EOSC Technical Specification

Common Services

Annotation Service

Version:	1
Status:	EOSC-hub Proposal
Dissemination Level:	Public
Document Link:	

Abstract

The EOSC Annotation service comprises the ability for end-users to create, manage and search for annotations on data resources.

OSC-HUB RECEIVES FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER RANT AGREEMENT NO. 777536.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	***Name***	***Partner/Activity***
From:	Yann Le Franc Heinrich Widmann	e-STF DKRZ
Reviewed by:	Bartosz Kryza Sara Ramezani	CYFRONET SURFSARA

DOCUMENT LOG

Issue	***Date***	***Comment***	***Author***
v.1	24/01/2020	First release ready for public consultation	Heinrich Widmann (DKRZ)

TERMINOLOGY

https://wiki.eosc-hub.eu/display/EOSC/EOSC-hub+Glossary

Terminology/Acronym	***Definition***

Contents

1 Introduction

The Annotation Service should enable end-users to extend descriptions of datasets or parts of datasets with user-defined content, without modifying the underlying dataset, e.g. adding comments, free text keywords or semantic tags (keywords from ontologies). These annotations can be used to search and aggregate datasets or parts of datasets into user-defined datasets, either localised in a unique data repository or throughout a heterogeneous and distributed set of data repositories.

2 High-level Service Architecture

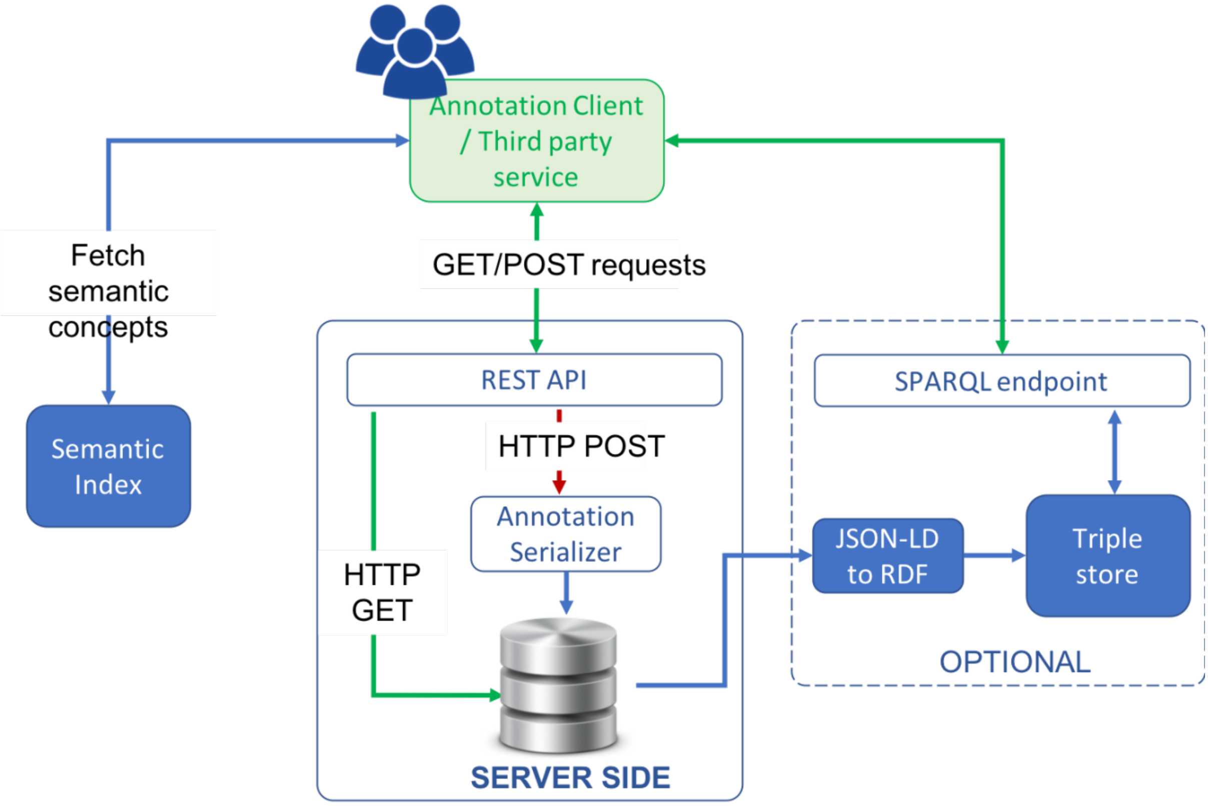

Figure 1: High-level Architecture of the EOSC Annotation Service

The technical implementation of an annotation service enabling semantic annotation should comprise the following components:

1. A client providing a User Interface to create, manage and use annotations
2. A database backend to store the annotations
3. An API with functionalities to retrieve and create annotations (for third party clients)

4. An index of Semantic resources to provide access to the wealth of existing semantic resources
5. A serializer to create annotations in JSON-LD following the Web Annotation standard.
6. A conversion microservice to transform JSON-LD into RDF (optional) and
7. An RDF store to provide access to annotation via a SPARQL endpoint (optional).

A schematic representation of this proposed architecture is shown in the diagram above.

3 Adopted standards

Standard	Short description	References
JSON-LD	JSON representation of an RDF graph	https://www.w3.org/TR/json-ld/
W3C Web Annotation data model	Data model and ontology to build annotation of documents and content	https://www.w3.org/TR/annotation-model/
OpenAPI specification	Standard, programming language-agnostic interface description for REST APIs, which allows both humans and computers to discover and understand the capabilities of a service	https://github.com/OAI/OpenAPI-Specification
RDF	A W3C standard to create graphs of data	https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/
SPARQL	SPARQL is a query language developed to query RDF graphs	https://www.w3.org/TR/sparql11-overview/
SOLR	SOLR is highly reliable, scalable and fault tolerant, providing distributed indexing, replication and load-balanced querying, automated failover and recovery, centralized configuration and more. SOLR powers the search and navigation features of many of the world's largest internet sites.	https://lucene.apache.org/solr/

Protocol/API	Short description	References
RESTful API	API based on OpenAPI principles to initialize the annotations and retrieve stored annotations. B2NOTE v1 API is not compliant with OpenAPI. This will be changed in later version.	https://restfulapi.net/ http://b2note.eudat.eu/api
User Interface integration (Iframe, webcomponent,…)	a widget for the integration within the User Interface of partner services	https://www.w3.org/TR/2011/WD-html5-20110525/the-iframe-element.html https://www.djangoproject.com/
SPARQL endpoint	URL giving access to the SPARQL query engine of a triple store	https://www.w3.org/TR/rdf-sparql-query/

4 Interoperability guidelines

Interoperability with similar service should be achieved through common functionalities of the REST API and the usage of JSON-LD as serialization format.

5 Examples of solutions implementing this specification

While there are many services for annotating web content (e.g. Pundit http://www-old.thepund.it/, Hypothesis https://web.hypothes.is/, etc…), there is only one service for annotating research data and for 'sematic tagging and discovery', theB2NOTE service offered by EUDAT:

- EUDAT-B2NOTE (http://b2note.eudat.eu) allows users to easily create, search and manage annotations. An annotation is a keyword or commentary attached to a digital object (data collection, file) that explains or classifies it. B2NOTE is a standalone service for annotating data content hosted within the EUDAT CDI. For example,B2NOTE is already integrated in B2SHARE and integration within B2FIND is work in progress and in the roadmap of EOSC-OpenAire cooperation. Furthermore B2NOTE can be integrated with community services such as the CLARIN Virtual Language Observatory and with OpenAire data services such as Zenodo and the Research Community Dashboard.

5.1 Procedure for integrating a service with the EOSC-hub *Annotation Service*

EUDAT-B2NOTE has been designed to be integrated with any data services through an easy integration with services web UI. The process is rather simple:

1. Create an HTTP iFrame calling the service UI
2. Enable Cross-Origin Resource Sharing (CORS, see (https://en.wikipedia.org/wiki/Cross-origin_resource_sharing)
3. Create a button in your UI to provide access to the iFrame
4. When activated, initialize the service with the PID and URL of the file to be annotated