

BETA { **Semantics.gr** }

A self-improving
service to repositories
& aggregators for
massively enriching
their content

DHC 2016 workshop

22.11.2016

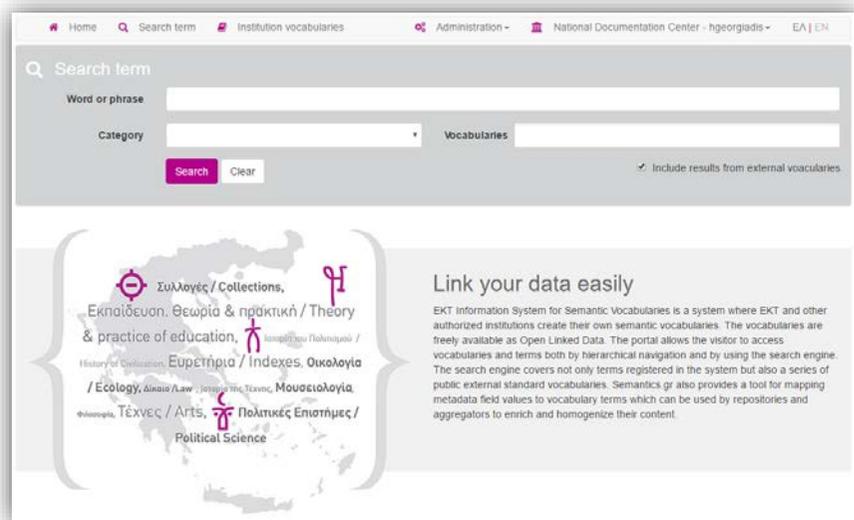
Haris Georgiadis PhD |

Senior Software Engineer/Researcher
National Documentation Centre | EKT



Συλλογές / Collections, 
Εκπαίδευση. Θεωρία & πρακτική / Theory
& practice of education, 
History of Civilization,  Ιστορία του Πολιτισμού /
Ευρετήρια / Indexes, Οικολογία
/ Ecology, Δίκαιο / Law,  Ιστορία της Τέχνης, Μουσειολογία,
Φιλοσοφία, Τέχνες / Arts,  Πολιτικές
Επιστήμες / Political Science

Semantics.gr: Information system for vocabularies & semantic enrichment



- A pilot system where EKT and other **authoritative institutions** will create, establish & link their own semantic **vocabularies and thesauri**
 - hosts vocabularies of virtually any schema
 - supports hierarchies and multilingualism
 - supports links to external vocabularies & thesauri
- Bilingual Open portal
 - Published vocabularies as Open Linked Data
 - Search engine
 - Hierarchical navigation
- **Metadata enrichment tool**

Creating a vocabulary in semantics.gr

Semantic class

Schema namespace	skos	Name	Concept
Property for preference label	Qualified name	skos:prefLabel	<input type="button" value="Choose"/> X
	Label	Preferable label	
Property for alternative labels	Qualified name	skos:altLabel	<input type="button" value="Choose"/> X
	Label	Alternative labels	
Property for equivalence	Qualified name	skos:exactMatch	<input type="button" value="Choose"/> X
	Label	Exact match	

Properties

Qualified name	Label	Type	Values
skos:definition	Definition	RDF object	Multiple multilingual literals
skos:broader	Broader (from this vocabulary)	RDF object	Reference
skos:narrower	Narrower (from this vocabulary)	RDF object	Multiple references
skos:related	Related concept (from this vocabulary)	RDF object	Multiple references
skos:notation	Notation	RDF object	Multiple multilingual literals

Item types by EKT Content

393 terms

Δισδιάστατα γραφικά : 2D Graphics

- Αγιογραφία : Icon
- Άλλο : Other
- Αντίγραφο : Replica
- Απόκομμα : Clipping

- Authorized users can create RDF based vocabulary schemata.
 - Create parametric **owl properties**
 - Group owl properties in **owl classes**
 - Successfully modelled **skos:Concept, edm:Place, edm:Timespan, edm:Agent**
- Institutions** will obtain user accounts and create their own vocabularies and thesauri.
 - The system provides with a **dynamic form** to create vocabulary entries.
 - The form reflects the parameters set.
- Institutions can **publish** their vocabularies. Those vocabularies are accessible through the portal as **Open Linked Data**.



