ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΜΗΜΑ ΠΟΛΙΤΙΣΜΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΕΝΟΠΟΙΗΣΗ ΠΟΛΙΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΕΤΕΡΟΓΕΝΕΙΣ ΠΗΓΕΣ

ΣΩΤΗΡΙΟΣ ΑΓΓΕΛΗΣ
ΑΕΜ: 18025

ΕΠΙΒΛΕΠΩΝ
ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ
ΚΩΝΣΤΑΝΤΙΝΟΣ ΚΩΤΗΣ

ΜΥΤΙΛΗΝΗ, ΦΕΒΡΟΥΑΡΙΟΣ 2020

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον Κωνσταντίνο Κώτη για την εμπιστοσύνη και την καθοδήγηση του καθ' όλη τη διάρκεια της συνεργασία μας.

Ευχαριστώ τον Γιώργο Σαντιπατάκη και το εργαστήριο Τεχνητής Νοημοσύνης του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιά, για την συνεργασία και υποστήριξη στη χρήση του εργαλείου RDF-Gen.

Θα ήθελα να ευχαριστήσω ιδιαίτερα την Αλεξάνδρα που βρίσκεται πάντα στο πλευρό μου και με στηρίζει.

Περίληψη

Καθώς ο Παγκόσμιος Ιστός εξελίσσεται και προσαρμόζεται προκειμένου να διατηρήσει την ικανότητα διαχείρισης του συνεχώς αυξανόμενου όγκου δεδομένων, οι εφαρμογές Ιστού τείνουν να αυτοματοποιούνται και να γίνονται πιο έξυπνες. Σε αυτό συμβάλουν οι τεχνολογίες του Σημασιολογικού Ιστού, που παρέχουν τη δυνατότητα στους υπολογιστές να κατανοήσουν το περιεχόμενο των εγγράφων, και να τα διαχειριστούν ως έγγραφα και σύνολα δεδομένων.

Η παρούσα εργασία διερευνά την εφαρμογή των τεχνολογιών του Σημασιολογικού Ιστού σε πολιτιστικά δεδομένα. Σκοπός της είναι η σημασιολογική ενοποίηση και διασύνδεση δεδομένων που έχουν προκύψει από διαδικασίες τεκμηρίωσης έργων τέχνης και αντικειμένων πολιτιστικής κληρονομιάς. Αυτό επιτυγχάνεται με χρήση των τεχνολογιών που αποτελούν τα σύγχρονα πρότυπα του σημασιολογικού ιστού (RDF, OWL, SPARQL) και ευρέως διαδεδομένων μοντέλων και λεξιλογίων σχετικών με το πεδίο ενδιαφέροντος (Dublin Core, SKOS, Europeana Data Model).

Στα πλαίσια της εργασίας γίνεται χρήση, αξιολόγηση και σύγκριση των εργαλείων KARMA, OpenRefine, RDF-Gen, τα οποία παρέχουν τη δυνατότητα μετατροπής δεδομένων από ετερογενείς πηγές σε RDF, καθώς και των εργαλείων ανακάλυψης συνδέσμων μεταξύ οντοτήτων (Silk και OpenRefine) που διασυνδέουν τοπικά σύνολα δεδομένων μεταξύ τους, αλλά και με εξωτερικές πηγές, όπως DBpedia και Wikidata. Τέλος, σχεδιάστηκε και υλοποιήθηκε εφαρμογή για την προβολή και εξερεύνηση των ενοποιημένων δεδομένων, αλλά και την περαιτέρω διασύνδεση τους με τις πηγές δεδομένων DBpedia και Europeana. Η εφαρμογή αναπτύχθηκε με χρήση του JavaScript framework React.js και η υποδομή της βασίζεται στον εξυπηρετητή Apache Jena Fuseki και το TDB triple store.

Λέξεις-κλειδιά: πολιτιστικά δεδομένα, σημασιολογική ενοποίηση δεδομένων, ανοιχτά διασυνδεδεμένα δεδομένα, RDF, SPARQL, EDM, Europeana, KARMA, OpenRefine, RDF-Gen, Silk, Apache Jena Fuseki, TDB, React.js.

Σωτήριος Αγγελής

sotirios.angelis@gmail.com

Τμήμα Πολιτισμικής Τεχνολογίας

Πανεπιστήμιο Αιγαίου

Φεβρουάριος 2020

Abstract

As the Web is evolving in order to manage the continuously increasing volume of data, web applications are becoming more automated and smarter. Due to the contribution of Semantic Web technologies the computers are able to understand the context of documents and handle them as documents and as datasets.

The current thesis is engaging with the implementation of Semantic Web technologies to cultural data. Main goal is the semantic integration and interconnection of data that are developed through documentation process of artworks and cultural heritage objects. This is accomplished by using technologies which are current standards of the semantic web (RDF, OWL, SPARQL) and widely accepted models and vocabularies, relevant to the cultural domain (DC, SKOS, Europeana Data Model).

A set of specialized tools which contains, KARMA, OpenRefine and RDF-Gen is being used and evaluated in order to achieve semantic integration of data from heterogeneous sources. Interconnection is achieved by the use of specialized tools Silk and OpenRefine, that discover links between entities, that interconnect local dataset and external data sources, such as DBpedia and Wikidata. Finally, a web application was designed and developed in order to explore the integrated data and interlink them further with the data sources of DBpedia και Europeana. The application was developed with JavaScript framework React.js and the infrastructure is based on the Apache Jena Fuseki server and TDB triple store.

Keywords: cultural data, semantic data integration, linked open data, RDF, SPARQL, EDM, Europeana, KARMA, OpenRefine, RDF-Gen, Silk, Apache Jena Fuseki, TDB, React.js

Sotirios Angelis
sotirios.angelis@gmail.com
Cultural Technology Department
University of Aegean
February 2020

Περιεχόμενα

Κεφάλαιο 1ο Εισαγωγή 1
1.1. Εισαγωγικά στοιχεία ... 1
1.2. Σκοπός Διπλωματικής Εργασίας ... 1
1.3. Οργάνωση κεφαλαίων ... 2
Κεφάλαιο 2ο Θεωρητικό και τεχνολογικό υπόβαθρο ... 3
2.1. Πόρος (Resource) & URI ... 3
2.2. Μεταδεδομένα ... 3
2.3. RDF ... 3
2.4. Τεχνολογίες Σημασιολογικού Ιστού ... 6
RDFS ... 6
OWL ... 7
DC ... 7
SKOS ... 7
SPARQL ... 7
2.5. Σημασιολογικός Ιστός ... 8
2.5.1. Αρχιτεκτονική του Σημασιολογικού Ιστού ... 9
2.6. Ανοικτά Διασυνδεδεμένα Δεδομένα ... 10
2.6.1. Ανοικτά δεδομένα ... 10
2.6.2. Διασυνδεδεμένα Δεδομένα ... 11
2.6.3. Ανοικτά Διασυνδεδεμένα Πολιτιστικά Δεδομένα ... 11
2.6.4. LOD CLOUD ... 12
2.6.5. 5star LOD ... 14
2.7. Διαδικτυακοί πληροφοριακοί χώροι πολιτιστικών δεδομένων ... 15
2.7.1. Europeana ... 15
2.7.2. Ελληνικός Συσσωρευτής Ψηφιακού Περιεχομένου Πολιτιστικής Κληρονομιάς ... 15
2.8. Πολιτιστικά δεδομένα που χρησιμοποιήθηκαν ... 15
2.8.1. Συλλογή Μουσείου Μοντέρνας Τέχνης ... 16
2.8.2. Συλλογή Μουσείου Τέχνης Carnegie ... 16
Κεφάλαιο 3ο Μετατροπή πολιτιστικών δεδομένων σε RDF και διασύνδεση τους ... 17
3.1 Μοντέλο EDM ... 17
3.2 Εποπτεία, καθαρισμός δεδομένων και RDFization ... 18
3.3 Εργαλεία μετατροπής και διασύνδεσης δεδομένων ... 24
3.3.1 RDFization ... 24

3.3.2 Links Discovery (LD) ... 40
Κεφάλαιο 4ο Περιγραφή και Ανάπτυξη Εφαρμογής ... 51
4.1. Λειτουργικές απαιτήσεις ... 51
4.2. Αρχιτεκτονική Εφαρμογής ... 52
4.3. Αποθήκευση και δημοσίευση Δεδομένων ... 53
4.3.1. Apache Jena Fuseki ... 54
4.3.2. TDB ... 56
4.4. Ανάπτυξη Εφαρμογής ... 62
App.js ... 62
Inputs.js ... 63
ResultTable.js ... 65
ResourceDetails.js ... 66
MakeHttpReq.js ... 67
To Curve.js ... 67
Στοιχεία κώδικα από άλλες πηγές ... 68
Κεφάλαιο 5ο Σενάρια Χρήσης Εφαρμογής ... 70
5.1. Σενάριο Χρήσης #1 ... 70
5.2. Σενάριο χρήσης #2 ... 75
Κεφάλαιο 6ο Συμπεράσματα ... 81
6.1. Συμπεράσματα Μετατροπής και Διασύνδεσης Δεδομένων ... 81
6.1.1. RDFization ... 81
6.1.2. Διασύνδεση ... 84
6.2. Συμπεράσματα χρήσης ανεπτυγμένης εφαρμογής ... 86
6.3. Μελλοντικές ενέργειες ... 86
Βιβλιογραφία ... 88

Λίστα εικόνων

Εικονα 1 RDF δηλωση 4
Εικονα 2 Παραδειγμα RDF 5
Εικονα 3 Αρχιτεκτονικη Σημασιολογικου Ιστου 9
Εικονα 4 LOD Cloud το 2007 12
Εικονα 5 LOD Cloud το 2019 13
Εικονα 6 ανοικτά διασυνδεδεμένα δεδομένα 5 αστέρων 14
Εικονα 7 Διαμορφωμενο μοντελο σημασιολογικης αντιστοιχισης δεδομενων μουσειων 24
Εικονα 8 Διαδικασια μοντελοποιησης δεδομενων στο KARMA 24
Εικονα 9 Εισαγωγη δεδομενων στο KARMA 26
Εικονα 10 Προεπισκοπηση δεδομενων στο KARMA 26
Εικονα 11 Δημιουργια νεας στηλης και εφαρμογη μετασχηματισμου 28
Εικονα 12 Σημασιολογικος χαρακτηρισμος δεδομενων 29
Εικονα 13 Δημιουργια μοντελου στο KARMA 29
Εικονα 14 Εξαγωγη μοντελου του KARMA σε μορφη εικονας 30
Εικονα 15 Προβολη δεδομενων στο OpenRefine 31
Εικονα 16 Παραδειγμα 1ο μετατροπης δεδομενων στο OpenRefine 32
Εικονα 17 Παραδειγμα 2ο μετατροπης δεδομενων στο OpenRefine 33
Εικονα 18 Προσθηκη χωρων ονοματων στο OpenRefine 33
Εικονα 19 Επιλογη τυπου δεδομενων OpenRefine 34
Εικονα 20 Εφαρμογη μετατροπης κατα την δημιουργια URI OpenRefine 35
Εικονα 21 Μοντελο που δημιουργηθηκε OpenRefine 35
Εικονα 22 Προεπισκοπιση RDF γραφου για υποσυνολο των δεδομενων 36
Εικονα 23 Διαδικασια μετατροπης δεδομενων σε RDF απο το RDF-Gen 37
Εικονα 24 Εισοδοι παραμετρων, μοντελου και αποτελεσμα διαδικασιας RDF-Gen 38
Εικονα 25 Αρχειο προδιαγραφων RDF-Gen 39
Εικονα 26 Προτυπο μοντελου RDF-Gen 40
Εικονα 27 Διεργασιες για τη δημιουργια συνδεσμων Silk 41
Εικονα 28 Εισαγωγη χωρων ονοματων Silk 42
Εικονα 29 Εισαγωγη δεδομενων Silk 42
Εικονα 30 Δημιουργια 2ης διεργασιας διασυνδεσης Silk 43
Εικονα 31 Δημιουργια 1ης διεργασιας διασυνδεσης Silk 43
Εικονα 32 Ορισμος συγκρισεων για την 1η διεργασια διασυνδεσης Silk 44
Εικονα 33 Ορισμος συγκρισεων για την 2η διεργασια διασυνδεσης Silk 45
Εικονα 34 Αποτελεσματα ευρεσης συνδεσμων διεργασιας 1 Silk 46
Εικονα 35 Αποτελεσματα ευρεσης συνδεσμων διεργασιας 2 Silk 47

Εικονα 36 Δημιουργια reconciliation service για τοπικο αρχειο OpenRefine 47
Εικονα 37 Ορισμος μεταβλητων προς συγκριση OpenRefine 48
Εικονα 38 Αποτελεσματα ανακαλυψης συνδεσμων OpenRefine 49
Εικονα 39 Αποτελεσματα ανακαλυψης συνδεσμων με Wikidata, OpenRefine 50
Εικονα 40 Αρχιτεκτονικο σχεδιο εφαρμογης 53
Εικονα 41 Αρχιτεκτονικη Apache Jena 54
Εικονα 42 HTTP POST αιτημα προς Fuseki 55
Εικονα 43 Αποκριση Fuseki σε HTTP αιτημα 56
Εικονα 44 Αρχειο Διαμορφωσης TDB χωρις συμπερασμο 57
Εικονα 45 Τριπλετες περιγραφης των πορων «Guitar» και «Pablo Picasso» 59
Εικονα 46 Αρχειο Διαμορφωσης TDB με συμπερασμο 60
Εικονα 47 Τριπλετες περιγραφης των πορων «Guitar» και «Pablo Picasso» μετα τη διαδικασια συμπερασμου 62
Εικονα 48 Αρχικη Οθονη Εφαρμογης 62
Εικονα 49 Επιλεγμενο πεδιο αναζητητσης χωρις ορισμο τιμης 64
Εικονα 50 Επιλεγμενο πεδιο αναζητησης με επιλογη τιμης 64
Εικονα 51 Προτεινομενες τιμες κατα την πληκτρολογηση 64
Εικονα 52 Αναπαρασταση αποτελεσματων αναζητησης σε μορφη πινακα 66
Εικονα 53 Αναπαρασταση αποτελεσματων αναζητησης σε μορφη κατευθυνομενου γραφου 68
Εικονα 54 Αλληλεπιδραση δομικων στοιχειων εφαρμογης 69
Εικονα 55 Αναζητηση και δημιουργια SPARQL ερωτηματος, Σεναριο χρησης #1 70
Εικονα 56 Αποτελεσματα αναζητησης, Σεναριο χρησης #1 72
Εικονα 57 Επιλογη πορου για προβολη, Σεναριο χρησης #1 73
Εικονα 58 Εναλλακτικη προβολη λεπτομερειων πορου RDF-XML, Σεναριο χρησης #1 74
Εικονα 59 Εναλλακτικη προβολη λεπτομερειων πορου JSON-LD, Σεναριο χρησης #1 74
Εικονα 60 Προβολη λεπτομερειων πορου συναθροισης, Σεναριο χρησης #1 75
Εικονα 61 Αναζητηση και δημιουργια SPARQL ερωτηματος, Σεναριο χρησης #2 76
Εικονα 62 Προβολη αποτελεσματων αναζητησης, Σεναριο χρησης #2 77
Εικονα 63 Προβολη λεπτομερειων πορου, Σεναριο χρησης #2 77
Εικονα 64 Αναπαρασταση αποτελεσματων αναζητησης σε κατευθυνομενο γραφο, Σεναριο χρησης #2 78
Εικονα 65 Προβολη λεπτομερειων πορου καλλιτεχνη, Σεναριο χρησης #2 79
Εικονα 66 Συνδεσμος πορου με εξωτερικη πηγη, DBpedia, Σεναριο χρησης #2 79
Εικονα 67 Συνδεσμος πορου με εξωτερικη πηγη, Wikidata, Σεναριο χρησης #2 80

Λίστα Πινάκων

Πινακας 1 Τυποι δεδομενων της συλλογης MoMA 20
Πινακας 2 Τυποι δεδομενων της συλλογης CMoA 21
Πινακας 3 Τυποι δεδομενων των συλλογων που χρησιμοποιηθηκαν 22
Πινακας 4 Αντιστοιχιση μεταβλητων MoMA, DBpedia για ανακαλυψη συνδεσμων 45
Πινακας 5 Πινακας συγκρισης Επαναχρησιμοποιηση Μοντελου 81
Πινακας 6 Πινακας συγκρισης Ταχυτητα Αναπτυξης Μοντελου 82
Πινακας 7 Πινακας συγκρισης Ταχυτητα Αποκρισης RDFization 83
Πινακας 8 Πινακας συγκρισης Ευκολια Χρησης RDFization 84
Πινακας 9 Πίνακας σύγκρισης Αποφυγή Λαθών 84
Πινακας 10 Πινακας συγκρισης Ταχυτητα Ανακαλυψης Συνδεσμων 85
Πινακας 11 Πινακας συγκρισης Ταχυτητα Αναπτυξης Διαδικασιας Διασυνδεσης 85
Πινακας 12 Πινακας συγκρισης Ευκολια Χρησης, Link Discovery 86

Κεφάλαιο 1ο
Εισαγωγή

1.1. Εισαγωγικά στοιχεία

Αναλογιζόμενοι την έννοια της πληροφορίας, συνειρμικά καταλήγουμε στο μεγαλύτερο αποθετήριο της που δεν είναι άλλο από τον παγκόσμιο ιστό. Στον παγκόσμιο ιστό μέσω της δομής του διαδικτύου, αποθηκεύεται, διαμοιράζεται και αναπαρίσταται πληροφορία σε πληθώρα μορφών και εκφάνσεων. Ιστότοποι, ηλεκτρονικές διευθύνσεις και διαδικτυακές υπηρεσίες καθιστούν προσβάσιμα και ελέγξιμα αντικείμενα του ψηφιακού και του φυσικού κόσμου, από την προβολή μιας εικόνας ή την αναπαραγωγή ενός ήχου, έως την περιγραφή της κατάστασης μιας ηλεκτρονικής συσκευής και τον έλεγχό της. Κάθε ένα από αυτά τα αντικείμενα αποτελεί στοιχείο του ιστού και αναφέρεται σαν πόρος.

Ο παγκόσμιος ιστός δημιουργήθηκε αρχικά για χρήση από ανθρώπους και παρόλο που οτιδήποτε διακινείται σε αυτόν είναι μηχανικά αναγνώσιμο και προσπελάσιμο δεν είναι ωστόσο μηχανικά κατανοητό. Ο όγκος των διεργασιών που λαμβάνουν χώρα στον παγκόσμιο ιστό, επιβάλει την αυτοματοποίηση τους, προκειμένου να διεκπεραιώνονται σε χρονικά πλαίσια που να δικαιολογούν την εκτέλεση τους.

Η λύση που προτάθηκε ήταν η χρήση μεταδεδομένων για περιγραφή των δεδομένων που περιέχονται στον ιστό. Η οργανωμένη χρήση μεταδεδομένων αποτελεί θεμέλιο λίθο για τις σύγχρονες και συνεχώς αναπτυσσόμενες τεχνολογίες του σημασιολογικού ιστού και των ανοιχτών διασυνδεδεμένων δεδομένων.

1.2. Σκοπός Διπλωματικής Εργασίας

Η παρούσα διπλωματική αναπτύσσεται στο πεδίο του Σημασιολογικού Ιστού και πιο ειδικά, τα ανοιχτά διασυνδεδεμένα δεδομένα και την εφαρμογή τους σε δεδομένα πολιτιστικής κληρονομιάς. Σκοπός της εργασίας είναι η σημασιολογική ενοποίηση πολιτιστικών δεδομένων που προέρχονται από ετερογενείς πηγές, η διασύνδεση τους με εξωτερικές πηγές και η ανάπτυξη εφαρμογής για την προβολή και εξερεύνηση των ενοποιημένων διασυνδεδεμένων δεδομένων. Για την επίτευξη του σκοπού απαιτείται:

- η εξέταση της υπάρχουσας σχετικής βιβλιογραφίας,

- η χρήση διαδεδομένων τεχνολογιών,
- η χρήση, αξιολόγηση και σύγκριση εργαλείων δημιουργίας και διασύνδεσης σημασιολογικών δεδομένων
- ανάλυση των απαιτήσεων και σχεδιασμός της προτεινόμενης εφαρμογής
- ανάπτυξη και αξιολόγηση της εφαρμογής

1.3. Οργάνωση κεφαλαίων

Το 2ο κεφάλαιο αποτελεί το θεωρητικό υπόβαθρο της εργασίας. Σε αυτό παρατίθενται οι τεχνολογίες και που χρησιμοποιήθηκαν στην εργασίας και αναλύονται έννοιες όπως, σημασιολογικός ιστός, διασυνδεδεμένα δεδομένα, RDF, οντολογία και μοντέλα δεδομένων.

Το 3ο κεφάλαιο περιγράφει την μεθοδολογία που ακολουθήθηκε για την μετατροπή πολιτιστικών δεδομένων σε ανοιχτά διασυνδεδεμένα δεδομένα. Παρουσιάζονται τα εργαλεία που χρησιμοποιήθηκαν για την σημασιολογική ενοποίηση και διασύνδεση τους και αναφέρονται τα αποτελέσματα της επεξεργασίας και της μετατροπής τους.

Το 4ο κεφάλαιο παρουσιάζει την αρχιτεκτονική της εφαρμογής του αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας. Αναφέρονται οι λειτουργικές απαιτήσεις και περιγράφεται η υποδομή της εφαρμογής. Στη συνέχεια, αναλύεται η λειτουργία του κάθε δομικού στοιχείου της εφαρμογής και παρουσιάζεται ο τρόπος που αλληλοεπιδρούν.

Το 5ο κεφάλαιο περιλαμβάνει την παρουσίαση δύο σεναρίων χρήσης που δημιουργήθηκαν για την αξιολόγηση της εφαρμογής.

Το 6ο κεφάλαιο περιλαμβάνει τα συμπεράσματα της διπλωματικής εργασίας. Συγκρίνονται τα εργαλεία που χρησιμοποιήθηκαν με βάση τα χαρακτηριστικά τους και τα αποτελέσματα της χρήσης τους. Τέλος γένεται συνολική εκτιμητή της εργασίας και αναφέρονται οι μελλοντικές ενέργειες.

Κεφάλαιο 2ο
Θεωρητικό και τεχνολογικό υπόβαθρο

2.1. Πόρος (Resource) & URI

Πόρος, είναι οτιδήποτε μπορεί να εκτεθεί ή να χρησιμοποιηθεί στο διαδίκτυο και οτιδήποτε με το οποίο μπορεί να υπάρξει αλληλεπίδραση. Για να χρησιμοποιηθεί ένας πόρος στον παγκόσμιο ιστό πρέπει πρώτον είναι εφικτό να εντοπιστεί και δεύτερον να υπάρχουν μέσα ώστε να είναι διαχειρίσιμος. Βάσει της προτυποποίησης του W3c, για το σκοπό αυτό γίνεται χρήση των Uniform Resource Identifiers ή εν συντομία URIs[1]

Ένα URI ταυτοποιεί έναν πόρο και ταυτόχρονα τον διευθυνσιοδοτεί, ώστε να είναι δυνατή η προσπέλαση και η διαχείριση του με βάση κάποιο διαδικτυακό πρωτόκολλο όπως το HTTP. Το URI ενός πόρου τον διαχωρίζει από οποιονδήποτε άλλον ενώ μέσω αυτού καθίστανται δυνατές και όλες οι αλληλεπιδράσεις μαζί του. Η σχέση μεταξύ URIs και πόρων είναι πολλά προς έναν, δηλαδή ένα URI μπορεί να καταδεικνύει μόνο έναν πόρο, αλλά ένας πόρος μπορεί να αναγνωρίζεται από διαφορετικά URIs[2].

2.2. Μεταδεδομένα

Τα μεταδεδομένα είναι δομημένη πληροφορία που περιγράφει, επεξηγεί, προσδιορίζει έναν πόρο και επιτρέπει την ευκολότερη ανάκληση, χρήση και διαχείριση του. Καλούνται συχνά και δεδομένα για τα δεδομένα.

Τα μεταδεδομένα διευκολύνουν την οργάνωση των πληροφοριακών πόρων ομαδοποιώντας εκείνους με κοινά χαρακτηριστικά και διακρίνοντας εκείνους με ασυσχέτιστα. Παρέχουν επίσης αναγνωριστικά και εξυπηρετούν στην αποτελεσματικότερη πρόσβαση στους πόρους, επιτρέποντας να βρεθούν μέσω αναζήτησης σχετικών με αυτούς γνωρισμάτων[3].

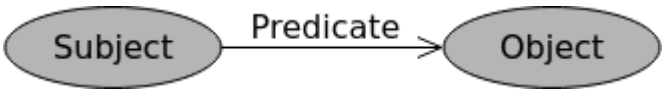
2.3. RDF

Η προσπάθεια για αποτελεσματικότερη διαχείριση του τεράστιου όγκου πόρων και δεδομένων που παράγονται, μετά τον μονοσήμαντο καθορισμό των πόρων από τα URIs, οδήγησε ήδη από τα τέλη του προηγούμενου αιώνα στην ανάγκη αρχικά για περιγραφή τους και στη συνέχεια στην απόπειρα απόδοσης νοήματος τόσο για τους τελικούς αποδέκτες όσο και για τα ίδια τα μηχανήματα που τους διαχειρίζονται.

Έτσι, για την επίτευξη του στόχου της σημασιολογικής αναπαράστασης της πληροφορίας, προτείνεται το 1999 το μοντέλο περιγραφής πόρων RDF [4]. Ένα μοντέλο δεδομένων που περιγράφει τον τρόπο διαχείρισης των μεταδεδομένων ώστε να παρέχουν, στις εφαρμογές που διαχειρίζονται την πληροφορία, δυνατότητα κατανόησης του περιεχομένου της. Ο ευρύτερος στόχος του RDF είναι ο ορισμός ενός μηχανισμού περιγραφής πόρων που δεν αφορά συγκεκριμένο πεδίο ενώ παράλληλα θα είναι ικανός να περιγράψει πληροφορίες οποιουδήποτε πεδίου[5].

Η στοιχειώδης ολοκληρωμένη μονάδα πληροφορίας στο μοντέλο RDF είναι η πρόταση. Η πρόταση συνδέει πόρους με ιδιότητες μέσω δηλώσεων. Πόροι (Resources) όπως ορίστηκαν παραπάνω, είναι οι οντότητες που περιγράφονται από το μοντέλο . Οι ιδιότητές (Properties) είναι τα κατηγορήματα που αποδίδουν γνωρίσματα στους πόρους[6].

Οι δηλώσεις (Statements) είναι ο τρόπος με τον οποίο επιτυγχάνεται η προαναφερθείσα αντιστοίχιση. Είναι της μορφής <υποκείμενο> – <κατηγόρημα> – <αντικείμενο> και αναπαρίσταται ως κατευθυνόμενος γράφος όπου υποκείμενο και αντικείμενο αποτελούν κόμβους του ενώ η ιδιότητα είναι η ακμή που τα συνδέει με κατεύθυνση από το υποκείμενο προς το αντικείμενο όπως φαίνεται στο παρακάτω σχήμα.



Εικόνα 1 RDF δήλωση

Το υποκείμενο της τριάδας είναι ένα URI που καταδεικνύει στον πόρο προς περιγραφή. Το αντικείμενο είναι είτε μία απλή τιμή (literal), για παράδειγμα μια συμβολοσειρά, ένας αριθμός ή μια ημερομηνία, είτε το URI ενός άλλου πόρου. Το κατηγόρημα εκφράζει την σχέση μεταξύ υποκειμένου και αντικειμένου για παράδειγμα η ημερομηνία γέννησης κάποιου, σε περίπτωση απλής τιμής, ή η γνωριμία μεταξύ ατόμων σε περίπτωση άλλου πόρου. Το κατηγορήματα αναγνωρίζονται από URIs που προέρχονται από λεξιλόγια. Τα λεξιλόγια αποτελούν συλλογές που παρέχουν πληροφορίες για συγκεκριμένα πεδία γνώσης.

Καθώς οι RDF γράφοι είναι ομάδες από τριπλέτες, μπορούν εύκολα να συνδυάσουν δεδομένα από πολλαπλές πηγές. Παρόλα αυτά συχνά είναι επιθυμητή η χρήση πολλών διακριτών γράφων και διατήρηση των περιεχομένων τους χωριστά.

Παράδειγμα RDF

Στη συνέχεια παρουσιάζεται ένα παράδειγμα που αφορά την περιγραφή του έργου “War and Peace” του Pablo Picasso η οποία χρησιμοποιήθηκε στα πλαίσια εκπόνησης της παρούσας διπλωματικής εργασίας.

```
<http://localhost:3000/moma/8973/artwork> <http://purl.org/dc/elements/1.1/title>
"War and Peace" .

<http://localhost:3000/moma/8973/artwork> <http://purl.org/dc/elements/1.1/date>
"1955" .

<http://localhost:3000/moma/8973/artwork> <http://purl.org/dc/elements/1.1/creator>
<http://localhost:3000/moma/artist/4609> .

<http://localhost:3000/moma/artist/4609>
<http://www.w3.org/2004/02/skos/core#prefLabel> "Pablo Picasso" .
```

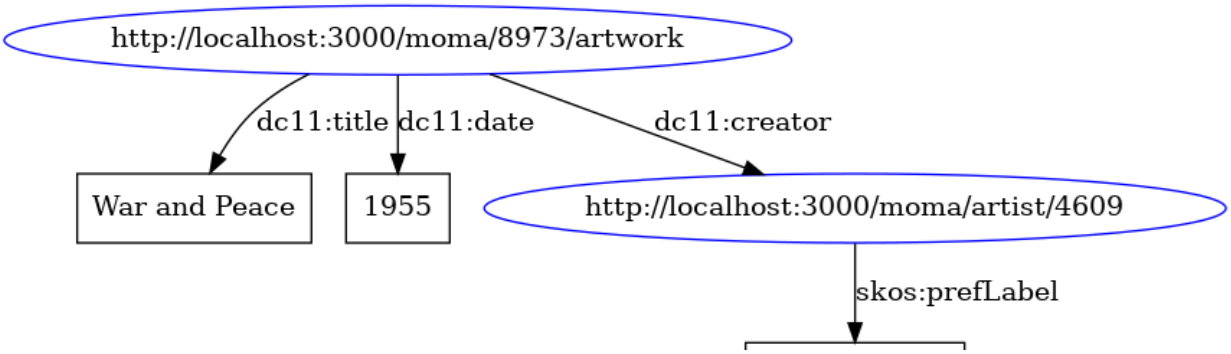
Σε ελεύθερη απόδοση ο γράφος βάσει των συσχετίσεων, αποδίδει τα παρακάτω νοήματα:

<Artwork 8973> <has title> <War and Piece>

<Artwork 8973> <date of creation> <1953>

<Artwork 8973> <was created by> <Artist 4609>

<Artist 4609> <is called> <Pablo Picasso>



Εικόνα 2 Παράδειγμα RDF

Πλεονεκτήματα RDF

- Η χρήση URIs σαν μοναδικά αναγνωριστικά τόσο για δεδομένα όσο και για όρους λεξιλογίων, επιτρέπει σε οποιονδήποτε να αναφερθεί σε οτιδήποτε.
- Κάθε χρήστης ενός RDF γράφου που βρίσκεται στον παγκόσμιο ιστό, μπορεί να ανατρέξει σε οποιοδήποτε URI και να λάβει πληροφορίες για τον πόρο, με αποτέλεσμα κάθε τριπλέτα ενός γράφου να αποτελεί εναρκτήριο σημείο για την εξερεύνηση του συγκεκριμένου πεδίου δεδομένων.
- Το μοντέλο επιτρέπει την διασύνδεση δεδομένων από διαφορετικές πηγές.
- Πληροφορία από διαφορετικές πηγές μπορεί εύκολα να συγχωνευθεί, συνδυάζοντας ομάδες τριπλετών σε ένα γράφο.
- Το RDF επιτρέπει τη χρήση διαφορετικών σχημάτων και λεξιλογίων σε έναν γράφο αυξάνοντας την εκφραστικότητα στην περιγραφή των πόρων[7].

Σε γενικές γραμμές η δομή του RDF είναι αρκετά απλή και δεν παρουσιάζει κάποια δυσκολία στην κατανόηση. Η χρήση του όμως σε πολύπλοκα σενάρια καθιστά δύσκολη την αποσφαλμάτωση τυχών λογικών λαθών. Σε συνδυασμό με την πρώτη της αναπαράσταση μέσω του προτύπου RDF/XML [8], η δομή αποθάρρυνε την υιοθέτησή της από τους προγραμματιστές και έτσι δεν χρησιμοποιήθηκε ευρέως. Δεν υπήρξε λοιπόν, άμεσα η επιθυμητή ανάπτυξη πόρων που συμπεριλαμβάνουν μεταδεδομένα. Για την αντιμετώπιση αυτού και άλλων προβλημάτων δημιουργήθηκαν οι αναπαραστάσεις Turtle [9], N-Triples [10], N-Quads [11], N3 [12], JSON-LD [13].

2.4. Τεχνολογίες Σημασιολογικού Ιστού

RDFS

Το RDF Schema[14] (RDFS) παρέχει ένα λεξιλόγιο μοντελοποίησης RDF δεδομένων. Είναι ένα σύνολο προκαθορισμένων πόρων που παρέχει τη δυνατότητα περιγραφής συσχετιζόμενων πόρων αλλά και την μεταξύ τους σχέση. Αποτελεί επέκταση του RDF και χρησιμοποιείται συχνά από οντολογίες για την περιγραφή βασικών εννοιών. Παράδειγμα χρήσης του RDFS είναι ο καθορισμός του πεδίου ορισμού και του πεδίου τιμών των ιδιοτήτων των πόρων.

OWL

Η Web Ontology Language[15] (OWL) είναι μια γλώσσα δημιουργίας οντολογιών.. Επεκτείνει το μοντέλο RDF και το RDFS παρέχοντας δυνατότητα ορισμού κλάσεων από την ένωση, την τομή ή την άρνηση μιας ή και περισσοτέρων άλλων κλάσεων. Βασίζεται στην περιγραφική λογική και έχει σχεδιαστεί για την αναπαράσταση σύνθετης γνώσης σχετικά με πράγματα ή ομάδες πραγμάτων και της σχέσης μεταξύ τους.

DC

Το λεξιλόγιο DCMI metadata terms [16] ή (Dublin Core, DC) περιέχει όρους που σκοπός τους είναι να χρησιμοποιηθούν σε συνδυασμό με όρους άλλων συμβατών λεξιλογίων για την δημιουργία νοηματικού πλαισίου εφαρμογών. Το Dublin Core είναι ένα σύνολο δεκαπέντε γενικών και ευρέως χρησιμοποιούμενων στοιχείων: Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights.

SKOS

Το Simple Knowledge Organization System [17] (SKOS) είναι ένα πρότυπο διαμοιρασμού δεδομένων που γεφυρώνει διαφορετικά πεδία γνώσης και τεχνολογίας. Χρησιμοποιείται για την αναπαράσταση συστημάτων οργάνωσης γνώσης (knowledge organization systems - KOSs), όπως θησαυροί, ταξονομίες και σχήματα ταξινόμησης.

SPARQL

Η SPARQL [18] (SPARQL Protocol and RDF Query Language) είναι μία γλώσσα για δημιουργία ερωτημάτων σε δεδομένα της μορφής RDF. Επιτρέπει την ανάκτηση και διαχείριση RDF γράφων . Η χρήση της αποτελεί πρότυπο και έχει χαρακτηριστεί ως ένα από τα βασικά στοιχεία του σημασιολογικού ιστού. Με τη χρήση της SPARQL είναι εφικτή η δημιουργία ερωτημάτων σε διαφορετικές πηγές δεδομένων και μπορούν να οριστούν απαραίτητα και προαιρετικά κριτήρια στα ερωτήματα. Οι επιστρεφόμενες τιμές σε SPARQL ερωτήματα μπορεί να είναι είτε σύνολα αποτελεσμάτων σε μορφή πίνακα είτε RDF γράφοι.

Οι βασικές εντολές της περιγράφονται από τους παρακάτω όρους:

- SELECT: Όπου επιστρέφει σε μορφή πίνακα τα αναγνωριστικά που επιλύουν τις δοσμένες προτάσεις.
- ASK: Μία πρόταση ή σειρά από αυτές θέτεται προς επαλήθευση και επιστρέφει αληθές/ψευδές.
- DESCRIBE: Θέτοντας ένα ερώτημα DESCRIBE αναμένεται να επιστραφούν τα αναγνωριστικά των πόρων που επαληθεύουν τις προτάσεις, αλλά και να περιγραφτούν από όλες οι σχέσεις – προτάσεις που σχετίζονται με τους πόρους αυτούς. Επιστρέφεται δηλαδή ένα κομμάτι του γράφου.
- CONSTRUCT: Χρησιμοποιείται για να επιστρέψει πληροφορίες σε RDF μορφή.
- WHERE: Ο όρος αυτός υποδηλώνει τις συνθήκες οι οποίες οφείλουν να ισχύουν.

Παράδειγμα SPARQL

```
PREFIX ex: http://example.com/exampleOntology#
SELECT *
WHERE{
        ?culturalHeritageObject ex:title ?title.
        ?culturalHeritageObject ex:creator ?artist.
        ?artist ex:name “Pablo Picasso”.
}
```

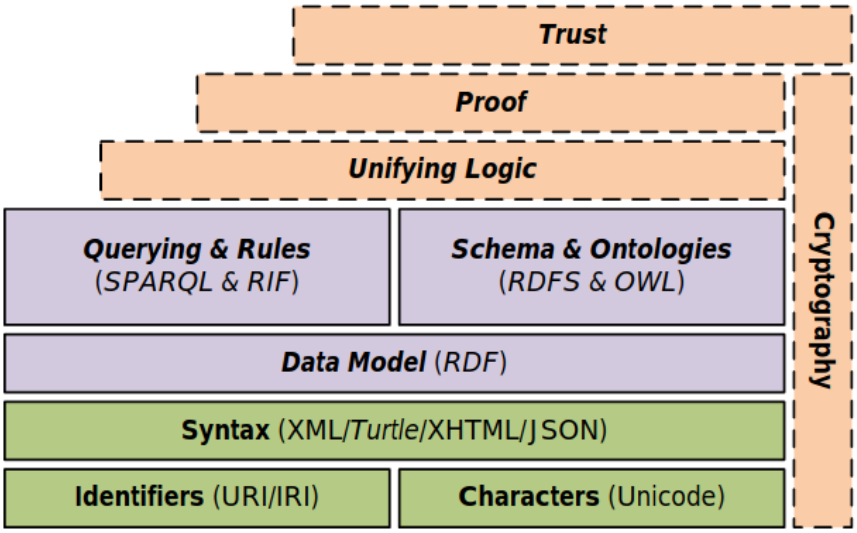
2.5. Σημασιολογικός Ιστός

Ο σημασιολογικός ιστός, Web 3.0 αποτελεί μια εξέλιξη του παγκόσμιου ιστού, που περιλαμβάνει και εφαρμόζει τεχνολογίες, μέρος των οποίων αναφέρθηκαν παραπάνω, και καλείται να μεγιστοποιήσει την αξία της πληροφορίας παρέχοντας δυνατότητα στους υπολογιστές να την κατανοούν και να την αξιοποιούν.

