

Εισαγωγή

ΤΟΣΟ ΣΤΗΝ ΚΑΘΗΜΕΡΙΝΗ ΖΩΗ όσο και στα επιστημονικά εγχειρήματα, η αναζήτηση πληροφοριών στηρίζεται σε δύο βασικά ερωτήματα: «Τι είναι αυτό το “πράγμα”» και «Πώς συνδέεται αυτό το πράγμα με άλλα πράγματα;» Στις κοινότητες Βιβλιοθηκών, Αρχείων και Μουσείων (BAM) (καθιερωμένες στην αγγλική γλώσσα ως LAM από τις λέξεις Libraries, Archives and Museums) αλλά και στους κλάδους της πληροφόρησης, η χρήση μεταδεδομένων —με την έννοια των δομημένων, κωδικοποιημένων δεδομένων, τα οποία περιγράφουν χαρακτηριστικά οντοτήτων που φέρουν πληροφορίες (ήτοι, *πράγματα*)— βοηθά στην αναγνώριση, την ανακάλυψη, την αξιολόγηση και τη διαχείριση των περιγραφόμενων οντοτήτων τις οποίες αναζητούμε (CC:DA, 2000). Από την πρώτη εμφάνιση χειρόγραφων και έντυπων καταλόγων και ευρετηρίων έως τη σύγχρονη εποχή των διαδικτυακών υπηρεσιών και κινητών εφαρμογών, η φύση αλλά και ο στόχος της περιγραφής πληροφοριακών οντοτήτων δεν έχουν μεταβληθεί σχεδόν καθόλου. Ωστόσο, οι μέθοδοι και οι τεχνολογίες έχουν αλλάξει σημαντικά. Όσο ο κόσμος γύρω μας γίνεται ολοένα και πιο περίπλοκος ως συνέπεια των τεχνολογικών εξελίξεων και των επιστημονικών ανακαλύψεων, και όσο συνεχίζουμε να ερχόμαστε αντιμέτωποι με υπερβολικό φόρτο πληροφοριών (υπερπληροφόρηση), η οργάνωση και διαχείριση πληροφοριών καθίσταται ένα καθοριστικής σημασίας έργο, τόσο για οργανισμούς όσο και για τον καθένα ξεχωριστά. Το κεφάλαιο αυτό δημιουργεί ένα πλαίσιο χρήσεων μεταδεδομένων στην καθημερινότητα και την εργασία, και κάνει μια σύντομη ιστορική αναδρομή στο κίνημα των μεταδεδομένων. Στη συνέχεια, γίνεται μια ανασκόπηση των βασικών εννοιών, περιγράφοντας τους διάφορους τύπους μεταδεδομένων, τις κατηγορίες προτύπων μεταδεδομένων και τις αρχές χρήσης μεταδεδομένων. Τέλος, δίνονται επιπλέον παραδείγματα με περιγραφές μεταδεδομένων.

1.1 Γενικές πληροφορίες

Τα στιγμιότυπα (instances) δομημένων δεδομένων που περιγράφουν οντότητες οι οποίες φέρουν πληροφορίες μπορεί να είναι τόσο απλές όσο οι ετικέτες σε συσκευασίες τροφίμων ή εμφιαλωμένων ποτών, που ενημερώνουν τους καταναλωτές για τα συστατικά και τα διατροφικά στοιχεία που περιέχονται. Τέτοιες πληροφορίες μπορεί να είναι ζωτικής σημασίας εάν, για παράδειγμα, κάποιος είναι αλλεργικός σε ένα συγκεκριμένο συστατικό του προϊόντος. Μια περιγραφή μεταδεδομένων χρησιμοποιεί δηλώσεις με ζεύγη *ιδιότητας-τιμής* (*property-value*), προκειμένου να περιγράψει με σαφήνεια τα χαρακτηριστικά ενός συγκεκριμένου πράγματος, όπως φαίνεται σε αυτή την ετικέτα τροφίμων (Εικόνα 1.1). Μια ιδιότητα προσδιορίζει το νόημα μιας τιμής δεδομένων. Για παράδειγμα, η ιδιότητα «θερμίδες» βοηθά τον καταναλωτή να καταλάβει τι σημαίνει η τιμή «230» στη συγκεκριμένη ετικέτα. Η διατήρηση μιας συνεπούς δομής και μορφότυπου για την επισήμανση περιεχομένου τροφίμων καθιστά το συγκεκριμένο περιεχόμενο εύκολα κατανοητό και φιλικό προς τον χρήστη. Στις Ηνωμένες Πολιτείες, η δομή και το μορφότυπο επισήμανσης τροφίμων ακολουθούν τις προδιαγραφές που έχει θέσει η Υπηρεσία Τροφίμων και Φαρμάκων (Food and Drug Administration – FDA).

Διατροφικά στοιχεία	
Περιέχει 8 μερίδες των 55 g (2/3 της κούπας)	
Ανά μερίδα	
Θερμίδες	230
% ημερήσια αξία*	
Ολικά λιπαρά 8g	10%
Κορεσμένα λιπαρά 1g	5%
Τρανς λιπαρά 0g	
Χοληστερόλη 0mg	0%
Νάτριο 160mg	7%
Ολικοί υδατάνθρακες 37g	13%
Φυτικές ίνες 4g	14%
Σάκχαρα 12g	
Περιέχει 10g πρόσθετα σάκχαρα	20%
Πρωτεΐνη 3g	
Βιταμίνη D 2mcg	10%
Ασβέστιο 260mg	20%
Σίδηρος 8mg	45%
Κάλιο 240mg	6%

ΙΔΙΟΤΗΤΕΣ
(δομικό στοιχείο)

ΤΙΜΕΣ
(συγκριμένο στοιχείο)

Εικόνα 1.1 Μια ετικέτα τροφίμων ως περιγραφή μεταδεδομένων για ένα μεμονωμένο αντικείμενο

Πηγή: Δημιουργήθηκε με βάση τη νέα σήμανση «New Label Image» από το «What's New with the Nutrition Facts Label?» της αμερικανικής Υπηρεσίας Τροφίμων και Φαρμάκων, <http://www.fda.gov/media/135197/download>

Ένα άλλο παράδειγμα είναι οι πληροφορίες στα συνταγογραφούμενα φάρμακα. Όταν ένας ασθενής ή κάποιο μέλος της οικογένειας ενός ασθενούς αναζητά λεπτομέρειες σχετικά με τις δράσεις και τις πιθανές παρενέργειες ενός φαρμάκου σε μια βάση δεδομένων για συνταγογραφούμενα φάρμακα, είναι ζωτικής σημασίας να γνωρίζει εάν τα ευρήματα είναι αξιόπιστα. Οι περιγραφές μεταδεδομένων σε σχέση με βάσεις δεδομένων συνταγογραφούμενων φαρμάκων μπορούν να παρέχουν πληροφορίες σχετικά με το ποιος διατηρεί τη βάση δεδομένων, πόσο συχνά ενημερώνονται τα δεδομένα και από πού προέρχονται οι πληροφορίες για κάθε φάρμακο, επιτρέποντας έτσι στους καταναλωτές να επιλέξουν την πιο αξιόπιστη και ποιοτικά κατάλληλη βάση δεδομένων την οποία θα εμπιστευτούν για την αναζήτησή τους.

