

	 <h1>Triple</h1> <p>Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration</p>
[SEPTEMBER 2022]	Advancing Open Scholarship
	D6.6 – APIs DEVELOPMENT - RP3 Version 1.0 – Final PUBLIC
	H2020-INFRAEOSC-2019 Grant Agreement 863420

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 863420

Disclaimer- "The content of this publication is the sole responsibility of the TRIPLE consortium and can in no way be taken to reflect the views of the European Commission. The European Commission is not responsible for any use that may be made of the information it contains."

This deliverable is licensed under a Creative Commons Attribution 4.0 International License



D6.6 - APIs Development - RP3

Project Acronym:	TRIPLE
Project Name:	Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration
Grant Agreement No:	863420
Start Date:	1/10/2019
End Date:	31/03/2023
Contributing WP	WP6
WP Leader:	CNR
Deliverable identifier	D6.6
Contractual Delivery Date: 09/2022	Actual Delivery Date: 31/10/2022
Nature: Report	Version 1.0 Final
Dissemination level	PU

Revision History

Version	Created/Modifier	Comments
0.0	Luca De Santis (Net7)	Structure revision from previous version (RP2)
0.1	Luca De Santis (Net7)	First complete version
0.2	Matthew Morris (CESSDA), Haris Georgiadis (EKT)	Revisions
1.0	Luca De Santis (Net7)	Final version

Table of Contents

1. The TRIPLE API ecosystem	6
2. Search APIs	6
2.1. Elasticsearch mapping	7
2.2. Access	17
2.3. Schema	17
2.4. Authentication	17
2.5. Endpoints	17
3. Semantic API	22
3.1. Classify API	22
3.1.1. Access	22
3.1.2. Schema	22
3.1.3. Authentication	23
3.1.4. Endpoints	23
3.2. Annotate API	24
3.2.1. Access	24
3.2.2. Schema	24
3.2.3. Authentication	26
3.2.4. Endpoints	26
4. OAI-PMH Document Endpoint	27
4.1. Access	27
4.2. Schema	27
4.3. Authentication	28
4.4. Endpoint	28
5. References	28

List of tables

<i>Table 1 - Documents index structure</i>	7
<i>Table 2 - Projects index structure</i>	12
<i>Table 3 - Profiles index structure</i>	14
<i>Table 4 - GET /documents</i>	18
<i>Table 5 - GET /documents/{id}</i>	19
<i>Table 6 - GET /projects</i>	19
<i>Table 7 - GET /projects/{id}</i>	20
<i>Table 8 - GET /authors</i>	20
<i>Table 9 - GET /authors/{id}</i>	21
<i>Table 10 - POST /classify</i>	23
<i>Table 11 - POST /annotate</i>	26

Acronyms

API	Application Programming Interface
DC	Dublin Core
JSON	JavaScript Object Notation
MORESS	Mapping of Research in European Social Sciences and Humanities
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting
REST API	An API which respects the REST architecture constraints (Representational State Transfer).
SCRE	Semantic Content Retrieval Engine
SOA	Service Oriented Architecture

DRAFT

Publishable Summary

We present here the public APIs that the GoTriple platform offers to software developers, system integrators and data aggregators alike.

They can be used to access and reuse the content acquired from various sources and enriched with multiple strategies by the TRIPLE Core platform named SCRE.

There are in fact the Search REST APIs which expose the metadata indexed in the GoTriple index for the three main classes of objects managed by the platform: documents, projects and researcher profiles.

Document data is also available for data harvesters through the standard OAI-PMH protocol.

The content is indexed using Elasticsearch¹, which is an open-source search engine that allows simple queries to be built in order to search and retrieve the content required.

Besides the APIs used to access GoTriple content, the present deliverable also documents advanced and reusable semantic services developed in the course of the TRIPLE project, which, when given a text, can return the most suitable SSH categories to classify it or associate to it a set of concepts from the TRIPLE SSH Vocabulary.

This deliverable consists of an updated and expanded version² of a previous document, released in March 2021 .