Ο σημασιολογικός ιστός προτείνεται και ορίζεται από τον ίδιο τον δημιουργό του παγκόσμιου ιστού Tim Berners-Lee [19] ως ακολούθως:

> “*The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. [...] A web of data that can be processed directly and indirectly by machines*”

2.5.1. Αρχιτεκτονική του Σημασιολογικού Ιστού



Εικόνα 3 Αρχιτεκτονική Σημασιολογικού Ιστού

Στο σχήμα απεικονίζεται η στοίβα του σημασιολογικού ιστού, που αρχικά δημιουργήθηκε από τον Tim Berners-Lee. Στην στοίβα εμφανίζεται η ιεραρχία των τεχνολογιών του σημασιολογικού ιστού και το πως κάθε μια βασίζεται στις προηγούμενες. Οι τεχνολογίες σύμφωνα με τον A. Hogan [20] αναλύονται ως εξής:

Κωδικοποίηση χαρακτήρων (Characterset): Αρχική απαίτηση είναι ο καθορισμός ενιαίας κωδικοποίησης για όλη την κειμενική πληροφορία. Για το σκοπό αυτό επιλέγεται η ευρέως αποδεκτή κωδικοποίηση Unicode.

Αναγνωριστικά (Identifiers): Η χρήση των URI εξυπηρετεί την ανάγκη για μονοσήμαντη αναγνώριση πόρων, μιας και χρησιμοποιείται ήδη στον παγκόσμιο ιστό. Νεότερα πρότυπα έχουν αρχίσει να υιοθετούν την χρήση Internationalized Resource Identifiers (IRI) που αποτελούν γενίκευση των URIs και υποστηρίζουν διευρυμένα πρότυπα Unicode.

Σύνταξη (Syntax): Για να είναι εφικτή η αυτοματοποιημένη προσπέλαση και ανάλυση του περιεχομένου από τους υπολογιστές, στον σημασιολογικό ιστό απαιτείτε χρήση σύνταξης με αυστηρά ορισμένη γραμματική. Έτσι μπορούν να επιλεχθούν οι XML, JSON ή κάποια σύνταξη παράγωγη της Turtle.

Μοντέλα δεδομένων (Data Model): Ο ορισμός ενός κοινού και επαρκώς γενικού μοντέλου για την δόμηση περιεχομένου διευκολύνει την διακίνηση μηχαναγνώσιμων δεδομένων

Σχήματα και Οντολογίες (Schema & Ontologies): Αυστηρά ορισμένες γλώσσες παρέχουν καλά ορισμένες έννοιες μπορούν να χρησιμοποιηθούν σε συνδυασμό με RDF μοντέλο για να δημιουργήσουν σχήματα και οντολογίες. Τα κοινά πρότυπα στον παγκόσμιο ιστό είναι σήμερα το RDF Schema (RDFS)[14] και η οντολογία Web Ontology Language (OWL)[15]

Ερωτήματα και κανόνες (Querying & Rules): Το τρέχον πρότυπο για δημιουργία ερωτημάτων στον σημασιολογικό ιστό είναι η SPARQL Protocol and RDF Query Language (SPARQL)[18], που αποτελεί μια ώριμη γλώσσα για ερωτήματα σε περιεχόμενο της μορφής RDF. Το πρότυπο για τους κανόνες στο σημασιολογικό ιστό είναι το Rule Interchange Format (RIF) που συγκεντρώνει την εκφραστικότητα δίφορων γλωσσών και προσφέρει μια βιβλιοθήκη ενσωματωμένων λειτουργιών.

Ενωτική λογική (Unifying Logic): Στόχος της είναι η δημιουργία ενός επιπέδου διαλειτουργικότητας που θα συνδυάζει όλες τις τεχνολογίες των κατώτερων επιπέδων..

Τεκμήρια (Proof): Δεδομένου του ότι ο σημασιολογικός ιστός θα επιτρέπει σε πράκτορες λογισμικού την εκτέλεση αυτοματοποιημένων εργασιών σε διάφορες πηγές δομημένων δεδομένων, είναι απαραίτητο να παρέχονται τεκμήρια ώστε να μπορεί να αξιολογηθεί διαδικασία ή ο σκοπός χρήσης των δεδομένων.

Εμπιστοσύνη (Trust): το επίπεδο της εμπιστοσύνης απαιτείται για την εξακρίβωση της αξιοπιστίας των πηγών δεδομένων αλλά και σε συστήματα πελατών – διακοσμητών ως μηχανισμός παροχής πρόσβασης.

Κρυπτογραφία (Cryptography): Ο σημασιολογικός ιστός απαιτεί τεχνικές κρυπτογράφησης για την επαλήθευση ταυτότητας και την παροχή πρόσβασης σε μηχανισμούς.

2.6. Ανοικτά Διασυνδεδεμένα Δεδομένα

2.6.1. Ανοικτά δεδομένα

Ανοιχτά είναι τα δεδομένα που μπορούν ελεύθερα να χρησιμοποιηθούν, και να αναδιανεμηθούν από οποιονδήποτε – υπό τον όρο να γίνεται αναφορά στους δημιουργούς και να διατίθενται, με τη σειρά τους, υπό τους ίδιους όρους.

Οι σημαντικότερες ιδιότητες των ανοικτών δεδομένων είναι :

- Διαθεσιμότητα και Προσβασιμότητα
- Επαναχρησιμοποίηση και Αναδιανομή
- Καθολική Συμμετοχή σε αυτά

Κύριο πλεονέκτημα των ανοικτών δεδομένων είναι η διαλειτουργικότητα, δηλαδή η δυνατότητα διαφορετικών συστημάτων να λειτουργούν μαζί. Αυτή η δυνατότητα διαμοίρασης και σύνδεσης συνιστωσών έχει θεμελιώδη σημασία για τη δόμηση μεγαλύτερων και πιο πολύπλοκων συστημάτων [21].

2.6.2. Διασυνδεδεμένα Δεδομένα

Η διασύνδεση δεδομένων είναι απλώς η χρήση του παγκόσμιου ιστού για τη δημιουργία συνδέσμων μεταξύ δεδομένων από διαφορετικές πηγές. Ο όρος διασυνδεδεμένα δεδομένα αναφέρεται σε δεδομένα που έχουν δημοσιευθεί στον ιστό με τέτοιο τρόπο ώστε να είναι μηχαναγνώσιμα, να έχει οριστεί το νόημα του λεπτομερώς, να έχουν συνδεθεί με εξωτερικά σετ δεδομένων και αντίστροφα να μπορούν να συνδεθούν από εξωτερικά σετ δεδομένων. Ενώ τα βασικά στοιχεία του παγκόσμιου ιστού είναι αρχεία HTML που συνδέονται με υπερσυνδέσμους, τα διασυνδεδεμένα δεδομένα βασίζονται σε έγγραφα που περιέχουν δεδομένα σε μορφή RDF. Έτσι ο όρος Web of Data μπορεί να ερμηνευθεί ως ιστός των πραγμάτων που υπάρχουν στον κόσμο, τα οποία περιγράφονται από δεδομένα στον παγκόσμιο ιστό [22].
Τα διασυνδεδεμένα δεδομένα αποτελούν την ουσία του σημασιολογικού ιστού, δηλαδή την ενοποίηση σε μεγάλη κλίμακα και την εξαγωγή συμπερασμάτων για δεδομένα στον παγκόσμιο ιστό. Σχεδόν όλες οι εφαρμογές και τα σενάρια χρήσης του σημασιολογικού ιστού βασίζονται στην προσβασιμότητα και την ενοποίηση δεδομένων [23].

2.6.3. Ανοικτά Διασυνδεδεμένα Πολιτιστικά Δεδομένα

Όπως ορίζεται από το Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) [24], όντας θεσμικά αρμόδιο για τη συλλογή και την προτυποποιημένη οργάνωση και τεκμηρίωση της επιστημονικής και πολιτιστικής πληροφορίας στη χώρα, η διαδικασία της τεκμηρίωσης είναι από τις πλέον σημαντικές λειτουργίες ενός πολιτιστικού φορέα και εξυπηρετεί στον εντοπισμό και τη διαχείριση τόσο των αντικειμένων όσο και των ψηφιακών τους υποκατάστατων, στην ερμηνεία και την παροχή πρόσβασης σε αυτά.

Η τεκμηριωτική πολιτιστική πληροφορία αντιπροσωπεύει έναν τεράστιο όγκο γνώσης γύρω από την προέλευση και την ερμηνεία των πολιτιστικών αντικειμένων, και συγκροτεί ένα από τα πλέον κατάλληλα σύνολα δεδομένων για δημοσίευση ως

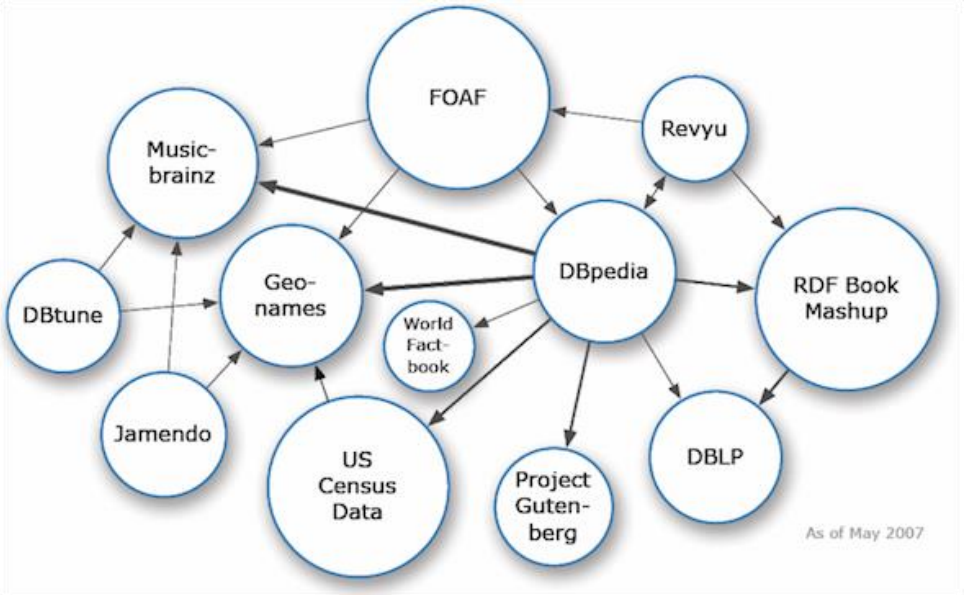
διασυνδεδεμένα δεδομένα, καθώς παραδοσιακά οι πολιτιστικοί φορείς χρησιμοποιούν πρότυπα για την οργάνωση και την ταξινόμηση της πληροφορίας, και άλλες δομημένες γλωσσικές πηγές όπως λεξιλόγια, θησαυρούς, κ.λπ., για την περιγραφή του υλικού και άυλου πολιτισμικού τους πλούτου. Παρέχουν, δηλαδή, έτοιμες σημασιολογικές συνδέσεις μεταξύ διαφορετικών δεδομένων.

Με το άνοιγμά της πληροφορίας αυτής, καθίσταται δυνατή η διασύνδεσή της με άλλες πηγές δομημένης πληροφορίας όπως ημερομηνίες γεννήσεως και θανάτου δημιουργών, γεωγραφικές συντεταγμένες τοπωνυμίων, πληροφορίες από τη Βικιπαίδεια και άλλες έγκριτες πηγές πληροφόρησης, καθιστώντας πολύ πιο πλούσια την εμπειρία του χρήστη [25].

2.6.4. LOD CLOUD

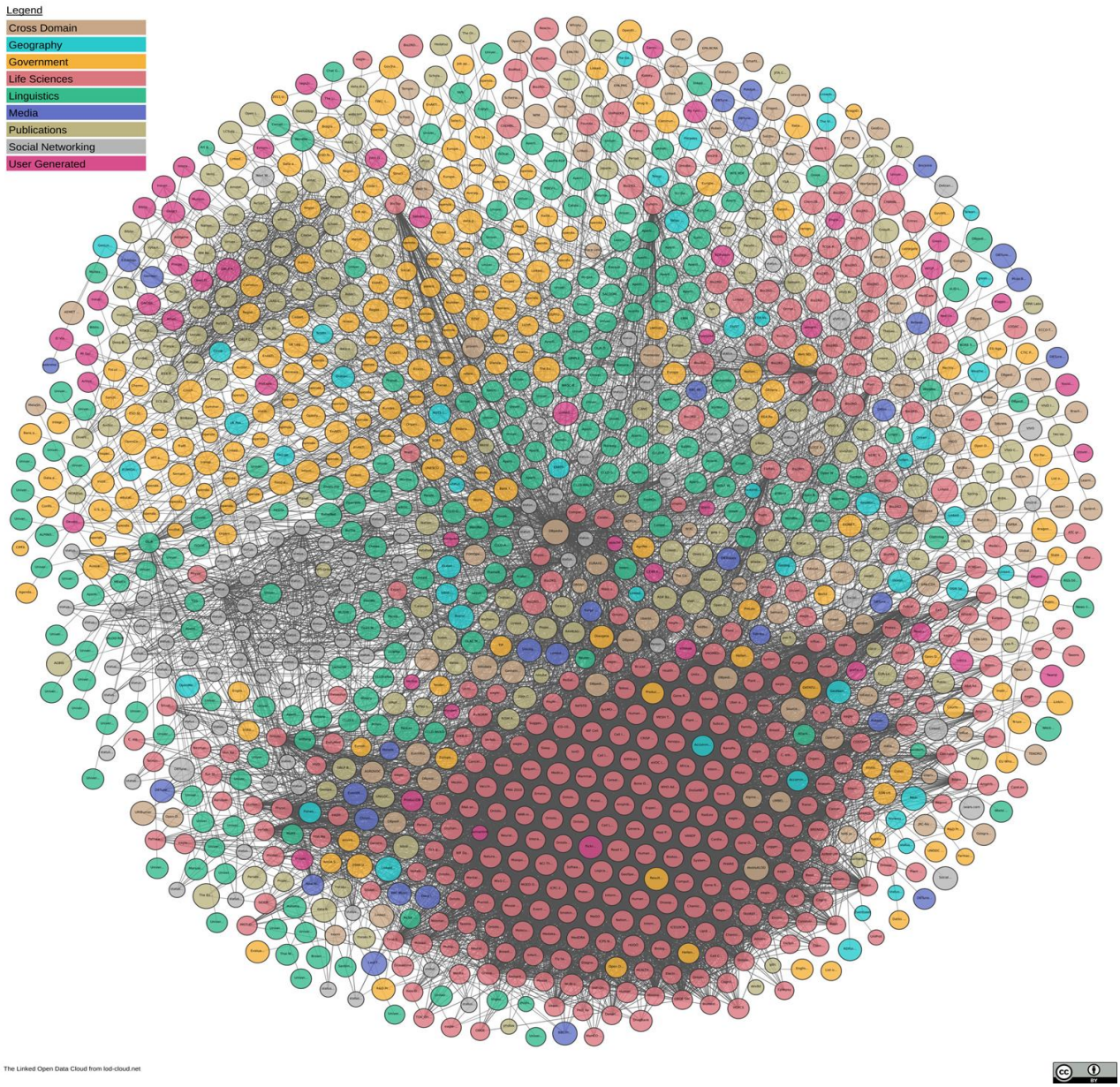
Η ιστοσελίδα https://lod-cloud.net/ φιλοξενεί το διάγραμμα του νέφους των ανοικτών διασυνδεδεμένων δεδομένων ή LOD cloud. Οι εικόνες δείχνουν τα σετ δεδομένων που έχουν δημοσιευθεί ως ανοικτά διασυνδεδεμένα δεδομένα.

Απεικόνιση του LOD Cloud τον Μάιο του 2007 με μόλις 12 σετ δεδομένων.



Εικόνα 4 LOD Cloud το 2007

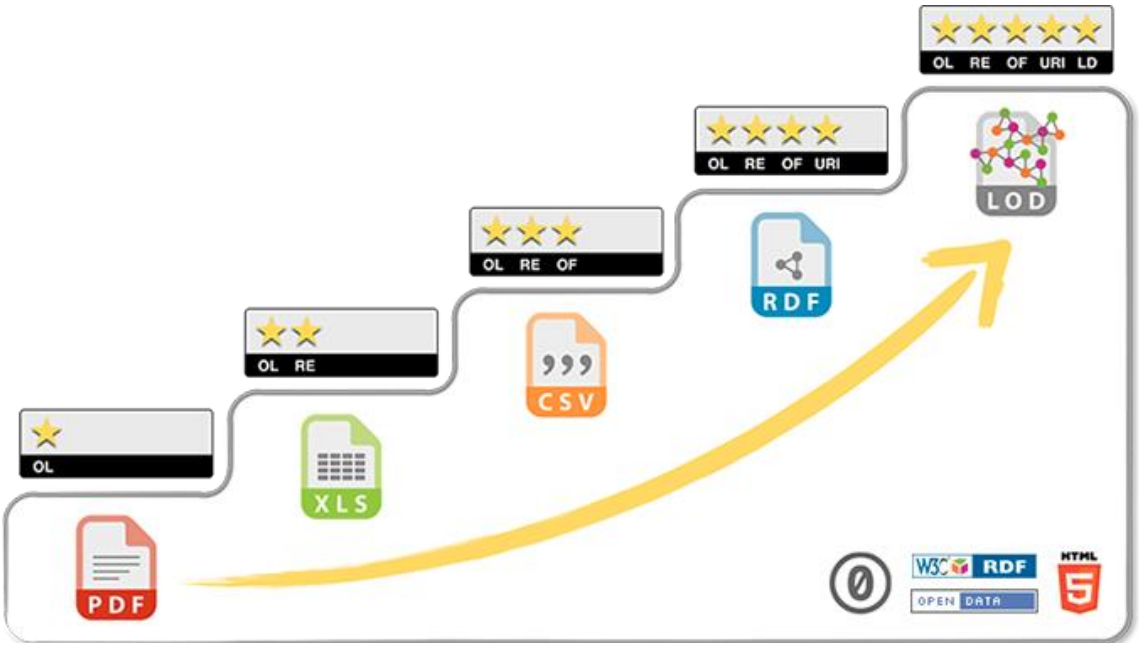
Απεικόνιση του LOD Cloud τον Μάρτιο του 2019 το οποίο περιέχει 1.239 σετ δεδομένων.



Εικόνα 5 LOD Cloud το 2019

2.6.5. 5star LOD

Τα ανοικτά διασυνδεδεμένα δεδομένα 5 αστέρων είναι ένα πρότυπο δημοσίευσης δεδομένων στον παγκόσμιο ιστό [26].



Εικόνα 6 ανοικτά διασυνδεδεμένα δεδομένα 5 αστέρων

Οι κανόνες του προτύπου προϋποθέτουν ο καθένας τους προηγούμενους και περιγράφονται από τον Tim Berners-Lee [27]ως εξης:

★	Διαθέσιμα ανοικτά δεδομένα στον παγκόσμιο ιστό σε οποιαδήποτε μορφή.
★★	Μηχαναγνώσιμα δομημένα δεδομένα
★★★	Δεδομένα σε γενική, μη ιδιωτική μορφή
★★★★	Χρήση ανοικτών προτύπων του W3C (RDF και SPARQL) για χαρακτηρισμό και αναγνώριση των δεδομένων
★★★★★	Διασύνδεση των δεδομένων

2.7. Διαδικτυακοί πληροφοριακοί χώροι πολιτιστικών δεδομένων

2.7.1. Europeana

Η Europeana [28] είναι ένας οργανισμός που ορίστηκε από την Ευρωπαϊκή Επιτροπή με σκοπό την δημιουργία μίας ψηφιακής πλατφόρμας για την πολιτιστική κληρονομιά της Ευρώπης. Η Europeana παρέχει μέσα και τεχνογνωσία για την συνεχή εξάπλωση της πολιτιστικής κληρονομιάς σε ευρύτερο κοινό. Οι συλλογές της Europeana είναι αποτέλεσμα συνεργασίας και αφορούν περισσότερα από 58.490.000 έργα τέχνης, βιβλία, ταινίες και μουσικά κομμάτια από περισσότερα από 3.500 μουσεία, βιβλιοθήκες και γκαλερί.

2.7.2. Ελληνικός Συσσωρευτής Ψηφιακού Περιεχομένου Πολιτιστικής Κληρονομιάς

Το SearchCulture.gr [29] είναι ο Ελληνικός Συσσωρευτής Ψηφιακού Περιεχομένου Πολιτιστικής Κληρονομιάς, ο ενιαίος πολιτιστικός πληροφοριακός χώρος που ανέπτυξε το Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) για την συγκέντρωση και ενοποίηση του συνόλου του Ελληνικού Ψηφιακού Πολιτιστικού Αποθέματος που παράγεται από φορείς με Δημόσια χρηματοδότηση. Μέσω της υποδομής SearchCulture.gr, το ΕΚΤ είναι ο διαπιστευμένος εθνικός συσσωρευτής πολιτιστικών δεδομένων για τη Europeana.
Στόχοι του εγχειρήματος είναι η ελεύθερη κεντρική πρόσβαση και ενιαία διάθεση ψηφιακών πολιτιστικών πόρων -στο πλαίσιο των πολιτικών που ακολουθούν οι φορείς παραγωγής τους- η ανάδειξη και διασύνδεση τους, η ασφαλής διαφύλαξή τους και η ένταξή τους στην Ευρωπαϊκή Ψηφιακή Βιβλιοθήκη Europeana σε συμφωνία με εθνικές και ευρωπαϊκές προτεραιότητες.
Το SearchCulture.gr φιλοξενεί σήμερα περισσότερα από 488.000 τεκμήρια από 59 φορείς και έχει διαθέσει πάνω από 160.000 από αυτά στην Europeana.

2.8. Πολιτιστικά δεδομένα που χρησιμοποιήθηκαν

Αρχική πρόθεση ήταν η χρήση δεδομένων ελληνικών μουσείων, αλλά καθώς η αναζήτηση δημοσιευμένων συνόλων δεδομένων από ελληνικές πηγές δεν ήταν επιτυχής, για την επίτευξη του στόχου της εργασίας χρησιμοποιήθηκαν δεδομένα από συλλογές δυο μεγάλων αμερικάνικων μουσείων τέχνης, του Museum of Modern Art και

του Carnegie Museum of Art. Οι συλλογές αυτές έχουν δημοσιευθεί σε αρχεία της μορφής CSV και JSON στο διαδίκτυο και είναι ελεύθερες προς χρήση υπό την άδεια Creative Commons zero licence (CC0).

2.8.1. Συλλογή Μουσείου Μοντέρνας Τέχνης

Το μουσείο σύγχρονης τέχνης (Museum of Modern Art – MoMA), απέκτησε τα πρώτα του εκθέματα το 1929. Σήμερα το μουσείο διαθέτει σχεδόν 200.000 έργα τέχνης από όλο τον κόσμο, που δημιουργήθηκαν τα τελευταία 150 χρόνια. Η συνεχώς αυξανόμενη συλλογή εκτείνεται στο πεδίο της οπτικής έκφρασης και περιλαμβάνει πίνακες ζωγραφικής, γλυπτά, σκίτσα, φωτογραφίες, αρχιτεκτονικά σχέδια, ταινίες κ.α.

Το MoMA στην προσπάθεια του να συμβάλει στην περαιτέρω διάδοση και κατανόηση και χρήση της συλλογής του δημιούργησε ένα ερευνητικό σύνολο δεδομένων που περιλαμβάνει περισσότερες από 139.000[1] εγγραφές. Το ερευνητικό σύνολο δεδομένων αναπαριστά όλα τα έργα που αποκτήθηκαν από το μουσείο και έχουν καταλογοποιηθεί στη βάση δεδομένων του. Περιέχει βασικά μεταδεδομένα για κάθε έργο, συμπεριλαμβανομένων: τίτλου, καλλιτέχνη, ημερομηνίας δημιουργίας, μέσου και διαστάσεων. Το σύνολο δεδομένων είναι διαθέσιμο σε μορφή CSV και JSON και έχει δημοσιευθεί υπό την άδεια CC0 [30].

2.8.2. Συλλογή Μουσείου Τέχνης Carnegie

Το μουσείο τέχνης Carnegie (Carnegie Museum of Art – CMoA) σαν εορτασμό της 120της επετείου λειτουργίας του προχώρησε στην δημοσίευση όλων των έργων τέχνης που υπήρξαν στην κατοχή του. Η έκδοση περιλαμβάνει πληροφορίες για περισσότερα από 28.000 αντικείμενα από όλα τα τμήματα του μουσείου, που χωρίζονται σε καλές τέχνες, φωτογραφία, σύγχρονη τέχνη και το αρχιτεκτονικό κέντρο Heinz.

Τα δεδομένα περιλαμβάνουν βασικές πληροφορίες και μεταδεδομένα για σχεδόν κάθε έργο που έχει καταλογοποιηθεί στην βάση δεδομένων του μουσείου. Το σύνολο δεδομένων είναι διαθέσιμο σε μορφή CSV και JSON και παρέχεται χωρίς περιορισμούς υπό την άδεια CC0 [31].

[1] το σύνολο που χρησιμοποιήθηκε είναι μια παλαιότερη έκδοση που περιλαμβάνει 128.000 εγγραφές

Κεφάλαιο 3ο
Μετατροπή πολιτιστικών δεδομένων σε RDF και διασύνδεση τους

Στο παρόν κεφάλαιο αναλύεται η διαδικασία που ακολουθήθηκε για την μετατροπή σε μορφή RDF και για την διασύνδεση πολιτιστικών δεδομένων που προέρχονται από τις συλλογές των μουσείων.
Για την μετατροπή των δεδομένων σε RDF (διαδικασία RDFization) επιλέχθηκε η χρήση του μοντέλου δεδομένων της Europeana (EDM) , για συμβατότητα με την Europeana και τον Εθνικό συσσωρευτή πολιτιστικών δεδομένων του Εθνικού Κέντρου Τεκμηρίωσης (SearchCulture.gr).

3.1 Μοντέλο EDM

Το Μοντέλο δεδομένων της Europeana [32] (EDM) έχει σχεδιαστεί με στόχο να αποτελεί ένα μέσο για συλλογή, σύνδεση και εμπλουτισμό των περιγραφών για τα αντικείμενα που παρέχονται στην συλλογή της Europeana. Καθώς δεν είναι εφικτή η διαχείριση ενός ανοιχτού συνόλου που κάθε πάροχος περιεχομένου της Europeana θα προσθέτει τις δικές του περιγραφές, έχει δημιουργηθεί ένα καλά ορισμένο σύνολο στοιχείων που χρησιμοποιείται από το EDM για την επίτευξη του στόχου αυτού. Τα στοιχεία μπορούν να διακριθούν σε δύο κύριες κατηγορίες:

1. Στοιχεία που ορίζονται από το EDM, και
2. Στοιχεία που έχουν δηλωθεί σε άλλους χώρους ονομάτων και επαναχρησιμοποιούνται από το EDM.

Οι χώροι ονομάτων (namespaces) από τους οποίους δανείζεται στοιχεία το EDM είναι οι παρακάτω:

- Resource Description Framework (RDF) και RDF Schema (RDFS) (http://www.w3.org/2000/01/rdf-schema#/)
- OAI Object Reuse and Exchange (ORE) (http://www.openarchives.org/ore/terms)
- Simple Knowledge Organization System (SKOS) (http://www.w3.org/2004/02/skos/core)

- Dublin Core (http://purl.org/dc/elements/1.1, http://purl.org/dc/terms και http://purl.org/dc/dcmitype/)
- W3C Data Catalog Vocabulary (DCAT) (http://www.w3.org/tr/vocab-dcat/)
- Creative Commons (CC) (http://creativecommons.org/ns)
- SIOC Services Ontology Module (http://rdfs.org/sioc/services#3)

Οι βασικές κλάσεις αντικειμένων που ορίζονται στο EDM είναι[33]:

- edm:ProvidedCHO - το παρεχόμενο αντικείμενο πολιτιστικής κληρονομάς
- edm:WebResource - ο διαδικτυακός πόρος που αποτελεί την ψηφιακή αναπαράσταση του αντικειμένου
- ore:Aggregation - η συνάθροιση που ομαδοποιεί τις παραπάνω κλάσεις

Οι κύριες κλάσεις συνάφειας περιεχομένου που χρησιμοποιούνται απαντούν στα εξής ερωτήματα:

- edm:Agent – Ποιος;
- edm:Place – Που;
- edm:TimeSpan – Πότε;
- skos:Concept – Τι;
- cc:License - Πρόσβαση και χρήση

3.2 Εποπτεία, καθαρισμός δεδομένων και RDFization

Τα δεδομένα που χρησιμοποιήθηκαν από τη συλλογή του MoMA βρίσκονται σε αρχείο μορφής CSV που περιλαμβάνει τα παρακάτω χαρακτηριστικά για τα εκθέματα και τους δημιουργούς τους.

MoMA Collection	
Χαρακτηριστικά εκθέματος	**Χαρακτηριστικά Καλλιτέχνη**
Title	Artist
Date	ConstituentID

MoMA Collection	
Medium	ArtistBio
Dimensions	Nationality
CreditLine	BeginDate
AccessionNumber	EndDate
Classification	Gender
Department	
DateAcquired	
Cataloged	
ObjectID	
URL	
ThumbnailURL	
Circumference (cm)	
Depth (cm)	
Diameter (cm)	
Height (cm)	
Length (cm)	
Weight (kg)	

MoMA Collection	
Width (cm)	
Seat Height (cm)	
Duration (sec.)	

Πίνακας 1 Τύποι δεδομένων της συλλογής MoMA

Τα δεδομένα που χρησιμοποιήθηκαν από την συλλογή CMoA είναι αποθηκευμένα σε μορφή JSON και είναι κατηγοριοποιημένα ως εξής:

CMoA Collection	
Χαρακτηριστικά εκθέματος	**Χαρακτηριστικά Καλλιτέχνη**
title	artist_id
creation_date	party_type
creation_date_earliest	full_name
creation_date_latest	cited_name
medium	role
accession_number	nationality
id	birth_date
credit_line	death_date
date_acquired	birth_place
department	death_place

CMoA Collection	
physical_location	
item_width	
item_height	
item_depth	
item_diameter	
web_url	
provenance_text	
classification	
image_url	

Πίνακας 2 Τύποι δεδομένων της συλλογής CMoA

Με σκοπό την ομογενοποίηση των δεδομένων και την διασύνδεσή τους κάτω από ένα κοινό RDF γράφο, επιλέχθηκαν τα χαρακτηριστικά που θα χρησιμοποιηθούν στην διαδικασία μετατροπής των δεδομένων σε RDF, με κριτήριο να είναι κοινά και στα δύο σύνολα και να αποδίδονται από το μοντέλο EDM.

Επίσης ακολουθώντας τις κατευθυντήριες του EDM και των κανόνων για τα ανοιχτά διασυνδεδεμένα δεδομένα, δημιουργήθηκαν κλάσεις και URIs βάσει των μοναδικών αναγνωριστικών κάθε εγγραφής. Το σύνολο των ιδιοτήτων κάθε συλλογής που τελικά μετατράπηκαν σε μορφή RDF αντιπαραβάλλεται στον παρακάτω πίνακα.

Κοινά Χαρακτηριστικά	
MoMA Collection	**CMoA Collection**

Κοινά Χαρακτηριστικά	
Title	title
Date	creation_date
Medium	medium
ObjectID	id
Classification	classification
URL	web_url
ThumbnailURL	image_url
ConstituentID	artist_id
Artist	full_name
ArtistBio	nationality
BeginDate	birth_date
EndDate	death_date
AggregationID	AggregationID
Aggregation_uri	Aggregation_uri
Artist_uri	Artist_uri
Artwork_uri	Artwork_uri

Πίνακας 3 Τύποι δεδομένων των συλλογών που χρησιμοποιήθηκαν

Σύμφωνα με τις προδιαγραφές του EDM, για την περιγραφή των πόρων έγινε χρήση των παρακάτω, κλάσεων, ιδιοτήτων και λεξιλογίων και αντιστοιχήθηκε με τα χαρακτηριστικά ως εξής:

Europeana Data Model edm: <http://www.europeana.eu/schemas/edm/>

- edm:ProvidedCHO --> Artwork_uri
- edm:EuropeanaAggregation --> Aggregation_uri
- edm:Agent --> Artist_uri
- edm:begin --> BeginDate, birth_date
- edm:end --> EndDate, death_date
- edm:hasView --> ThumbnailURL, image_url

Dublin Core dc: <http://purl.org/dc/elements/1.1/>

- dc:date --> Date, creation_date
- dc:title --> Title, title
- dc:type --> Classification, classification
- dc:creator --> Artist_uri

Open Archives Initiative Object Reuse and Exchange ore:
<http://www.openarchives.org/ore/terms/>

- ore:isAggregatedBy --> AggregationID

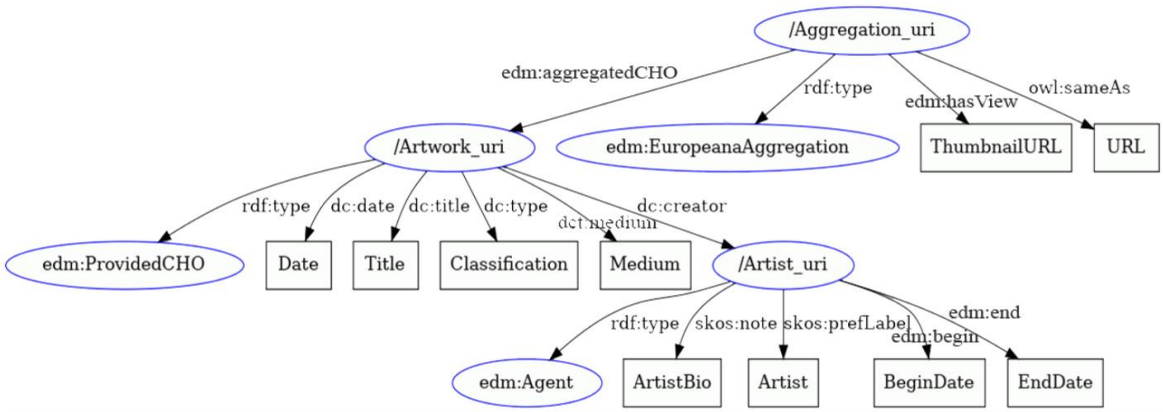
DCMI Metadata Terms dct: <http://purl.org/dc/terms/>

- dct:medium --> Medium, medium

Simple Knowledge Organization System skos:
<http://www.w3.org/2004/02/skos/core#>

- skos:note --> ArtistBio, nationality
- skos:prefLabel --> Artist, full_name

Τελικά το μοντέλο αντιστοίχισης παίρνει τη μορφή



Εικόνα 7 Διαμορφωμένο μοντέλο σημασιολογικής αντιστοίχισης δεδομένων μουσείων

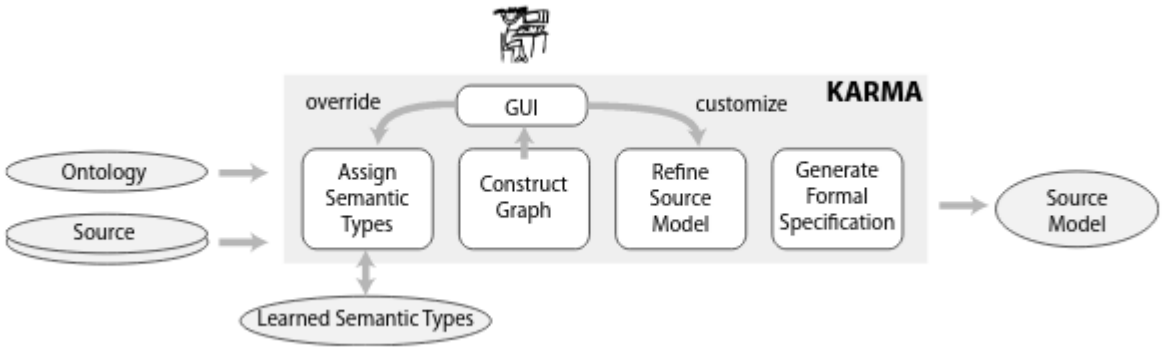
3.3 Εργαλεία μετατροπής και διασύνδεσης δεδομένων

3.3.1 RDFization

Ως RDFization ορίζεται η διαδικασία μετατροπής δεδομένων από μορφή πινάκων ή σχεσιακών βάσεων δεδομένων σε RDF και απαιτεί αρχικά την επιλογή των τιμών που θα περιέχονται στον παραγόμενο γράφο, τον χαρακτηρισμό του τύπου τους, για παράδειγμα εάν πρόκειται για αλφαριθμητικό ή URI και την δημιουργία ή μεταφόρτωση του μοντέλου αντιστοίχισης στην εκάστοτε εφαρμογή.