Στην καθημερινότητα αλλά και στην εργασία μας, υπάρχουν πολλά διαφορετικά είδη δομημένων δεδομένων, αρκετά πολύπλοκα κάποιες φορές, από προφίλ εργαζομένων, ιατρικά αρχεία, χρηματοοικονομικά αρχεία και ακαδημαϊκά αρχεία, έως δεδομένα που περιγράφουν κλινικές δοκιμές, γενικές ακολουθίες, χημικές δομές, κώδικες λογισμικού και πολλά άλλα. Διαφορετικά είδη πληροφοριακών οντοτήτων ενδέχεται μάλιστα να έχουν ορισμένες κοινές ιδιότητες, συνήθως όμως διαθέτουν και ένα πλήθος διαφορετικών άλλων γνωρισμάτων. Κάθε τομέας —εργαζόμενοι, ασθενείς,


πελάτες, κλινικές δοκιμές, γενετικές ακολουθίες κ.λπ.— έχει τα δικά του μοναδικά χαρακτηριστικά γνωρίσματα, τα οποία απαιτούν δομημένα δεδομένα ειδικά σχεδιασμένα για την περιγραφή τους. Επιπλέον, υπάρχουν πηγές πληροφόρησης που ενδείκνυνται για ποικίλες άλλες ανάγκες. Έστω ότι μας ενδιαφέρει να μάθουμε πληροφορίες σχετικά με την Πολιτεία της Νέας Υόρκης. Κατά την αναζήτηση του κατάλληλου χάρτη που θα ικανοποιεί αυτή την ανάγκη πληροφόρησης, οι απαιτούμενες πληροφορίες θα περιλαμβάνουν τοπωνύμια, γεωγραφική κάλυψη, γεωγραφικές συντεταγμένες, σκοπό και τύπο χάρτη. Τα συγκεκριμένα στοιχεία, τα οποία αποτελούν *ιδιότητες (properties)* του χάρτη, όταν συμπληρωθούν με τις κατάλληλες *τιμές (values)* δεδομένων, γίνονται *δηλώσεις (statements)*. Στο παραπάνω σενάριο, οι δηλώσεις σχετικά με έναν χάρτη με αναγνωριστικό «MAP12345» θα μπορούσαν να περιλαμβάνουν τα εξής:

Κάλυψη	MAP12345:	«Πολιτεία της Νέας Υόρκης»
Τύπος	MAP12345:	«Τοπογραφικός Χάρτης»

Τα συγκεκριμένα παραδείγματα κάλυψης και τύπου αποτελούν στην πραγματικότητα δύο δηλώσεις, καθεμία από τις οποίες περιγράφει ένα χαρακτηριστικό του χάρτη. Τα ζεύγη *ιδιότητας–τιμής* σχηματίζουν *δηλώσεις* που αναφέρουν ρητά ότι η κάλυψη του χάρτη είναι η Πολιτεία της Νέας Υόρκης και ο τύπος του χάρτη είναι τοπογραφικός. Μέσα από καλά δομημένες δηλώσεις, τα μεταδεδομένα γίνονται εύκολα κατανοητά, τόσο από τους χρήστες όσο και από τις μηχανές. Ένα σύνολο τέτοιων *δηλώσεων* που σχετίζονται με έναν πληροφοριακό πόρο συνθέτουν μια *περιγραφή* για τον συγκεκριμένο πόρο (Powell et al., 2005-2007). Οι σύγχρονες γλώσσες κωδικοποίησης μπορούν να μετατρέψουν αυτές τις δηλώσεις σε δεδομένα κατανοήσιμα αλλά και επεξεργάσιμα από τις μηχανές. Με άλλα λόγια, τα δεδομένα γίνονται πιο ενεργά και αλληλεπιδρώντα σε αυτές τις μορφές (Coyle, 2009). Στο σημερινό περιβάλλον που βασίζεται στο διαδίκτυο, δομημένα δεδομένα που περιγράφουν χαρακτηριστικά πραγμάτων μπορούν να διαμοιραστούν και να χρησιμοποιηθούν με πιο ευέλικτους τρόπους από μια περιορισμένη *εγγραφή (record)*. Σε αυτό το βιβλίο, θα επιχειρήσουμε να χρησιμοποιήσουμε τον πιο γενικό όρο *περιγραφή (description)* αντί του όρου *εγγραφή*, για να αναφερθούμε σε δεδομένα που περιγράφουν ένα συγκεκριμένο πράγμα. Μια *περιγραφή* μπορεί να περιλαμβάνει μία ή περισσότερες *δηλώσεις* με ζεύγη ιδιότητας–τιμής, όπως φαίνεται στο παράδειγμα με την ετικέτα τροφίμων.

Για να αντιληφθούμε περισσότερο την έννοια της δήλωσης και της περιγραφής, ας δούμε τα αποτελέσματα αναζήτησης για τον «Λεονάρντο ντα Βίντσι» στην Google και στη Wikipedia. Η Google παρουσιάζει τα αποτελέσματα σε ένα κατακόρυφο πλαίσιο γνώσεων (knowledge panel) στην πρώτη σελίδα αποτελεσμάτων (Εικόνα 1.2, αριστερά), ενώ στη σελίδα της Wikipedia οι πληροφορίες δίνονται σε ένα ορθογώνιο κουτί πληροφοριών (infobox) στην επάνω δεξιά γωνία (Εικόνα 1.2, δεξιά). Και οι δύο περιπτώσεις αποτελούν δύο καλά παραδείγματα *περιγραφής* μεταδεδομένων ενός καλλιτέχνη με πολλαπλές *δηλώσεις* του ζεύγους ιδιότητας–τιμής (π.χ. για τον Λεονάρντο ντα Βίντσι στα αποτελέσματα της Google, «Γεννήθηκε στις 15 Απριλίου 1452 στην πόλη Αντσιάνο της Ιταλίας»). Τα αποτελέσματα της Google μπορούν να οδηγήσουν τον χρήστη σε άλλα σχετιζόμενα πράγματα μέσω των *τιμών* που χρησιμοποιούν, για παράδειγμα, ποιος είναι ο ακριβής τόπος γέννησης του καλλιτέχνη ή ποια είναι τα πιο γνωστά του έργα. Η μηχανή αναζήτησης συγκεντρώνει περαιτέρω μεταδεδομένα και τα περιλαμβάνει στα αποτελέσματα της αναζήτησής μας. Για παράδειγμα, το πλαίσιο γνώσεων για τον καλλιτέχνη που αναζητήσαμε παρέχει επίσης πληροφορίες και για άλλους καλλιτέχνες υπό την ενότητα «Σχετικές αναζητήσεις».