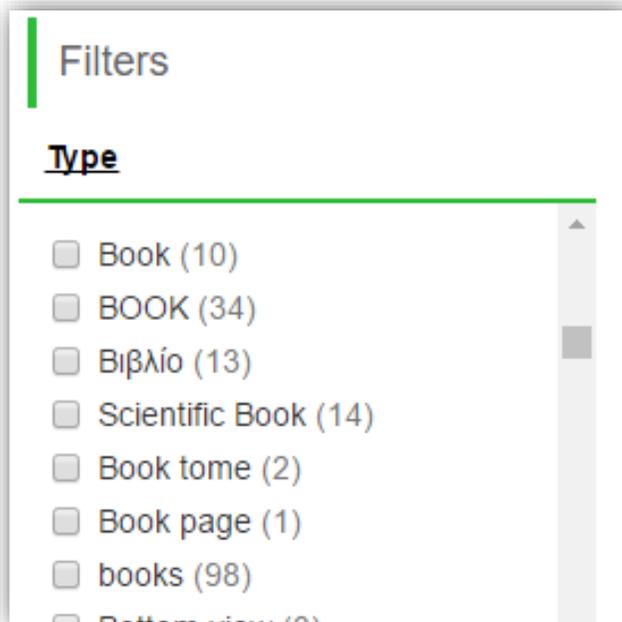
{ Search Culture.gr }

Searchculture.gr: aggregator of cultural heritage digital content

- **Aggregates digital content** from registered digital repositories
- **Public portal**
 - Modern & effective search engine
 - Dissemination as Open Linked Data
 - Multilingual (English / Greek)
 - Rights statements per item, mostly CC, defining copyright, accessibility and reusability status
 - Presents the participating institutions & their collections

150.000+ items | 53 collections | 42 institutions

Heterogeneity in Metadata



- Heterogeneity in **dc:type** metadata field:
 - **Representation-related** ex. different languages, synonyms, plural / singular numbers, different case styles
 - **Documentation-related** ex. use of very broad or very narrow (specialized) terms
- Negative impact of heterogeneity in user experience:
 - **Searchability & multilingual search**
 - **Discoverability & multilingual navigation**
 - **Visual presentation**



Four different types of repositories we came across

Documentation quality class

Documentation quality class description

A

Good documentation of **dc:type**

B

Extremely specialized documentation of **dc:type**

C

Insufficient documentation on **dc:type**, useful **dc:subject**

D

Insufficient documentation on **dc:type**, useful **dc:title**

Semantics.gr enrichment tool

- **Mapping rules** from metadata field values (ex. dc:type) to vocabulary terms per collection/repository
- **Automatic mapping suggestion** enhanced by a **self-improving mechanism**
- Use of **secondary metadata field** values (ex. dc:subject) as **filters**
 - Supports searching for **key-terms** inside **descriptive fields** (ex. dc:title) as **filters**
- Curated **complex expressions on filters** in order to create finer and more precise rules and avoid false positives
- A mapping rule set assigned to a repository/collection is available through a **REST API** in **json** format

Class A enrichment: automatic mapping metadata values to vocabulary entries

Vocabulary V1

```

→ http://scs.gr/sculpture
  skos:prefLabel "Sculpture"@en | "Γλυπτό"@el
    → http://scs.gr/figurine
      skos:prefLabel "Figurine"@en | "Ειδώλιο"@el
→ http://scs.gr/Jewellery
  skos:prefLabel "Jewellery"@en | "Κόσμημα"@el
→ http://scs.gr/vessel
  skos:prefLabel "Vessel"@en | "Σκεύος"@el
    → http://scs.gr/vase
      skos:prefLabel "Vase"@en | "Αγγείο"@el

```

dc:type value

Entry from vocabulary V1

sculpture art (120 items)

<http://scs.gr/sculpture>

auto

greek vases (230 items)

<http://scs.gr/vase>

auto

jewelleries (135 items)