Στη συνέχεια αναλύονται τρία εργαλεία που χρησιμοποιήθηκαν για αυτό το σκοπό και παρουσιάζονται τα αποτελέσματα της χρήσης τους. Όλες οι διαδικασίες εκτελέστηκαν σε φορητό υπολογιστή με επεξεργαστή i7 2ης γενιάς, μνήμη RAM 6GB και λειτουργικό σύστημα Ubuntu 18.04

KARMA



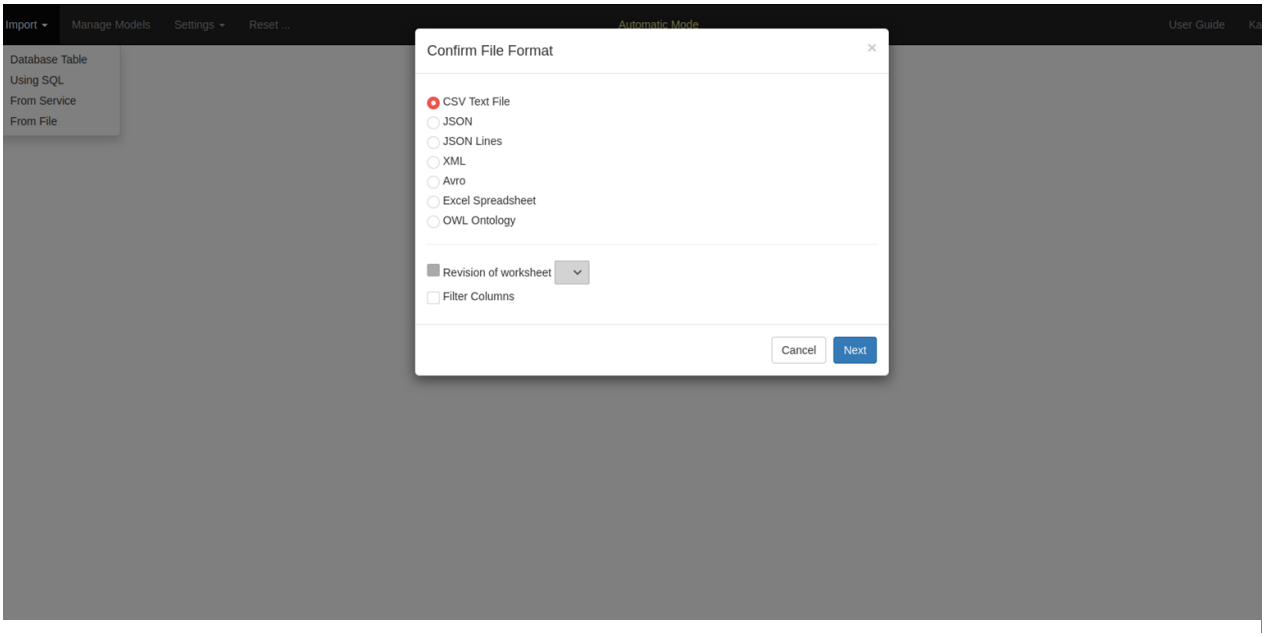
Εικόνα 8 Διαδικασία μοντελοποίησης δεδομένων στο KARMA

Το KARMA είναι ένα εργαλείο ανοιχτού κώδικα για την ενοποίηση δεδομένων/πληροφορίας. Δίνει τη δυνατότητα στους χρήστες του να επεξεργάζονται και να ενοποιούν σημασιολογικά, δεδομένα που προέρχονται από διάφορες πηγές όπως βάσεις δεδομένων, Web APIs ή αρχεία XML, CSV, JSON. Αποτελεί ερευνητικό έργο του Ινστιτούτου Επιστήμης της Πληροφορικής του Πανεπιστημίου Νότιας Καλιφόρνια και δημοσιεύθηκε στο Extended Semantic Web Conference το 2012 με τίτλο 'Semi-Automatically Mapping Structured Sources into the Semantic Web' [34]. Η πρακτική εφαρμογή του KARMA που δημοσιεύθηκε στο ESWC'2013 με θέμα τον σημασιολογικό χαρακτηρισμό του αρχείου του μουσείου τέχνης Smisthonian και την σύνδεση του με τα ανοιχτά διασυνδεδεμένα δεδομένα [35], αποτέλεσε μια από τις αφορμές για την εκπόνηση της παρούσας διπλωματικής εργασίας.

Το KARMA παρέχει μια γραφική διεπαφή που επιτρέπει αρχικά την εισαγωγή του συνόλου ή μέρους των δεδομένων στο σύστημα και στη συνέχεια την εύκολη προσαρμογή τους στην επιθυμητή κατάσταση. Μερικές από τις δυνατότητες προσαρμογής είναι η προσθήκη, αφαίρεση ή αναδιάταξη στηλών, ο διαχωρισμός δεδομένων ή η μετατροπή δεδομένων με χρήση εκφράσεων βασισμένων στην γλώσσα προγραμματισμού Python. Για παράδειγμα η στήλη που παρέχει το μοναδικό κωδικό id κάθε εγγραφής μπορεί να αντιγραφεί σε μια νέα στην οποία θα εφαρμοστεί ένας κανόνας που προσθέτει πρόθεμα ώστε να δημιουργηθεί μοναδικό URI για τον πόρο που θα δημιουργηθεί τελικά (για id:1 και προθέματος: 'example.org' προκύπτει το URI: 'example.org/1/').

Το σύστημα επιτρέπει την μεταφόρτωση και χρήση πολλών οντολογιών ταυτόχρονα για τον σημασιολογικό χαρακτηρισμό των δεδομένων. Ο χρήστης επιλέγει γραφικά την αντιστοίχιση κάθε στήλης με την κλάση ή ιδιότητα των οντολογιών, το σύστημα παράγει τον γράφο ενώ παράλληλα, με χρήση τεχνικών και αλγορίθμων εκμάθησης είναι σε θέση να προτείνει με επιτυχία σημασιολογικούς τύπους για τα δεδομένα, απλοποιώντας ακόμη περισσότερο την διαδικασία της αντιστοίχισης. Όταν ολοκληρωθεί το μοντέλο, μπορεί να εξαχθεί γραφικά ή σε αρχείο, και τα ενοποιημένα δεδομένα μπορούν να δημοσιευθούν σε μορφή RDF για διασύνδεση ή αποθήκευσή τους.

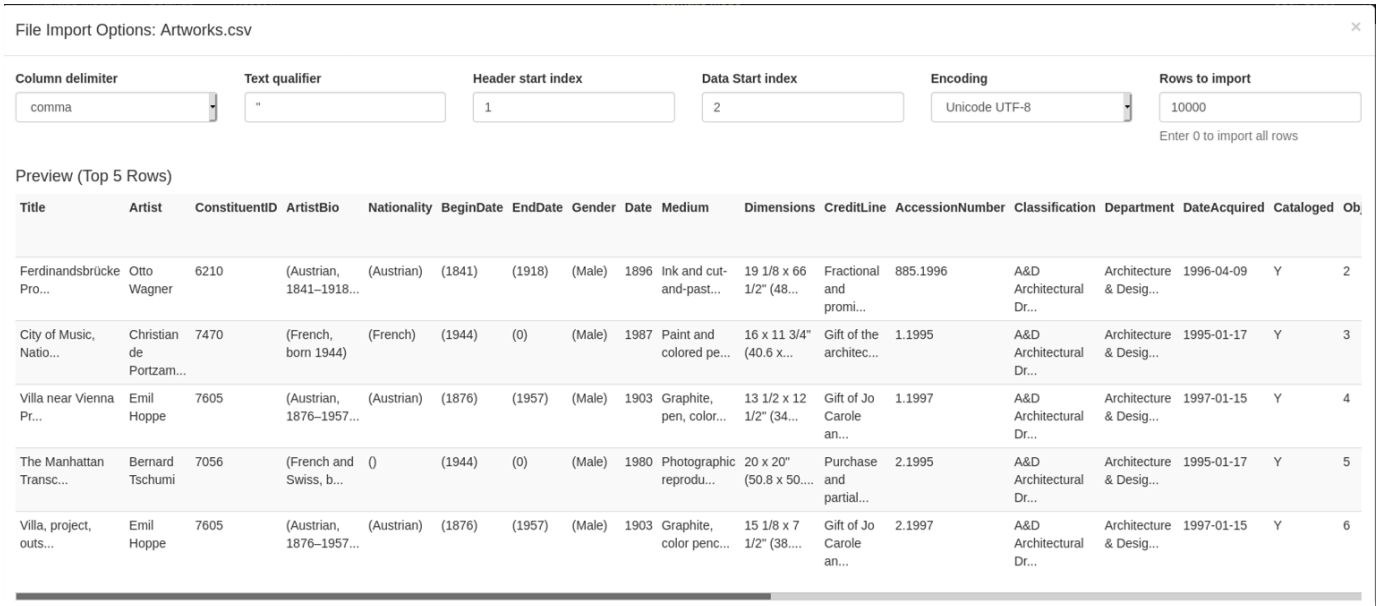
Χρήση του KARMA



Εικόνα 9 Εισαγωγή δεδομένων στο KARMA

Η πρώτη ενέργεια στην εφαρμογή του KARMA είναι η φόρτωση των δεδομένων. Μετά την επιλογή του αρχείου εμφανίζεται ένα παράθυρο διαλόγου που ζητάει την επιβεβαίωση του τύπου του αρχείου που πρόκειται να φορτωθεί στην εφαρμογή.

Στη συνέχεια σε επόμενο παράθυρο εμφανίζεται μια προεπισκόπηση των δεδομένων και ο χρήστης μπορεί να επιλέξει αν θα παραληφθούν κάποιες από τις εγγραφές ή αν θα συμπεριληφθούν οι κεφαλίδες των στηλών. Αρκετά εξυπηρετική ιδιότητα αποτελεί η



Title	Artist	ConstituentID	ArtistBio	Nationality	BeginDate	EndDate	Gender	Date	Medium	Dimensions	CreditLine	AccessionNumber	Classification	Department	DateAcquired	Cataloged	Ob
Ferdinandsbrücke Pro...	Otto Wagner	6210	(Austrian, 1841–1918...	(Austrian)	(1841)	(1918)	(Male)	1896	Ink and cut-and-past...	19 1/8 x 66 1/2" (48...	Fractional and promi...	885.1996	A&D Architectural Dr...	Architecture & Desig...	1996-04-09	Y	2
City of Music, Natio...	Christian de Portzam...	7470	(French, born 1944)	(French)	(1944)	(0)	(Male)	1987	Paint and colored pe...	16 x 11 3/4" (40.6 x...	Gift of the architec...	1.1995	A&D Architectural Dr...	Architecture & Desig...	1995-01-17	Y	3
Villa near Vienna Pr...	Emil Hoppe	7605	(Austrian, 1876–1957...	(Austrian)	(1876)	(1957)	(Male)	1903	Graphite, pen, color...	13 1/2 x 12 1/2" (34...	Gift of Jo Carole an...	1.1997	A&D Architectural Dr...	Architecture & Desig...	1997-01-15	Y	4
The Manhattan Transc...	Bernard Tschumi	7056	(French and Swiss, b...	()	(1944)	(0)	(Male)	1980	Photographic reprodu...	20 x 20" (50.8 x 50....	Purchase and partial...	2.1995	A&D Architectural Dr...	Architecture & Desig...	1995-01-17	Y	5
Villa, project, outs...	Emil Hoppe	7605	(Austrian, 1876–1957...	(Austrian)	(1876)	(1957)	(Male)	1903	Graphite, color penc...	15 1/8 x 7 1/2" (38....	Gift of Jo Carole an...	2.1997	A&D Architectural Dr...	Architecture & Desig...	1997-01-15	Y	6

Εικόνα 10 Προεπισκόπηση δεδομένων στο KARMA

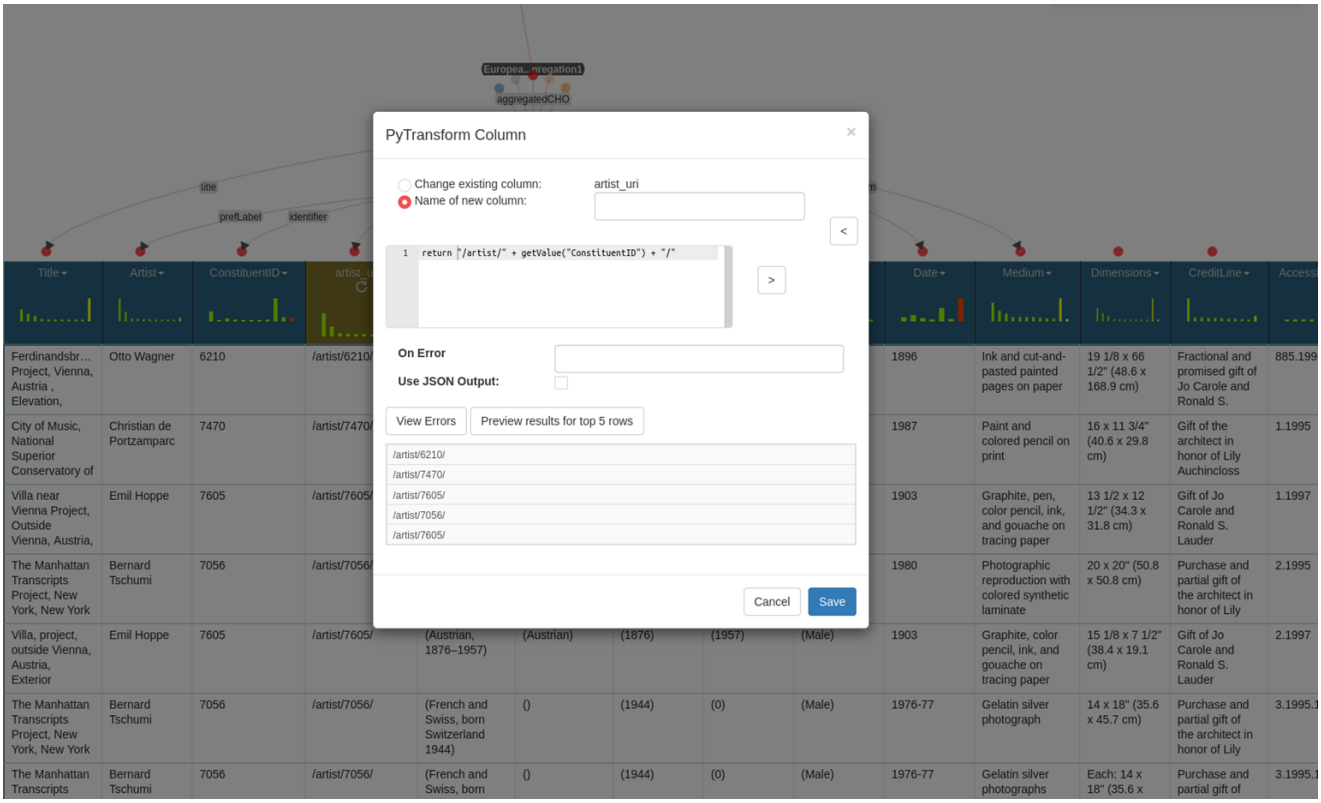
δυνατότητα επιλογής μέρους ή του συνόλου των δεδομένων προς επεξεργασία καθώς η διαδικασία μπορεί να επιταχυνθεί χρησιμοποιώντας ένα δοκιμαστικό υποσύνολο των δεδομένων. Έπειτα αφού ολοκληρωθεί η δημιουργία του μοντέλου υπάρχει δυνατότητα αποθήκευσης και εφαρμογής του σε ολόκληρο το σύνολο.

Επόμενο βήμα είναι η εξεργασία των δεδομένων και δημιουργία νέων στηλών με αναγνωριστικά και URIs που θα μπορούν να αντιστοιχηθούν στις κλάσεις που ορίστηκαν προηγουμένως. Στο παράδειγμα της εικόνας 11 φαίνεται η δημιουργία στήλης βασισμένη στο μοναδικό αναγνωριστικό του καλλιτέχνη, με χρήση της επιλογής 'PyTransform' της εφαρμογής. Ο μετασχηματισμός που εφαρμόζεται είναι

return "/artist/" + "getValue('ConstituentID')+ "/".

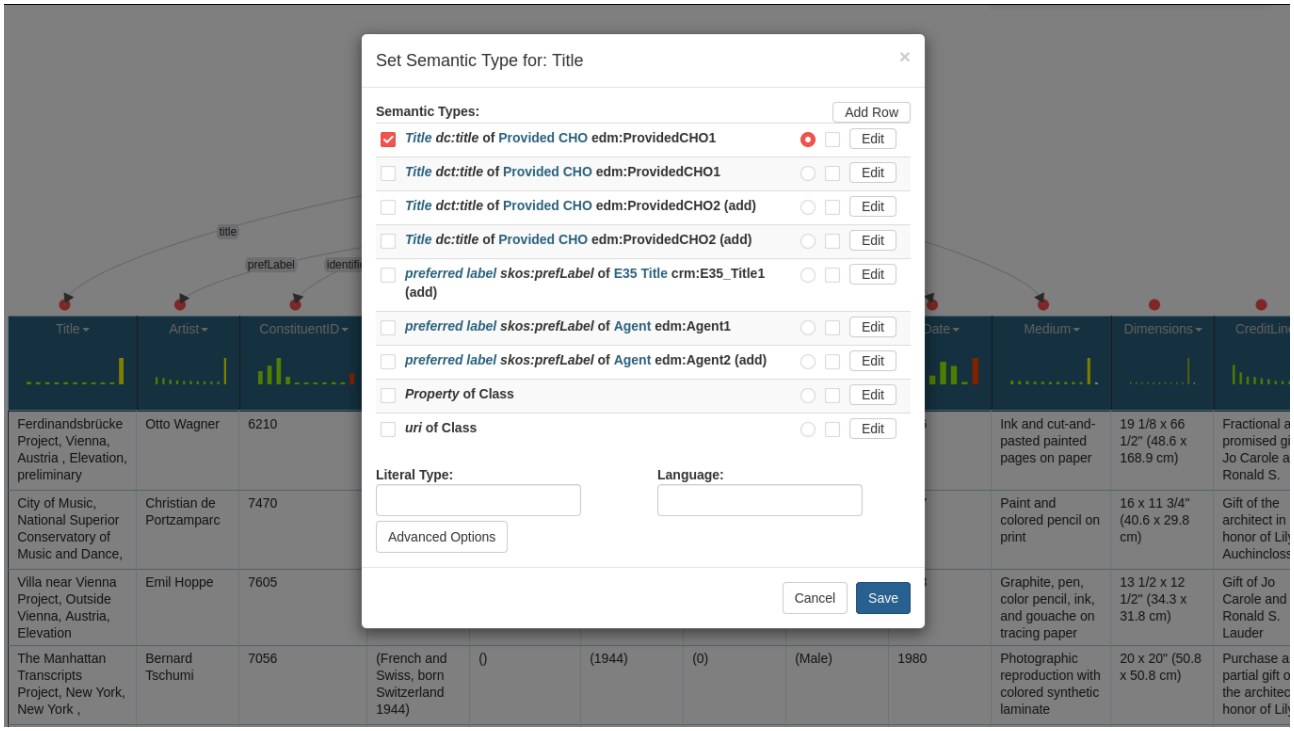
Έτσι οι τιμές της νέας στήλης θα περιλαμβάνουν το λεκτικό '/artist/' συνοδευόμενο από την αντίστοιχη τιμή της στήλης ConstituentID, όπως φαίνεται στα παραδείγματα εφαρμογής του μετασχηματισμού στο παράθυρο. Το KARMA επιτρέπει πιο πολύπλοκες εκφράσεις με χρήση της γλώσσα προγραμματισμού Python που εξυπηρετούν πιο σύνθετους μετασχηματισμούς, ανάλογα με το επιθυμητό αποτέλεσμα για τα δεδομένα. Στη συνέχεια αυτή η στήλη θα επισημανθεί σημασιολογικά ως URI που καταδεικνύει την κλάση του καλλιτέχνη. Κατά συνέπεια στο τελικό αρχείο που θα εξαχθεί, στην τιμή θα προστεθεί το πρόθεμα που περιλαμβάνει τον διαδικτυακό τομέα (domain) στον οποίο πρόκειται να δημοσιευθούν τα δεδομένα, πχ. *example.org/artist/6210* ή στην δικιά μας

περίπτωση *localhost:3000/artist/6210*, καθώς τα δεδομένα θα παραμείνουν τοπικά.



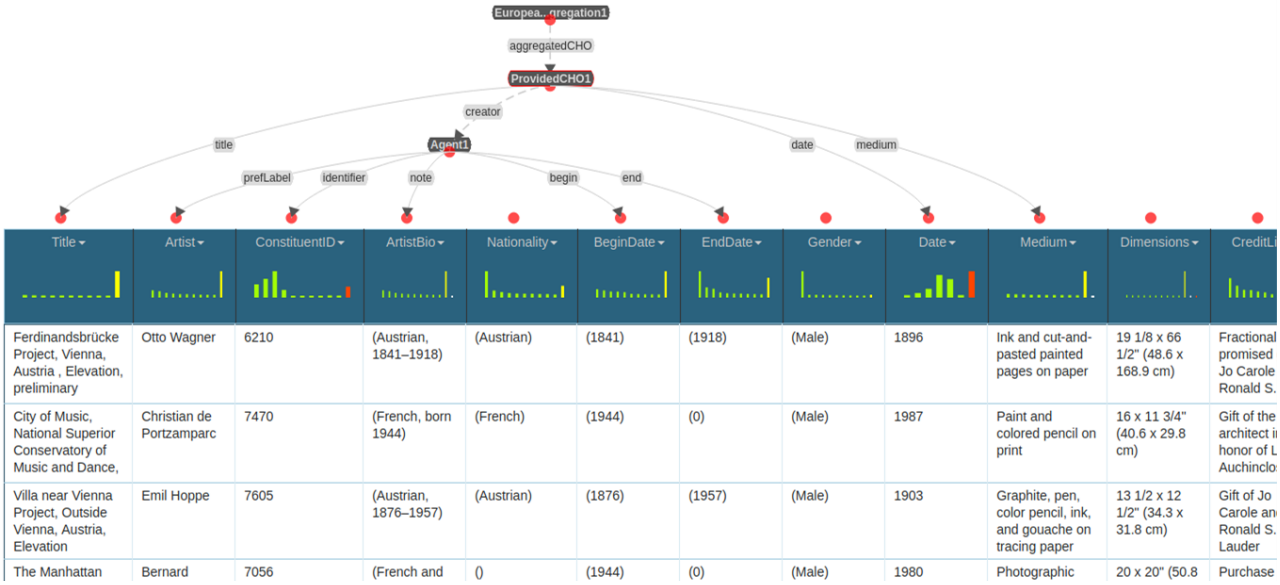
Εικόνα 11 Δημιουργία νέας στήλης και εφαρμογή μετασχηματισμού

Αφού ολοκληρωθεί η επεξεργασία των δεδομένων, επόμενο βήμα είναι ο σημασιολογικός χαρακτηρισμός τους. Το KARMA προτείνει σημασιολογικούς τύπους για τα δεδομένα βασισμένο σε προγενέστερες αντιστοιχίσεις ή βάση λεκτικής ομοιότητας των στοιχείων που περιγράφουν τα δεδομένα με τα στοιχεία του οντολογικού σχήματος, όπως φαίνεται στο παράδειγμα της εικόνας 12.



Εικόνα 12 Σημασιολογικός χαρακτηρισμός δεδομένων

Σε περίπτωση που καμία από τις προτάσεις δεν είναι ορθή, γίνεται επιλογή της ιδιότητας και της κλάσης, από τα λεξιλόγια και τις οντολογίες που έχουν μεταφορτωθεί από το χρήστη στην εφαρμογή. Σημειώνεται ότι για συχνά χρησιμοποιούμενες οντολογίες δίνεται η δυνατότητα αυτόματης φόρτωσης κατά την εκκίνηση.

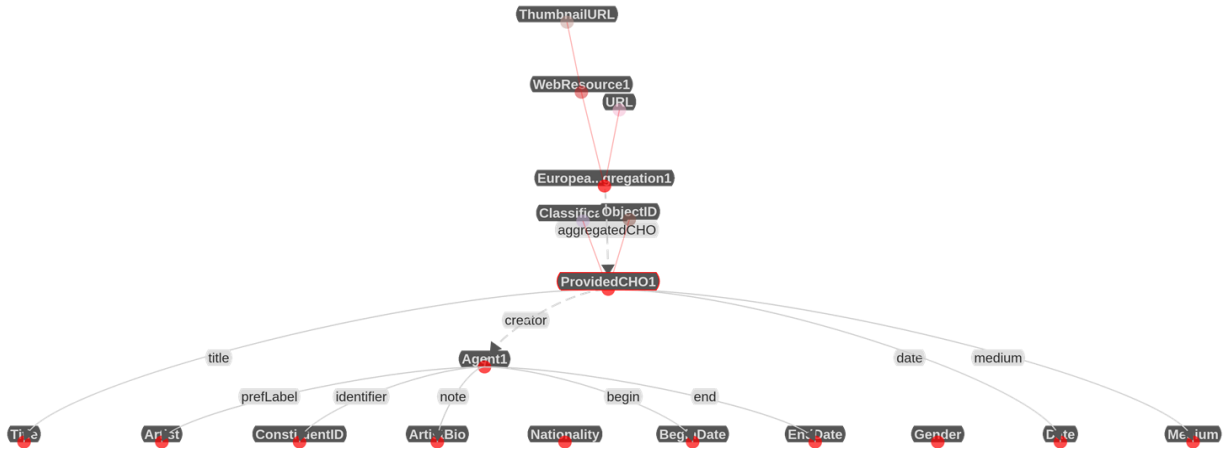


Εικόνα 13 Δημιουργία μοντέλου στο KARMA

Κατά τη διάρκεια χαρακτηρισμού των δεδομένων δημιουργείται γραφικά και το μοντέλο. Η εποπτεία του μοντέλου βοηθάει αποφυγή λαθών και στην καλύτερη κατανόησή του, ειδικά όταν αυτό χτίζεται με βάση τα διαθέσιμα δεδομένα.

Αφού ολοκληρωθεί η διαδικασία της αντιστοίχισης και έχει συμπληρωθεί το μοντέλο, υπάρχει επιλογή για εξαγωγή του χωριστά από τα δεδομένα για μετέπειτα χρήση και εφαρμογή σε άλλα σύνολα δεδομένων.

Κατά την ολοκλήρωση της διαδικασίας πραγματοποιήθηκε μετατροπή των δεδομένων του αρχείου MoMA.csv (~128.000 εγγραφών)σε RDF, η οποία διήρκεσε 1’ και 55”. Αντίστοιχη διαδικασία ακολουθήθηκε για το σύνολο δεδομένων της συλλογής του CMoA (~28.000 εγγραφών), από μορφή JSON σε RDF, η οποία διήρκεσε 50”.



Εικόνα 14 Εξαγωγή μοντέλου του KARMA σε μορφή εικόνας

OpenRefine

Το OpenRefine [36] είναι μια πλατφόρμα εξερεύνησης, καθαρισμού και διασύνδεσης δεδομένων. Είναι βασισμένο σε Java και εκτελείται τοπικά. Αναπτύχθηκε από την Metaweb Technologies Inc. (η οποία αγοράστηκε από την Google Inc. και το 2010 το προϊόν μετονομάστηκε σε Google Refine). Τελικά, η Google διέκοψε την συντήρησή του το 2012, έτσι μετονομάστηκε σε OpenRefine και πλέον υποστηρίζεται από διεθνή κοινότητα μηχανικών λογισμικού και χρηστών.

Προσφέρει πληθώρα τρόπων για εξερεύνηση, εφαρμογή προηγμένων μετασχηματισμών, ταξινόμηση, φιλτράρισμα και ομαδοποίηση δεδομένων. Η εκτέλεση εργασιών επεξεργασίας δεδομένων γίνεται με χρήση της ενσωματωμένης γλώσσας

General Refine Expression Language - GREL [37] Επίσης υποστηρίζεται η δυνατότητα διασύνδεσης των δεδομένων με χρήση της ενσωματωμένης υπηρεσίας reconciliation, η οποία επιτρέπει την δημιουργία συνδέσμων μεταξύ επιλεγμένων συνόλων δεδομένων.

Το εργαλείο OpenRefine υποστηρίζει τη χρήση διάφορων επεκτάσεων. Μια από τις ευρέως διαδεδομένες, η οποία χρησιμοποιήθηκε στην παρούσα εργασία, είναι η rdf-extension [38].

Με την προσθήκη της συγκεκριμένης επέκτασης καθίσταται δυνατή η εξαγωγή δεδομένων σε μορφή RDF. Συγκεκριμένα προστίθεται ένα γραφικό περιβάλλον με το οποίο ο χρήστης μπορεί να εισάγει οντολογίες και να αντιστοιχίσει τα δεδομένα σε γράφους. Ακόμη, παρέχεται η δυνατότητα σύνδεσης των δεδομένων με διαδικτυακές πηγές που προσφέρουν SPARQL endpoints όπως η DBpedia, Wikidata κα.

Χρήση OpenRefine

Η διαδικασία που ακολουθείται στο OpenRefine για μετατροπή των δεδομένων σε μορφή RDF είναι ανάλογη με εκείνη του που περιγράψαμε για το εργαλείο KARMA. Αρχικά γίνεται φόρτωση των δεδομένων στην εφαρμογή και παρέχονται οι πρώτες επιλογές καθαρισμού, όπως απαλοιφή στηλών ή κενών γραμμών από τους πίνακες.

28269 records

Show as: rows **records** Show: 5 10 **25** 50 records

All		id	AggregationID	title	medium	web_url	department
1.		692a68c5-af1e-4124-80f1-cbf38be51abe/artwork	692a68c5-af1e-4124-80f1-cbf38be51abe	Keith Haring	gelatin silver print	http://collection.cmoa.org/CollectionDetail.aspx?item=1	Contemporary Art
2.		871505af-a072-4a84-ab09-3abf932951b8/artwork	871505af-a072-4a84-ab09-3abf932951b8	Untitled	oil on canvas	http://collection.cmoa.org/CollectionDetail.aspx?item=2	Contemporary Art
3.		fe93f7e2-e4ca-455e-bdfa-ea9be22cce8a/artwork	fe93f7e2-e4ca-455e-bdfa-ea9be22cce8a	Trans East West (Tew) No. 5: The Attack of the Embassy (Reconstruction)	mixed media on paper	http://collection.cmoa.org/CollectionDetail.aspx?item=3	Contemporary Art
4.		9ee75ef7-d8a6-47bc-a0fe-dda1ec7d2984/artwork	9ee75ef7-d8a6-47bc-a0fe-dda1ec7d2984	Trans East West (Tew) No. 6: Rebuilding Beyrut	mixed media on paper	http://collection.cmoa.org/CollectionDetail.aspx?item=4	Contemporary Art
5.		d4e7057c-6f1f-4bce-be3e-579fe5ab2af6/artwork	d4e7057c-6f1f-4bce-be3e-579fe5ab2af6	Trans East West (Tew) No. 7: Bombing the Power Plant	mixed media on paper	http://collection.cmoa.org/CollectionDetail.aspx?item=5	Contemporary Art
6.		96622b88-ed5f-44cf-8493-852c4b53fbad/artwork	96622b88-ed5f-44cf-8493-852c4b53fbad	Trans East West (Tew) No. 8: Suddenly It was Completely Dark (Maybe Because I Closed My Eyes)	mixed media on paper	http://collection.cmoa.org/CollectionDetail.aspx?item=6	Contemporary Art
7.		41ee79b1-22cd-43c1-b112-31c63ba5cc68/artwork	41ee79b1-22cd-43c1-b112-31c63ba5cc68	Trans East West (Tew) No. 9: The "Meridien" Will Have a Nice View	mixed media on paper	http://collection.cmoa.org/CollectionDetail.aspx?item=7	Contemporary Art

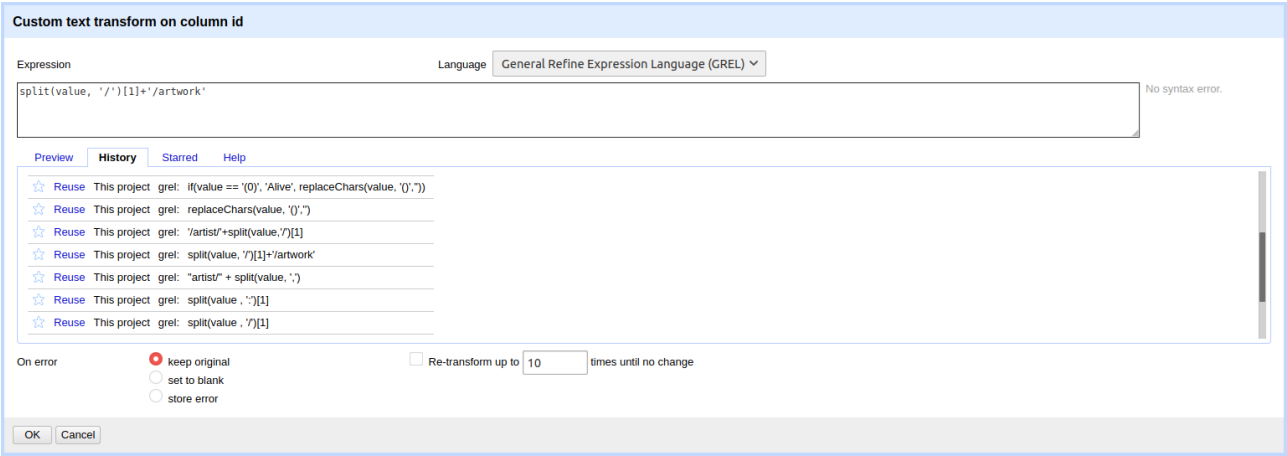
Εικόνα 15 Προβολή δεδομένων στο OpenRefine

Παραδείγματα καθαρισμού και μετασχηματισμού των δεδομένων με χρήση εκφράσεων στην γλώσσα GREL παρουσιάζονται στις εικόνες 16 και 17.

Αρχικά γίνεται εφαρμογή του τύπου στα δεδομένα της στήλης *id* που αποτελούν τα αναγνωριστικά των πόρων των έργων τέχνης

split(value, '/')[1]+'/artwork'

ο οποίος μετατρέπει τιμές της στήλης, από τη μορφή 'cmoa:things/871505af-a072-4a84-ab09-3abf932951b8' σε '871505af-a072-4a84-ab09-3abf932951b8/artwork', χωρίζοντας την τιμή της συμβολοσειράς με βάση την κάθετο και προσθέτοντας στο δεύτερο μισό το *'/artwork'*, ώστε να έρθει στην επιθυμητή μορφή για να μετατραπεί στην συνέχεια σε URI, εισάγοντας και το πρόθεμα του domain.

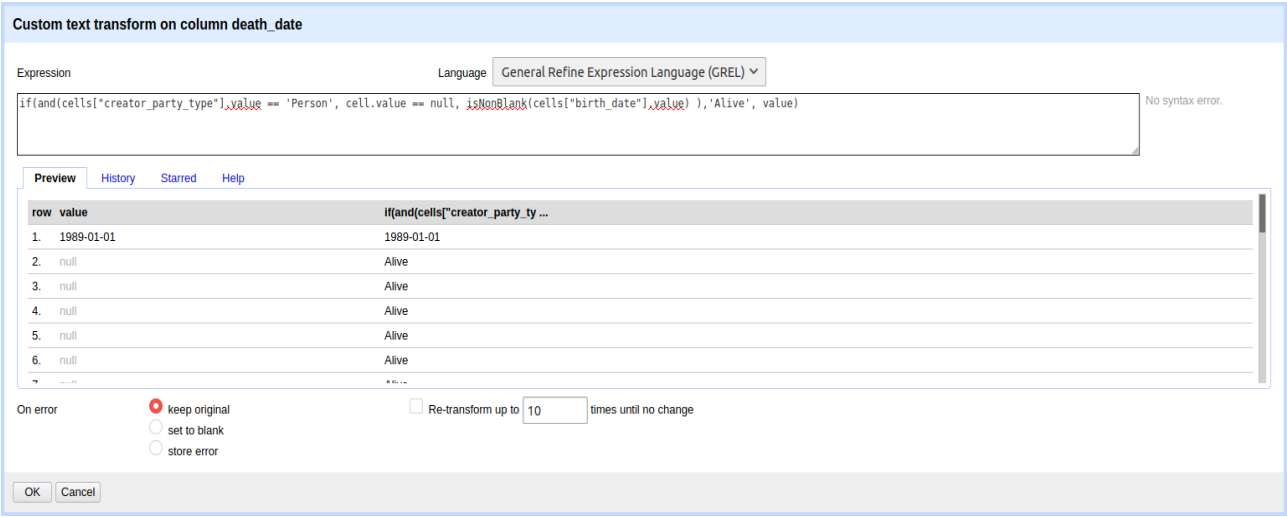


Εικόνα 16 Παράδειγμα 1ο μετατροπής δεδομένων στο OpenRefine

Κατά την εποπτεία των δεδομένων παρατηρήθηκε ότι στο πεδίο που αφορά στην ημερομηνία θανάτου του δημιουργού, η τιμή ήταν κενή στην περίπτωση που εκείνος είναι ακόμη εν ζωή, όταν σαν δημιουργός εμφανίζεται κάποιος οργανισμός αλλά και όταν δεν υπάρχουν στοιχεία. Έτσι εφαρμόστηκε μια πιο σύνθετη μετατροπή τιμών,

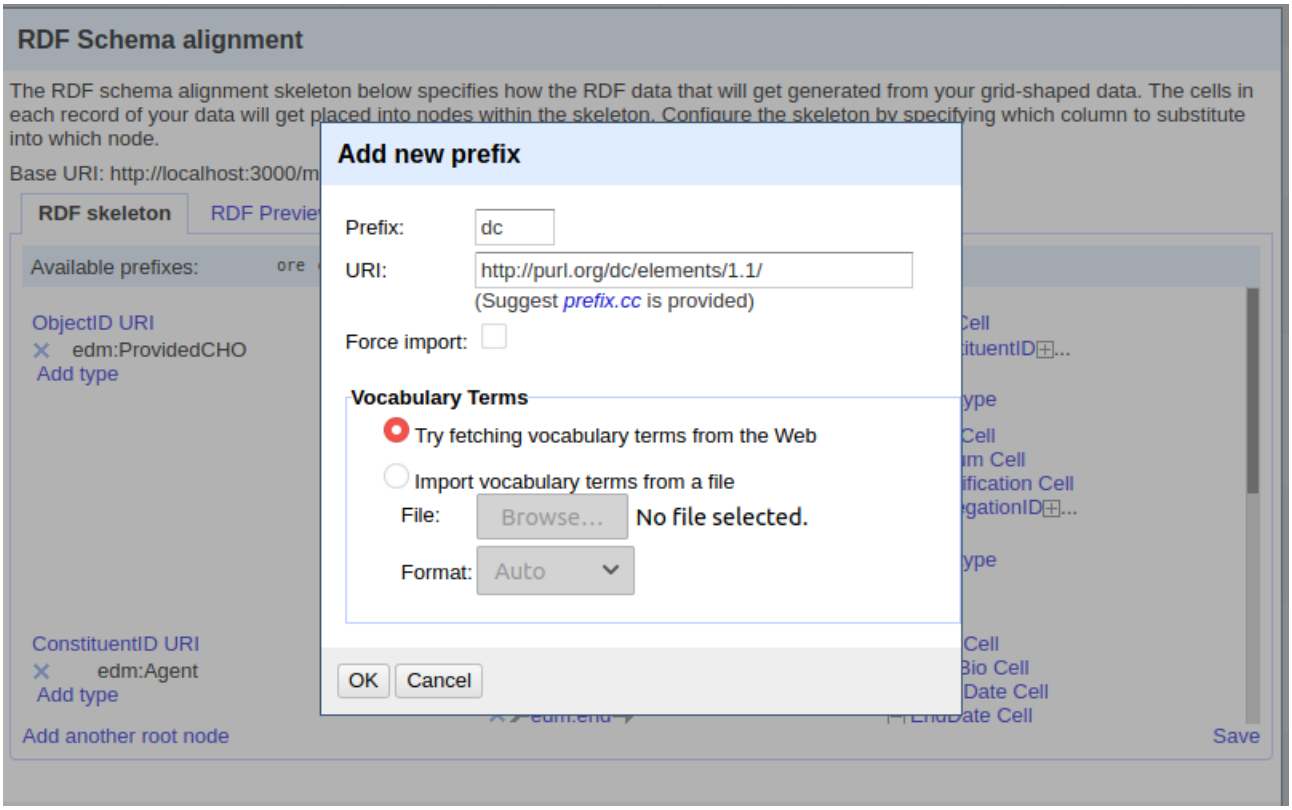
if(and(cells["creator_party_type"].value == 'Person', cell.value == null,
isNotBlank(cells["birth_date"],value)), 'Alive', value)

που τοποθετούσε την τιμή 'Alive' στα πεδία των οποίων η τιμή ήταν κενή, υπήρχαν πληροφορίες για την ημερομηνία γέννησης και όπου ο δημιουργός χαρακτηριζόταν ως άτομο, ενώ αν δεν επαληθευόταν κάποια από τις συνθήκες τη τιμή του πεδίου παρέμενε ως είχε.