¹ <https://www.elastic.co>

² D6.6 - APIs development-RP2.

1. THE TRIPLE API ECOSYSTEM

GoTriple's implementation follows closely a Service Oriented Architecture (SOA) paradigm, in which specialised components have been implemented in an independent fashion and expose their functionalities via REST APIs.

Some of them are very related to GoTriple's "inner logic", for example those related to users' registration and personalised functionalities.

Others are generic enough to be fruitfully reused in other developments, either within OPERAS or possibly by third party developers and system integrators.

The purpose of this deliverable is to provide ready to use documentation for the latter type of APIs. Three categories of services have been described in this document, in detail:

- Search APIs: public APIs, freely accessible, which allow anyone to search through the enriched TRIPLE objects for documents, projects and researcher profiles, i.e. the three main categories of data harvested by the platform.
- Semantic APIs, which consist of the Classify and Annotate services. Given a text, the former allows to classify it, that is associating it with one or more MORESS categories, which identify the main SSH disciplines³. The latter on the other hand allows to associate to the text a set of keywords from the TRIPLE controlled SSH vocabulary.
- OAI-PMH Document Endpoint: this allows third parties to harvest GoTriple's Documents index by using this standard protocol. The implementation supports the Dublin Core (DC) data format.

We provide herein a detailed description of these APIs.

2. SEARCH APIs

These are the public APIs that allow users to access the content harvested, curated and enriched by the SCRE's TRIPLE Core and indexed for the GoTriple platform.

From a technical perspective, these APIs query and retrieve data from the Elasticsearch [1] indexes where all GoTriple data is stored. Because of this, it is fundamental to present first the structure of these indexes, before providing the necessary instructions to interact with the search endpoints.

³ In the TRIPLE project a specific classification of SSH disciplines has been proposed. It consists of 27 categories, a selection of those initially identified in MORESS, a EU-funded project that ran from 2003 to 2005.

2.1. Elasticsearch mapping

Elasticsearch mapping⁴ of the TRIPLE objects are related to the TRIPLE data model presented in the WP2 D2.1 [2].

In the following table we provide the available fields for the TRIPLE object metadata (document, project, profile). The 'Properties' column contains the name of the actual Elasticsearch fields of each index. We also describe in the 'Structures' column the field data type⁵ of each property: the 'Multivalue' column specifies if the field is an array of elements. Some textual properties are copied into the *full_text* field to be globally indexed and accessible in a simple search (indicated in the 'Fulltext' column).

Finally, the fields added by the SCRE pipeline during the curation phase are presented and explained. Generally speaking, for every enrichment or adaptation applied, the original metadata are also maintained in separated properties (e.g. *original_document_types*, *discarded_keywords*,...). For a full understanding of the enrichment applied by the SCRE pipeline, the D2.5 deliverable [7] is a suggested reading.

Table 1 - Documents index structure

Metadata	Properties	Structures	multivalue	Fulltext
ID	id <i>This is the main "key" for identifying a piece of content.</i>	keyword		x
Identifiers	identifier	keyword	x	x
	doi <i>If amongst the identifiers a DOI is found, it is stored in this field.</i>	keyword	x	x
Title	headline	lang keyword	x	x

⁴ At the time of writing the Elasticsearch version used in GoTriple is 7.16.2. Documentation is available at <https://www.elastic.co/guide/en/elasticsearch/reference/7.16/mapping.html>

⁵ See <https://www.elastic.co/guide/en/elasticsearch/reference/7.16/mapping-types.html>

		<table><tr><td>text</td><td>text</td></tr><tr><td>original_lang</td><td>keyword</td></tr><tr><td>translated</td><td>boolean</td></tr></table>	text	text	original_lang	keyword	translated	boolean				
text	text											
original_lang	keyword											
translated	boolean											
Abstract	abstract	<table><tr><td>lang</td><td>keyword</td></tr><tr><td>text</td><td>text</td></tr><tr><td>original_lang</td><td>keyword</td></tr><tr><td>translated</td><td>boolean</td></tr></table>	lang	keyword	text	text	original_lang	keyword	translated	boolean	x	x
lang	keyword											
text	text											
original_lang	keyword											
translated	boolean											
Type	<div>additional_type</div> <div><i>This contains the code identifying the document type. The codes and their representation are presented in [7], chapter 2.2.4</i></div>	keyword	x									
	<div>original_document_types</div> <div><i>This contains the original type description of the harvested document. For details see [7], chapter 2.2.4</i></div>	keyword	x									