Πληροφορίες της Google



Πληροφορίες της Wikipedia

Λεονάρντο ντα Βίντσι



Πολυμαθής

Ο Λεονάρντο ντι σερ Πιέρο ντα Βίντσι, γνωστός ως Λεονάρντο ντα Βίντσι, ήταν Ιταλός πολυμαθής της Ύστερης Αναγέννησης. Ανάμεσα στα ενδιαφέροντά του ήταν η επιστήμη και η εφεύρεση, το σχέδιο, η ζωγραφική, η γλυπτική, η αρχιτεκτονική, η μουσική, τα μαθηματικά, η μηχανική, η λογοτεχνία, η ανατομία, η γεωλογία, η αστρονομία, η βοτανική, η παλαιοντολογία και η χαρτογραφία. Wikipedia

Γεννήθηκε: 15 Απριλίου 1452, Αντσιάνο, Ιταλία
Πέθανε: 2 Μαΐου 1519, Κλο-Λουσέ, Αμπουάζ, Γαλλία
Εκθέματα: Αμβροσιανή βιβλιοθήκη, Μουσείο του Λούβρου, ΠΕΡΙΣΣΟΤΕΡΑ
Περίοδοι: Ύστερη Αναγέννηση, Πρώιμη Αναγέννηση, Ιταλική Αναγέννηση, Φλωρεντινή Ζωγραφική
Γνωστός για: τέχνη (ζωγραφική, σχέδιο, γλυπτική), επιστήμη, μηχανική, αρχιτεκτονική, ανατομία
Εκπαίδευση: Αντρέα ντελ Βερόκκιο

Έργα τέχνης




Πορτραίτο που αποδίδεται στον Φραντζέσκο Μέλτζι, περίπου 1515-18

Όνομα στη μητρική γλώσσα: Leonardo da Vinci (Ιταλικά)
Γέννηση: 15 Απριλίου 1452, Αντσιάνο, Βίντσι, Δημοκρατία της Φλωρεντίας (σημερινή Ιταλία)
Θάνατος: 2 Μαΐου 1519, Κλο-Λουσέ, Αμπουάζ, Βασίλειο της Γαλλίας
Γνωστός για: τέχνη (ζωγραφική, σχέδιο, γλυπτική), επιστήμη, μηχανική, αρχιτεκτονική, ανατομία
Αξιοσημείωτο έργο: *Μόνα Λίζα*, *Ο Μυστικός Δείπνος*, *Η κυρία με την ερμίνα*, *Παναγία των Βράχων*, *Άνθρωπος του Βιτρούβιου*, *Salvator Mundi*
Κίνημα: Ύστερη Αναγέννηση

Υπογραφή



Εικόνα 1.2 Πλαίσιο γνώσεων για «Λεονάρντο ντα Βίντσι» στην Google (αριστερά) και κουτί πληροφοριών για «Λεονάρντο ντα Βίντσι» στη Wikipedia (δεξιά)

Σημειώσεις: Τα πλαίσια γνώσεων εμφανίζονται στην Google, όταν αναζητούμε οντότητες (ανθρώπους, μέρη, οργανισμούς, πράγματα) που είναι καταχωρημένες στο Γράφημα Γνώσεων της Google, μια βάση δεδομένων δισεκατομμυρίων εγγραφών για ανθρώπους, μέρη και πράγματα (Sullivan, 2020). Το κουτί πληροφοριών της Wikipedia είναι ένας πίνακας συγκεκριμένης μορφής που τοποθετείται συνήθως στην επάνω δεξιά γωνία άρθρων, με σκοπό την παρουσίαση με δομημένα δεδομένα μιας σύνοψης κοινών στοιχείων που εντοπίζονται σε σχετικά άρθρα (Wikipedia, χωρίς ημερομηνία).

Πηγή: Στιγμιότυπα οθόνης από τα αποτελέσματα της Google για «Λεονάρντο ντα Βίντσι» (<https://www.google.com/search?client=firefox-b-1-d&q=Leonardo+da+Vinci>) και από τις πληροφορίες της Wikipedia για «Λεονάρντο ντα Βίντσι» (https://en.wikipedia.org/wiki/Leonardo_da_Vinci). Ανάκτηση εικόνων στις 2 Ιουλίου 2020.

Οι περιγραφές μεταδεδομένων αποθηκεύονται συνήθως σε μια βάση δεδομένων για ανάκτηση, αλλά μπορούν επίσης να ενσωματωθούν στα αρχεία όπου βρίσκονται οι ψηφιακοί πόροι. Όπως όλα τα δεδομένα που είναι αποθηκευμένα σε συστήματα υπολογιστών, οι περιγραφές μεταδεδομένων μπορούν να μορφοποιηθούν και να δομηθούν με γλώσσες κωδικοποίησης, τις οποίες οι μηχανές επεξεργάζονται back-end, δηλαδή στο νωτιαίο άκρο, και εμφανίζουν front-end, δηλαδή στο μετωπιαίο άκρο, ήτοι στον τελικό χρήστη. Τα προϊόντα και οι υπηρεσίες που βασίζονται ή ενσωματώνουν δομημένα δεδομένα μπορούν να διαβιβασθούν σε ιστοσελίδες ή μέσω κινητών εφαρμογών που λειτουργούν σε διάφορα είδη συσκευών με ευελιξία και δυναμική.

Οι πληροφοριακές οντότητες είναι διαθέσιμες σε μια μεγάλη ποικιλία τύπων. Μπορεί να είναι φυσικά αντικείμενα (συμπεριλαμβανομένων πρωτογενώς ψηφιακών αντικειμένων και μη ψηφιακών αντικειμένων), ψηφιοποιημένα υποκατάστατα φυσικών αντικειμένων ή απλώς σύνολα πληροφοριών για ψηφιακά ή φυσικά αντικείμενα (π.χ. οποιοδήποτε σύνολο δεδομένων για μεταδεδομένα). Οι διαφορετι-

κές πληροφοριακές οντότητες απαιτούν διαφορετικά δομημένα δεδομένα για την περιγραφή τους, έτσι ώστε να μπορούν να αναγνωριστούν, να ανακαλυφθούν, να αποκτηθούν και να χρησιμοποιηθούν. Τα δομημένα, κωδικοποιημένα δεδομένα που περιγράφουν τα χαρακτηριστικά αυτών των πληροφοριακών οντοτήτων ονομάζονται «μεταδεδομένα».