Εικόνα 17 Παράδειγμα 2ο μετατροπής δεδομένων στο OpenRefine

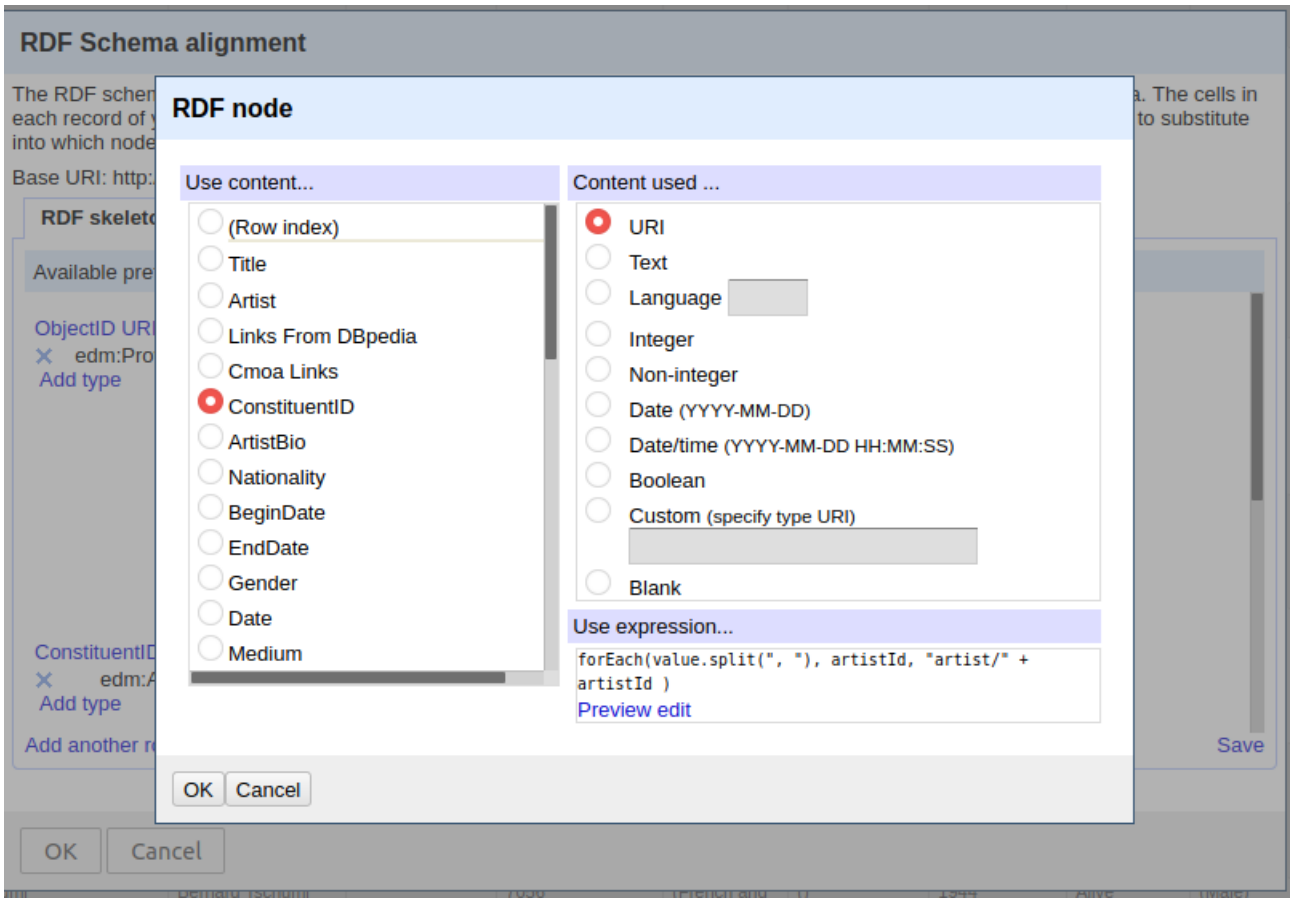
Αφού ολοκληρωθεί ο καθαρισμός και μετασχηματισμός των δεδομένων, επόμενο στάδιο είναι η χρήση της επέκτασης rdf-extension . Αρχικά εισάγονται τα λεξιλόγια και οι οντολογίες που θα χρησιμοποιηθούν. Για ευρέως γνωστές και διαδεδομένες οντολογίες αρκεί να επιλεχθεί μόνο το πρόθεμα, και η εφαρμογή θα παραθέσει αυτόματα το URI, ενώ η οντολογία θα αναζητηθεί και θα μεταφορτωθεί από το



Εικόνα 18 Προσθήκη χώρων ονομάτων στο OpenRefine

διαδίκτυο. Σε διαφορετική περίπτωση υπάρχει δυνατότητα μεταφόρτωσης της οντολογίας από τοπικά αποθηκευμένο αρχείο.

Στη συνέχεια αναπτύσσεται το μοντέλο που θα εφαρμοστεί στα δεδομένα για τη δημιουργία του RDF γράφου. Σε κάθε δυνητικό κόμβο του γράφου αντιστοιχίζεται μία τιμή από τις στήλες των δεδομένων και επιλέγεται ο τύπος της, που μπορεί να λαμβάνει τιμές όπως κείμενο, αριθμός, ημερομηνία, URI κ.α.

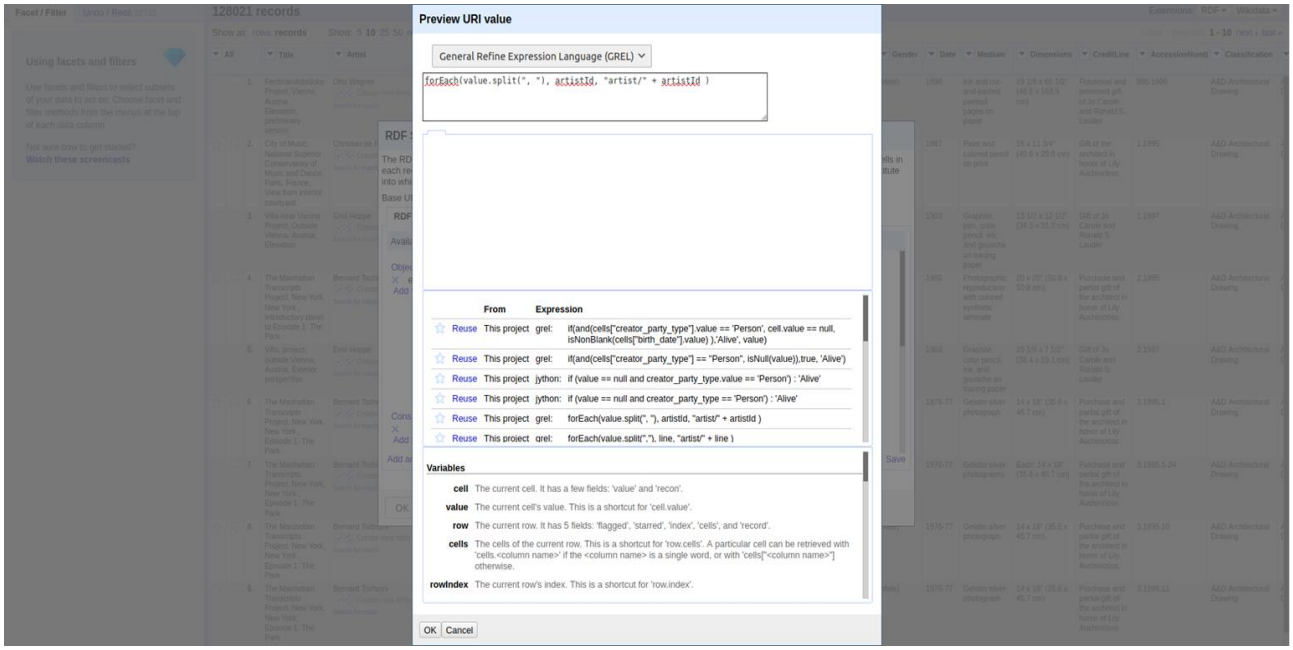


Εικόνα 19 Επιλογή τύπου δεδομένων OpenRefine

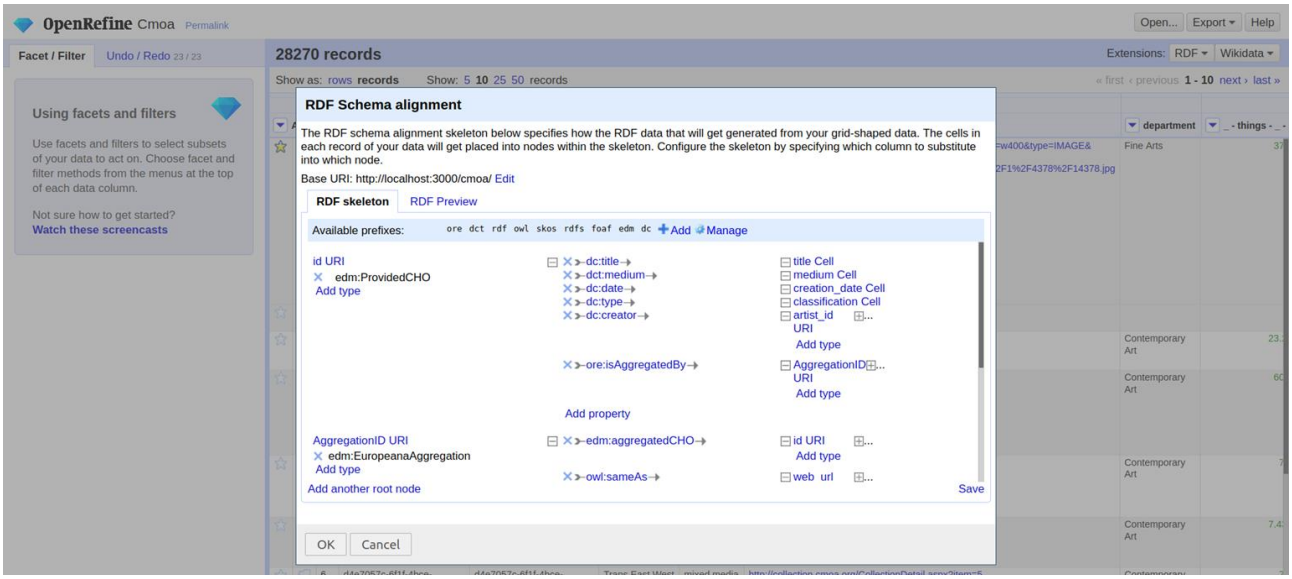
Οι τιμές των κόμβων μπορούν να λάβουν είτε την τιμή του πεδίου αυτούσια, είτε να εφαρμοστεί σε αυτήν κάποια μετατροπή. Συγκεκριμένα, στο αρχείο συλλογής του MoMA, σε περίπτωση που στην τιμή του δημιουργού καταγράφονται περισσότεροι του ενός καλλιτέχνες, το αποτέλεσμα που προέκυπτε κατά τη δημιουργία του URI ήταν αυτό να περιλαμβάνει τιμές και κόμματα και να μην μπορεί να αντιστοιχηθεί με συγκεκριμένο καλλιτέχνη. Έτσι για την αποφυγή διαχωρισμού της μεταβλητής σε στήλες ανάλογα με το πλήθος των καλλιτεχνών και την προσθήκη συνολικά τριάντα ενός νέων στηλών για να καλύπτονται όλες οι περιπτώσεις, χρησιμοποιήθηκε η εξής έκφραση:

forEach(value.split(“, “, artistId, “artist/” + artistId)

η οποία αρχικά δημιουργεί έναν πίνακα με πεδία τις τιμές των αναγνωριστικών κάθε δημιουργού του εκάστοτε έργου και στη συνέχεια επιστρέφει την κάθε τιμή χωριστά με προσθήκη του “*artist/*”, και καθώς έχει επιλεχθεί ο τύπος URI, του προθέματος του domain. Δηλαδή για την τιμή πεδίου “6969,8134” προκύπτει το αποτέλεσμα <http://localhost:3000/moma/artist/6969>, <http://localhost:3000/moma/artist/8134>



Εικόνα 20 Εφαρμογή μετατροπής κατά την δημιουργία URI OpenRefine



Εικόνα 21 Μοντέλο που δημιουργήθηκε OpenRefine

Όταν ολοκληρωθεί το μοντέλο, υπάρχει δυνατότητα προβολής του τελικού αποτελέσματος σε RDF με σύνταξη Turtle, για ένα δείγμα από τις πρώτες εγγραφές του συνόλου. Η συγκεκριμένη επιλογή κρίνεται ιδιαίτερα χρήσιμη καθώς βοηθά στην αποφυγή λαθών στο τελικό παραγόμενο αρχείο.

RDF Schema alignment

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .


<http://localhost:3000/moma/2/artwork> a edm:ProvidedCHO;
  dc:creator <http://localhost:3000/moma/artist/6210>;
  dc:date "1896";
  dc:title "Ferdinandsbrücke Project, Vienna, Austria , Elevation, preliminary version";
  dc:type "A&D Architectural Drawing";
  dct:medium "Ink and cut-and-pasted painted pages on paper";
  ore:isAggregatedBy <http://localhost:3000/moma/2> .

<http://localhost:3000/moma/artist/6210> a edm:Agent;
  edm:begin "1841";
  edm:end "1918";
  skos:note "(Austrian, 1841–1918)";
  skos:prefLabel "Otto Wagner" .

<http://localhost:3000/moma/2> a edm:EuropeanaAggregation;
  edm:aggregatedCHO <http://localhost:3000/moma/2/artwork>;
  edm:hasView <http://www.moma.org/media/W1siZiIsIjU5NDA1Il0sWyJwIiwiY29udmVydCIsIi1yZXNpemUgMzAweDMwMFx1MDAzZSJdXQ.jp
  owl:sameAs <http://www.moma.org/collection/works/2> .

<http://localhost:3000/moma/3/artwork> a edm:ProvidedCHO;
  dc:creator <http://localhost:3000/moma/artist/7470>;
  dc:date "1987";
  dc:title "City of Music, National Superior Conservatory of Music and Dance, Paris, France, View from interior courtyard";
  dc:type "A&D Architectural Drawing";
  dct:medium "Paint and colored pencil on print";
  ore:isAggregatedBy <http://localhost:3000/moma/3> .

<http://localhost:3000/moma/artist/7470> a edm:Agent;
  edm:begin "1944";
```

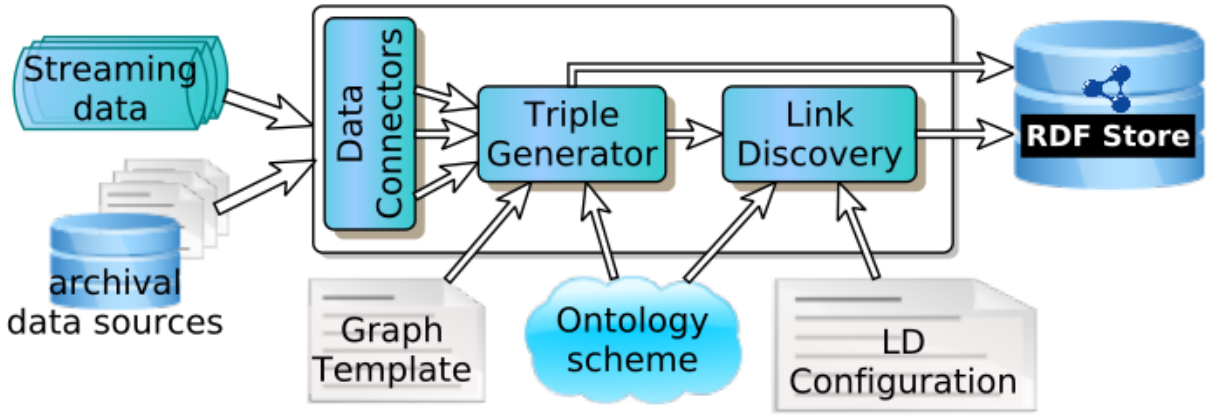
OK Cancel

Εικόνα 22 Προεπισκόπιση RDF γράφου για υποσύνολο των δεδομένων

Κατά την ολοκλήρωση της διαδικασίας πραγματοποιήθηκε εξαγωγή των δεδομένων του αρχείου MoMA.csv (~128.000 εγγραφών) σε RDF, η οποία διήρκεσε 1’ και 45”. Αντίστοιχη διαδικασία ακολουθήθηκε για την εξαγωγή του συνόλου δεδομένων της συλλογής του CMoA(~28.000 εγγραφών), από μορφή JSON σε RDF, η οποία διήρκεσε 35.

RDF-Gen

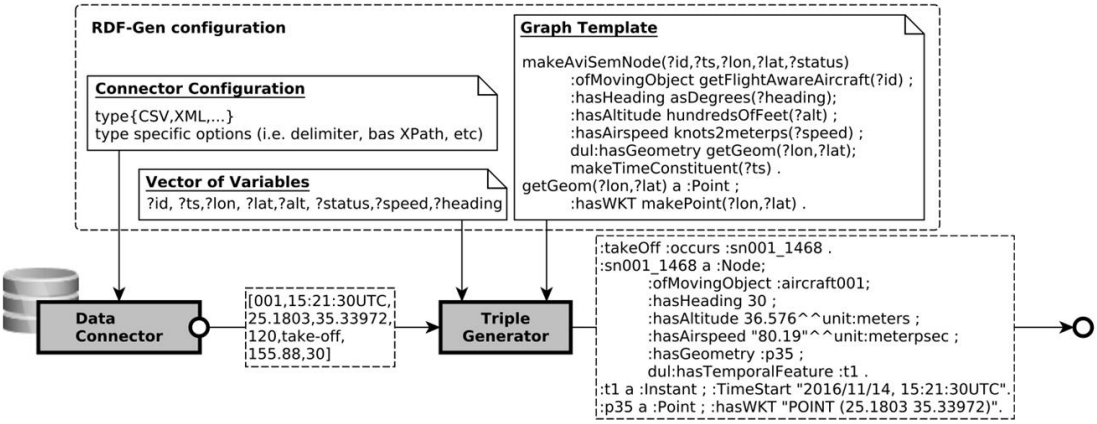
Το RDF-Gen αποτελεί ένα state-of-the-art εργαλείο ενοποίησης δεδομένων από ετερογενείς πηγές. Αποτελεί ερευνητικό έργο του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιά, και δημοσιεύθηκε αρχικά στο International Conference on Web Intelligence, Mining and Semantics το 2018 με τίτλο RDF-Gen: Generating RDF from Streaming and Archival Data [39]. Είναι υλοποιημένο σε Java και εκτελείται σε περιβάλλον κονσόλας. Το RDF-Gen επικεντρώνεται στην επεκτασιμότητα, την υπολογιστική αποτελεσματικότητα και την ελαχιστοποίηση της καθυστέρησης παραγωγής ενοποιημένων δεδομένων. Υποστηρίζει εξίσου στατικά/ιστορικά (static/historical) και δεδομένα δυναμικά/ροής (dynamic/streaming) και είναι σε θέση να διαχειριστεί διαφορετικούς τύπους δεδομένων καθώς βασίζεται σε συναρτήσεις και λειτουργίες που υλοποιούνται κατά περίπτωση.



Εικόνα 23 Διαδικασία μετατροπής δεδομένων σε RDF από το RDF-Gen

Οι δύο βασικές λειτουργίες του RDF-Gen είναι ο Data Connector και ο Triple Generator. Ο Data Connector λαμβάνει ως είσοδο τα δεδομένα και ένα αρχείο προδιαγραφών. Στο συγκεκριμένο αρχείο παρέχεται από το χρήστη πληροφορία για τα δεδομένα και των απαραίτητων αρχείων για την εκτέλεση της διαδικασίας μετατροπής των δεδομένων σε RDF. Συγκεκριμένα προσδιορίζονται οι θέσεις των αρχείων εισόδου/εξόδου και του προτύπου βάση του οποίου θα γίνει η μετατροπή, σημειώνεται ο τύπος των δεδομένων, ο τρόπος διαχωρισμού των εγγραφών, και τέλος γίνεται αναφορά στις μεταβλητές εισόδου και η αντιστοίχιση τους με τις μεταβλητές που υπάρχουν στο πρότυπο.

Οι εγγραφές που παράγονται από τον Data Connector παρέχονται στον Triple Generator, ο οποίος είναι υπεύθυνος για την μετατροπή του σε RDF τριπλέτες που βασίζονται σε δοσμένη οντολογία. Το πρότυπο αναπαριστά το μοντέλο πάνω στο οποίο θα δημιουργηθεί ο RDF γράφος και συντάσσεται με παρεμφερή τρόπο με τα ερωτήματα σε γλώσσα SPARQL. Στο πρότυπο γίνεται χρήση των υλοποιημένων συναρτήσεων που εξυπηρετούν στην προσπέλαση και την τροποποίηση των στοιχείων που περιέχονται στις μεταβλητές, για παράδειγμα asString(?x), asDate(?y), asURI(?z).



Εικόνα 24 Είσοδοι παραμέτρων, μοντέλου και αποτέλεσμα διαδικασίας RDF-Gen

Χρήση του RDF-Gen

Η χρήση του RDF-Gen για την μετατροπή των δεδομένων σε γράφους RDF, αποτελεί μια αρκετά απλή διαδικασία. Για τα δύο σύνολα δεδομένων συμπληρώθηκαν τα αρχεία των προδιαγραφών και του προτύπου με τον παρακάτω τρόπο.

Αρχικά στο αρχείο των προδιαγραφών αναφέρεται ότι η πηγή των δεδομένων είναι το αρχείο Artworks.csv, το οποίο είναι της μορφής csv και διαχωριστικό ανάμεσα στις τιμές είναι το κόμμα. Δηλώνεται επίσης ότι το πρότυπο θα αντληθεί από το αρχείο moma.q, οι μεταβλητές στα δύο αρχεία και τέλος το αρχείο output.ttl στο οποίο θα αποθηκευτούν τα αποτελέσματα της μετατροπής.

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dcf="http://www.datacron-project.eu/RDFGen_conf#">
        <!-- the absolute path to data source, this can be a folder to be recursively processed -->
        <dcf:DataSource rdf:about="Artworks.csv">
        <!-- Required: specifies the connector to be used -->
        <dcf:connector>csv</dcf:connector>
        <!-- Required: the template to be used -->
        <dcf:template>moma.q</dcf:template>
        <!-- Required: the variables to be used for the source in the order specified -->
        <dcf:delimiter>,</dcf:delimiter>
        <dcf:inputVariables>0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17</dcf:inputVariables>
        <!-- Required: the variables to be used for the source in the order specified -->
        <dcf:TemplateVariables>?Title,?Artist,?ConstituentID,?ArtistBio,?Nationality,?BeginDate,?En
        dDate,?Gender,?Date,?Medium,?Dimensions,?CreditLine,?AccessionNumber,?Classification,
        ?Department,?DateAcquired,?Cataloged,?ObjectID,?URL,?ThumbnailURL,?Circumference,?D
        epth,?Diameter, ?Height, ?Length,?Weight,?Width,?SeatHeight,?Duration
        </dcf:TemplateVariables>
        <!--optional: if provided the generated triples will be provided at this path -->
        <dcf:output>output.ttl</dcf:output>
        </dcf:DataSource>
</rdf:RDF>.
```

Εικόνα 25 Αρχείο προδιαγραφών RDF-Gen

Στο αρχείο moma.q περιγράφονται οι συσχετίσεις μεταξύ των μεταβλητών που δομούν το μοντέλο του RDF γράφου.

Κατά την ολοκλήρωση της διαδικασίας πραγματοποιήθηκε μετατροπή των δεδομένων του αρχείου MoMA.csv (~128.000 εγγραφών) σε RDF, η οποία διήρκεσε 25”. Αντίστοιχη διαδικασία ακολουθήθηκε για το σύνολο δεδομένων της συλλογής του CMoA (~28.000 εγγραφών), από μορφή JSON σε RDF, η οποία διήρκεσε 13.

```
asURI(?ObjectID) a edm:ProvidedCHO;
        dc:title asString(?Title);
        dc:creator asURI(?ConstituentID);
        dc:date ?Date;
        dct:medium asString(?Medium);
        dc:type asString(?Classification);
asURI(?AggregationID) a edm:EuropeanaAggregation;
        edm:aggregatedCHO asURI(?ObjectID);
        owl:sameAs asQuotedURI(?URL);
        edm:hasView asQuotedURI(?ThumbnailURL).
asURI(?ConstituentID) a edm:Agent;
        skos:prefLabel asString(?Artist);
        edm:begin ?BeginDate;
        edm:end ?EndDate ;
        skos:note asString(?ArtistBio) .
```

Εικόνα 26 Πρότυπο μοντέλου RDF-Gen

3.3.2 Links Discovery (LD)

Το πρότυπο των διασυνδεδεμένων δεδομένων υποδεικνύει αρχικά την μετατροπή των δεδομένων σε μηχαναγνώσιμη μορφή μέσω ενός εκφραστικού μοντέλου για την αναπαράσταση της πληροφορίας και στη συνέχεια την εύρεση συσχετιζόμενης πληροφορίας σε διαφορετικές πηγές και δημιουργία συνδέσμων με αυτήν.

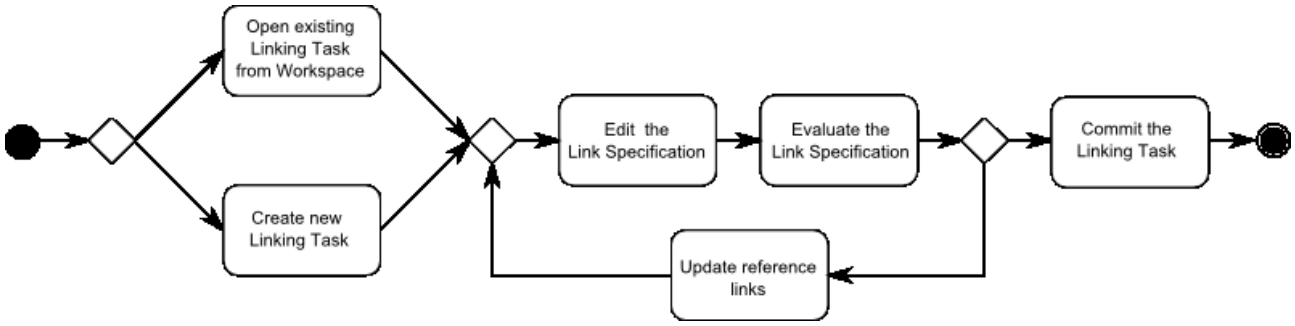
Για την ανακάλυψη συνδέσμων και κάλυψη αυτής της απαίτησης, έγινε επιλογή δύο εργαλείων που παρουσιάζονται στη συνέχεια, του Openrefine Reconciliation Service [36] και του Silk Framework [40]

SILK FRAMEWORK

Η πλατφόρμα SILK είναι ένα εργαλείο ανοιχτού κώδικα για ενοποίηση δεδομένων από ετερογενείς πηγές. Παρουσιάστηκε το 2009 στο 2nd Workshop about Linked Data on the Web με τίτλο 'Silk – A Link Discovery Framework for the Web of Data' [40].

Οι κύριες λειτουργίες του SILK περιλαμβάνουν

- την δημιουργία συνδέσμων μεταξύ συσχετιζόμενων δεδομένων από διαφορετικές πηγές.
- την διασύνδεση δημοσιευμένων ανοιχτών δεδομένων με άλλες πηγές στον παγκόσμιο ιστό
- την εφαρμογή μετασχηματισμών σε οργανωμένες δομές δεδομένων



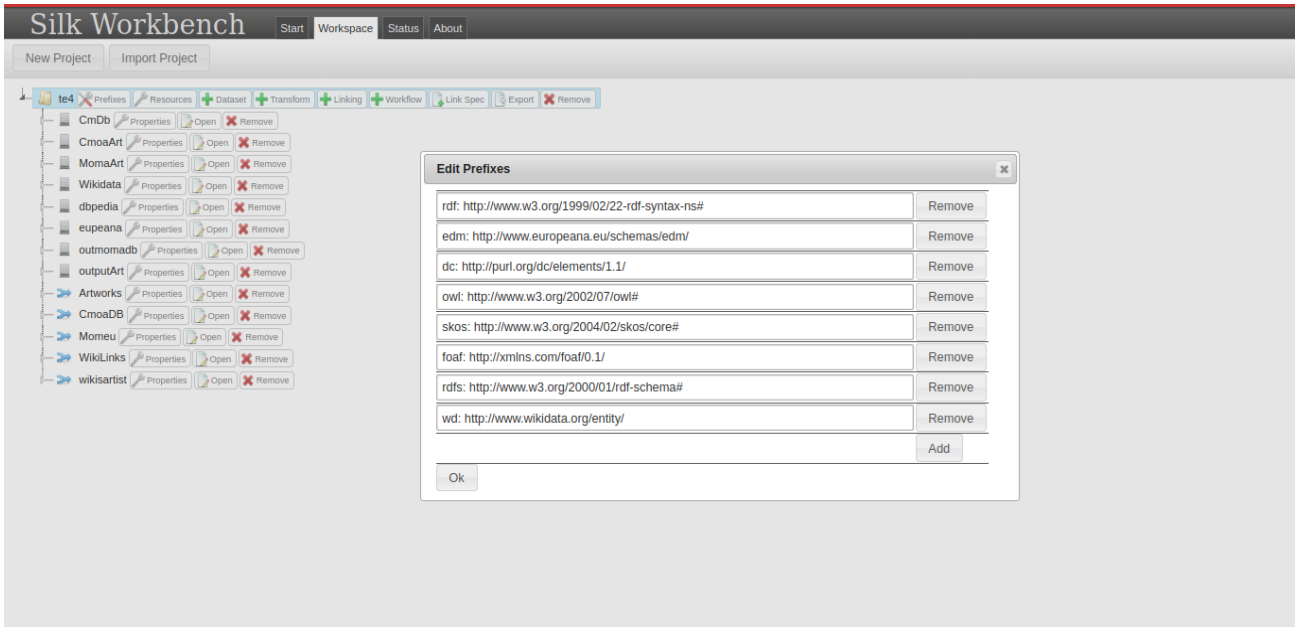
Εικόνα 27 Διεργασίες για τη δημιουργία συνδέσμων Silk [49]

Το Silk διαθέτει μια δηλωτική γλώσσα για τον ορισμό του τύπου των συνδέσμων που θα ανακαλυφθούν ανάμεσα στις πηγές δεδομένων αλλά και τον ορισμό των προϋποθέσεων κάτω από τις οποίες θα είναι δυνατόν να ολοκληρωθεί η σύνδεση τους. Οι προϋποθέσεις αυτές μπορεί να βασίζονται σε διάφορες μετρικές ομοιότητας (similarity metrics) οι οποίες περιλαμβάνονται στην πλατφόρμα. Για την πρόσβαση στα δεδομένα των αρχείων, το Silk κάνει χρήση του πρωτοκόλλου SPARQL και έτσι μπορεί να αποστείλει ερωτήματα εξίσου σε τοπικά και σε απομακρυσμένα σημεία. Οι προδιαγραφές για τους συνδέσμους δημιουργούνται είτε μέσω αρχείων XML είτε με χρήση του Silk Workbench που αποτελεί την γραφική διεπαφή του Silk.

Το Silk Workbench είναι μια δικτυακή εφαρμογή που καθοδηγεί τον χρήστη κατά τη διαδικασία διασύνδεσης διαφορετικών πηγών δεδομένων. Επιτρέπει τη διαχείριση διαφόρων συνόλων δεδομένων μέσω διαδικασιών μετασχηματισμού και διασύνδεσης και προσφέρει ένα γραφικό περιβάλλον στο χρήστη για την επίτευξη των διαδικασιών αυτών. Το Silk Workbench στο χρήστη δίνει τη δυνατότητα να αξιολογεί τους παραγόμενους συνδέσμους και τελικά την αποδοχή ή απόρριψή τους [41].

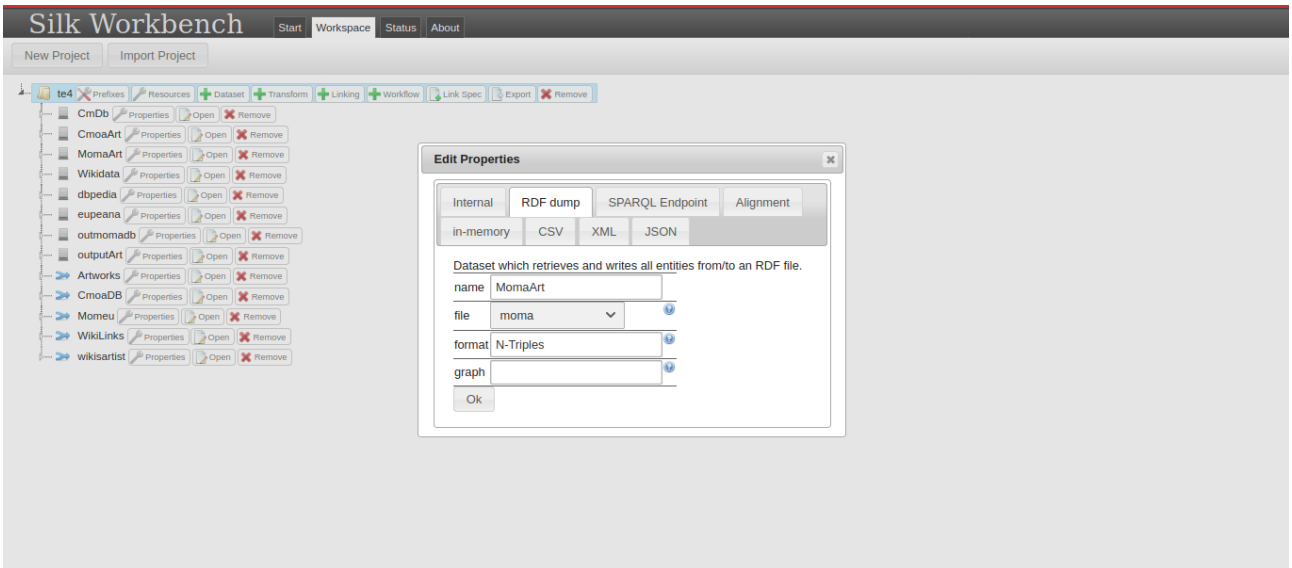
Χρήση του Silk Workbench

Τα βήματα που απαιτούνται για την ανακάλυψη των συνδέσμων είναι διακριτά και δομημένα με τρόπο που καθοδηγούν τον χρήστη στην ολοκλήρωση της διαδικασίας. Αρχικά επιλέγονται τα λεξιλόγια και οι οντολογίες που θα χρησιμοποιηθούν.



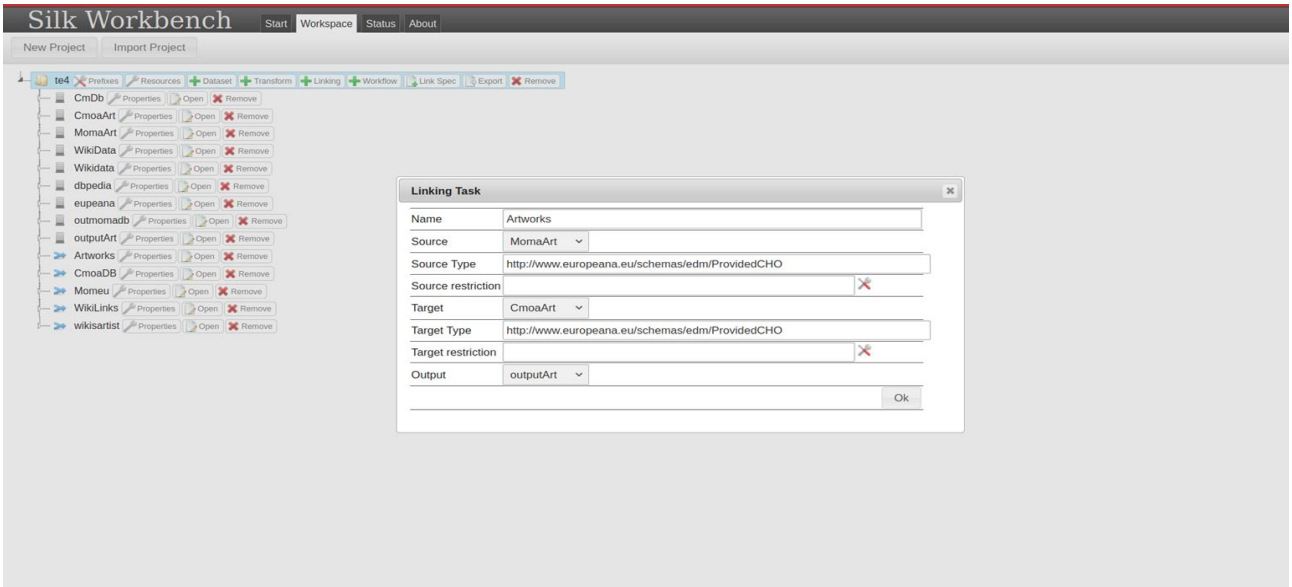
Εικόνα 28 Εισαγωγή χώρων ονομάτων Silk

Στη συνέχεια μεταφορτώνονται τα σύνολα δεδομένων ως RFD γράφοι και ορίζονται οι εξωτερικές πηγές, υπό την μορφή SPARQL endpoints, με τις οποίες θα γίνει προσπάθεια διασύνδεσης.