Authors	author	<div>fullname text</div> <div>id keyword</div> <div>agg keyword</div>	x	x
	discarded_authors <i>This field is populated when in a publication authors are written in multiple languages. We select only those belonging to the same language: the rest is stored here. For details see [7], chapter 2.2.6</i>	<div>lang keyword</div> <div>fullname text</div>	x	x
Contributors	contributor	text	x	x
Dates <i>For details see [7], chapter 2.2.1</i>	date_published	date		
	original_date_published	keyword	x	
	date_facets	keyword		
Timestamp	datestamp	date		
Full Text Language	in_language	keyword	x	
	original_languages <i>For details see [7], chapter 2.2.2</i>	keyword	x	
Category	topic		x	

	For details see [7], chapter 2.4.1	<table><tr><td>id</td><td>keyword</td></tr><tr><td>confidence</td><td>float</td></tr></table>	id	keyword	confidence	float						
id	keyword											
confidence	float											
Keyword	keywords	<table><tr><td>lang</td><td>keyword</td></tr><tr><td>text</td><td>text</td></tr><tr><td>original_lang</td><td>keyword</td></tr></table>	lang	keyword	text	text	original_lang	keyword	x	x		
lang	keyword											
text	text											
original_lang	keyword											
	discarded_keywords The whole element of a discarded keyword is stored as a string here (e.g. '<dc:subject xsi:type="dcterms:LCC">S 1-972</dc:subject>'). For details see [7], chapter 2.2.3.	<table><tr><td>keyword</td><td>keyword</td></tr></table>	keyword	keyword	x							
keyword	keyword											
TRIPLE concepts	knows_about For details see [7], chapter 2.4.2	<table><tr><td>uri</td><td>keyword</td></tr><tr><td>labels</td><td><table><tr><td>lang</td><td>keyword</td></tr><tr><td>text</td><td>text</td></tr></table></td></tr></table>	uri	keyword	labels	<table><tr><td>lang</td><td>keyword</td></tr><tr><td>text</td><td>text</td></tr></table>	lang	keyword	text	text	x	x
uri	keyword											
labels	<table><tr><td>lang</td><td>keyword</td></tr><tr><td>text</td><td>text</td></tr></table>	lang	keyword	text	text							
lang	keyword											
text	text											
Data provider	provider E.g. DOAJ, Isidore...	keyword	x									

Primary producer	producer	text	x	x
Publisher	publisher	text	x	x
Source (text)	mentions <i>e.g. a journal issue</i>	text	x	x
Source (http)	is_based_on_url <i>e.g. a publishing platform</i>	text		
URL of the landing page	main_entity_of_page	keyword	x	x
URL of the full text version	url	keyword	x	x
Cluster <i>These fields have been introduced to manage duplicate publications. For details see [7], chapter 2.5</i>	is_cluster	boolean		
	is_duplicate	boolean		
	cluster_id	keyword	x	
	cluster_children_count	short		
Licence <i>For details see [7], chapter 2.2.5</i>	license	keyword	x	x
	original_license	keyword	x	
Conditions of Access	conditions_of_access	keyword	x	

<i>For details see [7], chapter 2.2.5</i>	original_conditions_of_access	keyword	x	
Spatial location	spatial_coverage	keyword	x	x
Temporal period	temporal_coverage	keyword	x	x

Table 2 - Projects index structure

Metadata	Properties	Structures	multivalue	Fulltext	
Identifier	identifier	keyword		x	
Name	name	lang	keyword		x
		text	text		
		detected_lang	boolean		
		translated	boolean		
alternate_name		lang	keyword		x
		text	text		
		detected_lang	boolean		
		translated	boolean		

