



inDICES

# Measuring the Impact of Digital Culture

## Deliverable 4.2

### Database and Service Layers Architecture

VI



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870792.

The sole responsibility for the content of this publication lies with the authors. It does not necessarily represent the opinion of the European Union. Neither the EASME nor the European Commission is responsible for any use that may be made of the information contained therein.

The research leading to these results has received funding from the European Commission's Horizon 2020 Programme (H2020-DT-GOVERNANCE-13-2019) under grant agreement n° 870792.



**inDICES**

Measuring the Impact  
of Digital Culture

## D4.2 – Database and Service Layers Architecture V1

Version 1.0  
31 May 2021

**Grant Agreement Number:** 870792  
**Project Acronym:** inDICES  
**Project Title:** Measuring the impact of Digital Culture  
**Funding Scheme:** H2020-DT-GOVERNANCE-13-2019  
**Project Coordinator Name,  
Title and Organisation:** Simonetta Buttò, Director of the Central Institute for the  
Union Catalogue of the Italian Libraries (ICCU)  
**Tel:** +39 06 49210425  
**E-mail:** [simonetta.butto@beniculturali.it](mailto:simonetta.butto@beniculturali.it)  
**Project Website:** [www.indices-culture.eu](http://www.indices-culture.eu)

**Authors:** Katinka Böhm, Arno Scharl  
webLyzard technology (WLT)  
**Contributing Partners:** Nadia Nadesan, Ivan Verges - Platoniq  
**Reviewers:** Rasa Bocyte - Netherlands Institute for Sound and Vision  
Alba Irollo - Europeana

## Document History

- 25.04.2021 – Document Created from Template (WLT)
- 07.05.2021 – First Deliverable Draft (WLT)
- 14.05.2021 – Extended *Embeddable Widgets* Section (WLT)
- 18.05.2021 – Added Section *CHI Knowledge Graph* (WLT) and reworked other parts
- 19.05.2021 – Restructuring of Sections (WLT),  
*Participatory Space* Section & Additions to Architecture (Platoniq)
- 20.05.2021 – Added *Introduction* and *Summary* (WLT)
- 21.05.2021 – Clean-Up, Additions and Suggestions (WLT & Platoniq)
- 21.05.2021 – Revisions, additional figures and minor restructuring (WLT)
- 26.05.2021 – Deliverable Review (NISV, Europeana)
- 30.05.2021 – Final edits and formatting; Version 1.0 (WLT)

## Glossary

***inDICES Open Observatory (Platform), also: inDICES Platform.*** An integrated digital space that operates as a participatory and monitoring platform, manages the inDICES open data collection and provides methodological tools to the different inDICES stakeholders (cultural institutions, policy-makers, funding agencies, researchers, practitioner networks). It consists of three interrelated components; the *Participatory Space*, the *Visual Analytics Dashboard* and the *CHI Self-Assessment Tool*. The *Data Repository* has been established as a fourth component to be integrated in the inDICES platform.

***Participatory Space, also: Decidim platform.*** The inDICES *Decidim* platform for community engagement; a participatory space and collaborative environment for dialog and experience exchange in the form of surveys, debates and collaborative projects, for experts and members of the cultural and creative sectors. It supports different participation modules in which users can open a *Debate* or create a *Proposal*. For more details on the modules see D4.1 Section 5.

***Visual Analytics Dashboard, also: inDICES dashboard.*** An inDICES dashboard that is part of the *webLyzard analytics platform* and allows users to explore and visualise data from the *webLyzard content and metadata repository* along multiple metadata dimensions, based on a multiple coordinated view approach using visual widgets.

***CHI Self-Assessment Tool.*** An inDICES online tool for the CH sector that allows organisations and institutions to assess their readiness to DSM based on the *CHIs DSM-readiness Assessment Methodology*.

***Methodological Toolbox.*** A set of techniques tailored to inDICES data gathering, analysis and presentation, consisting of a Theoretical Framework, datasets on CHI digitisation and socio-economic impact and a set of associated indicators.

***inDICES Data Repository.*** Collection of inDICES datasets and content resources accessible through the inDICES Open Observatory Platform.

## Acronyms

**API** - Application Programming Interface

**CH** - Cultural Heritage

**CHI** - Cultural Heritage Institution

**CCS** - Cultural and Creative Sectors

**DSM** - Digital Single Market

**HTML** - Hypertext Markup Language

**KG** - Knowledge Graph

**LOD** - Linked Open Data

**NLP** - Natural Language Processing

**REST** - Representational State Transfer

## Table of Contents

<b>1. Introduction</b>	<b>5</b>
<b>2. System Architecture</b>	<b>5</b>
2.1 Data Infrastructure and Platform Integration	6
2.2 Content Ingestion	7
Web and Social Media Sources	7
Content from the Participatory Space	8
2.3 Contextualisation Services	10
2.4 CHI Knowledge Graph	10
2.5 inDICES API Framework	14
webLyzard API Overview	14
Authentication	14
Initialisation	15
Request an Access Token (JWT)	15
Document API Specification	15
Statistical API Specification	16
Search API Specification	18
Decidim API Specification	20
<b>3. Participatory Space</b>	<b>22</b>
3.1 Decidim Framework	22
3.2 inDICES Decidim Implementation	23
3.3 Transparency and Traceability	24
<b>4. Visual Analytics Dashboard</b>	<b>25</b>
4.1 Dashboard Structure and Configuration	26
4.2 Embeddable Widgets	27
Template-Based Widgets	28
Thread-Based Widgets	28
Similarity-Based Widgets	29
Similarity Requests via Text Snippets	30
Similarity Requests via URLs	30
<b>5. Summary and Outlook</b>	<b>31</b>

# 1. Introduction

This deliverable provides the first of two reports of the project frontend applications, namely the *Participatory Space*, i.e., the collaborative environment developed in *Task 4.2*, and the *Visual Analytics Dashboard* developed in *Task 4.3*, as well as an overview of the inDICES API Framework that is used to integrate and synchronise the various components. This deliverable provides the technical details and initial implementations of the inter-platform communication, including the components required for the proof of concepts reported in *D4.1 User Scenarios and Wireframes Report*.