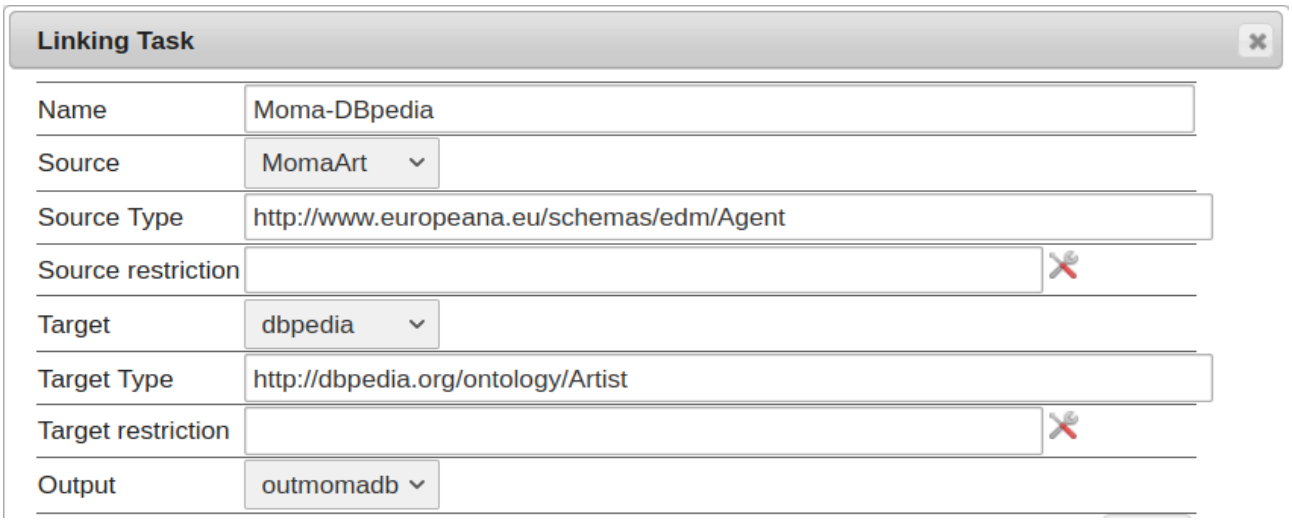
Εικόνα 29 Εισαγωγή δεδομένων Silk

Αφού ορισθούν όλες οι απαραίτητες πηγές, το επόμενο βήμα είναι η δημιουργία μιας διεργασίας διασύνδεσης (Linking Task). Σε κάθε διεργασία επιλέγονται οι δύο πηγές δεδομένων και ορίζεται ο σημασιολογικός τύπος, βάσει του οποίου θα γίνεται ο έλεγχος συσχέτισης. Ο σημασιολογικός τύπος μπορεί να ταυτίζεται αλλά χωρίς αυτό να είναι δεσμευτικό, καθώς τα διάφορα σύνολα δεδομένων μπορεί να έχουν εκφραστεί με διαφορετικά λεξιλόγια και με αυτόν το τρόπο επιτυγχάνεται η αντιστοίχιση τους.



Εικόνα 31 Δημιουργία 1ης διεργασίας διασύνδεσης Silk

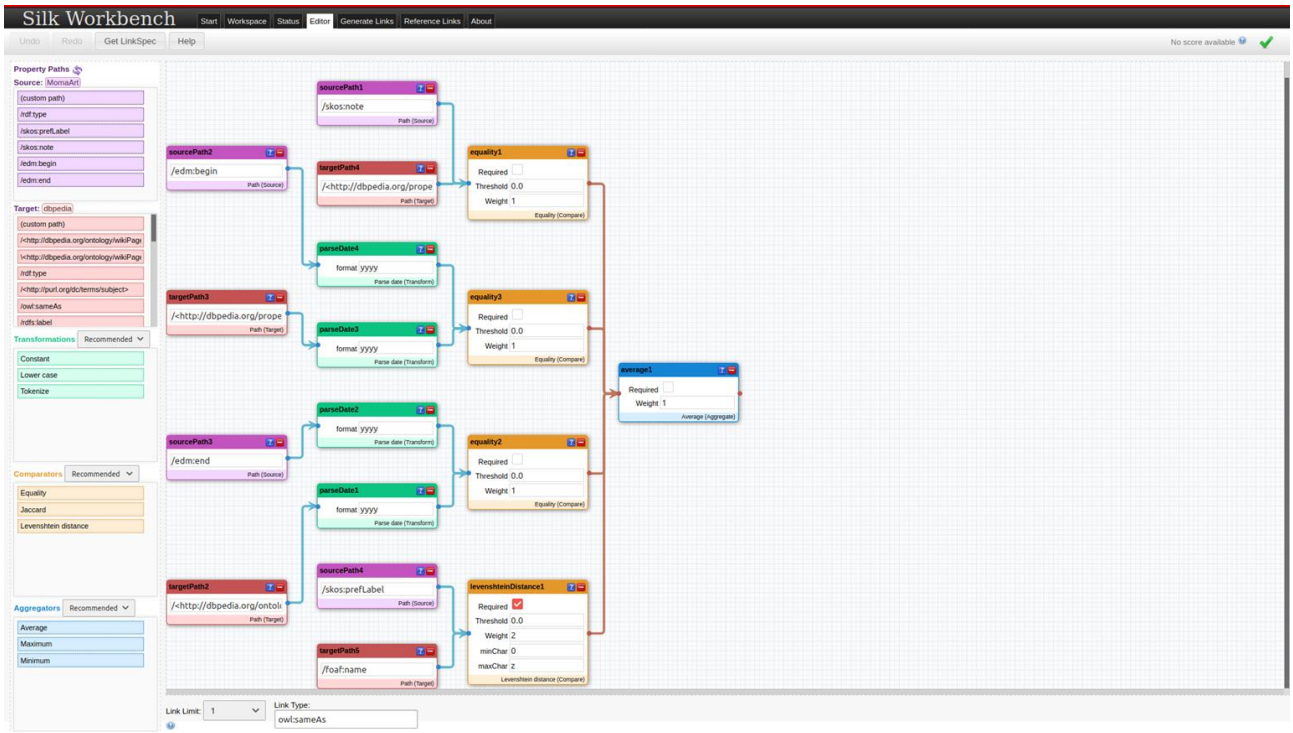
Στην περίπτωση της εικόνας 30, αναζητούνται σύνδεσμοι που αφορούν έργα τέχνης από τα σύνολα δεδομένων MoMA και CMoA Collection, τα οποία έχουν εκφραστεί σύμφωνα με το μοντέλο EDM, ο σημασιολογικός τύπος που επιλέγεται είναι *edm:ProvidedCHO.*



Εικόνα 30 Δημιουργία 2ης διεργασίας διασύνδεσης Silk

Στην περίπτωση που η ζητείται η εύρεση συνδέσμων ανάμεσα στο τοπικό σύνολο δεδομένων MoMA Collection και στον γράφο της Dbpedia, επιλέγονται οι τύποι *edm:Agent* και *dbo:Artist* αντίστοιχα.

Πριν την εκκίνηση της διαδικασίας εύρεσης συνδέσμων, πρέπει να ορισθούν τα χαρακτηριστικά των εξεταζόμενων πόρων που θα συγκριθούν, ώστε να προκύψει η ομοιότητα. Το Silk Workbench παρέχει ένα γραφικό περιβάλλον που καλείται Editor και διευκολύνει τον ορισμό των απαραίτητων συγκρίσεων.



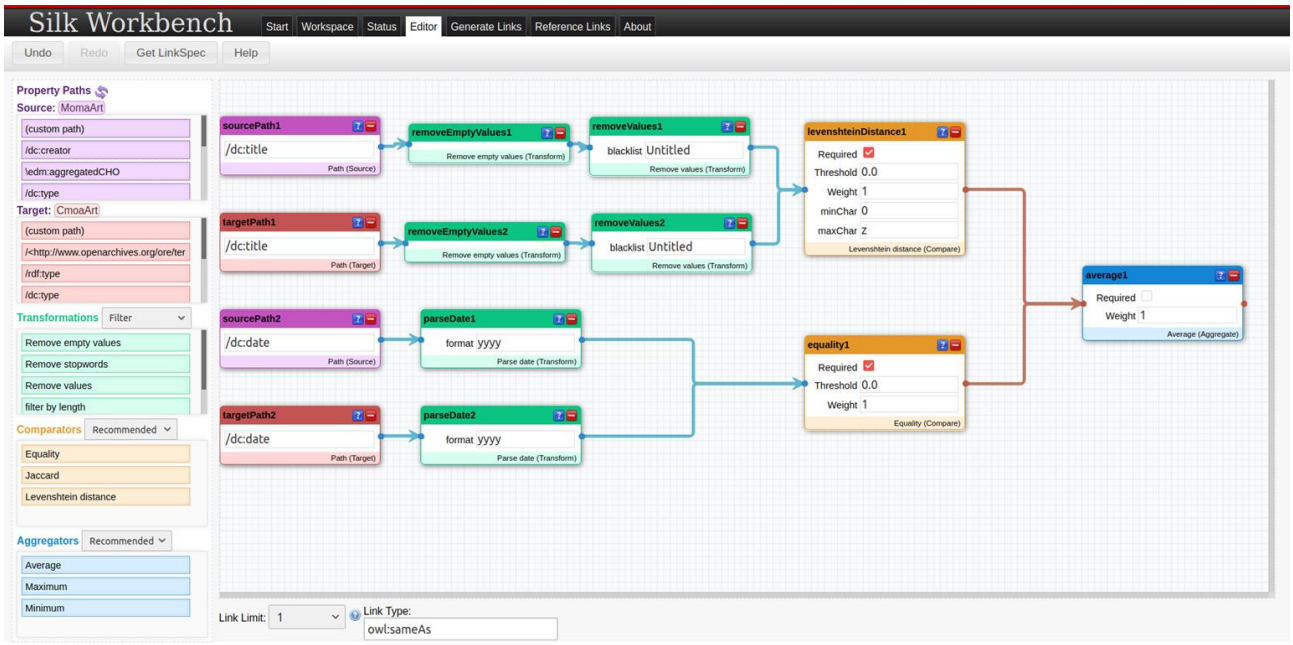
Εικόνα 32 Ορισμός συγκρίσεων για την 1η διεργασία διασύνδεσης Silk

Αρχικά επιλέγονται οι μεταβλητές που θα συγκριθούν από κάθε γράφο και δηλώνεται εάν οι τιμές του πρέπει να υποστούν κάποιο μετασχηματισμό. Στη συνέχεια επιλέγεται η μετρική βάσει της οποίας θα γίνει η σύγκριση, που μπορεί να αποτελούν μετρικές σύγκρισης συμβολοσειρών όπως η *Levenshtein distance*, η *Jaro–Winkler* ή μπορεί να είναι μετρικές ειδικού σκοπού όπως αριθμών, ημερομηνιών ή γεωγραφικών αποστάσεων. Τέλος επιλέγεται ο τρόπος με τον οποίο θα ληφθεί η απόφαση σύμφωνα με τις τιμές των συγκρίσεων.

Οι μεταβλητές που επιλέχθηκαν για την δημιουργία συνδέσμων μεταξύ των καλλιτεχνών της συλλογής MoMA και της DBpedia είναι το όνομα του καλλιτέχνη, οι ημερομηνίες γέννησης και θανάτου και η εθνικότητα του. Οι τύποι των τιμών που αντιστοιχίζονται στις μεταβλητές είναι για τη συλλογή MoMA:

MoMA Collection	**Dbpedia**
skos:prefLabel	foaf:name
edm:begin	dbo:birthDate
edm:end	dbo:deathDate
skos:note	dbo:nationality

Πίνακας 4 Αντιστοίχιση μεταβλητών MoMA, DBpedia για ανακάλυψη συνδέσμων



Εικόνα 33 Ορισμός συγκρίσεων για την 2η διεργασία διασύνδεσης Silk

Οι μεταβλητές που επιλέχθηκαν για την εύρεση συνδέσμων μεταξύ έργων τέχνης των συλλογών MoMA και CMoA είναι: Τίτλος έργου με τύπο dc:title και ημερομηνία δημιουργίας με τύπο dc:date.

Οι μετασχηματισμοί που εφαρμόστηκαν στα δεδομένα ήταν κατά περίπτωση οι παρακάτω:

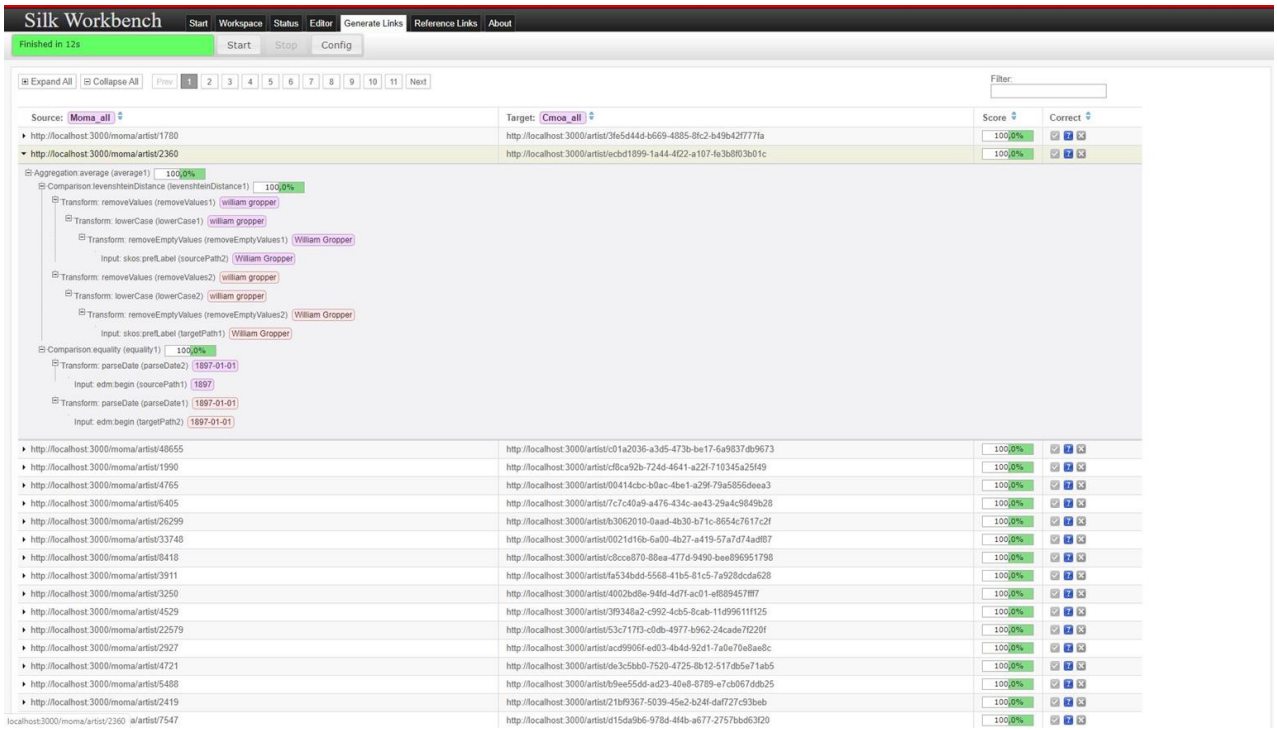
- Lower case : Μετατροπή όλων των γραμμάτων σε πεζά
- parseDate : Προσπέλαση στοιχείου ως ημερομηνίας
- removeEmptyValues : αφαίρεση κενών πεδίων

- removeValues : αφαίρεση τιμών που ισούνται με δοσμένη τιμή (χρήση της έγινε για απαλοιφή πεδίων με χωρίς τίτλο, δηλαδή με τιμή “Untitled”)

Οι μετρικές που χρησιμοποιήθηκαν για τις συγκρίσεις ήταν

- ισότητα για πεδία ημερομηνιών και
- Απόσταση Levenshtein για συμβολοσειρές

Η απόσταση Levenshtein μεταξύ δύο συμβολοσειρών ορίζεται ως αριθμός των αλλαγών ενός χαρακτήρα που απαιτούνται ώστε η μια να μετατραπεί σε στην άλλη. Αλλαγές θεωρούνται η εισαγωγή, η αφαίρεση ή η αντικατάσταση χαρακτήρα.

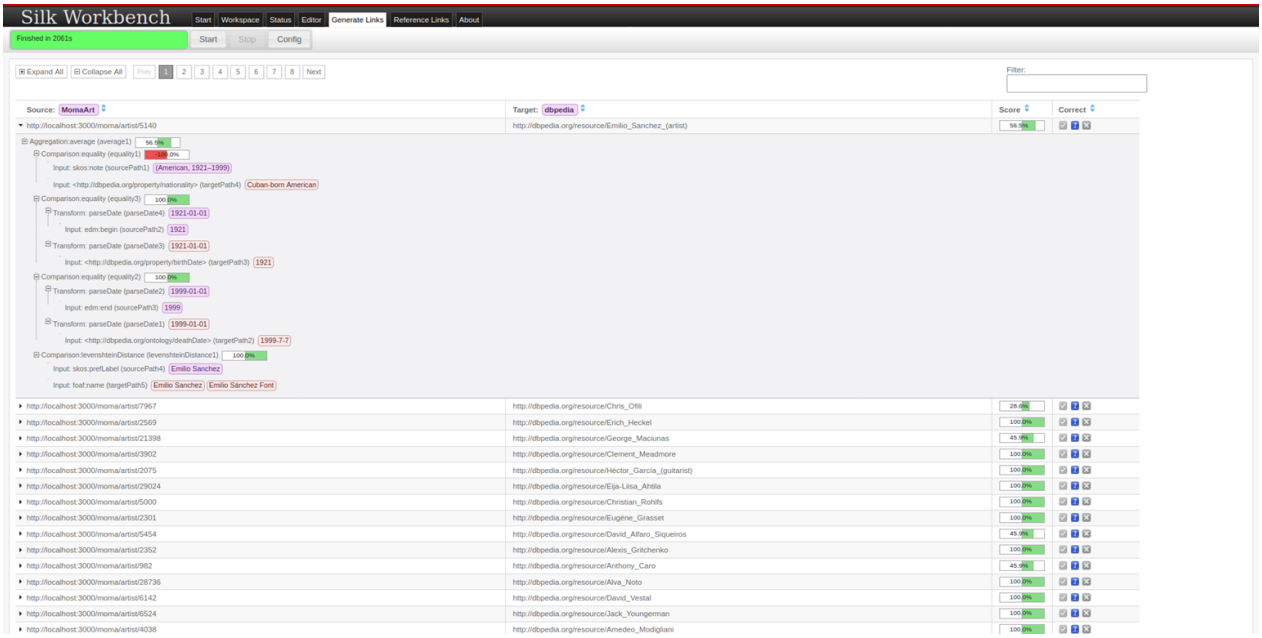


Εικόνα 34 Αποτελέσματα εύρεσης συνδέσμων διεργασίας 1 Silk

Η εύρεση συνδέσμων ανάμεσα στους δημιουργούς της συλλογής των MoMA και CMoA είχε ως αποτέλεσμα την δημιουργία 1042 συνδέσμων και διήρκεσε 12’’.

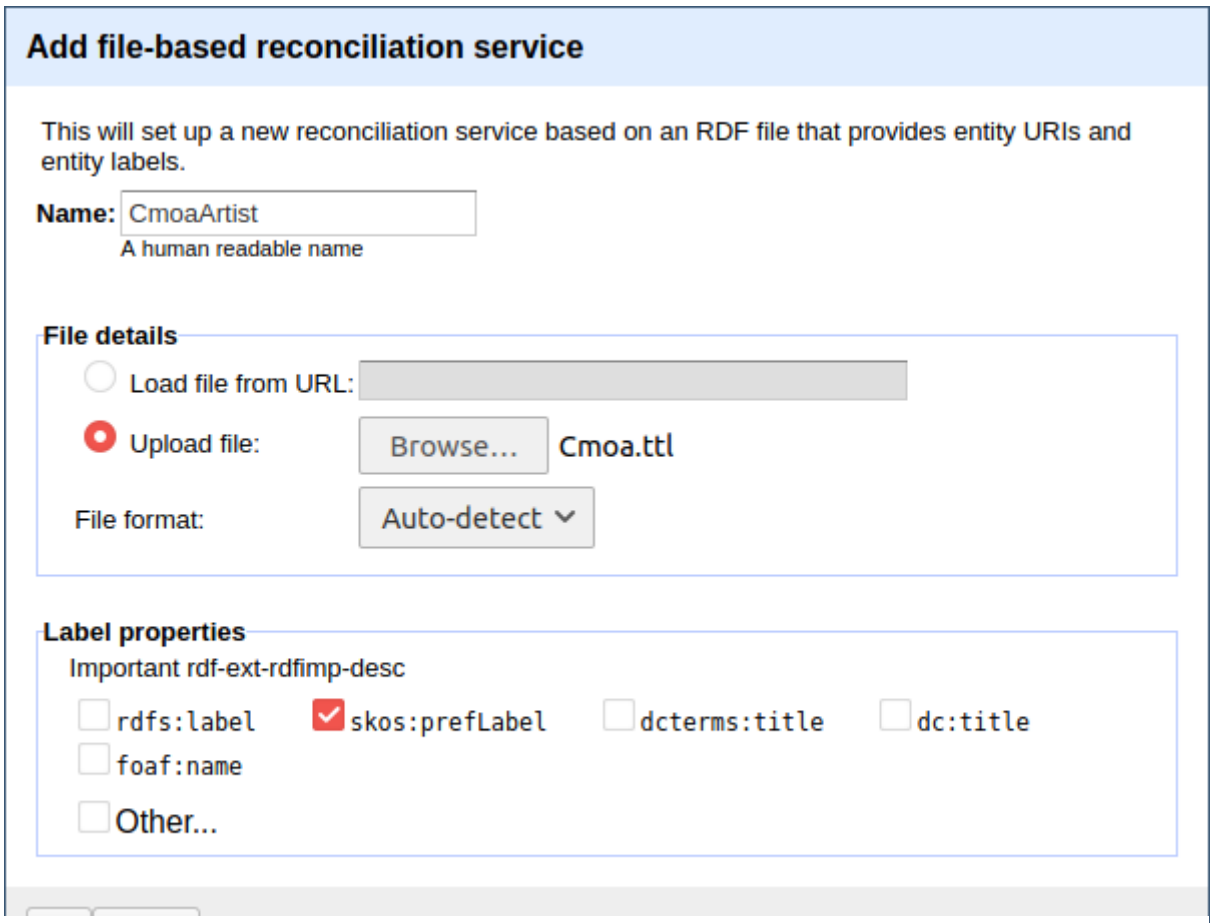
Η διαδικασία εύρεσης συνδέσμων των καλλιτεχνών από τον τοπικό γράφο της συλλογής του MoMA με την DBpedia είχε σαν αποτέλεσμα την διασύνδεση 760 οντοτήτων σε περίπου 35 λεπτά.

Αντίστοιχες διεργασίες εκτελέστηκαν και για τους καλλιτέχνες της συλλογής CMoA από όπου προέκυψαν 452 συνδέσεις σε 7’ και 25’’, ενώ για την διασύνδεση των έργων τέχνης των δύο τοπικών γράφων προέκυψαν 216 συνδέσεις σε διάστημα 40 λεπτών.



Εικόνα 35 Αποτελέσματα εύρεσης συνδέσμων διεργασίας 2 Silk

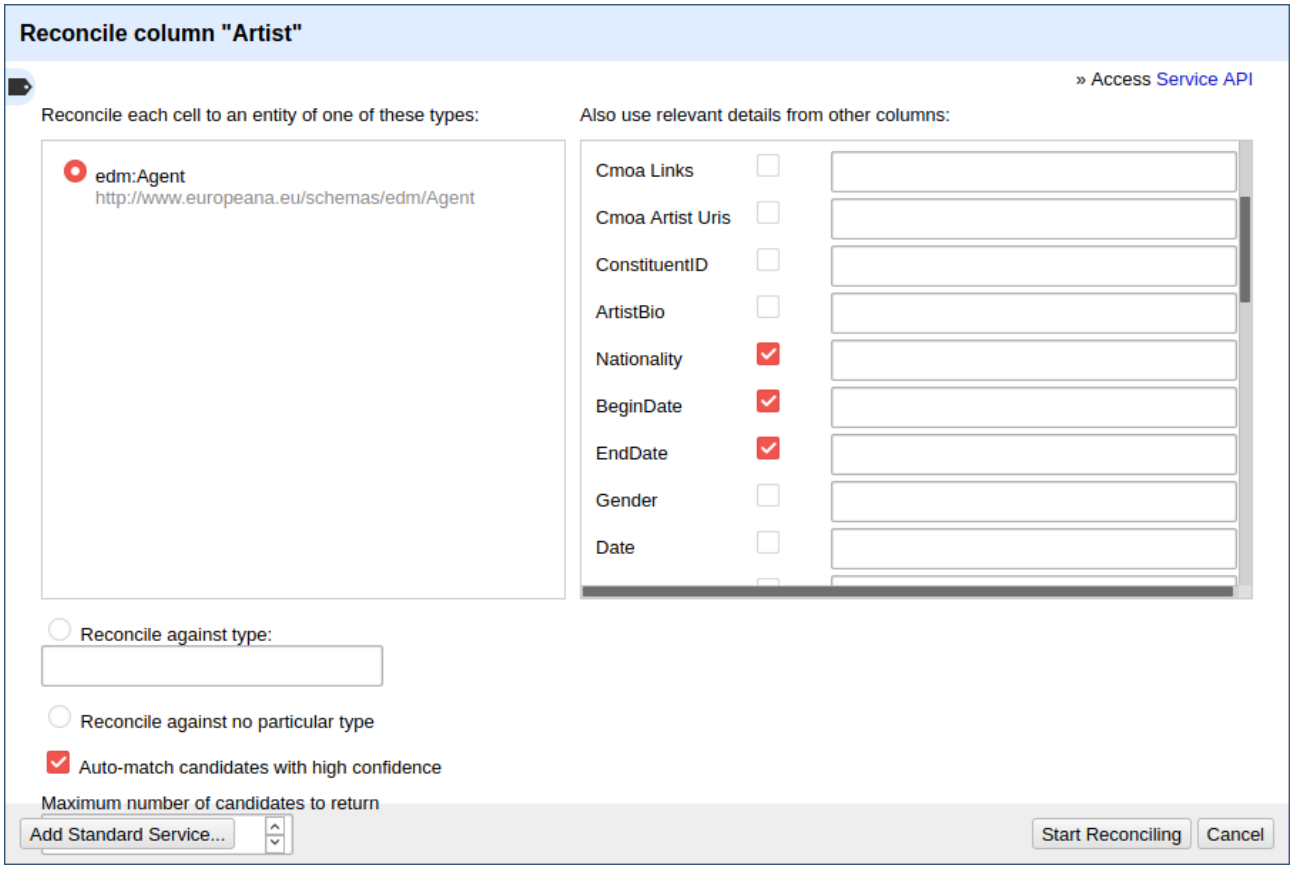
Openrefine Reconciliation Service



Εικόνα 36 Δημιουργία reconciliation service για τοπικό αρχείο OpenRefine

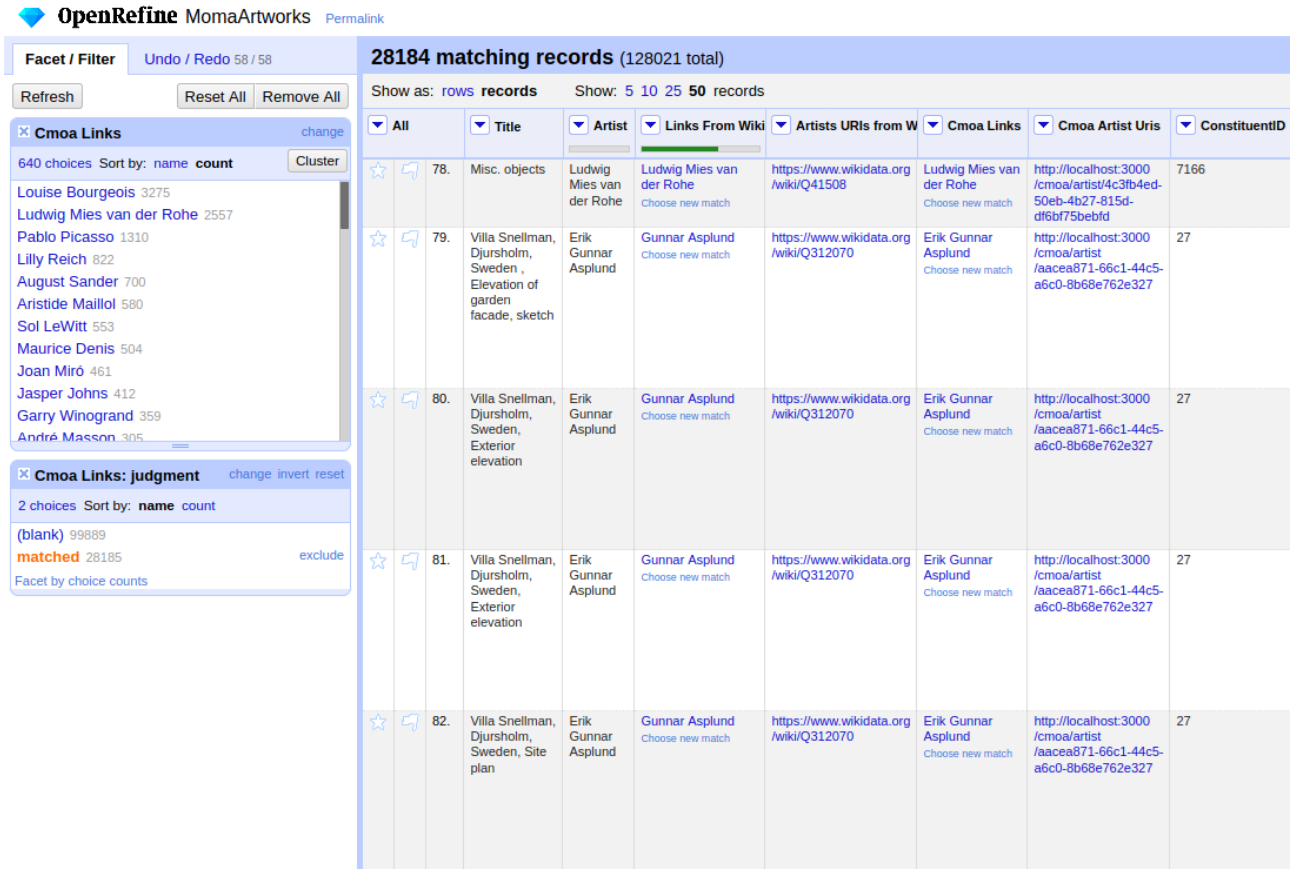
Μία ακόμη υπηρεσία του εργαλείου OpenRefine που εξειδικεύεται στην εύρεση συνδέσμων μεταξύ δεδομένων διαφορετικών πηγών είναι το Reconciliation Service που ενσωματώνεται στο rdf-extension. Η συγκεκριμένη υπηρεσία παρέχει τη δυνατότητα αναζήτησης σε τοπικά αρχεία RDF ή σε SPARQL endpoints.

Όπως και στο παράδειγμα του SILK, αρχικά πρέπει να επιλεχθεί ο τύπος της ιδιότητας για την οποία θα γίνουν οι συσχετίσεις και στη συνέχεια επιλέγονται η κλάση του πόρου και προαιρετικά κάποιες ακόμη ιδιότητες που μπορεί να εξυπηρετήσουν στην επιλογή των σωστών συνδέσμων. Για την εύρεση συνδέσμων ως προς την οντότητα «Δημιουργός» ανάμεσα στα σύνολα δεδομένων των δύο συλλογών, αρχικά επιλέχθηκε η ιδιότητα *rdfs:label* ως περιγραφή του πόρου και στη συνέχεια επιλέγεται η κλάση *edm:Agent*, η οποία χαρακτηρίζει τους καλλιτέχνες καθώς και οι ιδιότητες *εθνικότητα, ημερομηνία γέννησης* και *ημερομηνία θανάτου*.



Εικόνα 37 Ορισμός μεταβλητών προς σύγκριση OpenRefine

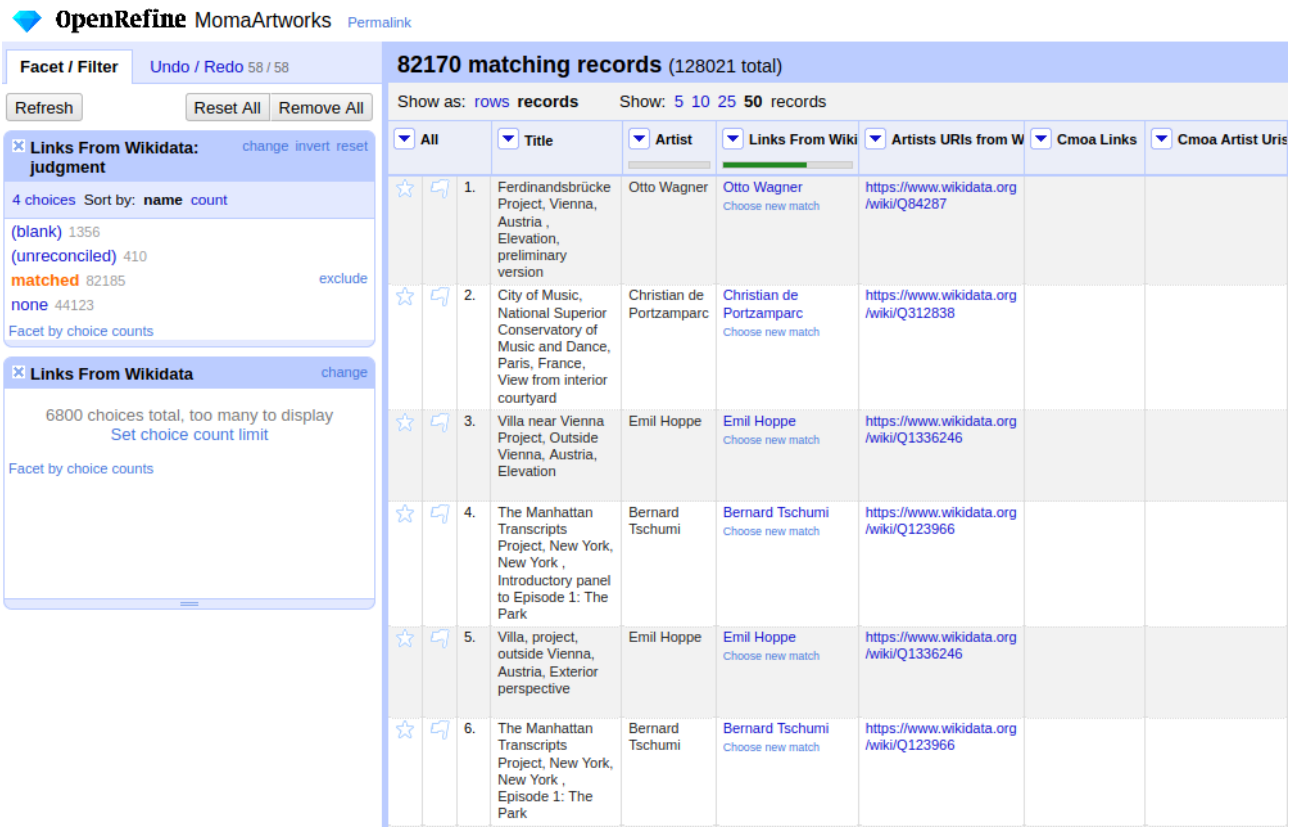
Τελικά, η ανακάλυψη συνδέσμων μεταξύ καλλιτεχνών των συλλογών MoMA και CMoA είχε σαν αποτέλεσμα τη δημιουργία 28.184 συνδέσμων για το σύνολο των εγγραφών, που αντιστοιχούν σε 640 μοναδικά URI πόρων καλλιτεχνών και διήρκεσε περίπου 2 ώρες.



Εικόνα 38 Αποτελέσματα ανακάλυψης συνδέσμων OpenRefine

Η διεργασία ανακάλυψης συνδέσμων μεταξύ της συλλογής MoMA και της DBpedia (μέσω του SPARQL endpoint), για την οντότητα «Καλλιτέχνης» ολοκληρώθηκε δίχως αποτελέσματα, εκτελώντας την αναζήτηση για τον τύπο dbo:artist με τον συνυπολογισμό των τιμών των κελιών ημερομηνίας γέννησης και θανάτου και εθνικότητας.

Η ανακάλυψη συνδέσμων καλλιτεχνών μεταξύ της συλλογής MoMA και του Wikidata για τον τύπο Q5 (Person) είχε σαν αποτέλεσμα τη δημιουργία 82185 συνδέσμων για το σύνολο των εγγραφών, που αντιστοιχούν σε 6800 μοναδικά URI πόρων καλλιτεχνών και διήρκεσε περίπου 12 ώρες.



Εικόνα 39 Αποτελέσματα ανακάλυψης συνδέσμων με Wikidata, OpenRefine

Κεφάλαιο 4ο
Περιγραφή και Ανάπτυξη Εφαρμογής

στο παρόν κεφάλαιο αναλύεται η ανάπτυξη της εφαρμογής που υλοποιήθηκε στα πλαίσια της παρούσας εργασίας. Σκοπός της εφαρμογής είναι η ανάκτηση και παρουσίαση των διασυνδεδεμένων πολιτιστικών δεδομένων από διαφορετικές πηγές (τοπικά και στον Ιστό).

Αρχικά περιγράφονται οι λειτουργικές απαιτήσεις που πρέπει να πληρούνται στην εφαρμογή. Στη συνέχεια παρουσιάζεται η αρχιτεκτονική της εφαρμογής και οι μέθοδοι που χρησιμοποιήθηκαν για την υλοποίησή της. Στη συνέχεια περιγράφεται ο τρόπος λειτουργίας κάθε δομικού στοιχείου και ο τρόπος που αυτό επικοινωνεί με τα υπόλοιπα.

Η υλοποίηση της εφαρμογής είναι αναρτημένη στο αποθετήριο GitHub στη διεύθυνση https://github.com/sotirisAng/search-on-cultural-lod

4.1. Λειτουργικές απαιτήσεις

1. Η εφαρμογή πρέπει να επιτρέπει την αναζήτηση πολιτιστικών διασυνδεδεμένων δεδομένων.
2. Η εφαρμογή πρέπει να παρέχει τη δυνατότητα εξειδίκευσης της αναζήτησης με βάση κριτήρια-πεδία του γράφου που επιλέγει ο χρήστης.
3. Τα κριτήρια της αναζήτησης πρέπει να αφορούν σε έργα τέχνης, καλλιτέχνες αλλά και στην σχέση τους.
4. Η εφαρμογή πρέπει να μετατρέπει τις αναζητήσεις του χρήστη σε ερωτήματα SPARQL.
5. Η εφαρμογή πρέπει να επικοινωνεί με τον εξυπηρετητή στον οποίο βρίσκεται το τοπικό triple store (με το αντίστοιχο SPARQL endpoint) και να αποστέλλει σε αυτόν ερωτήματα μέσω αιτημάτων HTTP .
6. Η εφαρμογή πρέπει να εμφανίζει το ερώτημα SPARQL που παράγεται δυναμικά για κάθε αναζήτηση του χρήστη βασισμένη στα επιλεγμένα κριτήρια.
7. Η εφαρμογή πρέπει να παρέχει προτάσεις αναζήτησης σχετικές με την οντότητα που αναζητεί ο χρήστης.
8. Τα αποτελέσματα της αναζήτησης πρέπει να προβάλλονται στο χρήστη με διαφορετικές μορφές (πίνακα, γράφος, RDF/XML, JSON-LD).