Description	description			x
		lang	keyword	
		text	text	
		detected_lang	boolean	
		translated	boolean	
Start date	start_date	date		
End date	end_date	date		
Organisation	organisation	keyword		x
Funding	funder	keyword		x
	funding_scheme	keyword		x
	sponsor	keyword		x
Category	topic		x	
		id	keyword	
		confidence	float	
Keyword	keywords		x	x
		lang	keyword	
		text	text	
		detected_lang	boolean	
		translated	boolean	
TRIPLE concepts	knows_about		x	x
		uri	text	

		<div>labels</div> <table><tr><td>lang</td><td>keyword</td></tr><tr><td>text</td><td>keyword</td></tr></table>	lang	keyword	text	keyword	
lang	keyword						
text	keyword						
Web page URL	url	text					

Table 3 - Profiles index structure

Metadata	Properties	Structures	multivalue	Fulltext
ID	id	keyword		x
Full Name	fullname	keyword		x
Duplicate of	AKA <i>When this field is not empty, it means that the current profile has been identified as a copy (when the same author appears in multiple documents in several forms, e.g. "Suzanne Dumouchel", "Dumouchel, Suzanne", "Dumouchel, S.",...). Its value is the identifier</i>	keyword		

	<i>of “main profile” of this person. For details see [7], chapter 2.6</i>			
Documents	author_of <i>This multivalue field contains the link of all the documents of which this profile is an author</i>	keyword	x	
	numberOfDocuments	int		
Topic	topic <i>These are the MORESS categories associated with the profile, either explicitly chosen by registered users or automatically identified from her/his publications.</i>	keyword	x	
<i>The following fields apply only for registered users in the GoTriple platform</i>				
Registered User	registeredUser	boolean		
GoTriple ID	goTripleId	keyword		
Other IDs	Identifier			x

	For example: { "orcid": "0000-...-.X" , "idref": "XXXXXX", }	type	value		
Given name (First name)	givenName	text			x
Family name (Last name)	familyName	text			x
Pronouns	pronouns	keyword			
Occupation	hasOccupation <i>Type of the user, (e.g. researcher); specified during the registration phase</i>	keyword			
Current position	currentOrganization	text			
	currentRole	text			
Open to collaboration	openToCollaboration	boolean			
Languages	knowsLanguage	keyword	x		x
URL	url	text	x		
Photo	photo <i>URL of the personal picture chosen by the user</i>	text			

2.2. Access

The APIs and their full description in Swagger/OpenAPI [6] format are available at the URL <https://api.gotriple.eu/>.

2.3. Schema

The API response uses the same schema of the Elasticsearch indexes described in chapter 2.1.

2.4. Authentication

The Search APIs are free to use.

2.5. Endpoints

Below we present the list of endpoints that can be used to access GoTriple's resources.

All of them can return results in three possible formats (JSON, JSON-LD, HTML) using content negotiation, thus, depending on the “accepts” HTTP Header (application/json, application/ld+json and text/html, respectively) passed as a parameter of the request.

The JSON-LD is the default in case no explicit “accept” header is specified: its schema automatically uses the Hydra Core Vocabulary [4] as automatically rendered by the API Platform [5], which is the framework used for implementing these APIs.