<http://scs.gr/Jewellery>

auto

Class B enrichment: the self-improving mechanism

Vocabulary V1

- <http://scs.gr/sculpture>
skos:prefLabel "Sculpture"@en | "Γλυπτό"@el
- <http://scs.gr/figurine>
skos:prefLabel "Figurine"@en | "Ειδώλιο"@el **keywords: "statuette"**
- <http://scs.gr/Jewellery>
skos:prefLabel "Jewellery"@en | "Κόσμημα"@el **keywords: "earring"**
- <http://scs.gr/vessel>
skos:prefLabel "Vessel"@en | "Σκεύος"@el
- <http://scs.gr/vase>
skos:prefLabel "Vase"@en | "Αγγείο"@el **keywords: "amphora", "oenochoe"**

dc:type values

Entry from vocabulary V1

amphora (100 items)	http://scs.gr/vase	manual
oenochoe (110 items)	http://scs.gr/vase	manual
earring (55 items)	http://scs.gr/Jewellery	manual
statuette (15 items)	http://scs.gr/figurine	manual

Class C enrichment: Using secondary fields as filters

Vocabulary V1

- <http://scs.gr/sculpture>
 skos:prefLabel "Sculpture"@en | "Γλυπτό"@el **keywords: "statue"**
 - <http://scs.gr/figurine>
 skos:prefLabel "Figurine"@en | "Ειδώλιο"@el **keywords: "statuette"**
- <http://scs.gr/Jewellery>
 skos:prefLabel "Jewellery"@en | "Κόσμημα"@el **keywords: "earring"**
- <http://scs.gr/vessel>
 skos:prefLabel "Vessel"@en | "Σκεύος"@el
 - <http://scs.gr/vase>
 skos:prefLabel "Vase"@en | "Αγγείο"@el **keywords: "amphora", "oenochoe"**

dc:type values

 Filters: **dc:subject**

Entry from vocabulary V1

ceramic objects (101 items)	amphora (↗), vase (↗), statuette (↗), figurine (↗), ...	http://scs.gr/vase	auto
		if filter in ["vase", "amphora"]	auto
exhibits (55 items)	earring (↗), amphora (↗), ...	http://scs.gr/figurine	auto
		if filter in ["statuette", "figurine"]	auto
		http://scs.gr/Jewellery	auto
		if filter in ["earring"]	auto
		http://scs.gr/vase	auto
		if filter in ["amphora"]	auto
		but NOT in ["earring"]	manual

Class D enrichment: Searching for terms inside descriptive fields

Vocabulary V1

- <http://scs.gr/sculpture>
 skos:prefLabel "Sculpture"@en | "Γλυπτό"@el **keywords: "statue"**
 - <http://scs.gr/figurine>
 skos:prefLabel "Figurine"@en | "Ειδώλιο"@el **keywords: "statuette"**
- <http://scs.gr/Jewellery>
 skos:prefLabel "Jewellery"@en | "Κόσμημα"@el **keywords: "earring"**
- <http://scs.gr/vessel>
 skos:prefLabel "Vessel"@en | "Σκεύος"@el
 - <http://scs.gr/vase>
 skos:prefLabel "Vase"@en | "Αγγείο"@el **keywords: "amphora", "oenochoe"**

Example of dc:title value: "An **amphora** from the Mycenaean period"

dc:type values	Filters (terms found in dc:title values)	Entry from vocabulary V1
3D objects (240)	amphora (↗), vase (↗), earring (↗), jewellery (↗)	http://scs.gr/vase auto if filter in ["vase", "amphora"] auto but NOT in ["statue"] manual
		http://scs.gr/Jewellery auto if filter in ["earring", "jewellery"] auto
art items (85)	sculpture (↗), statue (↗), figurine (↗)	http://scs.gr/sculpture auto if filter in ["sculpture", "statue"] auto http://scs.gr/figurine auto if filter in ["figurine"] auto

The GUI of the enrichment tool

The screenshot displays the 'Mapping' interface for the 'Delphi_Museum' repository. It features a left sidebar with 'Filters' and 'Automatic creation' options, and a central area with 'Harvest field values' and 'Automatic match' buttons. The main workspace shows two mapping rules:

- Rule 1:** Field `dc:type` with URI `http://repository.e-delphi.gr/dspace/authorities/type/LiturgicalObject` (12 items). The secondary field is `dc:subject`. The automatic match rule uses URI `ekt-item-types/glypto` with the value 'Γλυπτό' and filters for 'Γλυπτική' and 'Μικροτεχνία'.
- Rule 2:** Field `dc:type` with URI `http://repository.e-delphi.gr/dspace/authorities/type/ArchitecturalSculpture` (7 items). The secondary field is `dc:subject`. The automatic match rule uses URI `ekt-item-types/aggeio` with the value 'Αγγείο' and filter for 'Κεραμική'.