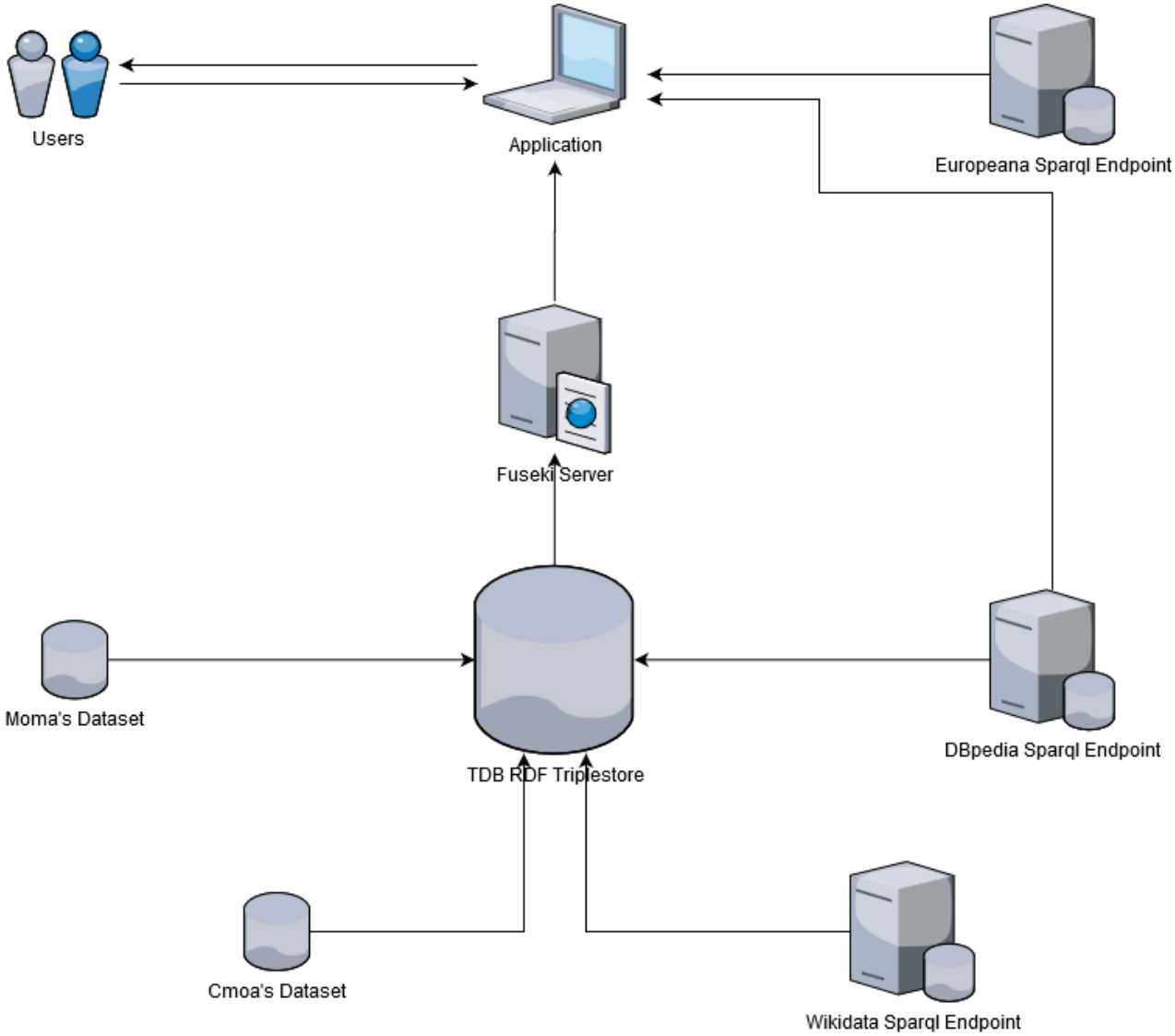
9. Η εφαρμογή πρέπει να δίνει τη δυνατότητα προβολής των αποτελεσμάτων σε γραφική αναπαράσταση του υποσυνόλου του γράφου.
10. Τα επιστρεφόμενα αποτελέσματα που είναι URIs πόρων ή διαδικτυακές διευθύνσεις πρέπει να προβάλλονται σε μορφή υπερσυνδέσμου.
11. Οι υπερσύνδεσμοι που αφορούν διευθύνσεις εκτός εφαρμογής πρέπει να οδηγούν το χρήστη σε αυτές.
12. Οι υπερσύνδεσμοι που αποτελούν URIs πόρων και αφορούν εσωτερικές διευθύνσεις πρέπει να προβάλουν στο χρήστη τις λεπτομέρειες του επιλεγμένου πόρου.
13. Η εφαρμογή πρέπει να προβάλει τη σελίδα λεπτομερειών κάθε πόρου από το URI του.
14. Οι λεπτομέρειες του πόρου πρέπει να παρουσιάζονται με τη μορφή πίνακα και κάθε ιδιότητα να αντιστοιχίζεται με τον σημασιολογικό τύπο της.
15. Αν κάποια από τις λεπτομέρειες του πόρου αφορά εικόνα, τότε αν υπάρχει, η εικόνα πρέπει να εμφανίζεται.
16. Η εφαρμογή θα πρέπει να μπορεί να αναζητά συνδέσμους σχετικούς με τον προβαλλόμενο πόρο, από εξωτερικές πηγές και να τους επιστρέφει στο χρήστη
17. Η εφαρμογή πρέπει να αποθηκεύει στο γράφο, αν δεν υπάρχουν ήδη, τους συνδέσμους που δημιούργησε.

4.2. Αρχιτεκτονική Εφαρμογής

Η εφαρμογή αποτελεί μια single page application συμβατή με σύγχρονους browser υπολογιστών και κινητών τηλεφώνων. Η υποδομή υλοποίησης της εφαρμογής περιλαμβάνει α) την αποθήκευση των ενοποιημένων και διασυνδεδεμένων δεδομένων από τα τοπικά σύνολα δεδομένων και τις συνδέσεις τους με τους γράφους των DBpedia και Wikidata, β) έναν εξυπηρετητή που εξυπηρετεί τα αιτήματα που αποστέλνονται και επιτρέπει την αλληλεπίδραση με τα δεδομένα, γ) τη γραφική διεπαφή με την ενσωματωμένη λειτουργικότητα και δ) την σύνδεσή της με την εξωτερική πηγή δεδομένων της Europeana και της DBpedia.

Όλη η λειτουργικότητα πραγματοποιείται στην πλευρά του client, με αποτέλεσμα να αποτελεί μια front-end εφαρμογή που ενώ στα πλαίσια της διπλωματικής εργασίας λειτούργησε συνδεδεμένη με έναν Fuseki εξυπηρετητή, μπορεί να λειτουργήσει αποτελεσματικά με οποιονδήποτε εξυπηρετητή φιλοξενεί το RDF γράφο των

ενοποιημένων-διασυνδεδεμένων δεδομένων και παρέχει την δυνατότητα εκτέλεσης ερωτημάτων SPARQL (SPARQL endpoint).



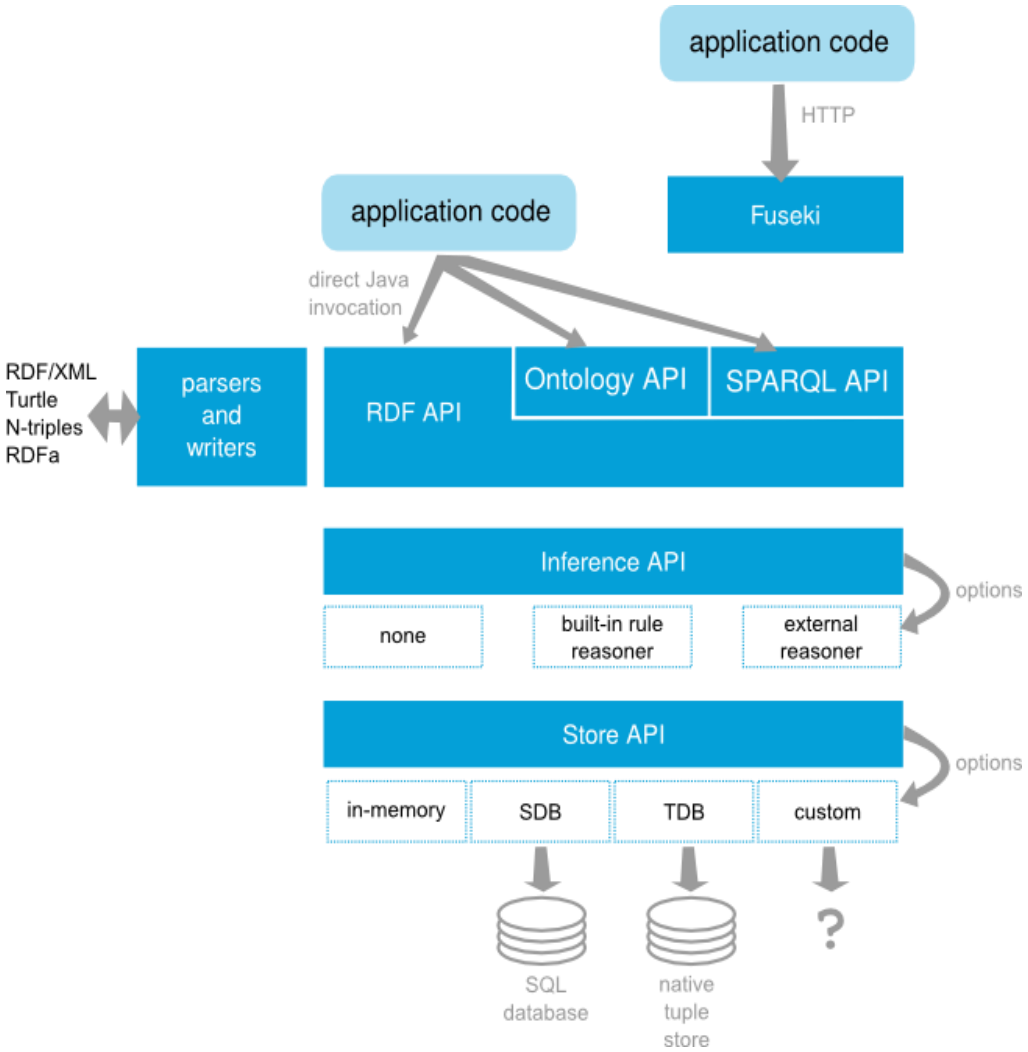
Εικόνα 40 Αρχιτεκτονικό σχέδιο εφαρμογής

4.3. Αποθήκευση και δημοσίευση δεδομένων

Η αποθήκευση των γράφων που παράχθηκαν από τις προηγούμενες διαδικασίες αποθηκεύονται στο TDB triple store της πλατφόρμας Apache Jena ενώ η αλληλεπίδραση με αυτούς επιτυγχάνεται μέσω του Fuseki server της ίδιας πλατφόρμας.

4.3.1. Apache Jena Fuseki

Η Apache Jena [42] είναι μια Java πλατφόρμα εργαλείων για δημιουργία εφαρμογών σημασιολογικού ιστού και διασυνδεδεμένων δεδομένων.



Εικόνα 41 Αρχιτεκτονική Apache Jena

Ο Apache Jena Fuseki [43] είναι ένας εξυπηρετητής (server) που δέχεται ερωτήματα SPARQL. Μπορεί να χρησιμοποιηθεί σαν υπηρεσία, διαδικτυακή εφαρμογή ή σαν αυτόνομος server. Διαθέτει περιβάλλον χρήστη για εποπτεία και διαχείριση της κατάστασης του server και είναι σχεδιασμένος για να λειτουργεί σε συνεργασία με το TDB [44] triple-store ώστε να παρέχουν ένα αξιόπιστο και λειτουργικό επίπεδο αποθήκευσης.

Ο Fuseki υποστηρίζει την υπηρεσία SOH (SPARQL over HTTP) δηλαδή να δέχεται και να εκτελεί ερωτήματα SPARQL μέσω αιτημάτων HTTP. Στη χρήση αυτής της λειτουργίας

βασίστηκε και αναπτύχθηκε η εφαρμογή που υλοποιήθηκε στα πλαίσια της παρούσας διπλωματικής. Ένα τυπικό αίτημα HTTP POST προς τον Fuseki έχει τη μορφή:

```
Host: localhost:3030
Accept: application/json,*/*;q=0.9
Accept-Language: en-US,el;q=0.7,en;q=0.3
Accept-Encoding: gzip, deflate
Content-Type: application/x-www-form-urlencoded; charset=UTF-8
Content-Length: 529
Origin: http://localhost:3000
DNT: 1
Connection: keep-alive
Referer: http://localhost:3000/
params: query = prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
        prefix edm: <http://www.europeana.eu/schemas/edm/>
        prefix dc: <http://purl.org/dc/elements/1.1/>
        prefix skos: <http://www.w3.org/2004/02/skos/core#>
        prefix dct: <http://purl.org/dc/terms/>
        prefix dbo: <http://dbpedia.org/ontology/>
        prefix foaf: <http://xmlns.com/foaf/0.1/>
        SELECT distinct * WHERE{
                ?artist skos:prefLabel ?name.
                ?artist edm:begin ?born.
                ?artist edm:end ?died.
                ?artist skos:note ?Nationality.
                FILTER regex(?name, "Salvador Dalí", "i")
         }
```

Εικόνα 42 HTTP POST αίτημα προς Fuseki

Η απάντηση αν δεν ορισθεί διαφορετικά, επιστέφεται σε μορφή JSON όπως παρακάτω:

```
 "head": {
  "vars": [ "artist" , "name" , "born" , "died" , "Nationality" ]
 } ,
 "results": {
  "bindings": [
   {
    "artist": { "type": "uri" , "value": "http://localhost:3000/cmoa/artist/cc25d328-36af-40e8-87a0-
cce899283526" } ,
    "name": { "type": "literal" , "value": "Salvador Dalí" } ,
    "born": { "type": "literal" , "value": "1904-01-01" } ,
    "died": { "type": "literal" , "value": "1989-01-01" } ,
    "Nationality": { "type": "literal" , "value": "Spanish" }
   } ,
   {
    "artist": { "type": "uri" , "value": "http://localhost:3000/moma/artist/1364" } ,
    "name": { "type": "literal" , "value": "Salvador Dalí" } ,
    "born": { "type": "literal" , "value": "1904" } ,
    "died": { "type": "literal" , "value": "1989" } ,
    "Nationality": { "type": "literal" , "value": "(Spanish, 1904\u20131989)" }
   }
  ]
 }
}
```


Εικόνα 43 Απόκριση Fuseki σε HTTP αίτημα

4.3.2. TDB

Το TDB [44] είναι το triple-store της Apache Jena που χρησιμοποιείται για την αποθήκευση των RDF γράφων. Το σύνολο δεδομένων που αποθηκεύεται μόνιμα (persistently) στο TDB αποτελείται από έναν βασικό, ανώνυμο γράφο (anonymous graph) σε συνδυασμό με άλλους, αν υπάρχουν, ονοματισμένους γράφους (named graphs).

Η χρήση του TDB παράλληλα με τον Fuseki server δίνει τη δυνατότητα μέσω της υπηρεσίας SOH για SPARQL ερωτήματα της μορφής:

- Select, ask, describe, construct μέσω των σημείων http://<uri>/<datasetName>/sparql ή /query και
- Insert, delete data μέσω του σημείου http://<uri>/<datasetName>/update

Παρατίθεται το αρχείο διαμόρφωσης (configuration file) που χρησιμοποιεί ο Fuseki για να δημιουργήσει ένα σύνολο δεδομένων στο TDB.

```
PREFIX :       <#>
PREFIX fuseki:  <http://jena.apache.org/fuseki#>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX tdb2:    <http://jena.apache.org/2016/tdb#>
PREFIX ja:      <http://jena.hpl.hp.com/2005/11/Assembler#>
[] rdf:type fuseki:Server ;
  fuseki:services (
    <#service_tdb2>
  ) .
<#service_tdb2> rdf:type fuseki:Service ;
   rdfs:label                    "MomaCmoa" ;
   fuseki:name                   "MomaCmoa" ;
   fuseki:serviceQuery           "query" ;
   fuseki:serviceQuery           "sparql" ;
   fuseki:serviceUpdate          "update" ;
   fuseki:serviceUpload          "upload" ;
   fuseki:serviceReadWriteGraphStore      "data" ;
   fuseki:serviceReadGraphStore    "get" ;
   fuseki:dataset          <#dataset> .
<#dataset> rdf:type      tdb2:DatasetTDB2 ;
   tdb2:location "/run/databases/MomaCmoa" .
```

Εικόνα 44 Αρχείο Διαμόρφωσης TDB χωρίς συμπερασμό

Με μετατροπή του configuration file της εικόνας 44 του TDB σε αυτό της εικόνας 46, ενεργοποιείται η λειτουργία συμπερασμού (reasoning) της Jena, η οποία προσθέτει νέες τριπλέτες που έχουν εξαχθεί από συμπερασμό (inferred triples) στον γράφο, που παράγονται από τον OWL Reasoner βάσει των ιδιοτήτων των οντολογιών που χρησιμοποιηθήκαν για την σημασιολογική αναπαράσταση των δεδομένων. Για παράδειγμα, οι τριπλέτες που περιγράφουν τον πόρο *<http://localhost:3000/moma/33508/artwork>* «Guitar» και τον καλλιτέχνη <http://localhost:3000/moma/artist/4609> «Pablo Picasso» χωρίς την ενεργοποίηση της διαδικασίας συμπερασμού αναγράφονται στην εικόνα 45, ενώ στην εικόνα 47 αναγράφονται οι τριπλέτες για τους ίδιους πόρους μετά την ολοκλήρωση της διαδικασίας συμπερασμού.

Στα πλαίσια της εργασίας εκτελέστηκαν ερωτήματα SPARQL με ενεργοποιημένη την λειτουργία συμπερασμού για δοκιμαστικά σύνολα δεδομένων με μικρό πλήθος τριπλετών και επιστράφηκαν σαν απάντηση επιτυχώς τα δεδομένα που υπήρχαν ήδη αποθηκευμένα σε συνδυασμό με αυτά που συνάχθηκαν από τον OWL Reasoner. Καθώς όμως ο Fuseki δεν είναι βελτιστοποιημένος για διαδικασίες συμπερασμού μεγάλων συνόλων δεδομένων, όταν έγινε απόπειρα εκτέλεσης των ίδιων ερωτημάτων για όλα τα δεδομένα, η ισχύς και η μνήμη των υπολογιστών στους οποίους έγιναν οι δοκιμές, δεν επαρκούσαν για να ολοκληρωθούν οι υπολογισμοί και να επιστραφούν απαντήσεις στα ερωτήματα.

```
<http://localhost:3000/moma/33508/artwork> a edm:ProvidedCHO;
  dc:creator <http://localhost:3000/moma/artist/4609>;
  dc:date "Paris, December 1912";
  dc:title "Guitar";
  dc:type "Drawing";
  dct:medium "Charcoal on paper";
  ore:isAggregatedBy <http://localhost:3000/moma/33508> .

<http://localhost:3000/moma/artist/4609> a edm:Agent;
  edm:begin "1881";
  edm:end "1973";
  owl:sameAs <http://dbpedia.org/resource/Pablo_Picasso> , <http://localhost:3000/cmoa/artist/b7eaca81-
9a3a-404a-866e-05844d616971>,
    <https://www.wikidata.org/wiki/Q5593>;
  skos:note "(Spanish, 1881–1973)";
  skos:prefLabel "Pablo Picasso" .

<http://localhost:3000/moma/33508> a edm:EuropeanaAggregation;
  edm:aggregatedCHO <http://localhost:3000/moma/33508/artwork>;
  edm:hasView
<http://www.moma.org/media/W1siZiIsIjc1NjQzIl0sWyJwIiwiY29udmVydCIsIi1yZXNpemUgMzAweDMwMFx1
MDAzZSJdXQ.jpg?sha=b340e3e0f4a77c83>;
  owl:sameAs <http://www.moma.org/collection/works/33508> .
```

Εικόνα 45 Τριπλέτες περιγραφής των πόρων «Guitar» και «Pablo Picasso»

```
PREFIX :       <#>
PREFIX fuseki:  <http://jena.apache.org/fuseki#>
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX tdb2:    <http://jena.apache.org/2016/tdb#>
PREFIX ja:      <http://jena.hpl.hp.com/2005/11/Assembler#>
[] rdf:type fuseki:Server ;
  fuseki:services (
    <#service_tdb2>
  ) .
<#service_tdb2> rdf:type fuseki:Service ;
   rdfs:label                    "MomaCmoa" ;
   fuseki:name                   "MomaCmoa" ;
   fuseki:serviceQuery           "query" ;
   fuseki:serviceQuery           "sparql" ;
   fuseki:serviceUpdate          "update" ;
   fuseki:serviceUpload          "upload" ;
   fuseki:serviceReadWriteGraphStore      "data" ;
   fuseki:serviceReadGraphStore    "get" ;
   fuseki:dataset          <#dataset> .
<#dataset> rdf:type       ja:RDFDataset ;
   ja:defaultGraph       <#model_inf>  .
<#model_inf> a ja:InfModel ;
    ja:baseModel <#tdbGraph> ;
    ja:reasoner [
       ja:reasonerURL <http://jena.hpl.hp.com/2003/OWLFBRuleReasoner>
    ] .
<#tdbDataset> rdf:type   tdb2:DatasetTDB2 ;
   tdb2:location "/run/databases/MomaCmoa" .
<#tdbGraph> rdf:type tdb2:GraphTDB ;
tdb2:dataset <#tdbDataset>  .
```

Εικόνα 46 Αρχείο Διαμόρφωσης TDB με συμπερασμό

Ο γράφος που περιέχει τα δεδομένα των συλλογών των δύο μουσείων μαζί με τους εξωτερικούς συνδέσμους και έχει αποθηκευτεί στο TDB περιλαμβάνει περισσότερες από 1.645.000 τριπλέτες.

<http://www.moma.org/collection/works/33508> a edm:EuropeanaAggregation , owl:Thing , owl:Resource ;
edm:aggregatedCHO <http://localhost:3000/moma/33508/artwork> ;
edm:hasView
<http://www.moma.org/media/W1siZiIsIjc1NjQzIl0sWyJwIiwiY29udmVydCIsIi1yZXNpemUgMzAweDMwMFx1MDAzZSJdXQ
.jpg?sha=b340e3e0f4a77c83> ;
owl:sameAs <http://www.moma.org/collection/works/33508> , <http://localhost:3000/moma/33508> .

<http://dbpedia.org/resource/Pablo_Picasso> a owl:Resource , owl:Thing , edm:Agent ;
edm:begin "1881" ;
edm:end "1973" ;
owl:sameAs <http://dbpedia.org/resource/Pablo_Picasso> , <https://www.wikidata.org/wiki/Q5593> ,
<http://localhost:3000/cmoa/artist/b7eaca81-9a3a-404a-866e-05844d616971> ,
<http://localhost:3000/moma/artist/4609> ;
skos:note "(Spanish, 1881–1973)" ;
skos:prefLabel "Pablo Picasso" .

<https://www.wikidata.org/wiki/Q5593> a owl:Thing , owl:Resource , edm:Agent ;
edm:begin "1881" ;
edm:end "1973" ;
owl:sameAs <http://dbpedia.org/resource/Pablo_Picasso> , <http://localhost:3000/cmoa/artist/b7eaca81-9a3a-404a-
866e-05844d616971> , <https://www.wikidata.org/wiki/Q5593> , <http://localhost:3000/moma/artist/4609> ;
skos:note "(Spanish, 1881–1973)" ;
skos:prefLabel "Pablo Picasso" .

<http://localhost:3000/cmoa/artist/b7eaca81-9a3a-404a-866e-05844d616971> a owl:Resource , owl:Thing , edm:Agent ;
edm:begin "1881" ;
edm:end "1973" ;
owl:sameAs <http://dbpedia.org/resource/Pablo_Picasso> , <https://www.wikidata.org/wiki/Q5593> ,
<http://localhost:3000/cmoa/artist/b7eaca81-9a3a-404a-866e-05844d616971> ,
<http://localhost:3000/moma/artist/4609> ;
skos:note "(Spanish, 1881–1973)" ;
skos:prefLabel "Pablo Picasso" .

<http://localhost:3000/moma/artist/4609> a edm:Agent , owl:Thing , owl:Resource ;
edm:begin "1881" ;
edm:end "1973" ;
owl:sameAs <http://localhost:3000/moma/artist/4609> , <http://dbpedia.org/resource/Pablo_Picasso> ,
<http://localhost:3000/cmoa/artist/b7eaca81-9a3a-404a-866e-05844d616971> , <https://www.wikidata.org/wiki/Q5593>
;
skos:note "(Spanish, 1881–1973)" ;
skos:prefLabel "Pablo Picasso" .

<http://localhost:3000/moma/33508> a edm:EuropeanaAggregation , owl:Thing , owl:Resource ;
edm:aggregatedCHO <http://localhost:3000/moma/33508/artwork> ;
edm:hasView
<http://www.moma.org/media/W1siZiIsIjc1NjQzIl0sWyJwIiwiY29udmVydCIsIi1yZXNpemUgMzAweDMwMFx1MDAzZSJdXQ
.jpg?sha=b340e3e0f4a77c83> ;
owl:sameAs <http://localhost:3000/moma/33508> , <http://www.moma.org/collection/works/33508> .

<http://localhost:3000/moma/33508/artwork>a edm:ProvidedCHO ;
dc:creator <http://localhost:3000/moma/artist/4609> , <https://www.wikidata.org/wiki/Q5593> ,
<http://dbpedia.org/resource/Pablo_Picasso> , <http://localhost:3000/cmoa/artist/b7eaca81-9a3a-404a-866e-
05844d616971> ;
dc:date "Paris, December 1912" ;
dc:title "Guitar" ;
dc:type> "Drawing" ;
dct:medium "Charcoal on paper" ;
ore:isAggregatedBy <http://localhost:3000/moma/33508> , <http://www.moma.org/collection/works/33508> .

Εικόνα 47 Τριπλέτες περιγραφής των πόρων «Guitar» και «Pablo Picasso» μετά τη διαδικασία συμπερασμού

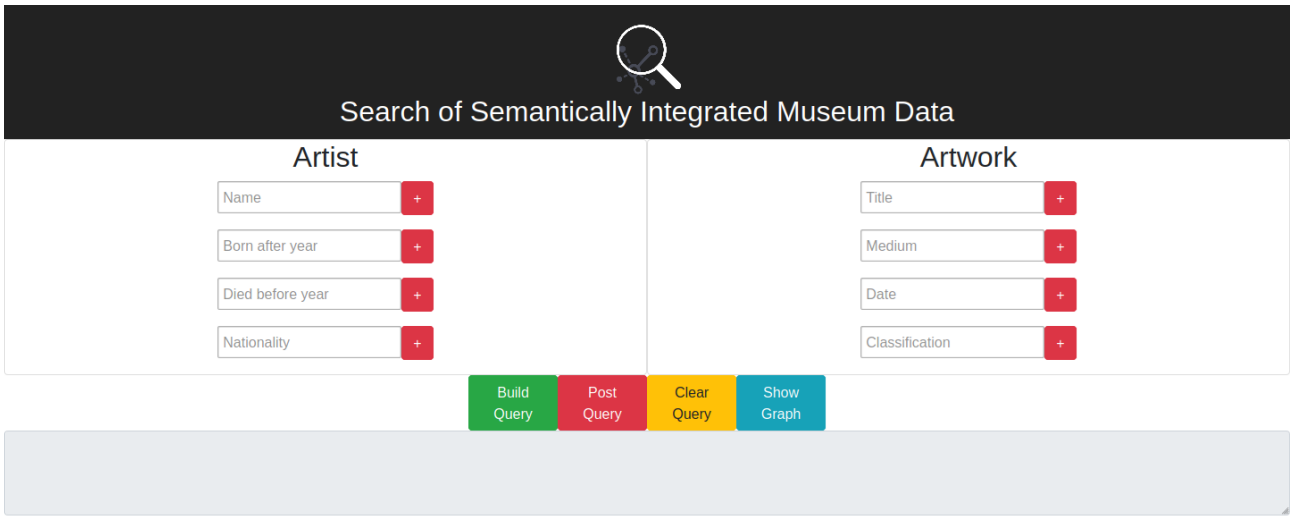
4.4. Ανάπτυξη Εφαρμογής

Για την ανάπτυξη της εφαρμογής χρησιμοποιήθηκαν τα δημοφιλή frameworks α) React.js[45] για την δημιουργία των διεπαφών και την υλοποίηση της λειτουργικότητας, και β) Bootstrap[46] για την μορφοποίηση των διεπαφών. Συγκεκριμένα χρησιμοποιήθηκε η έκδοση React v16.12 με χρήση της JSX[47] και σύνταξη ES6[48] της JavaScript και το Bootstrap 4.

Κατά την σχεδίαση ακολουθήθηκε η λογική που προτείνεται από την React για δημιουργία παραμετροποιήσεων και επαναχρησιμοποιούμενων δομικών στοιχείων που συντίθενται για το χτίσιμο της εφαρμογής. Συνεπώς έγινε προσπάθεια να διακριθούν οι λειτουργίες και να δημιουργηθεί ένα δομικό στοιχείο που θα καλύπτει τις απαιτήσεις κάθε μιας.

Έτσι πέρα από τον βασικό κορμό, δημιουργήθηκαν ακόμη πέντε δομικά στοιχεία της εφαρμογής.

App.js



Εικόνα 48 Αρχική Οθόνη Εφαρμογής

Αποτελεί τον κορμό και το υπόβαθρο της διεπαφής και όλα τα δομικά στοιχεία επιστρέφουν αποτελέσματα σε αυτό. Εδώ καλούνται τα αντικείμενα των πεδίων αναζήτησης και υπάρχει η υποδομή για το χτίσιμο των SPARQL ερωτημάτων. Ανάλογα

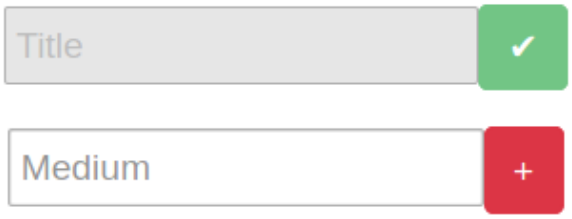
με τις τιμές που επιλέγει ο χρήστης επιστρέφονται οι μεταβλητές και δημιουργούνται τα κατάλληλα φίλτρα. Στη συνέχεια όλη η πληροφορία κωδικοποιείται (encoded/serialized) και δομείται το ερώτημα που αποστέλλεται στον εξυπηρετητή. Επίσης στο App.js ορίζονται οι διαδρομές (routes) και σε ποιο στοιχείο και με ποια δεδομένα θα οδηγούν αυτές.

Inputs.js

Το στοιχείο Inputs.js υλοποιεί τα πεδία εισόδου που χρησιμοποιούνται από τον χρήστη στην διαδικασία της αναζήτησης. Αποτελεί μια φόρμα ενός πεδίου που έχει παραμετροποιηθεί κατάλληλα ώστε ανάλογα με τα ορίσματα που δέχεται κατά την κλήση του, να αντιστοιχίζεται κάθε φορά σε διαφορετικό πεδίο αναζήτησης.

```
<Inputs
        passValue={this.passValue}
        passvalueType={'text'}
        passSubject={this.passSubject}
        triple={"?artist skos:prefLabel ?name. "}
        subject={'?name'}
        placeholder={'Name'}
        clear={this.state.clear_inputs}
/>
```

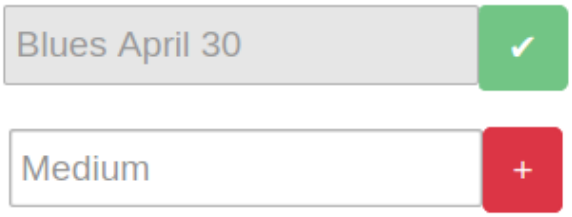
Κατά την υποβολή της φόρμας δομείται μέρος του SPARQL ερωτήματος που θα αποσταλεί όταν ολοκληρωθεί η αναζήτηση. Αν η φόρμα υποβληθεί κενή τότε θα επιστραφεί το πεδίο σαν ζητούμενη μεταβλητή. Αν η φόρμα υποβληθεί συμπληρωμένη τότε θα αποτελεί φίλτρο περιορισμού των αποτελεσμάτων. Για παράδειγμα η φόρμα του τίτλου που υποβάλλεται κενή παράγει το παρακάτω ερώτημα.



Εικόνα 49 Επιλεγμένο πεδίο αναζήτητσης χωρίς ορισμό τιμής

SELECT distinct * WHERE{ ?cho dc:title ?title. ?cho dc:creator ?artist. }

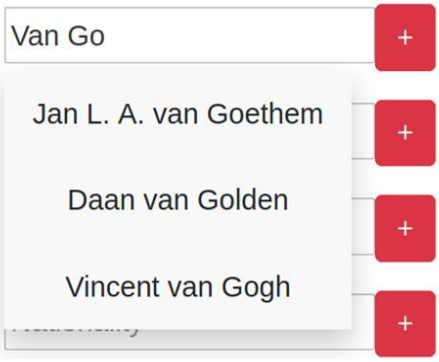
ενώ αν υποβληθεί συμπληρωμένη με τον τίτλο «Blues April 30» προσθέτει το ανάλογο φίλτρο



Εικόνα 50 Επιλεγμένο πεδίο αναζήτησης με επιλογή τιμής

SELECT distinct * WHERE{ ?cho dc:title ?title. ?cho dc:creator ?artist. FILTER regex(?title, "Blues April 30 ", "i") }

Για την διευκόλυνση της αναζήτησης, κατά την πληκτρολόγηση προβάλλονται προτάσεις για το πεδίο που συμπληρώνεται, που περιέχουν ότι έχει πληκτρολογήσει έως εκείνη τη στιγμή ο χρήστης. Αυτό επιτυγχάνεται αποστέλλοντας ένα ερώτημα για τον ανάλογο τύπο κάθε φορά που ο χρήστης προσθέτει ένα χαρακτήρα στο πεδίο της αναζήτησης και επιστρέφοντας όλα τα αποτελέσματα.



Εικόνα 51 Προτεινόμενες τιμές κατά την πληκτρολόγηση

ResultTable.js

Το συγκεκριμένο στοιχείο της εφαρμογής αναλαμβάνει να μετατρέψει τα δεδομένα που επιστρέφονται από τον εξυπηρετητή από μορφή JSON σε μια ευκολότερα αναγνώσιμη από τον χρήστη, δημιουργώντας δυναμικά έναν πίνακα αποτελεσμάτων . Σαν ορίσματα θα περαστούν η απόκριση του εξυπηρετητή αλλά και η δομή του SPARQL ερωτήματος που στάλθηκε. Αρχικά θα δημιουργηθεί ένας πίνακας αντικειμένων από μετατροπή της απόκρισης από την επιστρεφόμενη μορφή σε RDF τριπλέτες συνδυάζοντας τις μεταβλητές του ερωτήματος με τις τιμές που παίρνουν στην απάντηση και χρησιμοποιώντας την σημασιολογική ιδιότητα για τη σύνδεση τους, δηλαδή το στοιχείο της απόκρισης από

```
{
  "artist": {
    "type": "uri",
    "value": "http://localhost:3000/moma/artist/47318"
  },
  "name": {
    "type": "literal",
    "value": "Tomasz Lipiński"
  }
}
```

μετατρέπεται σε

```
{
  "subject": "http://localhost:3000/moma/artist/47318",
  "predicate": "skos:prefLabel",
  "object": "Tomasz Lipiński"
}
```

Η μετατροπή αυτή είναι απαραίτητη γιατί αποτελεί την απαιτούμενη σύνταξη που αναγνωρίζει το δομικό στοιχείο Curve.js για μετέπειτα προβολή των αποτελεσμάτων σε γραφική αναπαράσταση.

Στη συνέχεια, δημιουργούνται τα κελιά του πίνακα και τοποθετούνται μέσα σε αυτά οι τιμές των αποτελεσμάτων. Πριν γίνει η συμπλήρωση του κελιού ελέγχεται το είδος της τιμής και τοποθετείται αναλόγως. Αν πρόκειται για τιμή τύπου literal τότε αυτή τοποθετείται ως έχει, αν πρόκειται για URI τότε η τιμή μετατρέπεται σε υπερσύνδεσμο εκτός της περίπτωσης που ο σημασιολογικός τύπος του υπερσυνδέσμου αναφέρεται σε εικόνα, όπου γίνεται προβολή της.

ARTIST	NAME
http://localhost:3000/moma/artist/47318	Tomasz Lipiński

Εικόνα 52 Αναπαράσταση αποτελεσμάτων αναζήτησης σε μορφή πίνακα

ResourceDetails.js

Το ResourceDetails έχει υλοποιηθεί για να παρουσιάζει τις λεπτομέρειες ενός επιλεγμένου πόρου και να εκτελεί διαδικασίες διασύνδεσης του πόρου με εξωτερικές πηγές. Αρχικά κατά την φόρτωση γίνεται ανάλυση του URL της σελίδας, που αποτελεί και URI του πόρου, και με βάση αυτό χτίζεται ένα SPARQL ερώτημα προς τον Fuseki server. Μόλις ληφθούν τα αποτελέσματα, καλείται άλλη μια φορά ο ResourceTable για να τα μετατρέψει στην επιθυμητή μορφή και να τα παρουσιάσει.

Αν πληρούνται συγκεκριμένα κριτήρια, στο ερώτημα προστίθεται ένα *federated SPARQL query* το οποίο αναζητά όμοιους πόρους στους γράφους της Europeana και της Dbpedia βάσει του σημασιολογικού τύπου του πόρου και της ιδιότητας που τον περιγράφει. Αν υπάρχουν αποτελέσματα από την απομακρυσμένη αναζήτηση θα εκτελέσει μια σύγκριση της τιμής που αποστάλθηκε με τις αντίστοιχες τιμές του αποτελέσματος με χρήση μιας μετρικής ομοιότητας αλφαριθμητικού (string similarity metric). Στην υλοποίηση αυτή χρησιμοποιήθηκε επιλεκτικά η μετρική «Jaro-distance». Αν η τιμή της μετρικής υπερβαίνει ένα προκαθορισμένο όριο, τότε ορίζεται ομοιότητα και γίνεται ενέργεια καταγραφής του νέου συνδέσμου στο γράφο. Έτσι γίνεται πρώτα ένα ερώτημα τύπου *ask* στον γράφο για να εξακριβωθεί ότι δεν έχει ήδη αποθηκευτεί η ομοιότητα και

αν η απάντηση είναι αρνητική εκτελείται νέο ερώτημα για την εισαγωγή νέων δεδομένων.