The goal of the system integration between the two technical integration partners *Platoniq* and *webLyzard* is to make the visualisation and contextualisation functionalities of the *webLyzard* analytics platform available to the users of the *Decidim* participation platform in support of content creation in agreement with the user needs and functional requirements formulated in *Task 4.1 Co-Design the Observatory*. The Visual Analytics Dashboard, integrated as one of the core elements of the Open Observatory, provides data visualisations which allow users to observe and monitor trends on topics related to the digital transformation of CHIs. Additionally, the *participatory space* user benefits from recommendations and contextualisation during the ideation process, providing an extended information landscape both from real-world stories as well as from data shared by fellow *participatory space* users. The contextualisation process draws upon both internal as well external content sources - the current and previous discussions in the participatory space as well as the crawled Web content from news media, CHI and stakeholder sites.

After the general introduction, Section 2 gives an overview of the system architecture of both, the analytics components behind the inDICES dashboard, and the Decidim platform architecture. It describes the data infrastructure and integration between both platforms, as well as the contextualisation process and CHI knowledge graph that are being developed within the project. The final subsection *inDICES API Framework* provides a detailed description of the individual APIs used to make the different contents and visualisations available and shareable between the project partners. Section 3 and Section 4 focus on the frontend components, design principles and user experience of the inDICES dashboard and the participatory space, respectively. Finally, we give a summary and future outlook of further development.

## 2. System Architecture

The *inDICES Open Observatory* platform is the integration of all technical components that drive the inDICES applications, tightly coupling the inDICES *Visual Analytics Dashboard* with the *Decidim Participatory Platform*.

To facilitate this, the inDICES Open Observatory entails as webLyzard components (1) a data ingestion layer, in the form of a *Web Crawler* for ingestion of relevant Web content for contextualisation and a *Document API* and a *Statistical API* for ingestion of *Decidim* content,

(2) multi-modal content and metadata repository for data persistence, and (3) as exchange layer a *Search Index* with an API, and a *Visualisation Engine* with an API.

The backbone of the content and metadata repository are large Linux-based hardware clusters to run distributed, horizontally sharded *CockroachDB* and *ElasticSearch 7.x* databases. The individual services, built in *Python 3.7* and *Java 11*, are containerised through *Docker*, allowing for fully managed service provisioning via orchestration frameworks such as *Kubernetes* or *Portainer*. Load Balancing and Service Discovery are used both on microservice and database level where fitting.

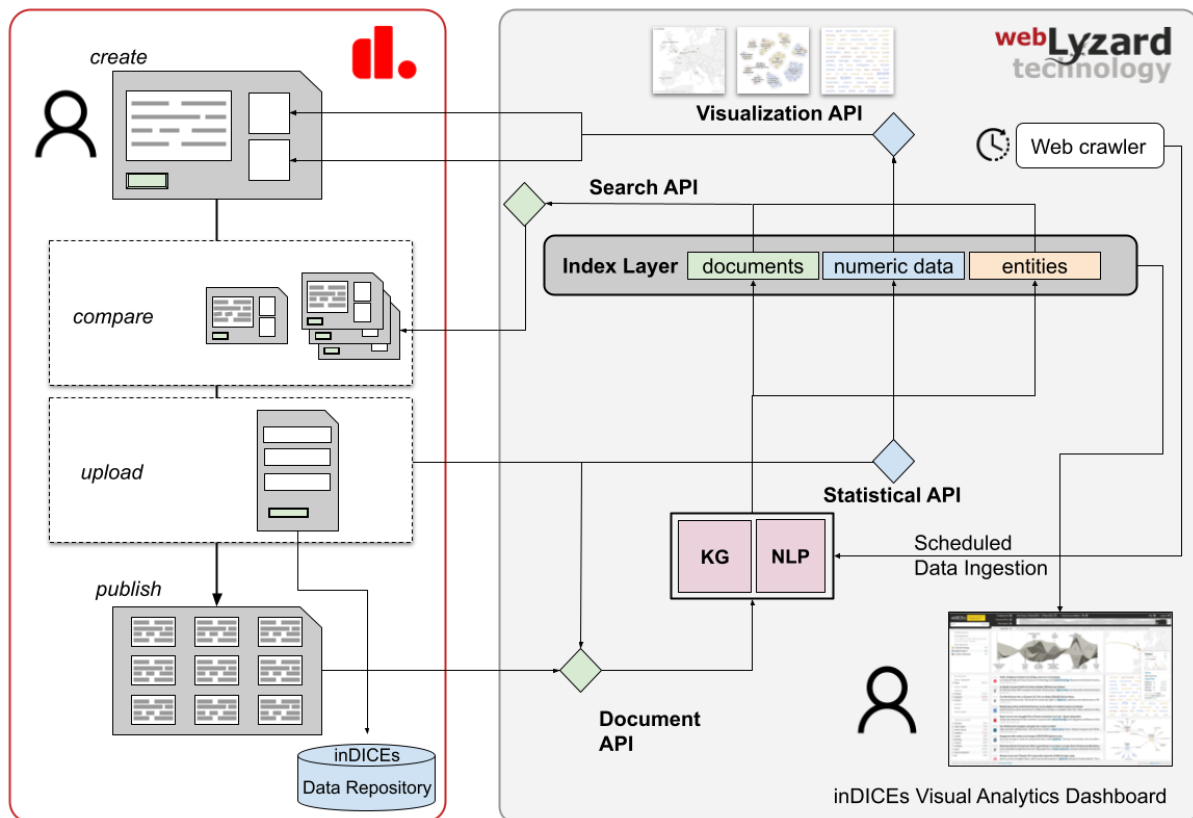
The second element of the inDICES Open Observatory is the participatory platform, which is based on the software Decidim. Decidim is a REST application that uses Ruby on Rails (v5.2) technology for serving the web application written in the Ruby language (v2.7). For permanent data storage uses an independent PostgreSQL server and the S3 disk storage service for user uploaded content. The SQL database (backed up every day automatically) is used for storing any user content except binary files or uploaded documents. A REDIS server is also in place as an ephemeral storage system used by queue processing activities (such as email sending or costing background operations). The system is deployed on Linux servers behind a load balancer and can be horizontally scaled in case of need.