Το σκεπτικό της ανάπτυξης και χρήσης μεταδεδομένων μπορεί να εξηγηθεί από διάφορες οπτικές γωνίες. Καθώς οι ψηφιακές πληροφορίες είναι πλέον παντού, ανακύπτουν αναπόφευκτα συνεχείς προκλήσεις ως προς τη διαχείριση όλων των τύπων και μορφότυπων ψηφιακής πληροφορίας και την παροχή εργαλείων που θα επιτρέπουν την αποτελεσματική, έγκαιρη και ακριβή ανακάλυψη πληροφοριών από οπουδήποτε και ανά πάσα στιγμή. Από τη δεκαετία του 1960 που η τεχνολογία των υπολογιστών άρχισε να χρησιμοποιείται ευρέως στις βιβλιοθήκες και στον κλάδο της πληροφόρησης, κύριοι δράστες οργάνωσης και παροχής πληροφοριών και εργαλείων για την αναζήτηση και χρήση των πληροφοριών ήταν οι βιβλιοθήκες και οι παραγωγοί βάσεων δεδομένων πληροφόρησης. Με την άνοδο του διαδικτύου, οι τεχνολογίες που βασίζονται στον παγκόσμιο ιστό έχουν επιτρέψει τη μαζική παραγωγή και δημοσίευση πληροφοριών μέσω μιας πλατφόρμας χαμηλών φραγμών εισόδου, όπερ σημαίνει ότι οποιοσδήποτε μπορεί να χρησιμοποιήσει ένα πρόγραμμα λήψης εικόνας ή κειμένου μπορεί πλέον να δημιουργήσει ψηφιακά αντικείμενα και να τα δημοσιεύσει απευθείας στον ιστό. Το γεγονός αυτό έχει εκδημοκρατίσει σε μεγάλο βαθμό τη δημοσίευση και διάδοση πληροφοριών, και έχει οδηγήσει σε μια εκθετική αύξηση του όγκου και της πολυπλοκότητας των ψηφιακών πόρων. Άτομα, οργανισμοί, κοινότητες, επιχειρήσεις και κυβερνήσεις καλούνται τώρα να οργανώσουν τις τεράστιες ποσότητες ψηφιακών πληροφοριών και δεδομένων στα πληροφοριακά τους συστήματα, προτού μπορέσουν να ανακαλύψουν, να χρησιμοποιήσουν και να επαναχρησιμοποιήσουν αυτές τις πληροφορίες, όταν χρειαστεί.

Σήμερα, το διαδίκτυο και ο παγκόσμιος ιστός εξελίσσονται με πολλούς τρόπους ως οι νέου τύπου κατάλογοι βιβλιοθηκών, βάσεις δεδομένων ευρετηρίασης, λεξικά, εγκυκλοπαιδίες, εφημερίδες, σχολεία, μουσεία, κέντρα ψυχαγωγίας, ταξιδιωτικά γραφεία, εμπορικά κέντρα, και πολλά άλλα πράγματα και μέρη στα οποία άλλοτε είχαμε μόνο *φυσική* πρόσβαση. Πώς βρίσκουμε τις πηγές που χρειαζόμαστε και τα μέρη που θέλουμε να επισκεφτούμε σε όλο το εύρος του ιστού ή σε ένα ενδοδίκτυο (intranet); Φυσικά, μέσα από τις μηχανές αναζήτησης. Πώς όμως οι μηχανές αναζήτησης μας οδηγούν εύστοχα στον αχανή κόσμο της ψηφιακής πληροφορίας; Τι τις κάνει να λειτουργούν; Κι ακόμα περισσότερο, τι τις κάνει να λειτουργούν *αποτελεσματικά*; Οι απαντήσεις σε αυτά τα ερωτήματα βρίσκονται στο αόρατο χέρι της αποδοτικής οργάνωσης της πληροφορίας, η οποία ενσωματώνεται στα μεταδεδομένα στο back-end (νωτιαίο άκρο) των υπηρεσιών πληροφόρησης. Όταν κάποιος εισάγει μια λέξη-κλειδί στο πλαίσιο κειμένου μιας μηχανής αναζήτησης ή, ακόμα καλύτερα, στο πλαίσιο αναζήτησης μιας ψηφιακής συλλογής, το πιο πιθανόν είναι η λέξη-κλειδί να είναι μία από τις χιλιάδες που συνθέτουν ένα ευρετήριο. Το ευρετήριο μπορεί να δημιουργηθεί από ένα αποθετήριο μεταδεδομένων ή με εξαγωγή λέξεων από έγγραφα πλήρους κειμένου. Οι πληροφορίες που εμφανίζονται στα αποτελέσματα αναζήτησης είναι μια τυπική περίπτωση μεταδεδομένων: περιγράφουν τι είναι ένα έγγραφο ιστού (τύπος) και ποια βασικά χαρακτηριστικά έχει (τίτλος, σύνδεσμος, χρόνος και περιγραφή). Η εμφάνιση πλαισίων γνώσεων στην Google για οποιαδήποτε αναζήτηση οντοτήτων —ανθρώπους, μέρη, οργανισμούς, πράγματα— που υπάρχουν στον Γράφο Γνώσης της Google (Google Knowledge Graph) (Εικόνα 1.2) ξεκίνησε το 2012 και πολύ σύντομα υιοθετήθηκε κι από άλλες μηχανές αναζήτησης. Είναι ένα βήμα παραπέρα στην ανακάλυψη πληροφοριών, εκθέτοντας και συνδέοντας δομημένες και λεπτομερείς περιγραφές για *πράγματα* (*things*), αντί να παρατίθεται απλώς μια λίστα συνδέσμων και βασικών δεδομένων σχετικά με ένα *έγγραφο ιστού* (*web document*).

