιείται για την περιγραφή τόπων και υλικών στοιχείων πολιτιστικής κληρονομιάς.	Ανθρωπιστικές Επιστήμες, Ιστορία, Γεωγραφία	Ανάλυση (Επισημείωση) Διατήρηση (Επιμέλεια) Παραγωγή (Δημιουργία, Ψηφιακή αφήγηση)
MODS		Πρότυπο περιγραφικών μεταδεδομένων για βιβλιοθηκονομικές εγγραφές. Αναπτύσσεται ως ένα υποσύνολο των στοιχείων του MARC, χρησιμοποιώντας λεκτικές αντί για αριθμητικές ετικέτες.	Επιστήμες της Πληροφορίας	Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση) Αναζήτηση (Εντοπισμός) Διατήρηση (Επιμέλεια)
OAI - ORE		Πρότυπο για την περιγραφή και τον διαμοιρασμό συσσωματωμένων ψηφιακών πηγών.	Ανθρωπιστικές και Κοινωνικές Επιστήμες, Επιστήμες της Πληροφορίας	Παραγωγή (Δημιουργία, Ψηφιακή αφήγηση) Διατήρηση (Επιμέλεια) Ανάλυση (Επισημείωση) Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση)

Σχήμα	Επεκτάσεις/ εξειδικεύσεις	Περιγραφή	Επιστημονικά πεδία	Τύποι δραστηριοτήτων (ΠΕ2)
OLAC		Αναπτύσσεται ως ένα ενιαίο σημασιολογικό πλαίσιο για την περιγραφή και την αναζήτηση γλωσσολογικών πόρων σε υποδομές και αποθετήρια.	Γλωσσολογία, Επιστήμες της Πληροφορίας	Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση) Αναζήτηση (Εντοπισμός) Ανάλυση (Επισημείωση)
QuDex		Εφαρμόζεται κατά την επισημείωση ποιοτικών δεδομένων, καθώς και της σχέσης τους με συναφή τεκμήρια.	Ανθρωπιστικές Επιστήμες	Οργάνωση (Διαχείριση, Δόμηση) Ανάλυση (Επισημείωση) Διατήρηση (Επιμέλεια)
SDMX		Πρότυπα και οδηγοί για τον διαμοιρασμό στατιστικών δεδομένων και μεταδεδομένων.	Κοινωνικές Επιστήμες	Τροποποίηση (Εμπλουτισμός, Οπτικοποίηση) Επικοινωνία (Διάδοση, Διαμοιρασμός, Κοινοποίηση)
TEI		Πρότυπο για την αναπαράσταση κειμένων σε ψηφιακή μορφή	Ανθρωπιστικές Επιστήμες	Ανάλυση (Επισημείωση) Διατήρηση (Επιμέλεια) Τροποποίηση (Εμπλουτισμός, Οπτικοποίηση)
	EpiDoc	Εξειδίκευση του προτύπου TEI, για την κωδικοποίηση ψηφιακών αντιγράφων αρχαίων κειμένων	Ανθρωπιστικές και Κοινωνικές Επιστήμες	Ανάλυση (Επισημείωση) Διατήρηση (Επιμέλεια) Τροποποίηση (Εμπλουτισμός, Οπτικοποίηση)