## 2.1 Data Infrastructure and Platform Integration

Figure 1 presents a visual schematic of the *inDICES* system integration as data flows from the *Decidim* platform to the *webLyzard* analytics platform via data ingestion services, Document API and Statistical API, and from the *webLyzard* analytics platform to the *Decidim* platform for embedded visualisations that are requested via the Visualisation API. It illustrates in more depth the inter-platform communication approach described in D4.1 (7.4.1 *Technology framework for inter-platform communication*).

The proposed integration concept includes a *similarity comparison* step between the created content and other already existing content on the Participatory Space Platform, e.g. after creating but before publishing a proposal, a new view shows other proposals with similar content. This makes the user aware of ongoing discussions and gives a chance to modify or abandon the current proposal.

During the creation process it is being considered to offer the possibility to propose and upload a dataset to the inDICES Data Repository. In this step, selected metadata of the datasets could be shared with the Document API (textual data) or the Statistical API (statistical observations). Linked open data (LOD) hosted through the open-source data management system CKAN, which is widely used by governments to provide public datasets, could also be ingested through a separate data mirroring process, thereby transforming LOD items into documents using the provided metadata as content and the upload date as a reference point in the dashboard.



**Fig 1: High-level view of the system architecture to integrate the *Decidim Participatory Space* and the *webLyzard Visual Analytics Platform*.**

To further connect both platforms, the *OAuth 2.0* authorisation protocol will be applied to allow for an easy “*Single-Sign-On*” functionality as outlined in the development roadmap in *D4.1 User scenarios and wireframes report*. This will make it possible for the user to switch from one platform’s front end application, i.e. the Participatory Space, respectively the Visual Analytics Dashboard, to the other. This will provide users with easy access to the inDICES dashboard while creating a proposal or making a hypothesis. It also opens the future possibility for the users to directly jump to the participatory space to create a proposal or start a discussion on findings made with the visual analytics components via permalinks.

## 2.2 Content Ingestion

### Web and Social Media Sources

WP1 has used and extended the digital content acquisition pipeline of webLyzard to capture topics and sources relevant for inDICES, with a special focus on CHI web sites and added support of Italian content. The inDICES document repository currently grows at a rate of approximately 20 million documents per month, including digital content in six different languages: English, French, German, Spanish, Dutch and Italian. Collected documents are being filtered for relevance and redundancy, pre-processed, and annotated along multiple semantic dimensions - including document keywords, geotags, sentiment, and referenced named entities.



A distributed Web crawler collects Web pages to be added to the inDICES content repository. We use the Java-based open-source *Apache Storm Crawler* to perform this task, using bandwidth limits to minimise the resulting load on third-party servers. The data collection process respects the Web site owner's *robots.txt* settings (a text file placed in the top directory, which is used by site administrators to restrict access to files and directories on a Web server). To gather social media content, we use the official APIs provided by the various networking platforms, strictly adhering to these platforms' usage restrictions and only accessing the public portion of the content. To ensure GDPR compliance, this includes the processing of status deletion notices and additional checks in batch mode to ensure that deleted content is removed from all storage systems ("right to be forgotten"). inDICES also uses strong encryption in conjunction with state-of-the-art authentication and security protocols. For API-based services, OAuth 2.0 authentication is provided.

### Content from the Participatory Space

As visualised in [Figure 1](#), content that is created in the participation space on the Decidim platform is converted into documents to be ingested by the Document API. After going through the webLyzard NLP pipeline to analyse the document and determine keywords, sentiment and contained entities, documents are available in the index layer.

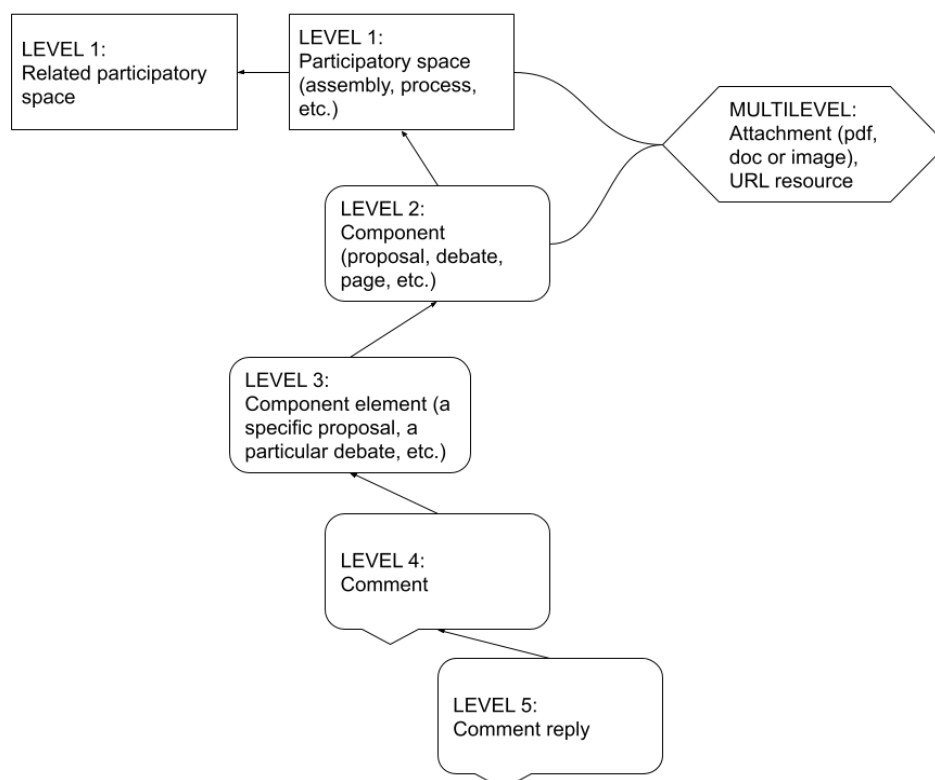
The Decidim platform is a modular system with different forms of participation spaces and components, as outlined in *D4.1 (5.2. Participatory Spaces and Main Features)*.