Table 4 - GET /documents

Description	Provides the full search on the Documents index.
Query Parameters	<ul style="list-style-type: none"> • q : The text to search for in all metadata aggregated in the <i>full_text</i> field. Search query elements must be separated by a whitespace. • include_duplicates (default false): when false (the default) duplicates, united in “clusters”⁶, are not returned by the API • fq : filter query. Available filters are: <i>topic, type, year, author, provider,, in_language,, conditions_of_access, license, is_cluster, cluster_id, is_duplicate</i>. Multiple filters are possible by using the “;” separator (the search is translated in AND between the filters). Multiple values for a filter are possible with the “,” separator (the search is translated in OR between the values). • aggs : aggregators. If specified, an additional element is added in the response with the number of objects returned for every single attribute indicated as “aggregator”. A typical use of “aggs” is to support filtering by facets. Available aggregators are: <i>topic, type, year, author, provider, in_language,, conditions_of_access, license, is_cluster, cluster_id, is_duplicate</i>. Multiple aggregators are possible by using the “;” separator. Aggregator length and order can be configured with <i>size, sort</i> and <i>order</i>. It is possible to filter the values for an aggregator with <i>include</i> or <i>exclude</i> based on strings. • page : The page number • size : The number of items per page (the default is 25)
Returns	An array of document objects which reflect the structure of the Documents Elasticsearch index described above.
Example	<pre>curl -X 'GET' \ 'https://api.gotriple.eu/documents?q=femicide&fq=year%3D2001%2C2013&aggs=topic%2Corder%3Ddesc&page=1&size=25' \ -H 'accept: application/json'</pre>

⁶ See D2.5 [7] chapter 2.5.

Table 5 - GET /documents/{id}

Description	Get a document resource
Path Parameter	A valid document Id
Returns	A document object which reflects the structure of the Documents Elasticsearch index described above
Example	<pre>curl -X 'GET' \ 'https://api.gotriple.eu/documents/50%7Cdedup_wf_001%3A%3Af0891a953ec33a65d6e17ac9fef03f1c' \ -H 'accept: application/json'</pre>

Table 6 - GET /projects

Description	Provides the full search on the Projects index.
Query Parameters	<ul style="list-style-type: none"> • q : the text to search for • fq : filter query. Available filters are: <i>topic</i>, <i>year</i>, <i>funder</i>, <i>funding_scheme</i>, <i>organization</i>, <i>sponsor</i>, <i>in_progress</i>. Multiple filters are possible by using the “;” separator (the search is translated in AND between the filters). Multiple values for a filter are possible with the “,” separator (the search is translated in OR between the values). • aggs : aggregators (see Table 4 for a full description of this parameter). Available aggregators are: <i>topic</i>, <i>year</i>, <i>funder</i>, <i>funding_scheme</i>, <i>organization</i>, <i>sponsor</i>, <i>in_progress</i>. Multiple aggregators are possible by using the “;” separator. Aggregator length and order can be configured with <i>size</i>, <i>sort</i> and <i>order</i>. It is possible to filter the values for an aggregator with <i>include</i> or <i>exclude</i> based on string. • page : the page number • size : the number of items per page
Returns	An array of project objects which reflect the structure of the Projects Elasticsearch index described above.
Example	<pre>curl -X 'GET' \ 'https://api.gotriple.eu/projects?q=triple&fq=topic%3Dsocio%3Byear%3D2001%2C2013&aggs=topic%2Cinclude%3Djohn%2Cexclude%3Dmartin%2Csize%3D10%2Csort%3Dcount%2Corder%3Ddesc&page=1&size=25' \ -H 'accept: application/json'</pre>

Table 7 - GET /projects/{id}

Description	Get a project resource
Path Parameter	A valid project Id
Returns	A project object which reflects the structure of the Projects Elasticsearch index described above
Example	curl -X 'GET' \ 'https://api.gotriple.eu/projects/fp7%3A322737' \ -H 'accept: application/json'

Table 8 - GET /authors

Description	Provides the full search on the Profiles index.
Query Parameters	<ul style="list-style-type: none"> • q : The text to search for • include_long_names : (default false): when false (the default) authors with full names (field <i>full_name</i>) longer than 30 characters will not be returned. This parameter has been introduced because sometimes Institutions are included as authors of a publication (e.g. "Department of Computer Science, University of Pisa, Italy"). The use of this parameter limits (but of course doesn't solve) the presentation's problems that this situation can generate on the front-end of GoTriple. • fq : filter query. Available filters are: <i>has_occupation</i>, <i>topic</i>, <i>knows_language</i>, <i>number_of_documents</i>⁷, <i>registered_user</i>. Multiple filters are possible by using the ";" separator (the search is translated in AND between the filters). Multiple values for a filter are possible with the "," separator (the search is translated in OR between the values). • aggs : aggregators (see Table 4 for a full description of this parameter). Available aggregators are: <i>open_to_collaboration</i>, <i>has_occupation</i>, <i>topic</i>, <i>knows_language</i>, <i>number_of_documents</i>, <i>registered_user</i>. Multiple aggregators are possible by using the ";" separator. Aggregator length and order can be configured with <i>size</i>, <i>sort</i> and <i>order</i>. It is possible to filter the values for an aggregator with <i>include</i> or <i>exclude</i> based on string.