At the bottom, a third rule is partially visible with URI `ekt-item-types/arxitektoniko-melos` and filter for 'Γλυπτική'.

Enriching the content
of Searchculture.gr: **The results**

Documentation Class	# of repositories	# of items
A: sufficient existing dc:type values	20	30.764
B: extremely specified dc:type values	5	11.102
C: insufficient dc:type values – useful dc:subject	24	60.181
D: insufficient dc:type – resorting to dc:title values	4	55.912
Total	53	157.959

Enriching the content of Searchculture.gr: The results

dc:type values before and after the enrichment process for a specific repository (Documentation class C)

Τύπος τεκμηρίου

- Slides (8643)
- Βίντεο (779)
- Ήχοι (867)
- Κείμενα (101)
- Μουσειακά αντικείμενα (1823)
- Φωτογραφίες (1110)



Homogenization based on dc:type and dc:subject (secondary field for filtering)

EKT τύπος τεκμηρίου

- Δισδιάστατα γραφικά (10155)
 - Αφιερωματικές κάρτες Αγίων (9)
 - Αφίσα (1)
 - Γραμματόσημο (45)
 - Έργο θρησκευτικής τέχνης (79)
 - Φορητή Εικόνα (77)
 - Ευχετήρια κάρτα (37)
 - Ζωγραφική (324)
 - Σχέδιο/Σκίτσο (28)
 - Ταχυδρομική κάρτα (202)
 - Τοιχογραφία (11)
 - Φιγούρα θεάτρου σκιών (522)
 - Φωτογραφία (1458)
 - Φωτογραφική διαφάνεια (8643)
 - Χαρτονόμισμα (5)
- Ηχογραφήσεις (867)
- Κείμενο (585)
 - Άρθρο (3)
 - Βιβλίο (29)
 - Διατριβή (2)
 - Έγγραφο (403)
 - Αλληλογραφία (141)
 - Έκθεση-Αναφορά (7)
 - Νομική υπόθεση και σημείωμα (3)
 - Εφημερίδα (9)
 - Ημερολόγιο/Σημειωματάριο (15)

...