Ακολουθώντας τα πρότυπα διαλειτουργικότητας,, εκτός από την παρουσίαση των δεδομένων σε μορφή πίνακα, δίνεται η δυνατότητα προβολής των λεπτομερειών του πόρου και σε μορφή RDF-XML και JSON-LD, αποστέλλοντας ένα SPARQL ερώτημα τύπου Describe για το URI του πόρου και ορίζοντας το ζητούμενο τύπο ανάκτησης των αποτελεσμάτων στην κεφαλίδα του HTTP αιτήματος που θα σταλεί στον εξυπηρετητή, συγκεκριμένα αποστέλλεται

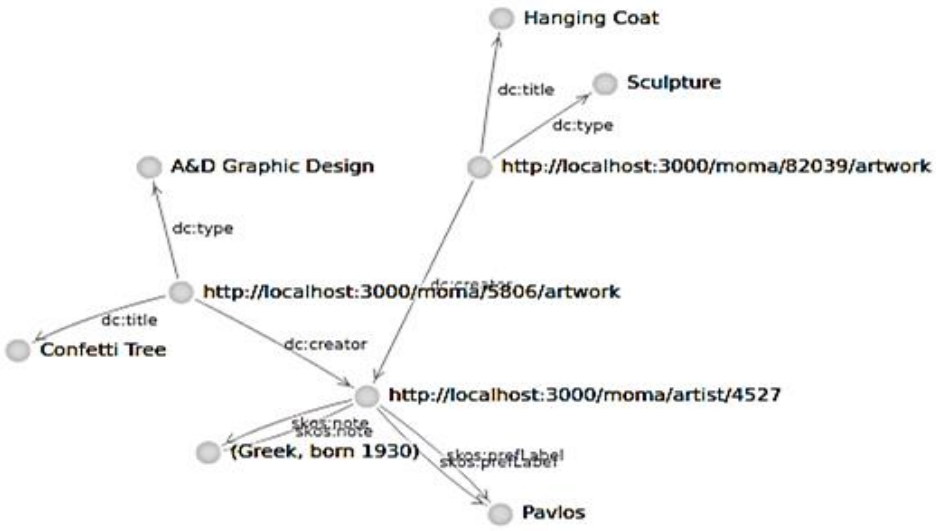
- Accept: application/json για λήψη περιεχομένων σε μορφή json, αποτελεί την προεπιλογή
- Accept: application/rdf+xml για λήψη δεδομένων σε μορφή RDF-XML
- Accept:application/ld+json για λήψη δεδομένων σε μορφή JSON-LD

MakeHttpReq.js

Το συγκεκριμένο στοιχείο χρησιμοποιείται από τα υπόλοιπα για την αποστολή ερωτημάτων στον Fuseki server δημιουργώντας και αποστέλλοντας ένα HTTP αίτημα. Δέχεται σαν όρισμα το τελικό σημείο αποστολής του ερωτήματος ανάλογα με τον τύπο του, δηλαδή θα λάβει την τιμή query αν πρόκειται για αναζήτηση στοιχείων ή ερώτηση ύπαρξης πληροφορίας, ενώ την τιμή update στην περίπτωση που πρόκειται να εισαχθούν πληροφορίες στο γράφο. Δεύτερο όρισμα είναι το ερώτημα καθαυτό που πρόκειται να αποσταλεί, και τρίτο η επιθυμητή μορφή στην οποία θα επιστραφούν τα δεδομένα. Η τιμή που επιστρέφεται είναι η απάντηση του εξυπηρετητή στην μορφή που ζήτησε η εκάστοτε κλήση.

Το Curve.js

Το στοιχείο Curve.js έχει ως σκοπό να αναπαραστήσει γραφικά τα αποτελέσματα της αναζήτησης. Αρχικά μετατρέπει τα αντικείμενα τύπου {subject, property, object} που δημιουργήθηκαν στο ResultTable σε αντικείμενα κόμβων και ακμών. Τα αντικείμενα αυτά χρησιμοποιούνται για να αναπαρασταθεί ο γράφος με χρήση της βιβλιοθήκης D3 , όπου αφού οριστούν τα χαρακτηριστικά τους θα δημιουργήσουν έναν κατευθυνόμενο γράφο με δυναμικές θέσεις σημείων, οι οποίες μπορούν να αλλάξουν από το χρήστη για καλύτερη εποπτεία.



Εικόνα 53 Αναπαράσταση αποτελεσμάτων αναζήτησης σε μορφή κατευθυνόμενου γράφου

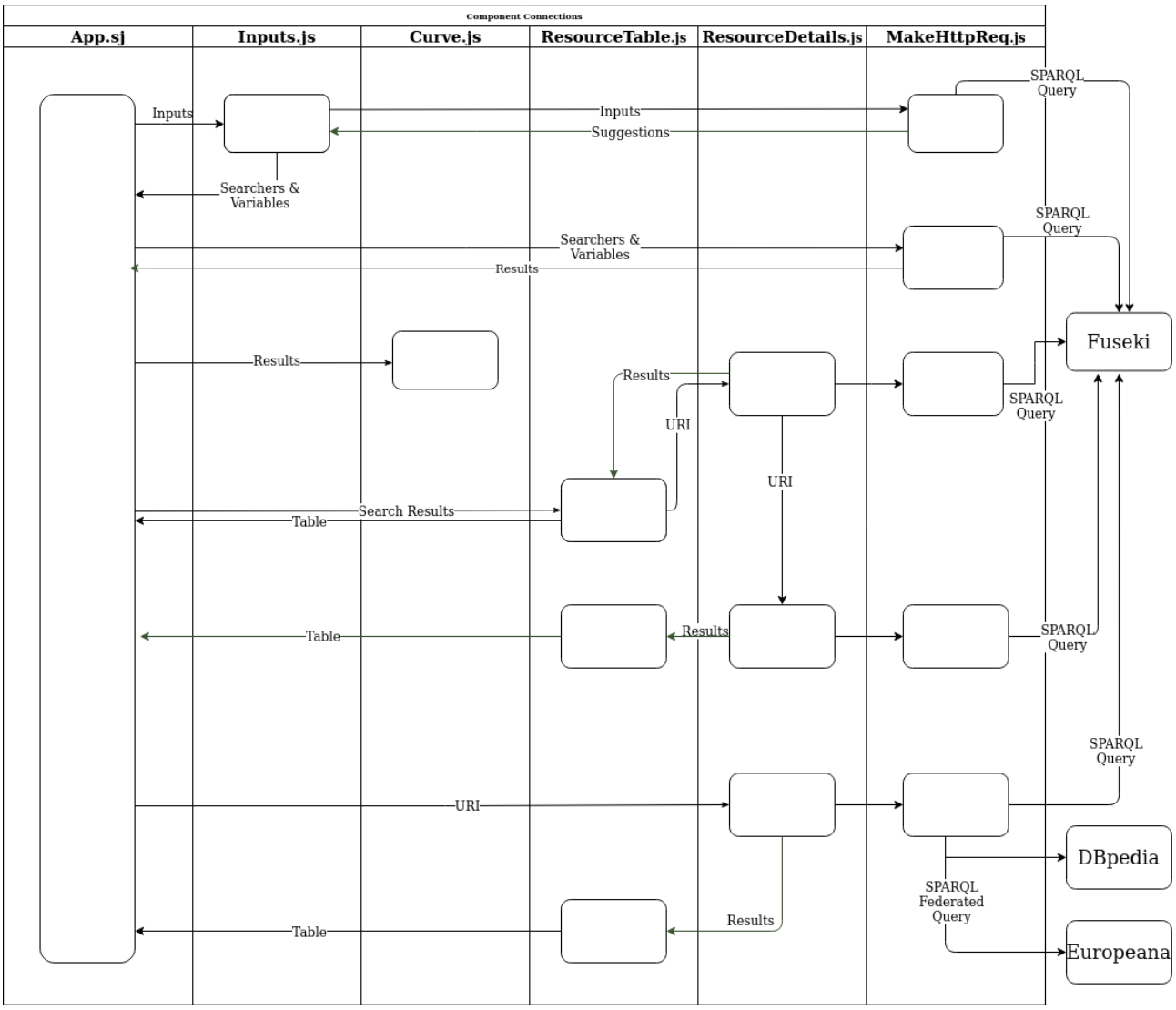
Τα τμήματα της εφαρμογής αλληλοεπιδρούν μεταξύ τους όπως φαίνονται στο παρακάτω σχήμα της εικόνας 54.

Στοιχεία κώδικα από άλλες πηγές

Κατά την υλοποίηση της εφαρμογής χρησιμοποιήθηκαν επίσης οι εξής προσθήκες:

- Η επέκταση της React, τον React Router για την μετάβαση σε στα διαφορετικά στοιχεία της εφαρμογής και την δυναμική χρήση των URL, δημιουργώντας ξεχωριστές σελίδες για κάθε πόρο.
- Την βιβλιοθήκη axios που χρησιμοποιήθηκε για την εκτέλεση των HTTP αιτημάτων στον εξυπηρετητή.
- Το στοιχείο της εφαρμογής Curve.js, αποτελεί τροποποίηση του εργαλείου D3RDF [https://github.com/Rathachai/d3rdf], σε σύνταξη της JavaScript ES6 και προσαρμογή του ώστε να είναι συμβατό με την λειτουργία της React καθώς επίσης και μετατροπή ορισμένων σημείων ώστε να είναι συμβατό με την νεότερη

έκδοση της βιβλιοθήκης D3.js που βρίσκεται στην 5η έκδοση, έναντι της 3ης στην οποία υλοποιήθηκε αρχικά.



Εικόνα 54 Αλληλεπίδραση δομικών στοιχείων εφαρμογής

Κεφάλαιο 5ο
Σενάρια Χρήσης Εφαρμογής

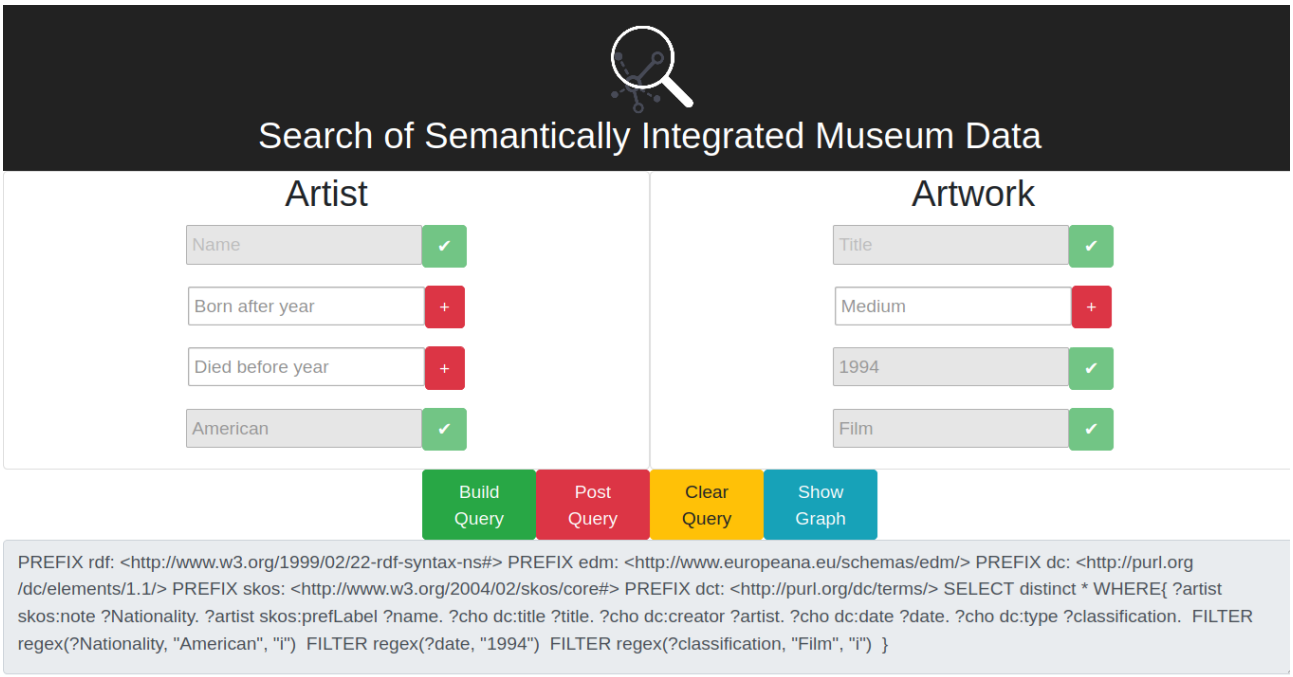
Στο παρόν κεφάλαιο παρουσιάζονται επιλεκτικά δύο από τα σενάρια χρήσης που δημιουργήθηκαν για την αξιολόγηση της εφαρμογής.

5.1. Σενάριο Χρήσης #1

Ο χρήστης αναζητά πληροφορίες για:

«*Κινηματογραφικές ταινίες Αμερικάνων καλλιτεχνών που γυρίστηκαν το 1994*»

Για τον ορισμό του ερωτήματος επιλέγονται οι μεταβλητές «όνομα καλλιτέχνη» και «τίτλος έργου» και ως φίλτρα, η « εθνικότητα του καλλιτέχνη», η «ημερομηνία δημιουργίας» και η «ταξινόμηση του έργου», που λαμβάνουν τις τιμές «American», «1994» και «Film» αντίστοιχα.



Εικόνα 55 Αναζήτηση και δημιουργία SPARQL ερωτήματος, Σενάριο χρήσης #1

Το SPARQL ερώτημα που παράγεται είναι :

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX edm: <http://www.europeana.eu/schemas/edm/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT distinct * WHERE {
    ?artist skos:note ?Nationality.
    ?artist skos:prefLabel ?name.
    ?cho dc:title ?title.
    ?cho dc:creator ?artist.
    ?cho dc:date ?date.
    ?cho dc:type ?classification.
    FILTER regex(?Nationality, "American", "i")
    FILTER regex(?date, "1994")
    FILTER regex(?classification, "Film", "i")
 }
```

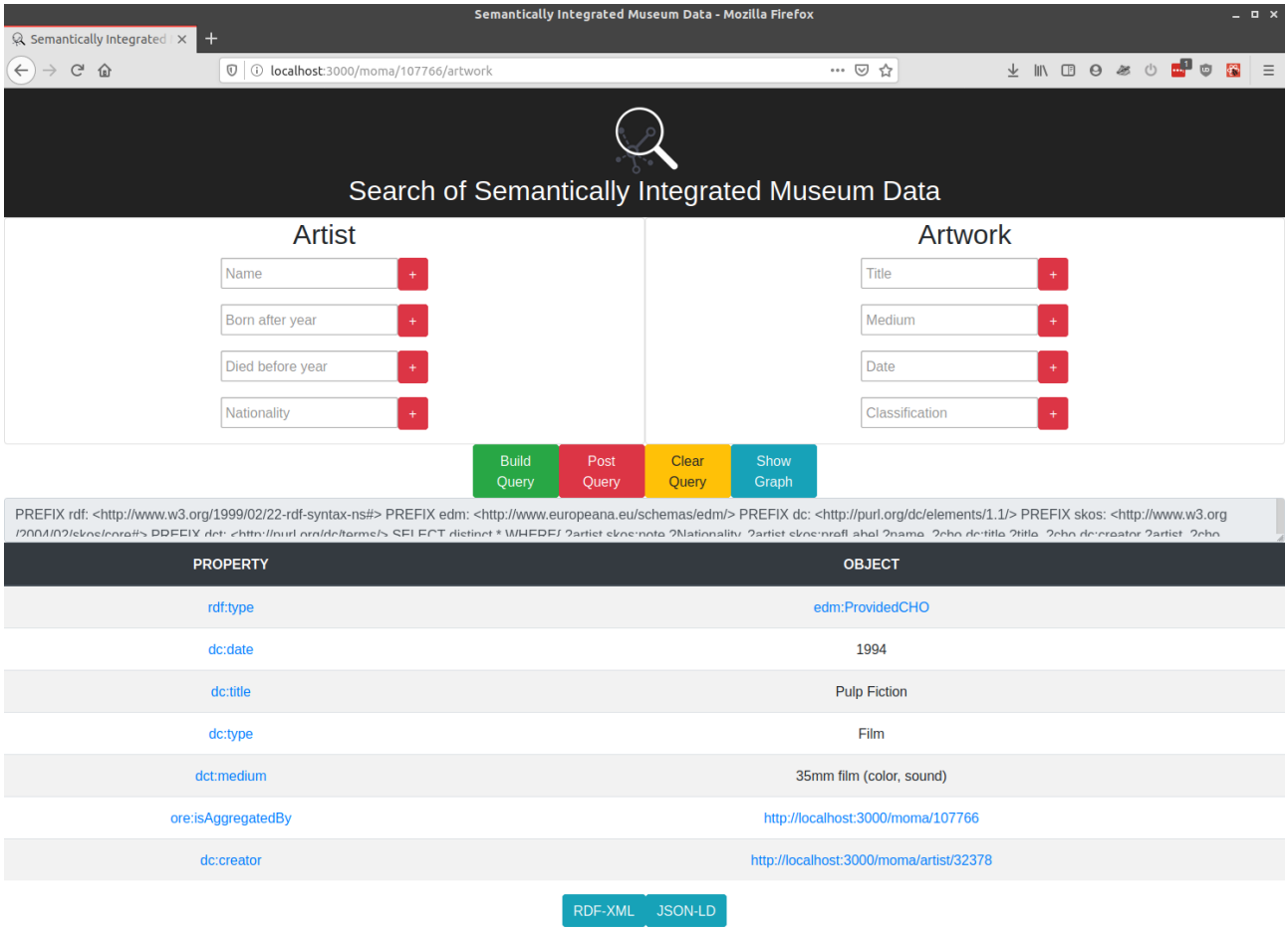
Αφού παραχθεί και στη συνέχεια αποσταλεί το ερώτημα στον εξυπηρετητή, επιστρέφονται τα αποτελέσματα, τα οποία προβάλλονται σε μορφή πίνακα. Ο πίνακας δομείται με την σειρά που επιλέχθηκαν οι μεταβλητές και τα φίλτρα, δηλαδή: αναγνωριστικό καλλιτέχνη, εθνικότητα, όνομα, αναγνωριστικό έργου, τίτλος έργου, ημερομηνία δημιουργίας και κατηγοριοποίηση, όπως παρουσιάζεται στην εικόνα 56.

ARTIST	NATIONALITY	NAME	CHO	TITLE	DATE	CLASSIFICATION
http://localhost:3000/moma/artist/41799	(American)	George Griffin	http://localhost:3000/moma/160266/artwork	A Little Routine	1994	Film
http://localhost:3000/moma/artist/42705	(American, born 1955)	Jay Rosenblatt	http://localhost:3000/moma/165665/artwork	The Smell of Burning Ants	1994	Film
http://localhost:3000/moma/artist/36155	(American, born 1958)	Tim Burton	http://localhost:3000/moma/173343/artwork	Ed Wood	1994	Film
http://localhost:3000/moma/artist/32360	(American, born 1958)	David O. Russell	http://localhost:3000/moma/157436/artwork	Spanking the Monkey	1994	Film
http://localhost:3000/moma/artist/41716	(American, born 1950)	Wheeler Winston Dixon	http://localhost:3000/moma/162017/artwork	Squatters	1994	Film
http://localhost:3000/moma/artist/41476	(American, born 1961)	Lynne Sachs	http://localhost:3000/moma/173214/artwork	Which Way is East: Notebooks from Vietnam	1994	Film
http://localhost:3000/moma/artist/32272	(American, 1924–2001)	Faith Hubley	http://localhost:3000/moma/160581/artwork	Seers and Clowns	1994	Film
http://localhost:3000/moma/artist/44277	(American, born 1946)	Michael Sporn	http://localhost:3000/moma/173202/artwork	Whitewash	1994	Film
http://localhost:3000/moma/artist/32206	(American, born 1954)	Joel Coen	http://localhost:3000/moma/120033/artwork	The Hudsucker Proxy	1994	Film
http://localhost:3000/moma/artist/34062	(American, born 1930)	Frederick Wiseman	http://localhost:3000/moma/125596/artwork	High School II	1994	Film
http://localhost:3000/moma/artist/37733	(American, born 1959)	Michael Almereyda	http://localhost:3000/moma/133333/artwork	Nadja	1994	Film
http://localhost:3000/moma/artist/32378	(American, born 1963)	Quentin Tarantino	http://localhost:3000/moma/107766/artwork	Pulp Fiction	1994	Film
http://localhost:3000/moma/artist/31268	(American, born 1970)	Kevin Smith	http://localhost:3000/moma/104629/artwork	Clerks	1994	Film
http://localhost:3000/moma/artist/32041	(American, 1925–2006)	Robert Altman	http://localhost:3000/moma/160745/artwork	Prêt-à-Porter (Ready to Wear)	1994	Film
http://localhost:3000/moma/artist/34037	(American, born ca. 1958)	Emily Hubley	http://localhost:3000/moma/160563/artwork	Enough	1994	Film

Εικόνα 56 Αποτελέσματα αναζήτησης, Σενάριο χρήσης #1

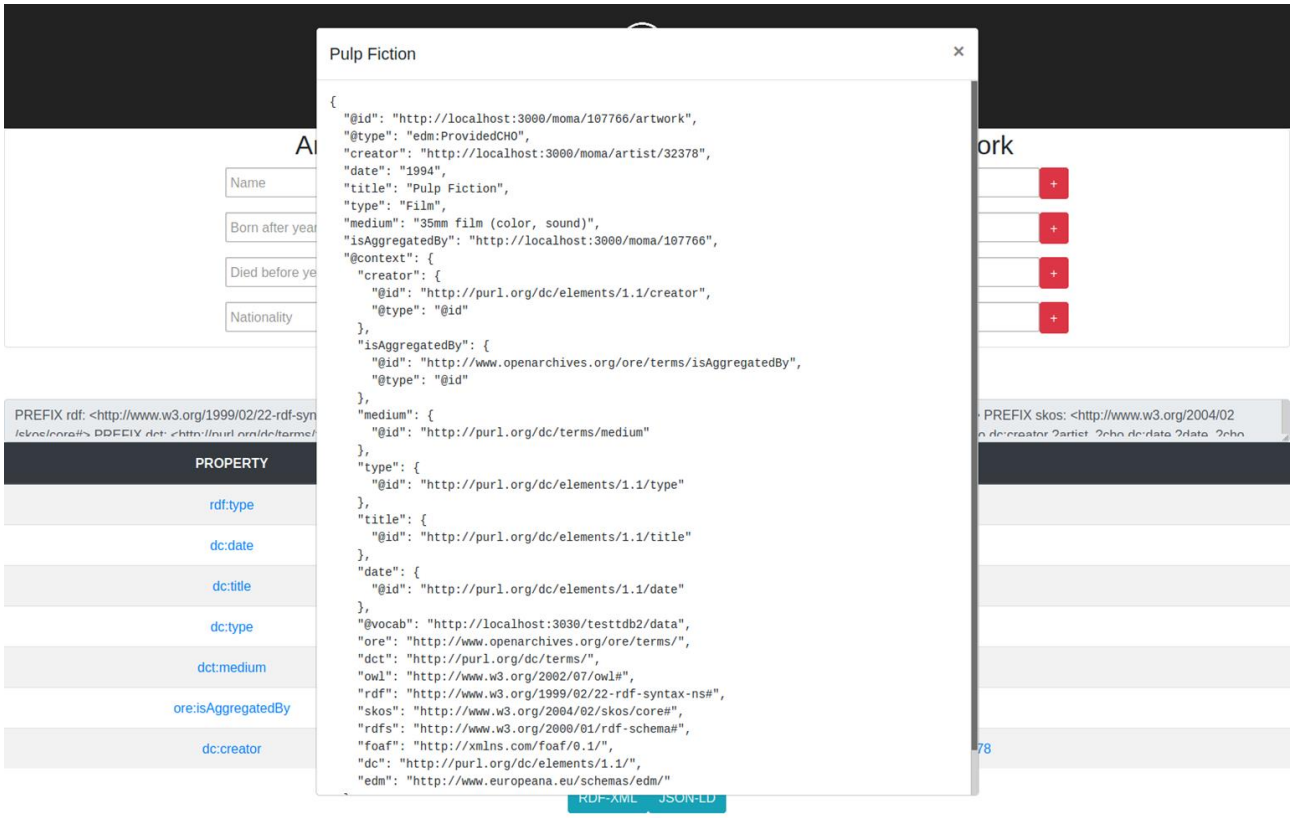
Ο χρήστης επιλέγει έναν από τους πόρους για να λάβει περισσότερες πληροφορίες σχετικά με αυτόν. Στο παρόν σενάριο χρήσης, επιλέγεται ο σύνδεσμος της ταινίας 'Pulp Fiction' του 'Quentin Tarantino'. Ο σύνδεσμος οδηγεί στην σελίδα 'localhost:3000/moma/107766/artwork' του συγκεκριμένου πόρου και φορτώνονται τα διαθέσιμα δεδομένα, δηλαδή δεδομένα που αφορούν στην περιγραφή του πόρου δηλαδή ότι ο πόρος είναι αντικείμενο πολιτιστικής κληρονομιάς, δημιουργήθηκε το 1994, έχει τίτλο 'Pulp Fiction', κατηγοριοποιείται ως ταινία, το μέσο δημιουργίας ήταν φιλμ 35 χιλιοστών, ο πόρος που συναθροίζει τις

πληροφορίες του έργου με την διαδικτυακή του αναπαράσταση είναι ο ‘localhost:3000/ moma/107766’ και ότι δημιουργός είναι ο καλλιτέχνης με αναγνωριστικό ‘localhost:3000/moma/artist/32378’.

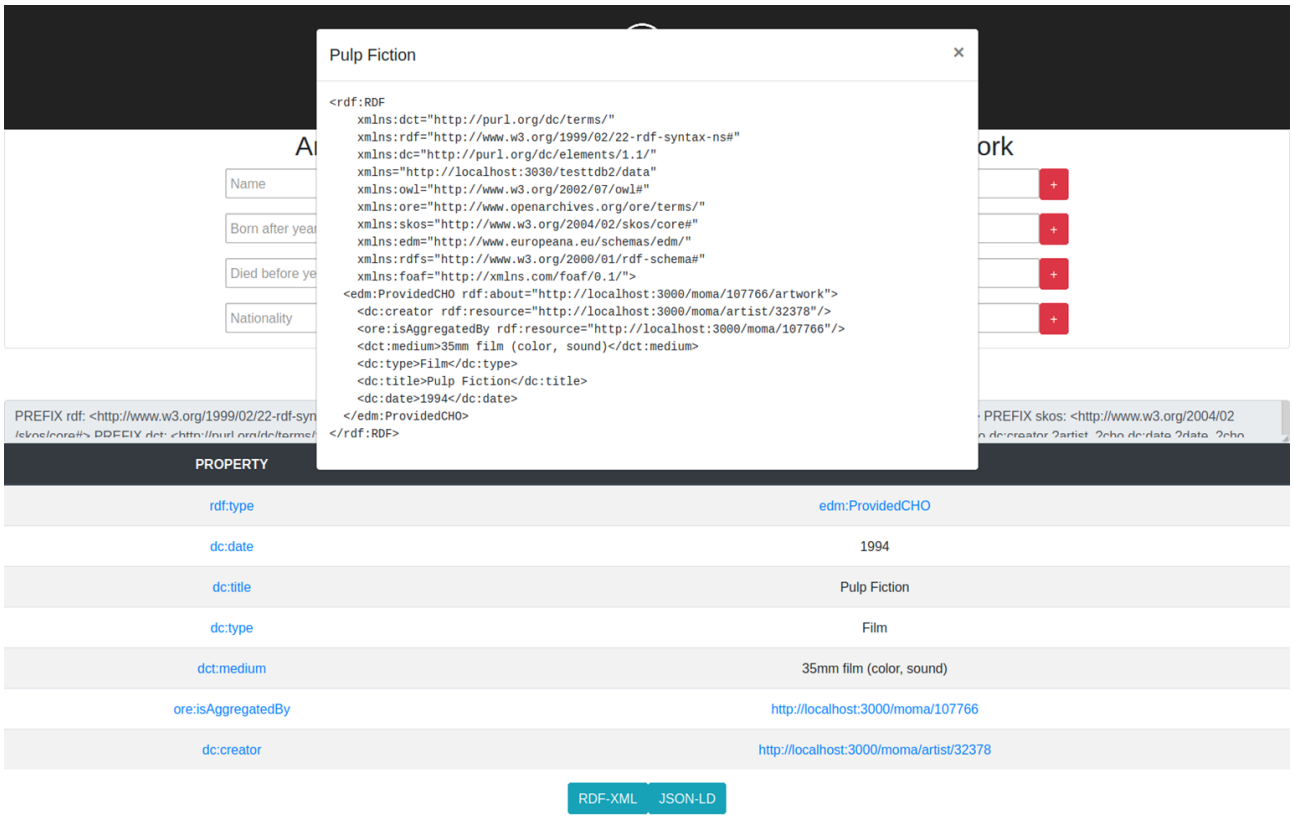


Εικόνα 57 Επιλογή πόρου για προβολή, Σενάριο χρήσης #1

Στη συνέχεια, επιλέγεται η εναλλακτική προβολή των πληροφοριών στις μορφές RDF-XML και JSON-LD, με το πάτημα του αντίστοιχου κουμπιού της διεπαφής.

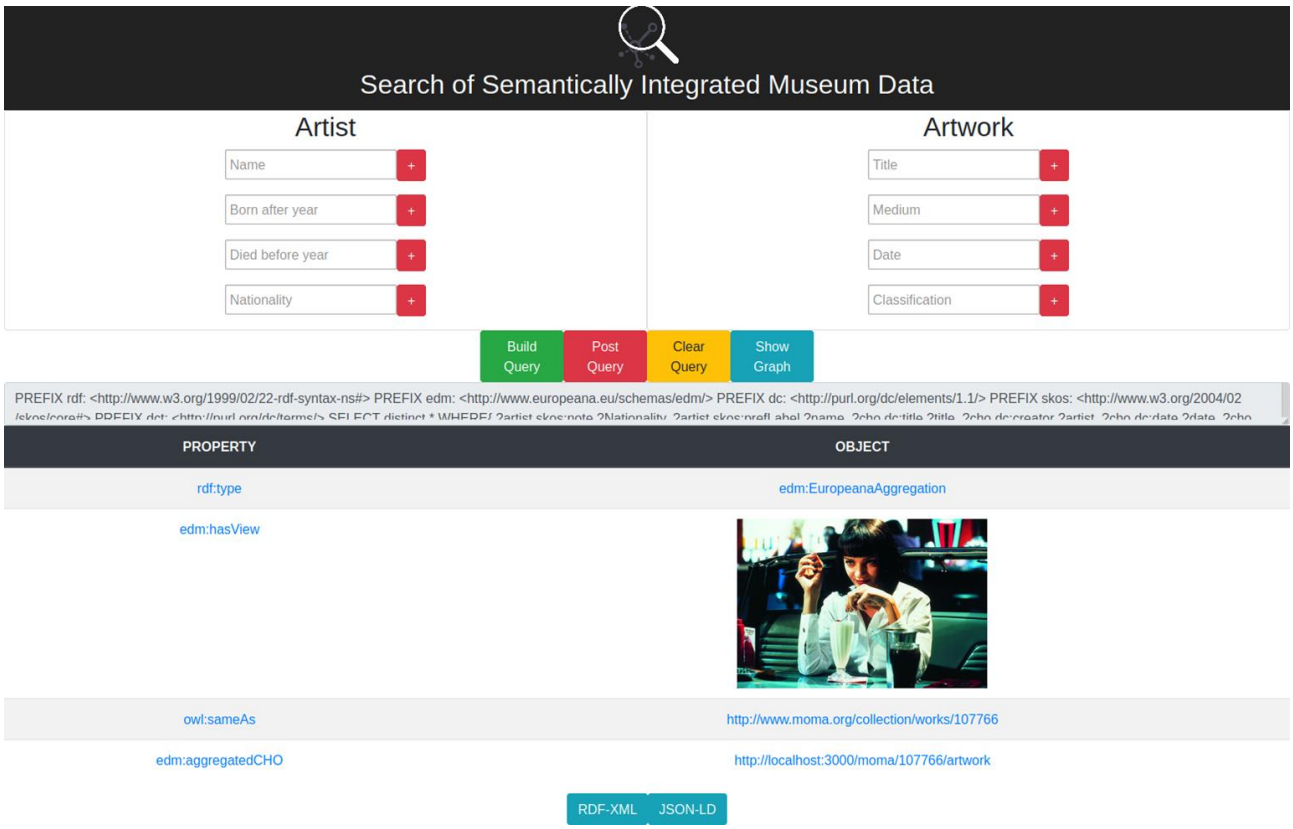


Εικόνα 59 Εναλλακτική προβολή λεπτομερειών πόρου JSON-LD, Σενάριο χρήσης #1



Εικόνα 58 Εναλλακτική προβολή λεπτομερειών πόρου RDF-XML, Σενάριο χρήσης #1

Τέλος, επιλέγεται η μετάβαση στη σελίδα προβολής λεπτομερειών του πόρου συνάθροισης για το συγκεκριμένο έργο τέχνης. Εκεί εμφανίζονται οι ιδιότητες όπως ο τύπος πόρου που ισούται με EuropeanaAggregation, το έργο στο οποίο αναφέρεται, ο εξωτερικός σύνδεσμος που περιγράφει τον πόρο στην ιστοσελίδα του μουσείου, αλλά και προβάλλεται η εικόνα που έχει ορίσει το μουσείο για το συγκεκριμένο έργο.

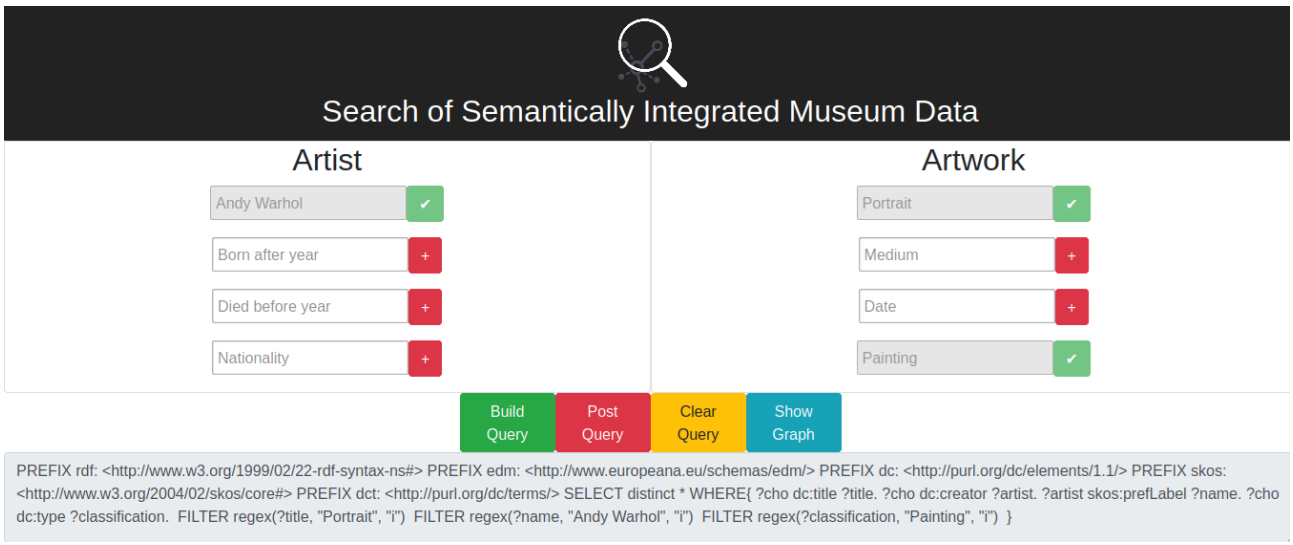


Εικόνα 60 Προβολή λεπτομερειών πόρου συνάθροισης, Σενάριο χρήσης #1

5.2. Σενάριο χρήσης #2

Ο χρήστης αναζητά πληροφορίες για:

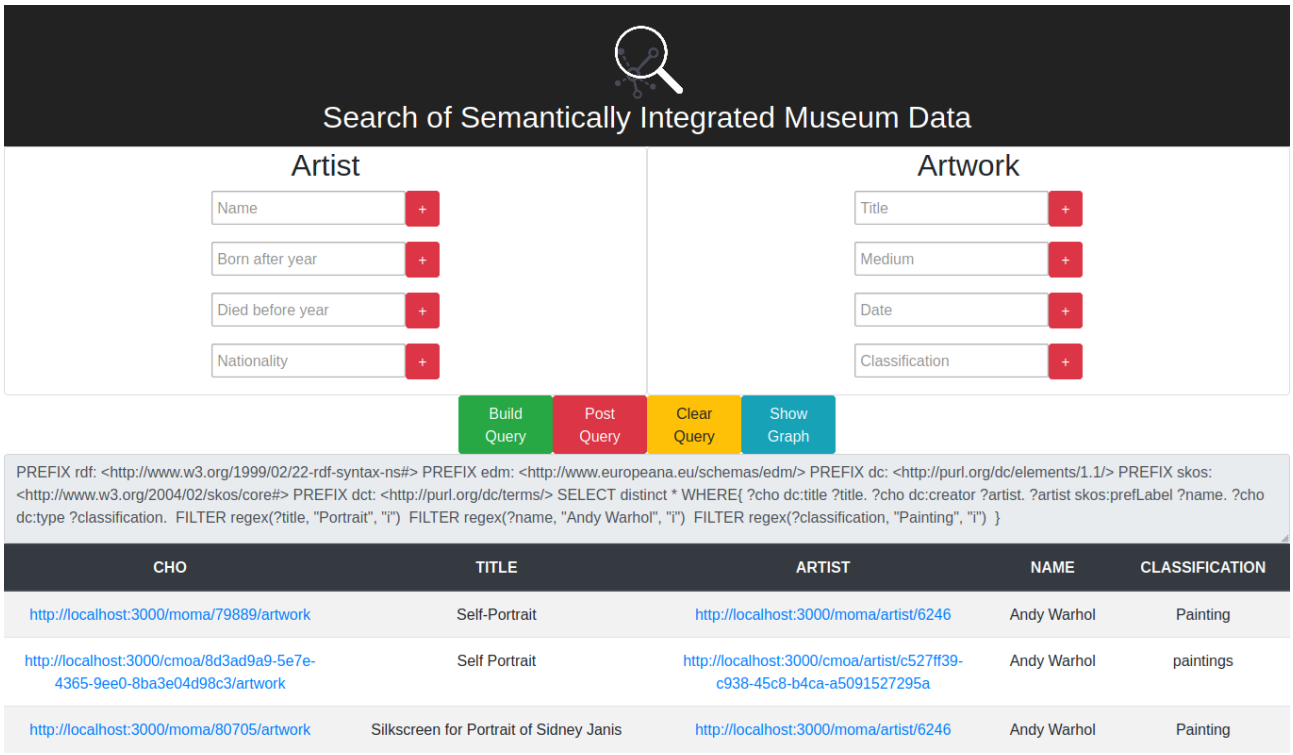
«*Πορτραίτα που έχει ζωγραφίσει ο Andy Warhol*»



Εικόνα 61 Αναζήτηση και δημιουργία SPARQL ερωτήματος, Σενάριο χρήσης #2

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX edm: <http://www.europeana.eu/schemas/edm/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dct: <http://purl.org/dc/terms/>
SELECT distinct * WHERE{ ?
    cho dc:title ?title.
    ?cho dc:creator ?artist.
    ?artist skos:prefLabel ?name.
    ?cho dc:type ?classification.
    FILTER regex(?title, "Portrait", "i")
    FILTER regex(?name, "Andy Warhol", "i")
    FILTER regex(?classification, "Painting", "i")
 }
```

Αφού παραχθεί και αποσταλεί το ερώτημα στον εξυπηρετητή, ο πίνακας αποτελεσμάτων δομείται από τις μεταβλητές: αναγνωριστικό έργου, τίτλος έργου, αναγνωριστικό καλλιτέχνη, όνομα καλλιτέχνη και κατηγοριοποίηση, όπως παρουσιάζεται στην επόμενη εικόνα.