Τα μεταδεδομένα χρησιμοποιούνται ευρέως εδώ και αρκετά χρόνια στην οργάνωση και τη διαχείριση πρωτογενώς ψηφιακών πόρων. Πολλά παραδείγματα εντοπίζονται σε ιστότοπους μέσω κοινωνικής δικτύωσης, όπου απλές ιδιότητες μεταδεδομένων λαμβάνονται αυτόματα και προστίθενται ετικέτες όταν υποβάλλεται μια ψηφιακή εικόνα, ένα βίντεο ή ένα ιστολόγιο (blog). Παρόμοιες προσεγγίσεις συναντώνται και σε συστήματα ψηφιακών βιβλιοθηκών καθώς και σε αποθετήρια δεδομένων που

απαιτούν από τους συντελεστές περιεχομένου να εισάγουν μεταδεδομένα για ψηφιακές πηγές που υποβάλλονται ή προστίθενται. Περισσότερα παραδείγματα πρωτογενώς ψηφιακών συλλογών με περιγραφές μεταδεδομένων παρουσιάζονται στην ενότητα 1.7.

Μεταδεδομένα χρησιμοποιούνται επίσης στην οργάνωση και τη διαχείριση ψηφιοποιημένων πόρων. Ας πάρουμε για παράδειγμα το βιβλίο *The Burgess Bird Book for Children* που εκδόθηκε πριν από έναν αιώνα. Η ψηφιοποίησή του πιθανότατα θα είχε ως αποτέλεσμα τη δημιουργία ενός μεγάλου αριθμού αρχείων εικόνας. Αυτό που χρειάζεται να γίνει είναι μια περιγραφή μεταδεδομένων όχι μόνο για τις ψηφιοποιημένες σελίδες του βιβλίου (ήτοι τα «ψηφιακά υποκατάστατα») αλλά και για τη σχέση μεταξύ των ψηφιοποιημένων και φυσικών πόρων, ώστε να ενημερώνονται οι χρήστες για ποιον σκοπό έχει δημιουργηθεί το ψηφιακό υποκατάστατο, ποιοι είναι οι δημιουργοί και οι συντελεστές του αρχικού έργου και ποιο είναι το θέμα, όπως φαίνεται στην Εικόνα 1.3. Οι γενικές κατηγορίες περιήγησης που βρίσκονται πάνω από την περιοχή αναζήτησης (τίτλος, συγγραφέας, ημερομηνία, συλλογή) καθίστανται δυνατές χάρη στα μεταδεδομένα που βρίσκονται πίσω από το front-end (μετωπιαίο άκρο), δηλαδή το περιβάλλον που βλέπει ο χρήστης (Εικόνα 1.3).



Εικόνα 1.3 Ψηφιοποιημένο βιβλίο που παρουσιάζεται με κατηγορίες περιήγησης στο επάνω μέρος της πρώτης οθόνης και επιλογές πλοήγησης σελίδων στην αριστερή πλευρά της δεύτερης οθόνης

Πηγή: Με την ευγενική άδεια της Biodiversity Heritage Library, «The Burgess Bird Book for Children», <http://www.biodiversitylibrary.org/bibliography/20184#/summary>.

Από την άλλη, η χρήση μεταδεδομένων για τη συλλογή και διαχείριση πληροφοριών για φυσικά αντικείμενα ήταν και συνεχίζει να είναι μια βασική και αποτελεσματική μέθοδος. Παραδείγματος χάριν, ένα πρότυπο μεταδεδομένων που ονομάζεται Object ID (Αναγνωριστικό Αντικειμένου) και

κυκλοφόρησε το 1997 (<https://icom.museum/en/resources/standards-guidelines/objectid/>) έχει υιοθετηθεί από τον Οργανισμό των Ηνωμένων Εθνών για την Εκπαίδευση, την Επιστήμη και τον Πολιτισμό (United Nations Educational, Scientific and Cultural Organization – UNESCO), το Διεθνές Συμβούλιο Μουσείων (International Council of Museums – ICON), την INTERPOL (τη μεγαλύτερη Διεθνή Οργάνωση Εγκληματολογικής Αστυνομίας), και από κυβερνήσεις σε όλο τον κόσμο, με σκοπό την αντιμετώπιση της παράνομης ιδιοποίησης αντικειμένων τέχνης, αφού καθιστά εύκολη την τεκμηρίωση πολιτιστικής περιουσίας. Το διεθνές αυτό πρότυπο στοχεύει στην περιγραφή έργων τέχνης, ρετρό ή βίντατζ αντικειμένων (αντίκες) και αρχαιοτήτων μέσα από ένα σύνολο ιδιοτήτων που απαντούν στα βασικά ερωτήματα για τον εντοπισμό ενός αντικειμένου που έχει κλαπεί: *Τύπος Αντικειμένου, Υλικά και Τεχνικές, Μετρήσεις, Επιγραφές και Σημάνσεις, Διακριτικά Χαρακτηριστικά, Τίτλος, Θέμα, Ημερομηνία ή Περίοδος, Κατασκευαστής και Περιγραφή*. Επισυνάπτοντας μια φωτογραφία του αντικειμένου, οι συγκεκριμένες περιγραφές μεταδεδομένων βοηθούν στην καταχώριση και την κοινοποίηση των απαραίτητων πληροφοριών για τον εντοπισμό κλεμμένων ή αγνοουμένων αντικειμένων σε όλο τον κόσμο (Εικόνα 1.4).

THE SCREAM

Type: **PAINTING / MODERN / FIGURATIVE / EXTERIOR SCENE WITH FIGURE(S)**

Period: **1893**

Artist(s): **MUNCH**

Technique: **PAINTING / TEMPERA, DRAWING / PASTEL**

Recovered on: **30/06/2006 in**

Height: **91 cm**

Width: **73.5 cm**

Signature state: **LEGIBLE SIGNATURE**

Materials: **CARDBOARD**

ADMINISTRATIVE INFORMATION

Case happened in: **Norway**

Folder: **2004/36031-1.1**

PHOTOS

IF YOU HAVE ANY INFORMATION PLEASE CONTACT

Your national or local police
General Secretariat of INTERPOL

ΙΔΙΟΤΗΤΕΣ (δομικό στοιχείο)