To accommodate the different interaction and communication options available for users, the following hierarchy is used to model the different Decidim components and reproduce content relations:

1. **First level:** a participatory space, i.e. an assembly, process etc.
2. **Second level:** the communication component, i.e. proposal, debate, page etc.
3. **Third level:** the component element, a specific proposal post, debate post etc.
4. **Fourth + Fifth level:** comments on a component element, nested one depth

The hierarchy levels and their connections are illustrated in [Figure 2](#). For the first content ingestion implementation, the proposals component has been selected, and upon creation individual proposal posts are pushed to the Document API and are stored in the index layer as documents.

The mechanism of ingestion is started by Decidim every time a user publishes or edits a proposal on the platform. Decidim provides a series of "hooks" - i.e., trigger mechanisms that specific instances such as inDICES can use to attach custom behaviours after some actions have been performed by end users. In this case, publishing a proposal triggers a call to the webLyzard API which, in return, answers back the ID of the published document. This ID is saved in a custom log entry for future use.



**Fig 2: Decidim hierarchy level overview**

Each created proposal document consists of a *title*, the main *contribution* as HTML-formatted text and its public *URL*, as well as generic metadata, such as the *timestamp* of creation, the *author* and - if known - the *language* of the proposal. A *category* is assigned based on the first level component (the participatory space). The dedicated relations.sioc:has\_container field of the Document API will be used to declare the URL of the proposal component in order to interconnect the individual proposals for thread-based visualisations (see section *Thread-Based Widgets*).

Regarding GDPR considerations, it is important to note that the content sent to webLyZard should not include personal data of any kind, unless some specific process is designed with the intention of asking the user to create proposals with personal data, but this kind of process is highly unlikely because it will mean that this type of data would be publicly available. Another possibility is that some user, by error, does precisely that and uploads some content that then wants to withdraw. But, even in this case, there are two mechanisms to control the data flow.

On one hand, the integration with the dashboard visualisations can be controlled (and removed) by the author of a proposal or the administrator when configuring the particular proposal component. On the other hand, every document sent to the webLyZard platform is kept in a database table log with its corresponding ID. This can be used to manually remove content, if needed, using the same API used to upload it in the first place. This mechanism can also serve the purpose of removing content automatically if the user decides to withdraw a proposal or in the event of admins blocking content that might be considered spam.

## 2.3 Contextualisation Services

The inDICES dashboard and processing components are continuously developed within the project and customised to the project needs. In the past months, further innovations have been initiated to provide improved contextualisation for the inDICES use cases by:

1. enlarging the set of *named entities* that are recognised in text to include cultural institutions. The initial entities and their metadata are extracted from Wikidata based on their categorisation as *GLAM* (galleries, libraries, archives and museums). The new entities will be part of a *CHI Knowledge Graph* that will store sites and organisations and associated metadata of relevance to the CH sector.
2. extending the NLP pipeline, i.e. POS (*part-of-speech*) tagging, keyword corpora, translations, stop words and sentiment analysis to Italian.
3. configuring automated source processing in collaboration with CHI use case partners to include *Italian News Outlets* and *Web Sources* in the CCS. As outlined in D1.3, further web sources will be gathered through project partners and an *Open Call for Open Sources*.

During the first reporting period, Platoniq and webLyzard have been working towards the integration goals and proof of concepts outlined in Section 7 of *D4.1 User Scenarios and Wireframes Report*, especially focusing on contextualisation through widget integrations for data driven proposals and debates in the participatory space. [Figure 3](#) shows an image of the first integration, where during the creation of a *Hypothesis Proposal* on the Decidim platform, the short text formulating the main hypothesis is taken to generate a visualisation widget, served via the webLyzard Visualisation API, based on similar documents from the context and metadata repository.

## 2.4 CHI Knowledge Graph

Next to the collected websites and social media accounts, e.g. Twitter handles and Facebook pages that have been activated for the inDICES dashboard (ref. *Configuration Template* introduced in D1.1), a CHI knowledge graph will be built. This KG will store curated metadata on CH institutions and their online channels.

The CHI Knowledge Graph will be part of the webLyzard Semantic Knowledge Base (SKB) that already stores roughly a million named entities, such as persons, organisations, locations and events. Metadata of those named entities is sourced from LOD (*Linked Open Data*) sources, such as *Wikidata*, *GeoNames* and *OpenStreetMap*, as well as fed from an in-house processing pipeline of third-party sources for events.

The screenshot displays the inDICES web interface. At the top, there is a navigation bar with links for Home, Processes, Assemblies, Calendars, Help, and Blog. A search bar and 'Sign Up / Sign In' buttons are also present. The main header area features the title 'Designing a Model for Community Participation' and a sub-header 'Building Community and Community Models'. Below this, a navigation menu includes 'THE PROCESS', 'HYPOTHESIS', 'PERSONAS', and 'GOVERNANCE'. The main content area shows a hypothesis proposal titled 'Gender Inequality in Cultural Heritage Institutions' by Nadia Nadesan, dated 15/04/2021. The hypothesis text states: 'There is a gender gap in terms gender equal leadership cultural heritage organizations and institutions.' It also lists 'Gaps in Knowledge', 'Resources needed', and '2 data sources/ sets:'. An 'AUTOMATED DATA CONTENT ANALYSIS' widget is integrated into the proposal, showing a world map with sentiment analysis color coding (Negative vs Positive) and a 'Geomap' tab. The widget also includes a 'Tag Cloud' and a 'Keyword Graph' tab. A 'Share' button and a 'Follow' button are visible on the right side of the hypothesis text. At the bottom, there is an 'ADD YOUR COMMENT' section with a sign-in prompt.

**Fig 3: Similarity-based visualisation widget integrated into a hypothesis proposal**

The dataset itself is expressed as RDF (*Resource Description Framework*) triples following the *W3C Semantic Web Standard*<sup>1</sup> and stored in a Apache Jena Fuseki<sup>2</sup> RDF triple store. The SKB drives the Natural Language Processing tasks, e.g. Named Entity Recognition, behind the inDICES dashboard and provides the metadata in the dashboard entities view.

<sup>1</sup> <https://www.w3.org/TR/rdf11-concepts/>

<sup>2</sup> <https://jena.apache.org/documentation/fuseki2/>

Through the CHI Knowledge Graph we want to leverage essential information on CH organisations that project partners can provide or will gather and curate during the project, and make a new sector of named entities available for dashboard search and tracking. In addition to the generic metadata available on Wikidata and social media accounts, the CHI Knowledge Graph will also support custom metrics provided by inDICES partners.