⁷ number_of_documents aggregates the number of publications in a range with these possible values: 0, 1-4, 5-10, 10+

	<ul style="list-style-type: none"> • page : the page number • size : the number of items per page
Returns	An array of profile objects which reflect the structure of the Profiles Elasticsearch index described above.
Example	<pre>curl -X 'GET' \ 'https://api.gotriple.eu/authors?q=Pierre%20Bourdieu&include_long_names=false&aggs=number_of_documents&page=1&size=25' \ -H 'accept: application/json'</pre>

Table 9 - GET /authors/{id}

Description	Get a profile resource
Path Parameter	A valid profile Id
Returns	A profile object which reflects the structure of the Profiles Elasticsearch index described above
Example	<pre>curl -X 'GET' \ 'https://api.gotriple.eu/authors/suzanne_dumouchel_PklafWmcVuoegTyUv5KFI' \ -H 'accept: application/json'</pre>

3. SEMANTIC API

3.1. Classify API

The purpose of this service is to assign one or more MORESS categories to the text passed as input. It takes the following parameters:

- the language of the text, codified in ISO-639-1
- the threshold (0 .. 1.0)
- the text to be classified.

The service is based on a machine learning model which has been trained with documents in twelve languages⁸ which represent all the 27 MORESS categories of the TRIPLE project.

If the threshold parameter is set, only the categories with a higher score are returned. In any case, the service only returns up to two categories.

The full description of this service is available in Swagger/OpenAPI [6] format at the following URLs:

- Swagger UI: <https://services-ca.gotriple.eu/triple-api/v1/ui/>
- Swagger JSON file: <https://services-ca.gotriple.eu/triple-api/v1/swagger.json>

3.1.1. Access

The API is available at the URL <https://services-ca.gotriple.eu/triple-api-dev/v1/classify>.

3.1.2. Schema

The response is returned in JSON in the format that follows:

```
[
  {
    "category": "category 1",
    "score": score category 1
  },
  {
    "category": "category 2",
    "score": score category 2
  }
]
```

The codification of the MORESS categories returned by the service is described in the TRIPLE D2.5 deliverable [7], chapter 2.4.1.

⁸ At the time of writing the service supports only nine languages: English, French, Spanish, German, Greek, Croatian, Italian, Polish and Portuguese.

3.1.3. Authentication

To use the API it is necessary to be authenticated via HTTP Basic Access Authentication (login and password) with credentials provided by the TRIPLE team responsible for this service.

3.1.4. Endpoints

Table 10 - POST /classify

Description	Given a text, this service returns up to two MORESS categories for classifying it
Parameters	<ul style="list-style-type: none"> • lang (string - as a query/URL parameter): The language of the text (possible values: "de", "el", "en", "es", "fr", "hr", "it", "pl", "pt") • threshold (number - as a query/URL parameter): The minimum score for the categories to be returned. It accepts a number between 0 and 1. Its default value is 0.05. • text (string - in the body of the HTTP Post request): The text to classify
Returns	An array with no more than two objects, one for each returned category. The precision score is also returned. For the detailed format, see 3.1.2
Example (curl)	<pre>curl -X 'POST' \ 'https://services-ca.gotriple.eu/triple-api/v1/classify?lang=es&threshold=0.2' \ -H 'accept: application/json' \ -H 'authorization: Basic AUTHENTICATION-TOKEN-HERE' \ -H 'Content-Type: text/plain; charset=utf-8' \ -d 'Desde una perspectiva dialógica entre la filosofía, las ciencias sociales y la realidad social, que desemboca en una epistemología renovada, el artículo busca comprender: el fenómeno de la</pre>