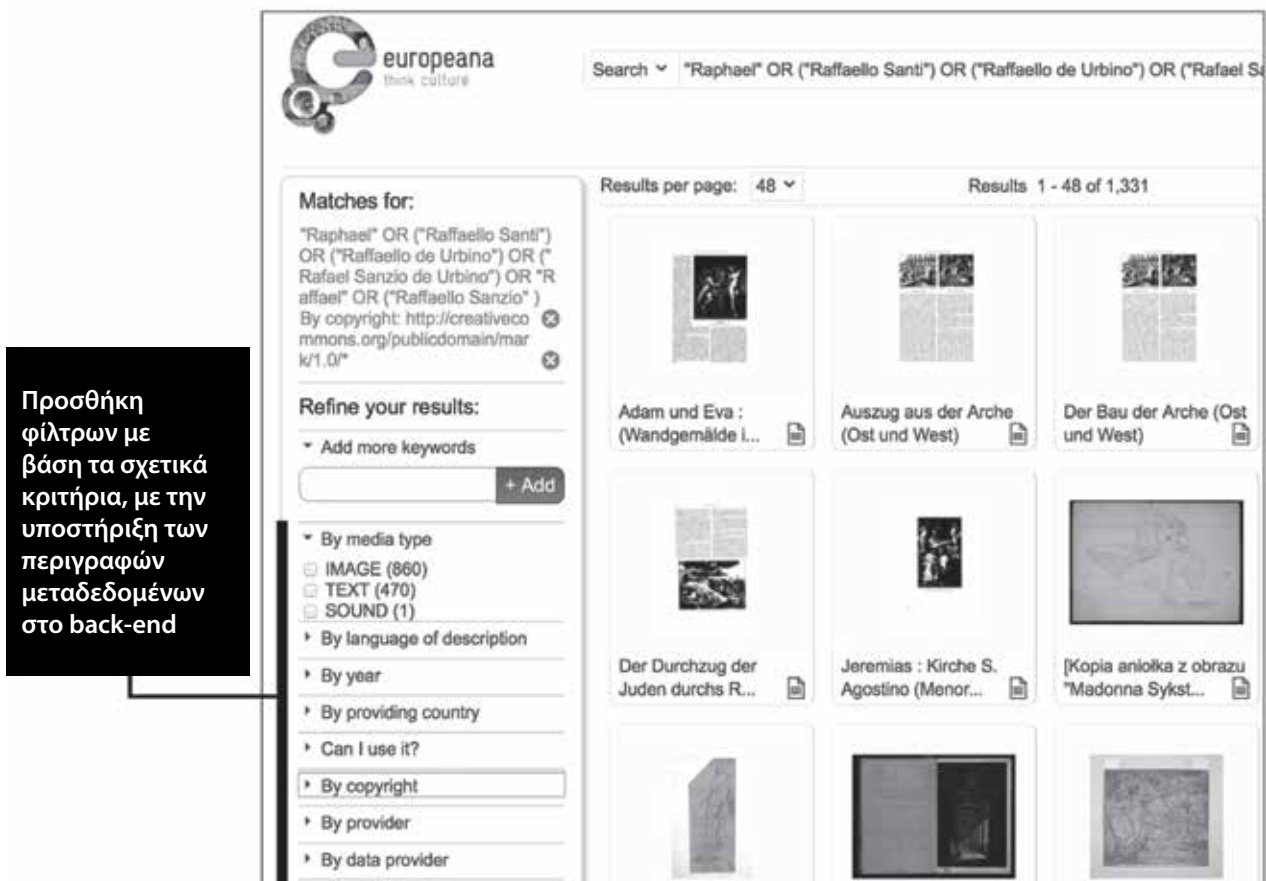
ΤΙΜΕΣ (συγκεκριμένο στοιχείο)

Εικόνα 1.4 Ανάκτηση κλεμμένου έργου τέχνης τεκμηριωμένου στην ηλεκτρονική βάση δεδομένων της INTERPOL

Πηγή: Σχολιασμός σε στιγμιότυπα οθόνης από την ηλεκτρονική βάση δεδομένων της INTERPOL «Stolen Works of Art». Τελευταία πρόσβαση το 2014, <http://www.interpol.int/notice/search/woa/1030880>. Αναφορά στη βάση δεδομένων <https://www.interpol.int/en/Crimes/Cultural-heritage-crime/Stolen-Works-of-Art-Database> (Με την ευγενική άδεια της INTERPOL).

Τα μεταδεδομένα είναι εξίσου χρήσιμα για την ανάλυση και εμφάνιση των αποτελεσμάτων αναζήτησης. Μια αναζήτηση σε μια ψηφιακή συλλογή αποφέρει συνήθως εκατοντάδες, ακόμα και χιλιάδες, αποτελέσματα. Οι χρήστες θα πρέπει να διατρέξουν όλο αυτό τον τεράστιο όγκο για να επιλέξουν τις πληροφορίες που αναζητούν (ή τελικά να εγκαταλείψουν, απογοητευμένοι). Η χρήση μεταδεδομένων,

ευτυχώς, επιτρέπει στο σύστημα να προβαίνει σε επεξεργασία μετά την αναζήτηση και να παρουσιάζει τα αποτελέσματα σε κατηγοριοποιημένες ομάδες. Ας πάρουμε για παράδειγμα την Europeana, την ψηφιακή βιβλιοθήκη, μουσείο και αρχείο της Ευρώπης, που από το 2020 συγκεντρώνει δεδομένα από πάνω από 4.000 ιδρύματα σε όλη την Ευρώπη (Europeana 2020). Στο παρακάτω παράδειγμα, οι επιλογές που εμφανίζονται μετά την αναζήτηση (αριστερά στην Εικόνα 1.5) επιτρέπουν στους χρήστες να περιορίσουν τα αποτελέσματα αναζήτησης ανά *τύπο μέσου, γλώσσα περιγραφής, έτος, χώρα διάθεσης, δικαιώματα χρήσης, κάτοχο πνευματικών δικαιωμάτων και πάροχο*. Κάνοντας κλικ πάνω στην εικόνα, οι χρήστες έχουν τη δυνατότητα να δουν τις λεπτομέρειες των δηλώσεων μεταδεδομένων, οι οποίες είναι απαραίτητες για την υποστήριξη όλων αυτών των φίλτρων και για την παροχή διαδρομών εντοπισμού των επιθυμητών πληροφοριών. Η αναζήτηση υποστηρίζεται επίσης από τις τιμές δεδομένων που ελέγχονται στις περιγραφές μεταδεδομένων, για παράδειγμα, από τυποποιημένους όρους που χρησιμοποιούνται για να υποδείξουν τύπους μέσων (εικόνα, κείμενο, ήχο), όπως φαίνεται στην Εικόνα 1.5.