CHO	TITLE	ARTIST	NAME	CLASSIFICATION
http://localhost:3000/moma/79889/artwork	Self-Portrait	http://localhost:3000/moma/artist/6246	Andy Warhol	Painting
http://localhost:3000/cmoa/8d3ad9a9-5e7e-4365-9ee0-8ba3e04d98c3/artwork	Self Portrait	http://localhost:3000/cmoa/artist/c527ff39-c938-45c8-b4ca-a5091527295a	Andy Warhol	paintings
http://localhost:3000/moma/80705/artwork	Silkscreen for Portrait of Sidney Janis	http://localhost:3000/moma/artist/6246	Andy Warhol	Painting

Εικόνα 62 Προβολή αποτελεσμάτων αναζήτησης, Σενάριο χρήσης #2

Τα αποτελέσματα που επιστρέφονται για το συγκεκριμένο ερώτημα περιλαμβάνουν πόρους και από τα δύο σύνολα δεδομένων. Συγκεκριμένα το έργο «Self Portrait» με αναγνωριστικό «http://localhost:3000/cmoa/8d3ad9a9-5e7e-4365-9ee0-8ba3e04d98c3/artwork» προέρχεται από την συλλογή του CMoA, ενώ τα άλλα δύο έργα προέχονται από τη συλλογή του MoMA.

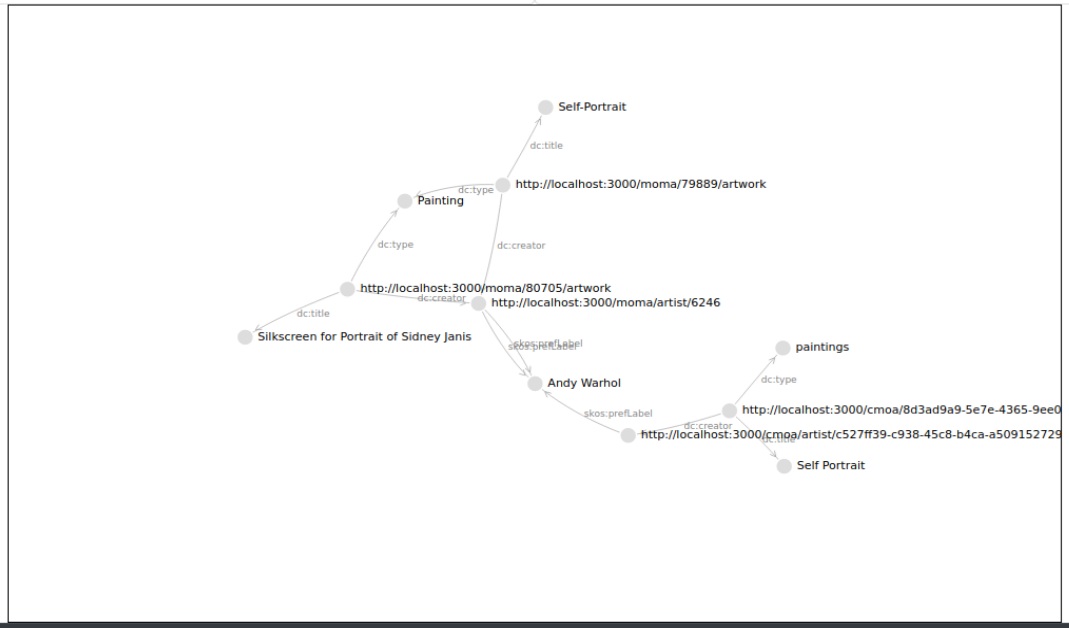
Επιλέγεται η γραφική αναπαράσταση των αποτελεσμάτων σε κατευθυνόμενο γράφο, όπου εμφανίζονται και οι συσχετίσεις μεταξύ των δεδομένων.

PROPERTY	OBJECT
rdf:type	edm:ProvidedCHO
dc:creator	http://localhost:3000/moma/artist/6246
dc:date	1967
dc:title	Silkscreen for Portrait of Sidney Janis
dc:type	Painting
dct:medium	Photosensitive gelatin on silk, with wood frame
ore:isAggregatedBy	http://localhost:3000/moma/80705

Εικόνα 63 Προβολή λεπτομερειών πόρου, Σενάριο χρήσης #2

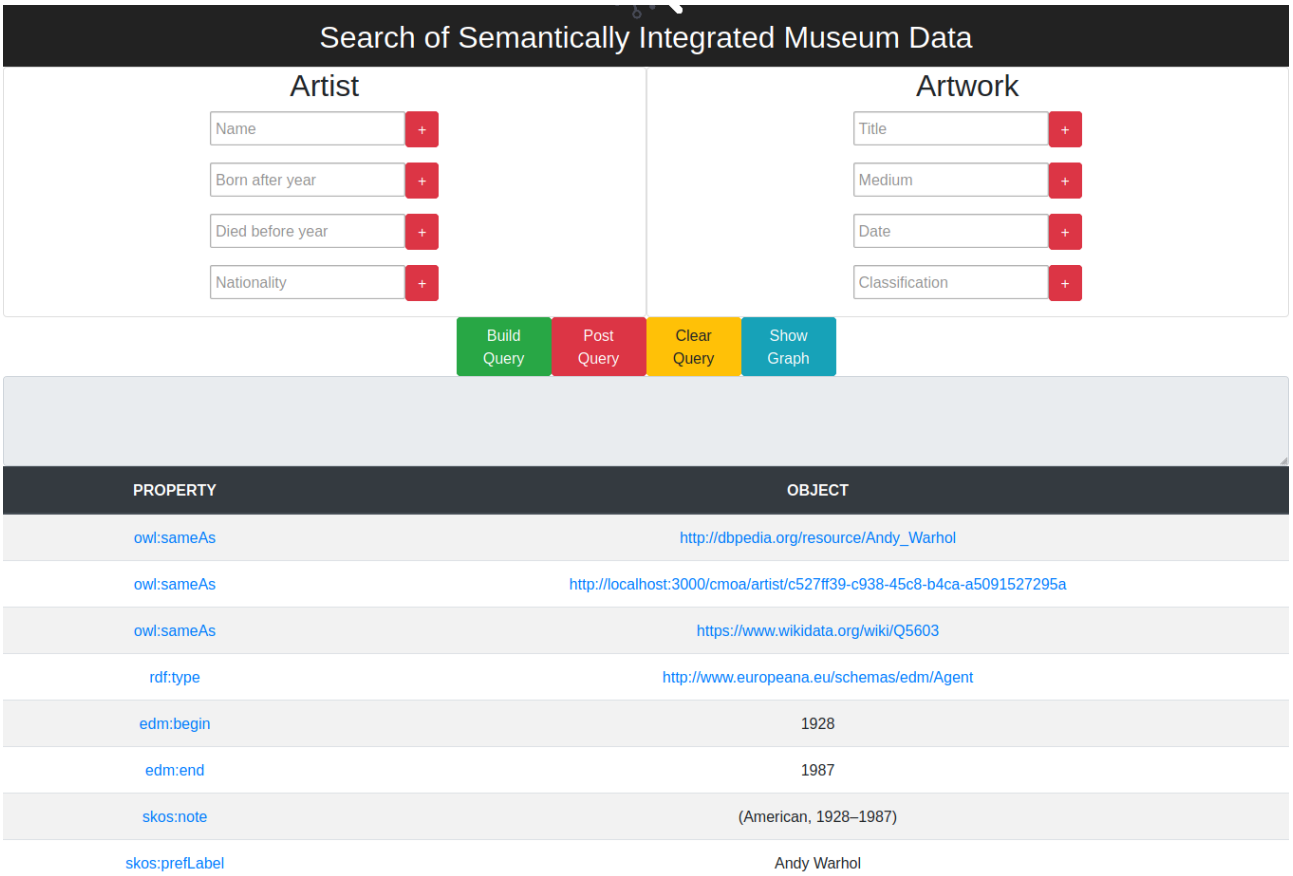
Επιλέγεται η μετάβαση στη σελίδα ιδιοτήτων του πόρου με τίτλο «Silkscreen for Portrait of Sidney Janis», οι οποίες προβάλλονται στον παρακάτω πίνακα.

Έπειτα, επιλέγεται η προβολή του πόρου του καλλιτέχνη και εμφανίζονται οι ιδιότητες που χαρακτηρίζουν τον πόρο, Αμερικάνος, έτος γέννησης 1928, έτος θανάτου 1987, όνομα Andy Warhol, τύπος πόρου Agent και τρεις σύνδεσμοι που εκφράζουν ομοιότητα με τον συγκεκριμένο πόρο, δύο σε εξωτερικές πηγές, της Wikidata και της DBpedia, και ένας στον πόρο της συλλογής CMoA.

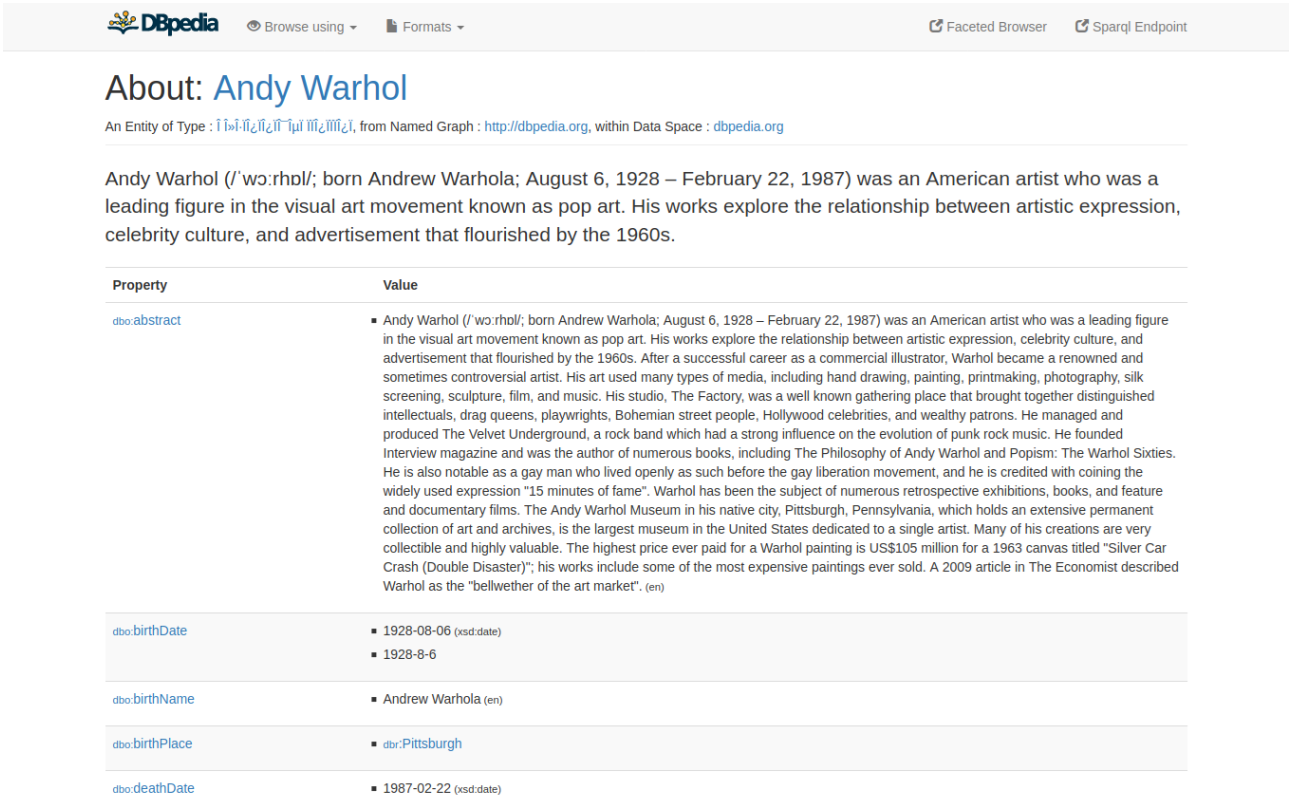


CHO	TITLE	ARTIST	NAME	CLASSIFICATION
http://localhost:3000/moma/79889/artwork	Self-Portrait	http://localhost:3000/moma/artist/6246	Andy Warhol	Painting
http://localhost:3000/cmoa/8d3ad9a9-5e7e-4365-9ee0-8ba3e04d98c3/artwork	Self Portrait	http://localhost:3000/cmoa/artist/c527ff39-c938-45c8-b4ca-a5091527295a	Andy Warhol	paintings
http://localhost:3000/moma/80705/artwork	Silkscreen for Portrait of Sidney Janis	http://localhost:3000/moma/artist/6246	Andy Warhol	Painting

Εικόνα 64 Αναπαράσταση αποτελεσμάτων αναζήτησης σε κατευθυνόμενο γράφο, Σενάριο χρήσης #2

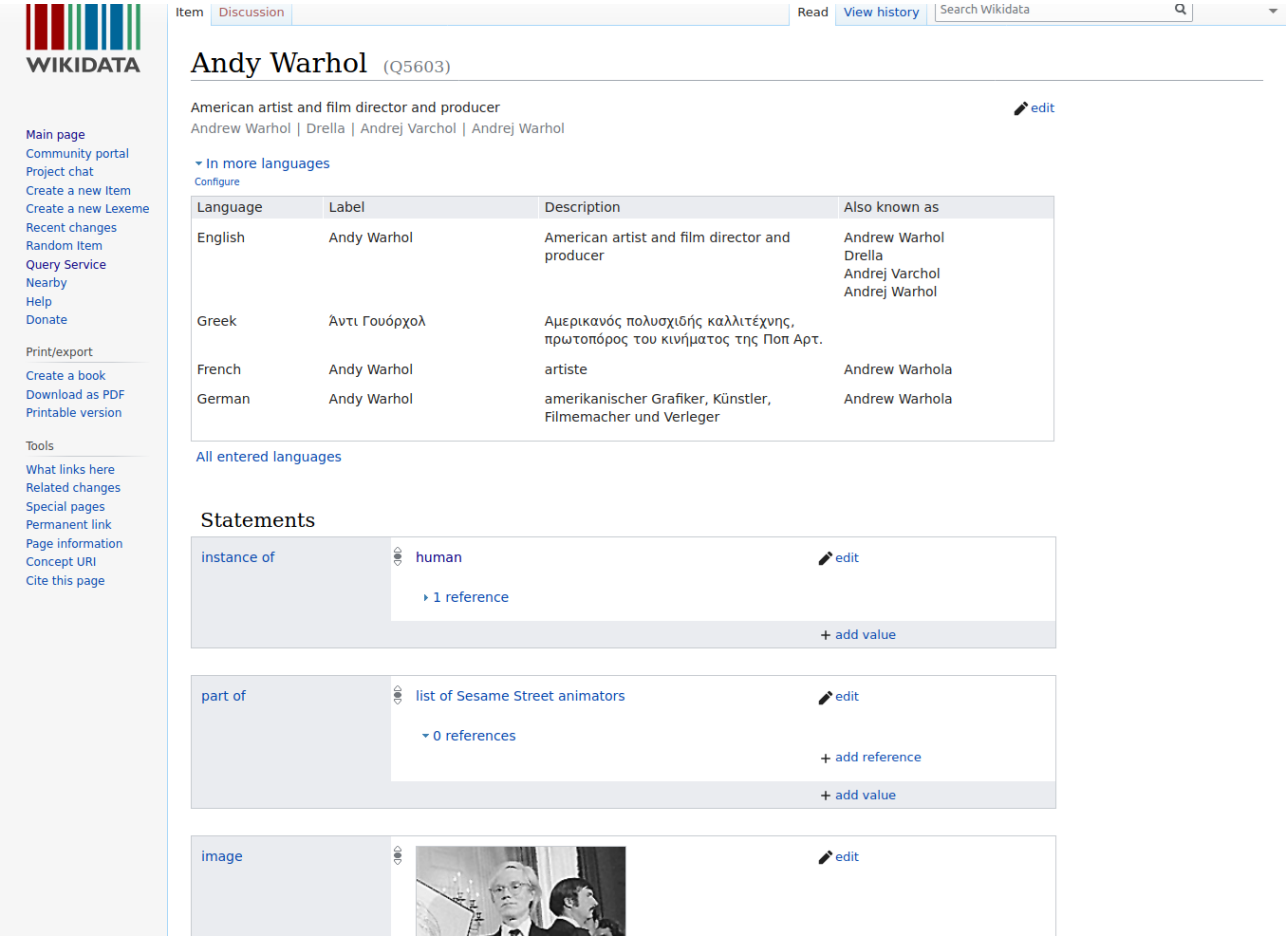


Εικόνα 65 Προβολή λεπτομερειών πόρου καλλιτέχνη, Σενάριο χρήσης #2



Εικόνα 66 Σύνδεσμος πόρου με εξωτερική πηγή, DBpedia, Σενάριο χρήσης #2

Τέλος επιλέγεται η μετάβαση στις εξωτερικές πηγές από όπου μπορεί να συνεχιστεί η περιήγηση και να ληφθούν περισσότερες πληροφορίες, όπς φαίνεται στις εικόνες 66 και 67.



Εικόνα 67 Σύνδεσμος πόρου με εξωτερική πηγή, Wikidata, Σενάριο χρήσης #2

Κεφάλαιο 6ο
Συμπεράσματα

6.1. Συμπεράσματα Μετατροπής και Διασύνδεσης Δεδομένων

Τα εργαλεία που χρησιμοποιούνται για την σημασιολογική ενοποίηση των δεδομένων και για την διασύνδεση τους, παίζουν σημαντικό ρόλο στο αποτέλεσμα, καθώς διευκολύνουν τη διαδικασία, καθοδηγούν το χρήστη και διευρύνουν τις δυνατότητες του. Στα πλαίσια της εργασίας χρησιμοποιήθηκαν τρία εργαλεία για το RDFization των δεδομένων και δύο για την διασύνδεσή τους. Η χρήση του καθενός από τα εργαλεία αποδείχθηκε χρήσιμη για την διεξαγωγή της παρούσας εργασίας και προσέφερε στην τελική διαμόρφωση των δεδομένων.

Ακολουθεί η συγκριτική παράθεση χαρακτηριστικών των εργαλείων. Τα κριτήρια που επιλέχθηκαν για την σύγκριση προκύπτουν από την βιβλιογραφία σχετική με την αξιολόγηση λογισμικού.

6.1.1. RDFization

Επαναχρησιμοποίηση Μοντέλου	
KARMA	Το παραγόμενο μοντέλο μπορεί να εξαχθεί και να χρησιμοποιηθεί σε άλλα δεδομένα με την ίδια δομή ή σε διαφορετικό σύνολο του ίδιου σύνολο δεδομένων
OpenRefine (rdf-extension)	Το μοντέλο είναι μέρος του project δεν εφαρμόζεται σε άλλο σύνολο δεδομένων.
RDF-Gen	Το μοντέλο περιγράφεται στο template file και μπορεί να χρησιμοποιηθεί σε άλλα σύνολα δεδομένων και σε δεδομένα συνεχόμενης ροής

Πίνακας 5 Πίνακας σύγκρισης Επαναχρησιμοποίηση Μοντέλου

Ταχύτητα Ανάπτυξης Μοντέλου	
KARMA	Η ταχύτητα ανάπτυξης του μοντέλου είναι απλή αν και στην αρχική ανάπτυξη μοντέλων για μη συσχετισμένα

	σύνολα δεδομένων απαιτούνται αρκετές ενέργειες για την συσχέτιση των οντοτήτων. Η διαδικασία επιταχύνεται σημαντικά καθώς το σύστημα μαθαίνει από προηγούμενες καταχωρήσεις και προτείνει σημασιολογικούς τύπους για τα δεδομένα.
OpenRefine (rdf-extension)	Η ανάπτυξη του μοντέλου είναι γρήγορη διαδικασία αντιστοίχισης αλλά επιβραδύνεται όσο αυξάνει ο αριθμός των κλάσεων.
RDF-Gen	Η ανάπτυξη του μοντέλου διαρκεί όσο η πληκτρολόγηση του στο template file, όπου περιγράφετε σε μορφή και σύνταξη παρόμοια των ερωτημάτων SPARQL. Ο χρόνος αυξάνεται σε εξειδικευμένες περιπτώσεις που απαιτείται ανάπτυξη συγκεκριμένων μεθόδων που θα χρησιμοποιηθούν στο μοντέλο.

Πίνακας 6 Πίνακας σύγκρισης Ταχύτητα Ανάπτυξης Μοντέλου

Ταχύτητα Απόκρισης RDFization	
KARMA	Η ταχύτητα δημιουργίας του γράφου είναι ικανοποιητική για τον όγκο των δεδομένων. Για παράδειγμα η εξαγωγή των δεδομένων του αρχείου MoMA.csv (~128.000 εγγραφών) σε RDF, διήρκεσε 1’ και 55”, ενώ η εξαγωγή για το σύνολο δεδομένων της συλλογής του CMoA (~28000 εγγραφών) από JSON σε RDF διήρκεσε 50’’ (σε φορητό Η/Υ με μνήμη RAM 6GB, Επεξεργαστή i7 2ης γενιάς και λειτουργικό σύστημα Ubuntu Linux 18.04)
OpenRefine (rdf-extension)	Η ταχύτητα δημιουργίας του γράφου είναι ικανοποιητική για τον όγκο των δεδομένων. Σε περίπτωση που προστεθούν πράξεις μετατροπής των δεδομένων κατά τη διαδικασία της εξαγωγής ο χρόνος δημιουργίας πολλαπλασιάζεται. Η εξαγωγή των δεδομένων του αρχείου MoMA.csv (~128.000 εγγραφών) σε RDF, διήρκεσε 1’ και 45” ενώ για το

	σύνολο δεδομένων της συλλογής του CMoA (~28000 εγγραφών) από JSON σε RDF διήρκεσε 35'' (σε φορητό Η/Υ με μνήμη RAM 6GB, Επεξεργαστή i7 2ης γενιάς και λειτουργικό σύστημα Ubuntu Linux 18.04)
RDF-Gen	Η επεκτασιμότητα και η ταχύτητα είναι οι καινοτομίες του εργαλείου. Ο χρόνος είναι σημαντικά μικρότερος. Η εξαγωγή των δεδομένων του αρχείου MoMA.csv (~128.000 εγγραφών) σε RDF, διήρκεσε 25'' ενώ για το σύνολο δεδομένων της συλλογής του CMoA (~28000 εγγραφών) από JSON σε RDF διήρκεσε 13''(σε φορητό Η/Υ με μνήμη RAM 6GB, Επεξεργαστή i7 2ης γενιάς και λειτουργικό σύστημα Ubuntu Linux 18.04)

Πίνακας 7 Πίνακας σύγκρισης Ταχύτητα Απόκρισης RDFization

	Ευκολία Χρήσης
KARMA	Το KARMA έχει αναπτυχθεί με για σημασιολογική ανάλυση δεδομένων από ειδικούς του εκάστοτε πεδίου γνώσης. Με αυτό ως μια των προϋποθέσεων, έχει δημιουργηθεί ένα φιλικό περιβάλλον εργασίας. Η γραφική αναπαράσταση του μοντέλου βοηθά στην κατανόηση της δομής του και οι αυτοματοποιημένες ενέργειες διευκολύνουν την διαδικασία ανάπτυξης. Επίσης υπάρχουν αναλυτικές οδηγίες για τις δυνατότητες που παρέχει το εργαλείο.
OpenRefine (rdf-extension)	Το OpenRefine καλύπτει ένα ευρύ πεδίο διαχείρισης δεδομένων και ο σημασιολογικός χαρακτηρισμός είναι μια από τις δυνατότητες του. Θετικό αντίβαρο στην πολυπλοκότητα του εργαλείου, είναι η ευρεία χρήση του και το πλήθος παραδειγμάτων και οδηγιών για τις διάφορες λειτουργίες.
RDF-Gen	Το RDF-Gen. αποτελεί ένα εξειδικευμένο εργαλείο που δεν περιλαμβάνει γραφική διεπαφή και

	απευθύνεται κυρίως σε πιο εξοικειωμένο με το αντικείμενο κοινό. Παρόλα αυτά η λειτουργία του είναι εύκολο να κατανοηθεί και να εκτελεστεί αν ακολουθηθούν οι οδηγίες του.

Πίνακας 8 Πίνακας σύγκρισης Ευκολία Χρήσης RDFization

Αποφυγή Λαθών	
KARMA	Το KARMA με την γραφική αναπαράσταση του μοντέλου, επιτρέπει μέσω της εποπτείας του, την αποφυγή λογικών λαθών στη δομή και στην σχέση μεταξύ των οντοτήτων.
OpenRefine (rdf-extension)	Η δυνατότητα του OpenRefine για προβολή του γράφου που παράγεται για ένα υποσύνολο των δεδομένων, εξυπηρετεί την αποφυγή λαθών στο τελικό παραγόμενο αποτέλεσμα. Επίσης η δυνατότητα φιλτραρίσματος των δεδομένων και προβολής συγκεκριμένου υποσυνόλου, βοηθάει στην διαχείριση πιθανών ιδιαίτερων ή οριακών τιμών των δεδομένων.
RDF-Gen	Ενδεχόμενα λάθη μπορούν να παρατηρηθούν στο τελικό αποτέλεσμα, καθώς η ταχύτητα παραγωγής του γράφου αντισταθμίζει την έλλειψη πρόληψης λαθών από το εργαλείο.

Πίνακας 9 Πίνακας σύγκρισης Αποφυγή Λαθών

6.1.2. Διασύνδεση

Ταχύτητα Ανακάλυψης Συνδέσμων	
Silk	Όπως είναι αναμενόμενο η διάρκεια ανακάλυψης συνδέσμων είναι ανάλογη του μεγέθους των γράφων και της πολυπλοκότητας των διαδικασιών σύγκρισης των πόρων. Γενικά όμως, το Silk σαν εξειδικευμένο εργαλείο στη διασύνδεση πόρων, έχει μικρό χρόνο

	απόκρισης. Η εύρεση συσχετίσεων ανάμεσα στους δημιουργούς της συλλογής των MoMA και CMoA διήρκεσε 12''. Η διαδικασία εύρεσης συνδέσμων των καλλιτεχνών από τον τοπικό γράφο της συλλογής του MoMA με την DBpedia διήρκεσε 35 λεπτά. Αντίστοιχα για τους καλλιτέχνες της συλλογής CMoA με την DBpedia η διάρκεια ήταν 7' και 25'', ενώ για την διασύνδεση των έργων τέχνης των δύο τοπικών γράφων η διάρκεια ήταν 40 λεπτά.
OpenRefine (rdf-extension)	Η διαδικασία ανακάλυψης συνδέσμων από την επέκταση rdf-extension του OpenRefine είναι σημαντικά πιο αργή σε σύγκριση με την προηγούμενη. Συγκεκριμένα για τη σύνδεση καλλιτεχνών των συλλογών MoMA, CMoA η διάρκεια ήταν περίπου 2 ώρες. Ενώ η σύνδεση καλλιτεχνών μεταξύ της συλλογής MoMA και του Wikidata διήρκεσε περίπου 12 ώρες.

Πίνακας 10 Πίνακας σύγκρισης Ταχύτητα Ανακάλυψης Συνδέσμων

Ταχύτητα Ανάπτυξης Διαδικασίας Διασύνδεσης	
Silk	Η διαδικασία περιλαμβάνει τη δήλωση του τύπου των μεταβλητών, τον ορισμό τω οντοτήτων προς σύγκριση καθώς και την επιλογή των μετρικών και των παραμέτρων τους. Έτσι μπορεί να αποτελέσει χρονοβόρα διαδικασία ειδικά για μη έμπειρους ή μη εξοικειωμένους με το εργαλείο χρήστες.
OpenRefine (rdf-extension)	Η λειτουργία είναι ενσωματωμένη στο εργαλείο και δεν δίνεται η δυνατότητα επιλογή των μετρικών σύγκρισης, οπότε απαιτείται μόνο ο ορισμός των οντοτήτων προς σύγκριση και η διάρκεια είναι ελάχιστη.

Πίνακας 11 Πίνακας σύγκρισης Ταχύτητα Ανάπτυξης Διαδικασίας Διασύνδεσης

	Ευκολία Χρήσης
Silk	Ο ορισμός της διαδικασίας μπορεί να φανεί περίπλοκος σε αρκετές περιπτώσεις λόγω της πληθώρας των διαθέσιμων επιλογών. Υπάρχει δημοσιευμένο ένα εγχειρίδιο χρήσης με βασικές πληροφορίες, αλλά μεγάλο μέρος της διαδικασίας βασίζεται στην δοκιμή του λογισμικού.
OpenRefine (rdf-extension)	Ο ορισμός και εκτέλεση της διαδικασίας είναι απλός καθώς δεν προσφέρονται πολλές επιλογές στον χρήστη.

Πίνακας 12 Πίνακας σύγκρισης Ευκολία Χρήσης, Link Discovery

6.2. Συμπεράσματα χρήσης ανεπτυγμένης εφαρμογής

Κατά την χρήση της εφαρμογής επιτυγχάνεται η παρουσίαση των πόρων και των λεπτομερειών τους που σχετίζονται με τις αναζητήσεις του χρήστη. Τα δεδομένα που επιστρέφονται στο χρήστη αποτελούν την ενοποιημένη και διασυνδεδεμένη αναπαράσταση των πόρων. Παρέχεται στον χρήστη η διαθέσιμη πληροφορία και από τα δύο μουσεία, αλλά επισημαίνεται και η σχέση μεταξύ τους. Επίσης επιτυγχάνεται η ανάδειξη της διασύνδεσης των πόρων με εξωτερικές πηγές, με την δυνατότητα επίσκεψης ιστοσελίδων που παρέχουν σχετικό περιεχόμενο, μέσω των συνδέσμων που ανακαλύπτονται. Συνολικά εκτιμάται ότι ο στόχος για σημασιολογική ενοποίηση, διασύνδεση και ανάδειξη πολιτιστικών δεδομένων μέσα από την επεξεργασία των δεδομένων και των αποτελεσμάτων που παρέχονται από την χρήση της εφαρμογής, έχει επιτευχθεί.

6.3. Μελλοντικές ενέργειες

Η μετατροπή των δεδομένων των μουσείων MoMA και CMoA σε ανοιχτά διασυνδεδεμένα δεδομένα και ανάπτυξη της εφαρμογής για την ανάδειξη τους έγιναν με χρήση των ερευνητικών συνόλων δεδομένων που είχαν δημοσιευθεί και παρέμειναν τοπικά.

Μελλοντικές ενέργειες είναι

- Η δημοσίευση των δεδομένων και της εφαρμογής στον Παγκόσμιο Ιστό

- Η διασύνδεση του γράφου με το LOD Cloud
- Η επικοινωνία με τα μουσεία MoMA και CMoA για παρουσίαση της εργασίας.
- Η ερώτηση για σύνολα δεδομένων με περισσότερες πληροφορίες, (όπως περιγραφές των έργων, διαχωρισμός τους σε θεματικές ενότητες, βιογραφικά στοιχεία για τους καλλιτέχνες κ.α.) με σκοπό ενσωμάτωση τους στην εφαρμογή που θα έχει σαν αποτέλεσμα την βελτίωση της εμπειρίας του χρήστη.
- Η παρουσίαση της εργασίας σε Μουσεία στην Ελλάδα που ενδιαφέρονται να μετατρέψουν τα δεδομένα του αρχείου τους σε ανοιχτά και διασυνδεδεμένα.
- Δημοσίευση της εργασίας σε επιστημονικό-ερευνητικό άρθρο, στο περιοδικό «Semantic Web Journal (IOS press)».

Βιβλιογραφία

[1] "RFC 1630 - Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web." [Online]. Available: https://tools.ietf.org/html/rfc1630.

[2] "Universal Resource identifiers in WWW." [Online]. Available: https://www.w3.org/Addressing/URL/uri-spec.html.

[3] "Understanding Metadata." [Online]. Available: https://groups.niso.org/apps/group_public/download.php/17446/UnderstandingMetadata.pdf.

[4] O. Lassila and R. R. Swick, "Resource description framework (RDF) model and syntax specification. World Wide Web Consortium Recommendation," no. October, 1999.

[5] C. Bizer, *Linked Data Evolving the Web into a Global Data Space*, vol. 31, no. 1. 2016.

[6] "RDF 1.1 Concepts and Abstract Syntax." [Online]. Available: https://www.w3.org/TR/rdf11-concepts/.

[7] T. Heath and C. Bizer, "Linked Data: Evolving the Web into a Global Data Space," *Synth. Lect. Semant. Web Theory Technol.*, vol. 1, no. 1, pp. 1–136, Feb. 2011, doi: 10.2200/S00334ED1V01Y201102WBE001.

[8] "RDF 1.1 XML Syntax." [Online]. Available: https://www.w3.org/TR/rdf-syntax-grammar/.

[9] D. Beckett, T. Berners-Lee, E. Prud'hommeaux, and G. Carothers, "RDF 1.1 Turtle," *W3C Recommendation*, 2014. [Online]. Available: http://www.w3.org/TR/2014/REC-turtle-20140225/.

[10] "RDF 1.1 N-Triples." [Online]. Available: https://www.w3.org/TR/n-triples/.

[11] "RDF 1.1 N-Quads." [Online]. Available: https://www.w3.org/TR/n-quads/.

[12] "Notation3 (N3): A readable RDF syntax." [Online]. Available: https://www.w3.org/TeamSubmission/n3/.

[13] "JSON-LD 1.0." [Online]. Available: https://json-ld.org/spec/FCGS/json-ld/20130328/.

[14] "RDF Schema 1.1." [Online]. Available: https://www.w3.org/TR/rdf-schema/.

[15] "OWL 2 Web Ontology Language Document Overview (Second Edition)." [Online]. Available: https://www.w3.org/TR/owl2-overview/.

[16] "DCMI: Using Dublin Core." [Online]. Available: https://www.dublincore.org/specifications/dublin-core/usageguide/.

[17] “SKOS Simple Knowledge Organization System Reference.” [Online]. Available: https://www.w3.org/TR/skos-reference/.

[18] “SPARQL 1.1 Query Language.” [Online]. Available: https://www.w3.org/TR/sparql11-query/.

[19] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” *Scientific American*. 2001, doi: 10.1038/scientificamerican0501-34.

[20] A. Hogan, “Linked Data & the Semantic Web Standards,” 2014.

[21] “Τι είναι τα Ανοιχτά Δεδομένα;” [Online]. Available: https://opendatahandbook.org/guide/el/what-is-open-data/.

[22] C. Bizer, T. Heath, and T. Berners-Lee, “Linked data - The story so far,” *Int. J. Semant. Web Inf. Syst.*, vol. 5, no. 3, pp. 1–22, 2009, doi: 10.4018/jswis.2009081901.

[23] “Data - W3C.” [Online]. Available: https://www.w3.org/standards/semanticweb/data.

[24] Μ. Η. Υπηρεσιών and Μ. Σ. Περιεχομένου, “Εθνικό Κέντρο Τεκμηρίωσης | ΕΙΕ.”

[25] “Ανοικτά Δεδομένα: Η πρώτη ύλη για την Κοινωνία της Γνώσης | Εθνικό Κέντρο Τεκμηρίωσης - ΕΚΤ.” [Online]. Available: https://www.ekt.gr/el/magazines/features/15726#3.

[26] “5-star Open Data.” [Online]. Available: https://5stardata.info/en/.

[27] “Linked Data - Design Issues.” [Online]. Available: https://www.w3.org/DesignIssues/LinkedData.html.

[28] “Europeana Collections.” [Online]. Available: https://www.europeana.eu/portal/en.

[29] “SearchCulture.gr | Ο Ελληνικός Συσσωρευτής Περιεχομένου Πολιτιστικής Κληρονομιάς.” [Online]. Available: https://www.searchculture.gr/.

[30] “MuseumofModernArt/collection: The Museum of Modern Art (MoMA) collection data.” [Online]. Available: https://github.com/MuseumofModernArt/collection.

[31] “cmoa/collection: The collection data of the Carnegie Museum of Art in Pittsburgh, Pennsylvania.” [Online]. Available: https://github.com/cmoa/collection.

[32] Europeana, “Definition of the Europeana Data Model,” 2017.

[33] E. Union, “EDM_Mapping_Guidelines_v2.4_102017,” 2017.

[34] C. A. Knoblock *et al.*, “Semi-automatically mapping structured sources into the semantic web,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7295 LNCS, pp. 375–390, 2012, doi: 10.1007/978-3-642-30284-8_32.

[35] P. Szekely *et al.*, “Connecting the smithsonian american art museum to the linked data cloud,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7882 LNCS, pp. 593–607, 2013, doi: 10.1007/978-3-642-38288-8-40.

[36] “OpenRefine.” [Online]. Available: https://openrefine.org/.

[37] “General Refine Expression Language · OpenRefine/OpenRefine Wiki.” [Online]. Available: https://github.com/OpenRefine/OpenRefine/wiki/General-Refine-Expression-Language.

[38] “stkenny/grefine-rdf-extension: An extension to OpenRefine that enables graphical mapping of OpenRefine project data to an RDF skeleton and then exporting it in RDF format.” [Online]. Available: https://github.com/stkenny/grefine-rdf-extension.

[39] G. M. Santipantakis, G. A. Vouros, K. I. Kotis, and C. Doulkeridis, “Rdf-gen: Generating rdf from streaming and archival data,” *ACM Int. Conf. Proceeding Ser.*, no. June, 2018, doi: 10.1145/3227609.3227658.

[40] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov, “Silk - A Link Discovery Framework for the Web of Data,” *CEUR Workshop Proc.*, vol. 538, 2009.

[41] “Silk - The Linked Data Integration Framework.” [Online]. Available: http://silkframework.org/.

[42] “Apache Jena - Getting started with Apache Jena.” [Online]. Available: https://jena.apache.org/getting_started/index.html.

[43] “Apache Jena - Apache Jena Fuseki.” [Online]. Available: https://jena.apache.org/documentation/fuseki2/.

[44] “Apache Jena - TDB.” [Online]. Available: https://jena.apache.org/documentation/tdb/index.html.

[45] “React – A JavaScript library for building user interfaces.” [Online]. Available: https://reactjs.org/.

[46] “Bootstrap · The most popular HTML, CSS, and JS library in the world.” [Online]. Available: https://getbootstrap.com/.

[47] “Introducing JSX – React.” [Online]. Available: https://reactjs.org/docs/introducing-jsx.html.

[48] “ECMAScript 2015 Language Specification – ECMA-262 6th Edition.” [Online]. Available: http://www.ecma-international.org/ecma-262/6.0/.

[49] “Silk-book - Junseok Yoon, 윤 준석.” [Online]. Available: https://phobyjun.github.io/2019/07/07/Silk-book.html.