For that, a pilot has been started in which candidate Wikidata types, shown in [Figure 4](#), based on the GLAM sector have been defined. Using the Netherlands as a pilot case, matching entities for the sector have been identified via a SPARQL query issued to the Wikidata API. Those entities and associated metadata, such as a *labels*, *description*, *inception*, *official website* and *social media information* have been converted into a spreadsheet for easy shareability and accessibility (see [Figure 5](#)). The spreadsheets will be refined and extended with websites and social media channels, through manual curation by expert partners. Based on the experience gained in the pilot, the spreadsheet creation process will be automated on a per country basis.

General classes of interest (all subclasses are included)		Specific classes of interest		Specifically excluded classes	
museum	wd:Q33506	memory institution	wd:Q1497649	exhibition	wd:Q464980
archives	wd:Q166118	heritage center	wd:Q5739135	arboretum	wd:Q272231
library	wd:Q7075	historical society	wd:Q5774403	park	wd:Q22698
		art collection	wd:Q7328910	bus stop (prevents public library bus stop)	wd:Q953806
		cultural institution	wd:Q3152824		
		cultural property	wd:Q2065736		
		library association	wd:Q856564		

**Fig 4: Wikidata classification applied for the NL Pilot**

In addition to the spreadsheets initially populated from Wikidata, existing data from externally conducted surveys, such as the *Survey of GLAM open access policy and practice* (dataset collected by *Douglas McCarthy and Dr. Andrea Wallace*) will be integrated and interconnected. Furthermore, the *Digital Heritage (re)Use Cases* collected in WP3 for the task *Policy analysis of value chains for CHIs in the Digital Single Market* contain data on digital heritage initiatives and projects that will be integrated into the CHI Knowledge Graph. [Figure 6](#) illustrates the different resources with example metadata that will flow into the knowledge graph.

InDICES Pilot Netherlands

File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

100% \$ % .0 .00 123

8 B I S A

http://www.wikidata.org/entity/Q2470853

1	A	B	C	D	E	F
uri	name	alternate name	description (en)	descriptions (nl)	class / category	
20	<a href="http://www.wikidata.org/entity/Q22809741">http://www.wikidata.org/entity/Q22809741</a>	Accordeonmuseum De Muse		accordion museum	museum in Heumen, Nederland	museum
21	<a href="http://www.wikidata.org/entity/Q28150595">http://www.wikidata.org/entity/Q28150595</a>	Accordeonmuseum Harte Meijer			museum in Aa en Hunze, Nederland	museum, musical instrument museum
22	<a href="http://www.wikidata.org/entity/Q59962286">http://www.wikidata.org/entity/Q59962286</a>	Achterhoekse Poort Public Library Foundation		Public library organisation in the municipalities of Aalten and Oude IJsselstreek, The Netherlands	Bibliotheekorganisatie in de gemeenten Aalten en Oude IJsselstreek, Nederland	public library
23	<a href="http://www.wikidata.org/entity/Q63385185">http://www.wikidata.org/entity/Q63385185</a>	Achternveld public library		Public library in the village of Achternveld, municipality of Leusden, The Netherlands	Openbare bibliotheekvestiging Achternveld, gemeente Leusden	library branch, public library
24	<a href="http://www.wikidata.org/entity/Q1863317">http://www.wikidata.org/entity/Q1863317</a>	Admiraliteitshuis, Dokkum	rijksmonumentnummer 13144, Museum Dokkum	museum in Dokkum, the Netherlands	museum in Friesland	museum
25	<a href="http://www.wikidata.org/entity/Q60020855">http://www.wikidata.org/entity/Q60020855</a>	Aduard public library		Public library in Aduard, province of Groningen, The Netherlands	Openbare bibliotheekvestiging in Aduard, Groningen	library branch, public library
26	<a href="http://www.wikidata.org/entity/Q2053602">http://www.wikidata.org/entity/Q2053602</a>	Affichemuseum			museum in Noord-Holland	museum
27	<a href="http://www.wikidata.org/entity/Q100342006">http://www.wikidata.org/entity/Q100342006</a>	Afghanistan, een nieuw begin		exhibition in 1989	Tentoonstelling uit 1989	exhibition
28	<a href="http://www.wikidata.org/entity/Q2470853">http://www.wikidata.org/entity/Q2470853</a>	Africa Museum	Nationaal Museum van Wereldculturen locatie Afrikamuseum, Afrikamuseum, Nationaal Museum van Wereldculturen locatie Afrikamuseum	museum in Berg en Dal, the Netherlands	museum in Gelderland	museum
29	<a href="http://www.wikidata.org/entity/Q62518728">http://www.wikidata.org/entity/Q62518728</a>	African Studies Centre Library	ASCL library	Library of the African Studies Centre Leiden	Bibliotheek in Leiden, Nederland	library
30	<a href="http://www.wikidata.org/entity/Q100342226">http://www.wikidata.org/entity/Q100342226</a>	Afrikaanse kunst in Nederland		exhibition in 1947	Tentoonstelling uit 1947	exhibition
31	<a href="http://www.wikidata.org/entity/Q100341437">http://www.wikidata.org/entity/Q100341437</a>	Afrikaanse werpmessen		exhibition in 2006	Tentoonstelling uit 2006	exhibition
32	<a href="http://www.wikidata.org/entity/Q81498600">http://www.wikidata.org/entity/Q81498600</a>	Afsluitdijk Wadden Center			expositiecentrum	exhibition
33	<a href="http://www.wikidata.org/entity/Q19765103">http://www.wikidata.org/entity/Q19765103</a>	AGA Museum			museum in Emmen, Nederland	museum
34	<a href="http://www.wikidata.org/entity/Q56460705">http://www.wikidata.org/entity/Q56460705</a>	Agarisch Museum Westerherm		museum in Middenbeemster, the Netherlands	museum in Middenbeemster	museum

Explore

Fig 5: Google Spreadsheet with extracted Wikidata metadata for the NL CHI pilot case