desconfianza del ciudadano con instituciones de la sociedad y del Estado chileno, la desconfianza del ciudadano frente a lógicas actuales del mercado como la mercantilización de las relaciones sociales y, por último, la desconfianza entre ciudadanos en espacios cotidianos. El trabajo es parte de los estudios de la sociología y de la antropología desde la perspectiva de los imaginarios sociales y se interesa en la deconstrucción de la desconfianza en tanto elemento característico central de los vínculos sociales en el Chile de la post-dictadura y de su relación potencial o real con el descontento.'

3.2. Annotate API

Given a text, this service identifies the concepts from the TRIPLE vocabulary that emerge from an automatic analysis of it.

The service accepts as parameters:

- the language of the text, codified in ISO-639-1
- the text to be annotated.

The response is very rich (see the Schema below). The full description of this service is available in Swagger/OpenAPI [6] format at the following URLs:

- Swagger UI: <https://services-ca.gotriple.eu/triple-api/v1/ui/>
- Swagger JSON file: <https://services-ca.gotriple.eu/triple-api/v1/swagger.json>

3.2.1. Access

The API is available at the URL <https://services-ca.gotriple.eu/triple-api/v1/annotate>

3.2.2. Schema

The response is returned in JSON in a very rich format. It consists of an array of JSON objects, one for every identified keyword, which contain:

- uri: the uri of the concept in the TRIPLE Vocabulary ontology
- pref_label: the preferred labels that describe the concept in all the various available languages supported by the TRIPLE Vocabulary (12 at the present date⁹)
- alt_label: the alternative labels of the concept, again in the multiple available languages

⁹ The languages in which the entities of the TRIPLE Vocabulary are described are: Croatian, Dutch, English, Finnish, French, German, Greek, Hungarian, Italian, Portuguese, Spanish, Ukrainian.

- matched: the parts of the text that have been identified to correspond to the current keyword
- exact_matches: the URIs of corresponding concepts in other external vocabularies
- close_matches: the URIs of similar concepts in other vocabularies
- broaders: the URIs of the more general concepts (if any) in the TRIPLE Vocabulary hierarchy
- narrower: the URIs of more specific concepts (if any) in the TRIPLE Vocabulary hierarchy.

An example of the output of the service is presented below:

```
[
  {
    "alt_label": [
      {
        "lang": "fr",
        "value": "L'étude des humains"
      },
      {
        "lang": "en",
        "value": "Study of humans"
      }
    ],
    "broaders": [
      "http://semantics.gr/authorities/SSH-LCSH/sh85028263",
      "http://semantics.gr/authorities/SSH-LCSH/sh85036229"
    ],
    "close_matches": [
      "http://www.ysc.fi/onto/ysc/p1560",
      "http://id.worldcat.org/fast/810196"
    ],
    "exact_matches": [
      "http://id.loc.gov/authorities/subjects/sh85005581"
    ],
    "matched": [
      "anthropology",
      "study human"
    ],
    "narrower": [
      "http://semantics.gr/authorities/SSH-LCSH/sh00002492",
      "http://semantics.gr/authorities/SSH-LCSH/sh2006007942"
    ],
    "pref_label": [
      {
        "lang": "fr",
        "value": "Anthropologie"
      },
      {
        "lang": "en",
        "value": "Anthropology"
      }
    ],
    "uri": "http://semantics.gr/authorities/SSH-LCSH/sh85005581"
  }
]
```

3.2.3. Authentication

To use the API it is necessary to be authenticated via HTTP Basic Access Authentication (login and password) with credentials provided by the TRIPLE team responsible for this service.