Εικόνα 1.5 Σχολιασμός σε στιγμιότυπο οθόνης για το αποτέλεσμα αναζήτησης «Raphael» με προσθήκη φίλτρων

Πηγή: Δημιουργήθηκε με βάση την εικόνα Europeana #OnThisDay, καταγραφή στις 19 Νοεμβρίου 2014 και επιβεβαίωση στις 8 Μαρτίου 2021, στην αρχική πύλη της Europeana portal, <http://bit.ly/Raphael-Europeana>. CC0

Οι περιγραφές μεταδεδομένων μεμονωμένων πόρων βοηθούν στην ομαδοποίηση παρόμοιων πόρων με βάση τις ανάγκες των χρηστών. Σε αυτό το παράδειγμα της Europeana, τα μεταδεδομένα που περιγράφουν τα έργα τέχνης που δημιουργήθηκαν από τον Raphael και άλλα έργα, σχετικά με τον Raphael, συγκεντρώνονται από διανεμημένα σύνολα δεδομένων που παρέχονται από ιδρύματα-μέλη της Europeana. Το φιλτράρισμα μετά την αναζήτηση, με τη βοήθεια μεταδεδομένων, έχει γίνει μια δημοφιλής υπηρεσία στα επιγραμμικά/ διαδίκτυα συστήματα καταλόγων. Μια άλλη περίπτωση είναι η

παρουσίαση μιας συλλογής στον οπτικοποιημένο τοίχο συλλογών ενός μουσείου, όπως στον τοίχο ArtLens του Μουσείου Τέχνης του Κλίβελαντ (<https://www.clevelandart.org/artlens-gallery/artlens-wall>). Εκατοντάδες ψηφιακές εικόνες πολιτιστικών αντικειμένων προωθούνται τακτικά στον τοίχο με βάση το θέμα, το χρώμα, τον πολιτισμό, τον δημιουργό και τον τύπο έργου, με τους επισκέπτες να έχουν τη δυνατότητα να επιλέξουν μία από αυτές τις εικόνες και να συγκεντρώσουν ένα σύνολο σχετικών αντικειμένων με βάση τη χρονική περίοδο, το θέμα και τον χώρο έκθεσης. Οι επισκέπτες μπορούν να δημιουργήσουν τη δική τους ξενάγηση με βάση αυτά τα συγκεντρωμένα αντικείμενα, να κατεβάσουν την ξενάγηση σε μια κινητή συσκευή με ένα άγγιγμα και να μεταβούν σε άλλο χώρο έκθεσης, για να δουν τα πραγματικά αντικείμενα. Όλες αυτές οι ξεναγήσεις που επιλέγονται εξατομικευμένα συνοδεύονται από λεπτομερείς περιγραφές, σχολιασμούς και οδηγούς πολυμέσων ενσωματωμένους στις εικόνες που αντιπροσωπεύουν τα επιλεγμένα αντικείμενα.

Τα παραπάνω παραδείγματα αποδεικνύουν ότι τα μεταδεδομένα είναι ικανά να εκτελούν τις ακόλουθες λειτουργίες, όπως αυτές συνοψίζονται στο φυλλάδιο του Εθνικού Οργανισμού Προτύπων Πληροφόρησης (National Information Standards Organization – NISO) με τίτλο *Understanding Metadata* (Riley, 2004):

- περιγράφουν τους πόρους και τι αφορούν, και τους οργανώνουν με βάση ελεγχόμενα κριτήρια,
- επιτρέπουν τον εντοπισμό των πόρων με βάση συναφή κριτήρια, συγκεντρώνουν παρόμοιους πόρους και παρέχουν διαδρομές εντοπισμού των επιθυμητών πληροφοριών,
- διευκολύνουν την ανταλλαγή μεταδεδομένων και επιτρέπουν τη διαλειτουργικότητα,
- παρέχουν ψηφιακή ταυτοποίηση και περιγραφή για την αρχειοθέτηση και διατήρηση των πόρων.

1.2 Ορισμοί

Ο όρος «μεταδεδομένα» (metadata) εμφανίζεται στη βιβλιογραφία και ως «μετα δεδομένα» (meta data) και «μετα-δεδομένα» (meta-data). Στην αγγλική γλώσσα, η λέξη μπορεί να χρησιμοποιηθεί είτε σε ενικό είτε σε πληθυντικό αριθμό, όπως και η λέξη «data». Στον ενικό αριθμό θα χρησιμοποιηθεί όταν θέλουμε να αναφερθούμε σε ένα είδος δεδομένων, ενώ στον πληθυντικό αναφέρεται σε πράγματα που μπορούν να μετρηθούν (Turner, Moal, & Desnoyers, 2004). Στην ελληνική γλώσσα προτιμάται ο πληθυντικός αριθμός. Οι ακαδημαϊκές και ερευνητικές κοινότητες, όπως η κοινότητα συστημάτων γεωγραφικών πληροφοριών (GIS) συνήθως χρησιμοποιούν τον όρο στον πληθυντικό αριθμό (FGDC 2000). Συνηγορώντας υπέρ των ανοικτών δεδομένων, τόσο τα δεδομένα (data) όσο και τα μεταδεδομένα (metadata) αναφέρονται σε πληθυντικό αριθμό στις αρχές FAIR (Findable, Accessible, Interoperable, and Reusable – Εντοπίσιμα, Προσβάσιμα, Διαλειτουργικά και Επαναχρησιμοποιήσιμα) (σχετικές πληροφορίες στην ενότητα 1.6 παρακάτω). Όταν θέλουμε να αναφερθούμε ταυτόχρονα τόσο σε δεδομένα όσο και σε μεταδεδομένα, ο όρος μπορεί να γραφεί ως «(μετα)δεδομένα» [(meta) data] [GO FAIR (2018)].

Ο απλούστερος ορισμός για τα μεταδεδομένα είναι «δεδομένα για δεδομένα» ή «πληροφορίες σχετικά με πληροφορίες» (FGDC, 2000; Hodge, 2001; Riley, 2004). Τα μεταδεδομένα ως έννοια έχουν χρησιμοποιηθεί σε διαφορετικά συγκείμενα, για να αναφερθούν σε πληροφορίες που αφορούν συγκεκριμένα πράγματα. Στην κοινότητα ΒΑΜ, τα μεταδεδομένα δημιουργούνται συνήθως ως κατάλογοι δημοσιευμένου υλικού ή μουσειακών αντικειμένων, κατάλογοι αρχειακού υλικού και ευρετήρια άρθρων ακαδημαϊκών περιοδικών. Τα μεταδεδομένα στον επιχειρηματικό κόσμο εμφανίζονται με τη μορφή καταλόγων προϊόντων ή εμπορευμάτων, αρχείων καταγραφής συναλλαγών ή γνωσιακών βάσεων υποστήριξης πελατών, και πολλά άλλα. Ο πλήρης ορισμός του όρου *μεταδεδομένα* έχει δοθεί από την Επιτροπή της Αμερικανικής Ένωσης Βιβλιοθηκών για την Καταλογογράφηση: Περιγραφή και Πρόσβαση (CC:DA) και συγκεκριμένα από την Ομάδα Εργασίας για Μεταδεδομένα (Task Force on Metadata) ως

δομημένα, κωδικοποιημένα δεδομένα που περιγράφουν τα χαρακτηριστικά πληροφοριακών οντοτήτων και ως τέτοια επιτρέπουν λειτουργίες για την αναγνώριση, την ανακάλυψη, την αξιολόγηση και τη διαχείριση των πληροφοριακών οντοτήτων (CC:DA, 2000).