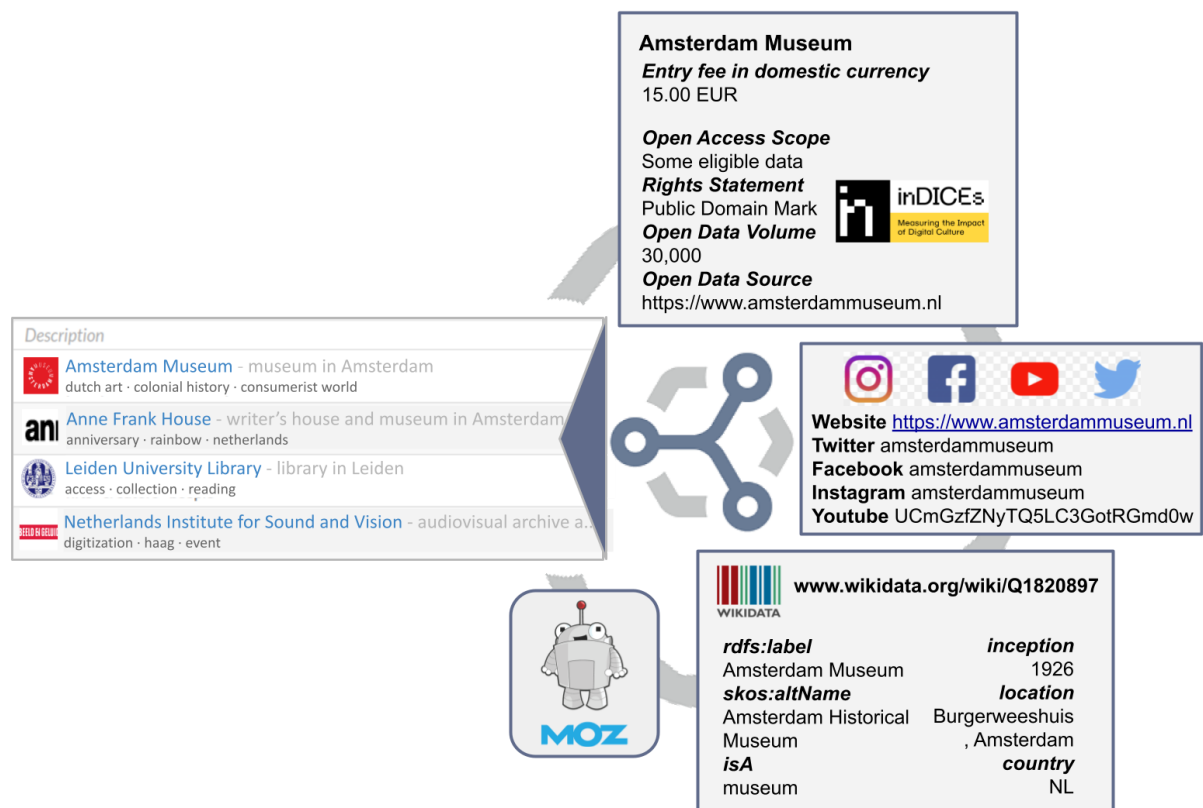
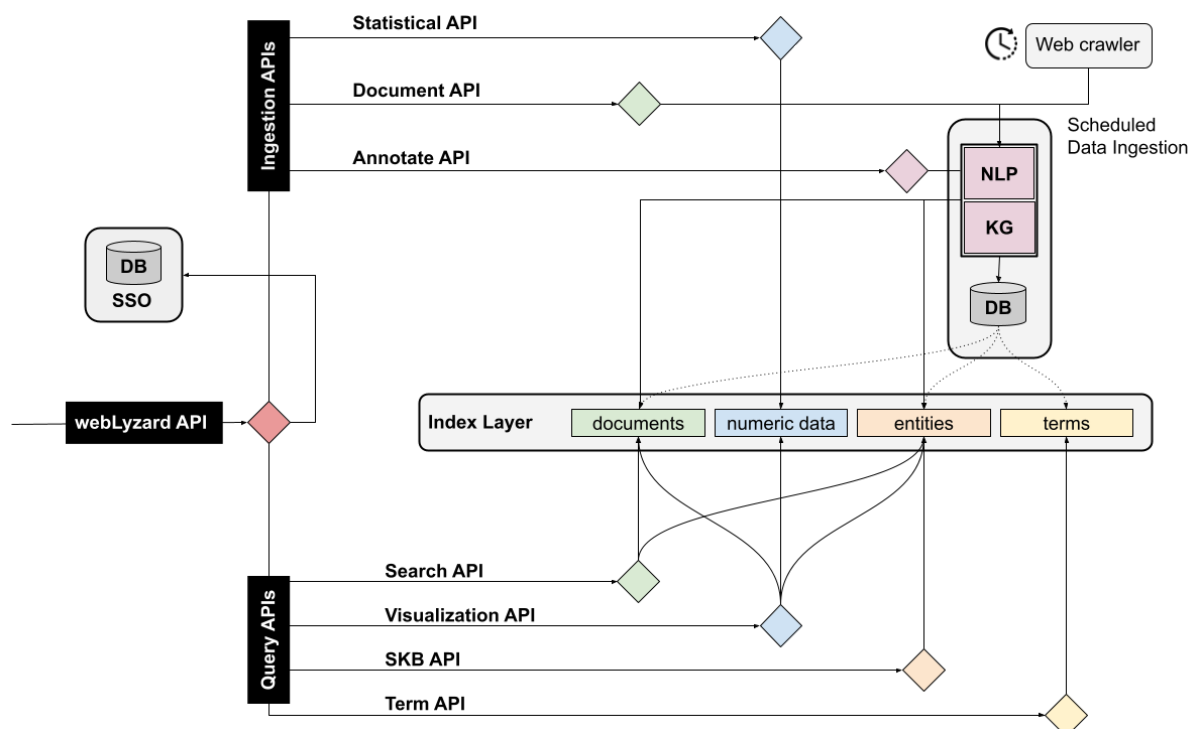


Fig 6: Conceptual overview of different metadata resources for the CHI Knowledge Graph feeding the Visual Analytics Dashboard Entities View

## 2.5 inDICES API Framework

In this chapter, the API specifications are described in more detail. The first part provides an overview of the different webLyzard APIs that are available, and subsequent sections provide detailed specifications with examples describing the individual ingestion and query APIs that are included in the Open Observatory API structure. The last section outlines the Decidim API that provides a comprehensive coverage of all the public content that can be found on the inDICES Participatory Space platform.