3.2.4. Endpoints

Table 11 - POST /annotate

Description	Given a text, this service returns the most suitable keywords from the TRIPLE Vocabulary that can be associated with the text sent as parameter.
Parameters	<ul style="list-style-type: none"> • lang (string - query/URL parameter): The language of the text (possible values: "de", "el", "en", "es", "fr", "hr", "it", "pl", "pt") • minimum_clique_size (integer - query/URL parameter): this parameter was introduced to improve the quality of the service. As the TRIPLE Vocabulary is defined as a hierarchy, for all identified keywords there is a top-level concept. Given a potential keyword identified in the text, it is returned if and only if there are at least other identified keywords which share the top-level concept and their global number is \geq minimum_clique_size. This improves the probability that a keyword is correct because other related concepts are present in the text. Its default value is 3. • text (string - in the body of the HTTP Post request): The text to annotate
Returns	An array of objects, one for each returned keyword. For the detailed format of the response see 3.2.2.
Example (curl)	curl -X 'POST' \ 'https://services-ca.gotriple.eu/triple-api/v1/annotate?lang=es&minimum_clique_size=3' \

```
-H 'accept: application/json' \  
-H 'authorization: Basic dHJpcGxlX3VzZXI6c2RlZkZKQVFaNDY4Nw==' \  
-H 'Content-Type: text/plain; charset=utf-8' \  
-d 'Desde una perspectiva dialógica entre la filosofía, las ciencias  
sociales y la realidad social, que desemboca en una epistemología  
renovada, el artículo busca comprender: el fenómeno de la  
desconfianza del ciudadano con instituciones de la sociedad y del  
Estado chileno, la desconfianza del ciudadano frente a lógicas actuales  
del mercado como la mercantilización de las relaciones sociales y, por  
último, la desconfianza entre ciudadanos en espacios cotidianos. El  
trabajo es parte de los estudios de la sociología y de la antropología  
desde la perspectiva de los imaginarios sociales y se interesa en la  
deconstrucción de la desconfianza en tanto elemento característico  
central de los vínculos sociales en el Chile de la post-dictadura y de su  
relación potencial o real con el descontento.'
```

4. OAI-PMH DOCUMENT ENDPOINT

It consists of a full implementation of the OAI-PMH [8] standard harvesting protocol in which data is formatted in Dublin Core (DC) and serialised in XML. All documents data of the Publications index are therefore available through this endpoint.

4.1. Access

The endpoint is available at the URL <https://api.gotriple.eu/oai2>.

4.2. Schema

As indicated above, the implementation only supports the OAI_DC (Dublin Core) metadata model, so all requests must include the parameter

```
metadataPrefix=oai_dc
```

All documents are associated to one or more “sets” (via the setSpec header) corresponding to the possible values of the field “provider” in the Publications index (e.g. “isidore”, “openaire”, “ekt”, etc.).

4.3. Authentication

No authentication is required. This endpoint is therefore free to use.

4.4. Endpoint

All OAI-PMH standard verbs have been implemented. We list below the supported verbs and some working examples for using them.

- Identify: <https://api.gotriple.eu/oai2?verb=Identify>
- ListSets: <https://api.gotriple.eu/oai2?verb=ListSets>
- ListMetadataFormats: <https://api.gotriple.eu/oai2?verb=ListMetadataFormats> (only oai_dc)
- ListRecords: https://api.gotriple.eu/oai2?verb=ListRecords&metadataPrefix=oai_dc
- ListIdentifiers: https://api.gotriple.eu/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc
- GetRecord:
https://api.gotriple.eu/oai2?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:gotriple.eu:50|doiboost::b10dfc7431bd5272b48332bb4892ad8f

5. REFERENCES

- [1] Elasticsearch - <https://www.elastic.co/elasticsearch/>
- [2] D2.1 - Data Acquisition Plan
- [3] TRIPLE Vocabulary: an SSH multilingual vocabulary based in LCSH - <https://www.semantics.gr/authorities/vocabularies/SSH-LCSH/?language=en>
- [4] Hydra Core Vocabulary: <https://www.hydra-cg.com/spec/latest/core/>
- [5] API Platform: <https://api-platform.com/>
- [6] Swagger/OpenAPI - <https://swagger.io/>
- [7] D2.5 - Report on data enrichment
- [8] OAI-PMH Protocol - <http://www.openarchives.org/OAI/openarchivesprotocol.html>