Τα μεταδεδομένα υπάρχουν όχι μόνο στον παραδοσιακό κόσμο των βιβλιογραφικών δεδομένων αλλά και στην καθημερινή μας ζωή, και μπορεί να εμφανίζονται σε μορφές και μέρη που πολλοί από εμάς δεν τα προσέχουν. Όπως εξηγείται στην αρχή αυτού του κεφαλαίου, για παράδειγμα, οι ετικέτες που είναι τοποθετημένες σε συσκευασίες τροφίμων μπορούν να θεωρηθούν ως μεταδεδομένα διατροφής και συστατικών. Ομοίως, οι κατάλογοι επιχειρήσεων περιέχουν δεδομένα ταυτότητας, επικοινωνίας, προϊόντων και υπηρεσιών σχετικά με τις συγκεκριμένες επιχειρήσεις. Και τα δύο αυτά αποτελούν χαρακτηριστικά παραδείγματα μη βιβλιογραφικών μεταδεδομένων. Όταν βγάζουμε ψηφιακές φωτογραφίες ή δημιουργούμε αρχεία βίντεο ή ήχου, οι τιμές των ιδιοτήτων μεταδεδομένων, όπως η χρονική σήμανση, η ανάλυση, το μέγεθος αρχείου και ο συνδυασμός χρωμάτων, καταγράφονται αυτόματα.

Καθώς η έρευνα και η εφαρμογή των μεταδεδομένων εξελίσσονταν, ο ορισμός τους βελτιώθηκε ώστε να είναι πιο συγκεκριμένος και σαφής, κι έτσι τα μεταδεδομένα χαρακτηρίστηκαν έπειτα ως «δομημένες πληροφορίες που περιγράφουν, εξηγούν, εντοπίζουν ή διευκολύνουν με άλλον τρόπο την ανάκτηση, τη χρήση ή τη διαχείριση ενός πληροφοριακού πόρου» (Hodge, 2001, 3) καθώς και «δεδομένα που σχετίζονται είτε με ένα πληροφοριακό σύστημα είτε με ένα αντικείμενο πληροφοριών για σκοπούς περιγραφής, διαχείρισης, νομικών απαιτήσεων, τεχνικής λειτουργικότητας, χρήσης, χρησιμοποίησης και διατήρησης» (DCMI Glossary, 2005). Οι μεμονωμένες δηλώσεις μεταδεδομένων και τα πακέτα περιγραφών ή εγγραφών (π.χ. σε μια συσκευασία) αναφέρονται συχνά ως «μεταδεδομένα» ενός πράγματος. «Τα μεταδεδομένα, οι πληροφορίες που δημιουργούμε, αποθηκεύουμε και μοιραζόμαστε για να περιγράψουμε πράγματα, μας επιτρέπουν να αλληλεπιδράσουμε με αυτά τα πράγματα, για να αποκτήσουμε τη γνώση που χρειαζόμαστε» (Riley, 2017, 1). Στο βιβλίο αυτό θα προσπαθούμε να χρησιμοποιούμε ακριβείς εκφράσεις για τη διασαφήνιση και καλύτερη κατανόηση της πληθώρας σχετικών, συχνά συγκεχυμένων, εννοιών που υπάρχουν.

Στην πάροδο του χρόνου, ένας μεγάλος αριθμός προτύπων λεξιλογίου μεταδεδομένων έχει αναπτυχθεί ή προταθεί από κοινότητες σε διαφορετικούς τομείς. Το πρότυπο (*standard*) είναι ένα επίσημο έγγραφο που καθορίζει ορισμένα ενιαία κριτήρια, μεθόδους, διαδικασίες και πρακτικές. Βασικό συστατικό των προτύπων είναι το *σύνολο στοιχείων* (*element set*) που καθορίζει τη δομή και τη σημασιολογία των στοιχείων. Για παράδειγμα, το διεθνές πρότυπο *Σύνολο Στοιχείων Μεταδεδομένων Dublin Core* (DCMES ή Dublin Core) ορίζει 15 βασικά στοιχεία που χρησιμοποιούνται για την περιγραφή πληροφοριακών πόρων. Η κοινότητα των μεταδεδομένων έχει χρησιμοποιήσει μια σειρά από όρους σε διαφορετικά πλαίσια, για να αναφερθεί σε *σύνολα στοιχείων* μεταδεδομένων, σε συνδυασμό με τα μηχανικά επεξεργάσιμα *σχήματα* (*schemas*) μέσω των οποίων κωδικοποιούνται. Οι όροι «πρότυπα μεταδεδομένων», «σύνολα στοιχείων», «λεξικά μεταδεδομένων», «σχήματα μεταδεδομένων», «λεξιλόγια ιδιοτήτων», «λεξικά δεδομένων» και «λεξικά μεταδεδομένων» έχουν μικρές διαφορές και χρησιμοποιούνται εναλλακτικά στη βιβλιογραφία. Οι φράσεις αυτές θα επεξηγηθούν και θα διαφοροποιηθούν στη συνέχεια στο βιβλίο, όπου αυτό κριθεί απαραίτητο.

Διαφορετικές κοινότητες αναφέρονται επίσης στα ίδια στοιχεία μεταδεδομένων με διαφορετική ορολογία. Για παράδειγμα, στην πληροφορική, τα μεταδεδομένα είναι πληροφορίες σχετικά με αντικείμενα βάσης δεδομένων ή/και αντικείμενα προγράμματος, όπως πίνακες και αποθηκευμένες διαδικασίες. Ένας σχεδιαστής βάσης δεδομένων μπορεί να ονομάσει τα στοιχεία μεταδεδομένων ως «πεδία δεδομένων» (*data fields*) αντί για «στοιχεία». Άλλοι όροι, όπως «ιδιότητες» (*properties*) και «ιδιοχαρακτηριστικά» (*attributes*), χρησιμοποιούνται επίσης συχνά στη βιβλιογραφία μεταδεδομένων, αλλά η σημασία τους αλλάζει όταν αλλάζει το πλαίσιο στο οποίο χρησιμοποιούνται. Μια οντολογική οπτική των μεταδεδομένων, για παράδειγμα, αντιμετωπίζει το Πρόσωπο (*Person*) ως μια *κλάση* (*class*) και το «όνομα», την «ηλικία», το «φύλο» και το «επίπεδο εκπαίδευσης» ως *ιδιότητες* (*properties*) της κλάσης Προσώπου. Παράλληλα, ένα μοντέλο σχέσης-οντότητας αποκαλεί το Πρόσωπο ως *οντότητα*