### webLyzard API Overview



**Fig 7: Overview of the webLyzard API Framework**

Figure 7 gives an overview of the APIs available to ingest and query data from the *webLyzard* analytics platform. All APIs are encapsulated by a singular endpoint, denoted in the left of Figure 7 by *webLyzard API*, that acts as the central point of user authentication. The API landscape can be divided into APIs that query the *webLyzard Index*, and Ingestion APIs that push new data into the *webLyzard Index*.

### Authentication

All API access to webLyzard APIs requires authorisation via JSON Web Tokens (JWT). An access token is valid across all configured webLyzard APIs, and it expires after eight hours. For the *inDICES* project, we have set up accounts and repositories for consortium partner *Platoniq* to push their content into the content repository, and update or delete individual content items as required. The authentication process is demonstrated in Listings 1 and 2.

## Initialisation

```
#!/bin/bash
REPOSITORY_ID=indices.weblyzard.com/api
USERNAME=api@indices.weblyzard.com
PASSWORD=R3QWqTPxMXip6
SERVER_URL=https://api.weblyzard.com
API_VERSION=1.0
```

**Listing 1:** Initialisation of the authentication process.

## Request an Access Token (JWT)

```
#!/bin/bash
TOKEN=$(curl -s -u $USERNAME:$PASSWORD $SERVER_URL/$API_VERSION/token)
```

**Listing 2:** Requesting (GET) an access token with the credentials provided in Listing 1.

## Document API Specification

The WLT Document API provides a simple HTTP wrapper for data ingestion into the WLT document (text) repository. It allows for standard CRUD functionality, and has been configured and activated for data ingestion from the *Decidim platform*. Upon request the setting for unique URLs was activated for Decidim content, prohibiting the creation of multiple documents that share the same URI. If content changes the respective document can be updated using the HTTP PUT method.

Document data to be pushed to the *webLyzard platform* must conform to the following specification. An example document is presented in [Listing 3](#).

- content, text-based content of the post
- content\_type, text/plain, text/html
- repository\_id, indices.weblyzard.com/api
- uri, required field to be populated with the URL of the post on the participatory platform (permalink)
- published\_date, required field to be populated with the date when post was created
- features.category, optional field for custom search integration in indices.weblyzard.com, here the assembly or process could be used if such a filter is desired
- relations.sioc:reply\_of, optional field containing the immediate parent post of a threaded dialogue, i.e. (i) a parent comment to which the current URL is a reply of, or (ii) the idea to which a top-level comment refers to, or (iii) the consultation of an idea page.
- relations.sioc:has\_container, optional field containing the root URL of a threaded dialogue, i.e. a consultation URL or an idea URL (a debate/proposal). This field determines the root of embedding as used via the webLyzard Visualisation API, so it should be populated in alignment with this use case.



```
{
  "content": "Coronavirus (COVID-19) and cultural and creative sectors: impact,
innovations and planning for post-crisis: a series of discussions followed by targeted
training for policy makers and practitioners. Video recordings posted online.
http://www.oecd.org/cfe/leed/culture-webinars.htm",
  "content_type": "text/plain",
  "repository_id": "indices.weblyzard.com/api",
  "uri":
"https://participate.indices-culture.eu/assemblies/COVID19/f/12/debates/2?commentId=1",
  "title": "Inspirations",
  "meta_data": {
    "published_date": "2020-04-22T12:00:00Z",
    "user_name": "Nadia Nadesan",
    "language_id": "en"
  },
  "features": {
    "category": "COVID-19"
  },
  "relations": {
    "sioc:reply_of":
"https://participate.indices-culture.eu/assemblies/COVID19/f/12/debates/2",
    "sioc:has_container":
"https://participate.indices-culture.eu/assemblies/COVID19/f/12/"
  }
}
```

**Listing 3:** A sample JSON document from the *Decidim* participation platform to be ingested by the WLT Document API.

Sample *cURL* requests for the Document API are shown in [Listings 4-7](#).

```
curl -H "Authorization: Bearer $TOKEN" -H "Content-Type: application/json" -X POST
-d @document.json https://api.weblyzard.com/1.0/documents/indices.weblyzard.com/api
```

**Listing 4:** Push (POST) a JSON document to the Metadata Repository.

```
curl -H "Authorization: Bearer $TOKEN" -X GET
https://api.weblyzard.com/1.0/documents/indices.weblyzard.com/api/44620231267414243
```

**Listing 5:** Retrieve (GET) a JSON document to the Metadata Repository.

```
curl -H "Authorization: Bearer $TOKEN" -H "Content-Type: application/json" -X PUT
--data @new_document.json
https://api.weblyzard.com/1.0/documents/indices.weblyzard.com/api/44620231267414243
```

**Listing 6:** Update (PUT) a JSON document to the Metadata Repository.

```
curl -H "Authorization: Bearer $TOKEN" -X DELETE
https://api.weblyzard.com/1.0/documents/indices.weblyzard.com/api/44620231267414243
```

**Listing 7:** Remove (DELETE) a JSON document to the Metadata Repository.

## Statistical API Specification

The WLT Statistical API provides a simple HTTP wrapper for data ingestion into the WLT numeric (structured) repository. It allows for standard CRUD functionality. It is planned to be configured and activated for data ingestion from the *Decidim platform* in the next steps of the project to share numeric and statistical data if requested.