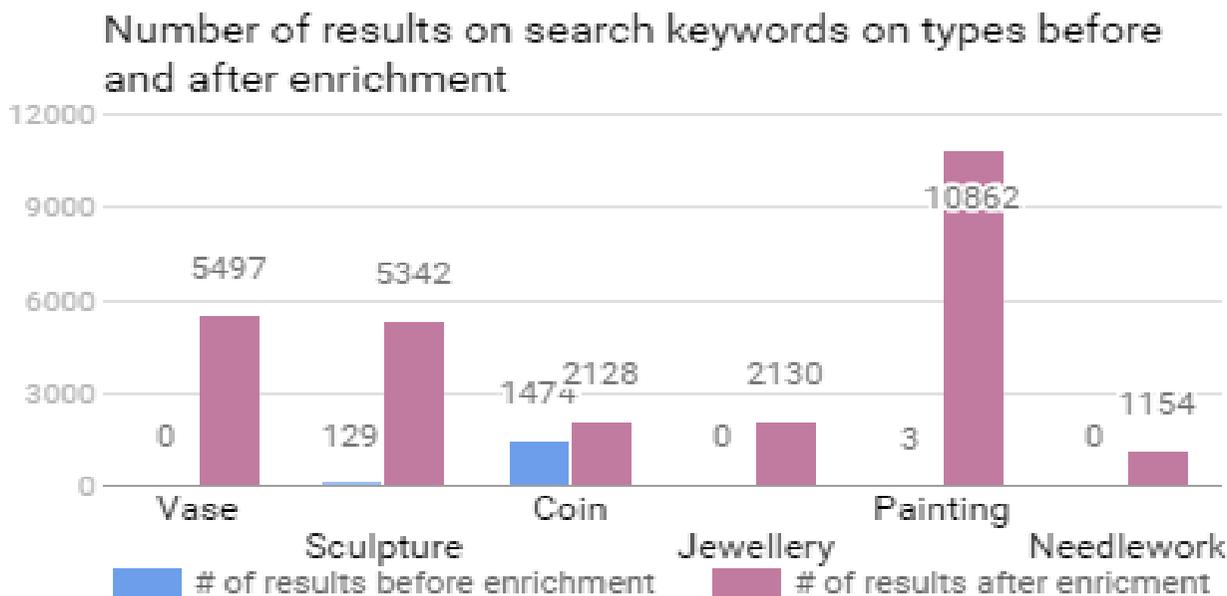
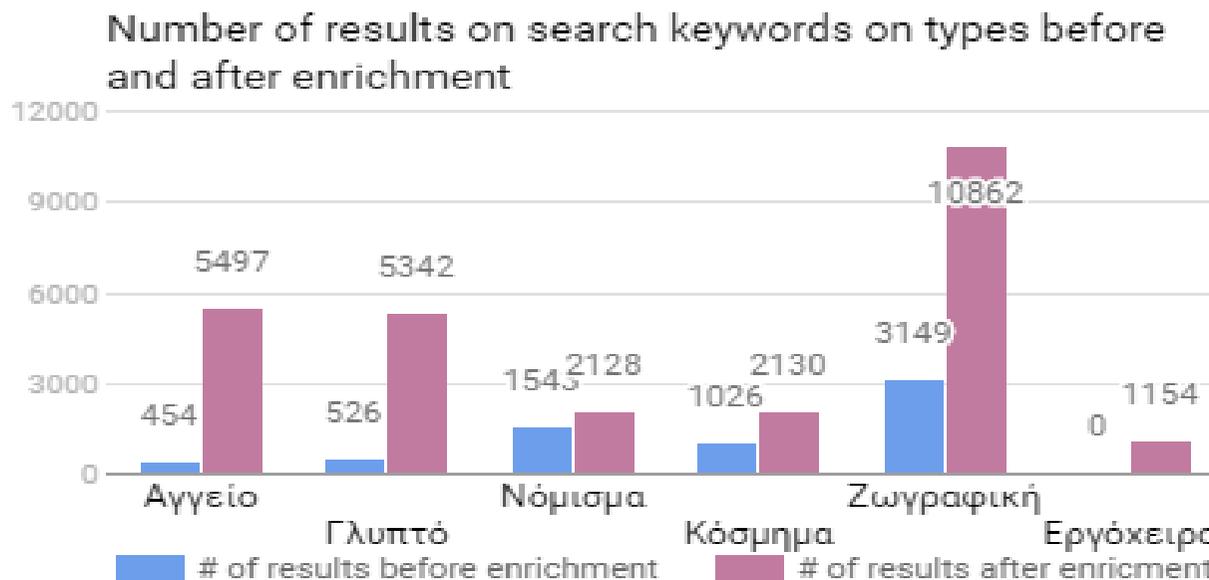
EKT type

- 2D Graphics (10155)
 - Banknote (5)
 - Drawing/Sketch (28)
 - Greetings card (37)
 - Mural painting (11)
 - Painting (324)
 - Photo (1458)
 - Postcard (202)
 - Poster (1)
 - Religious visual works (79)
 - Icon (77)
 - Santino (9)
 - Shadow puppet (522)
 - Slide (8643)
 - Stamp (45)
- 3D Artefacts and Realia (3872)
 - Arm (2)
 - Ceramic ware (1294)
 - Coin (17)
 - Costume (mode of fashion) (1107)
 - Furniture (17)
 - Household accessories (1700)
 - Jewellery (45)
 - Machine (11)
 - Medal (11)
 - Model (61)
 - Scale model (61)
 - Musical instrument (3)
 - Needlework (838)
 - Personal accessories (26)

...

Enriching the content of Searchculture.gr: The results

Improved
searchability
 & search
 multilingualism





Thank you!

Haris Georgiadis PhD |

Senior Software Engineer/Researcher
National Documentation Centre | EKT
e: hgeorgiadis@ekt.gr

Agathi Papanoti |

Archeologist
National Documentation Centre | EKT
e: apapano@ekt.gr

Special Thanks to

Dimitra Pelekanou |

Graphic Designer
National Documentation Centre | EKT
e: pelekanou@ekt.gr



National Documentation Centre

a: 48, Vas. Constantinou Av. GR-11635, Athens
e: ekt@ekt.gr | www.ekt.gr