Numerical data to be pushed to the *webLyzard platform* must conform to the following specification. An example document is presented in [Listing 8](#).

```
{
  "uri": "https://participate.indices-culture.eu/data/9245003270420487",
  "date": "2021-04-27T13:30:00",
  "indicator_id": "chi_distribution_2021",
  "value": "2110",
  "source_country": "IT",
  "source_type": "POI",
  "target_country": "IT",
  "target_type": "POI",
  "target_location": [
    {
      "name": "Rome",
      "point": {
        "lat": 47.0722,
        "lon": 15.417
      }
    }
  ],
  "metadata": {
    "institution": "Roma Natural History Museum"
  }
}
```

**Listing 8:** A sample JSON numerical observation to be ingested by the WLT Statistical API.

Sample cURL requests for the Statistical API are shown in Listings 9-12.

```
curl -H "Authorization: Bearer $TOKEN" -H "Content-Type: application/json" -X POST
-d @data.json https://api.weiblyzard.com/1.0/observations/indices.weiblyzard.com/statistics/user_repo01
```

**Listing 9:** Push (POST) a JSON observation to the repository/dashboard.

The above request returns an ID hash that can be used for the subsequent update/get/delete operations. For a sample data.json observation, see Listing 8 above. The POST method also accepts a list of such observations (batch mode) for convenience.

```
curl -H "Authorization: Bearer $TOKEN" -H "Content-Type: application/json" -X PUT
--data @updated_data.json https://api.weiblyzard.com/1.0/observations/indices.weiblyzard.com/statistics/user_repo01/315a40d1-b5a1-11e8-9dfe-0242ac110009
```

**Listing 10:** Update (PUT) a JSON observation in the repository/dashboard.

```
curl -H "Authorization: Bearer $TOKEN" -X GET
https://api.weiblyzard.com/1.0/observations/indices.weiblyzard.com/statistics/user_repo01/315a40d1-b5a1-11e8-9dfe-0242ac110009
```

**Listing 11:** Retrieve (GET) a JSON observation from the repository/dashboard.

```
curl -H "Authorization: Bearer $TOKEN" -X DELETE
https://api.weiblyzard.com/1.0/observations/indices.weiblyzard.com/statistics/user_repo01/315a40d1-b5a1-11e8-9dfe-0242ac110009
```

**Listing 12:** Remove (DELETE) a JSON observation from the repository/dashboard.

## Search API Specification

The WLT Search API has been made available for the inDICES project so that project partners have access to documents and their respective metadata (sentiment, geo location, keywords).

A sample *cURL* request to retrieve documents is shown in Listing 13 and Listing 14. Additionally, it is possible to provide a *query* or *filter* utilizing the Elasticsearch query language, to further restrict the search.

```
{
  "sources": [
    "indices.weblyzard.com/news_en",
    "indices.weblyzard.com/web_en",
    "indices.weblyzard.com/decidim_en"
  ],
  "count":1000,
  "beginDate": "2021-01-01",
  "endDate": "2021-01-07"
}
```

**Listing 13:** A sample JSON to retrieve documents from the Search API.

```
curl -H 'Authorization: Bearer <TOKEN>' -H 'Content-Type: application/json' -X POST
-d @search.json
https://api.weblyzard.com/1.0/search
```

**Listing 14:** Send (POST) a JSON search query to retrieve enriched documents from the repository.

The default response returns the specified number ("*count*") of document result hits, where each hit consists of the date, the document title and respective URL. In a separate field the total number of documents that match the search request is returned.

Additional document fields can be requested, which include a short summary and text snippet, the full document text, identified keywords, sentiment, location information, the associated domain as well as additional metadata and metrics.

Listing 15 and Listing 16 show exemplary response snippets for the default and extended response, respectively.

```
{
  "result": {
    "hits": [
      {
        "date": 1609632000000,
        "title": "The Best Christmas Markets To Visit In London",
        "url":
"https://sheerluxe.com/london/best-christmas-markets-to-visit-in-london"
      },
      {
        "date": 1609632000000,
```

```

        "title": "Introducing Quarantune, the blog celebrating pre-COVID
dancefloor memories",
        "url":
"https://www.dazeddigital.com/music/article/51558/1/introducing-quarantune-the-blog-celebrating-pre-covid-dancefloor-memories"
    },
    {
        "date": 1609632000000,
        "title": "Wham's 'Last Christmas' Completes 36-Year Journey to U.K.
Chart Summit",
        "url":
"https://www.billboard.com/articles/news/9506360/wham-last-christmas-36-year-journey-uk-no-1/"
    }
],
"total": 26158
}
}

```

**Listing 15:** A sample JSON result from the Search API.

```

{
  "date": 1609545300000,
  "contentid": 4951546561589178826,
  "title": "How Omoregie Osakporlor is using photography to preserve culture and
recreate stories of the average Nigerian - Ventures Africa",
  "url":
"https://venturesafrica.com/how-omoregie-osakporlor-is-using-photography-to-preserve-culture-and-recreate-stories-of-the-average-nigerian/",
  "domain": "venturesafrica.com",
  "sentiment": 0.3011693060398102,
  "targetlocation": {
    "name": "Abuja",
    "longitude": 7.49508,
    "latitude": 9.05785
  },
  "sourceid": 31414,
  "keywords": [
    {
      "key": "http://sws.geonames.org/2332459/",
      "name": {
        "de": "Lagos",
        "en": "Lagos",
        "fr": "Lagos",
        "es": "Lagos"
      }
    },
    {
      "key": "http://weblyzard.com/skb/keyword/en/noun/rights_activism",
      "name": {
        "de": "Rechtsaktivismus",

```

```

        "en": "rights activism",
        "fr": "activisme des droits",
        "es": "activismo de derechos"
    }
}
],
  "fulltext": "This article is the third in our '2020: Also the year for innovation'
series . \"What motivated me into photography was, as a teenager, I always believed that
I was going to change the world. I've always loved social activism because I was
introduced to some of the writings of great African crusaders like Wole Soyinka, Nelson
Mandela at a very early age. [...] The protests have lasted for 6 days with young
Nigerians asking that the government dissolve the Special Anti-Robbery Squad (SARS) over
brutality and violation of human rights. Here is the link to a trailer of a film that
documents the plight of pensioners in Nigeria.\"
}

```

**Listing 16:** A single document JSON result from the Search API (full text shortened).

## Decidim API Specification

Decidim<sup>3</sup> provides an API that follows the GraphQL<sup>4</sup> specification. GraphQL is a technology developed originally by the company Facebook in 2012 and liberated with an open source license in 2015. The main difference with respect to other API standards (such as the REST approach used by the webLyzard API) is that it requires only one URL endpoint in order to fetch any resource.

The Decidim API provides a read-only resource and has a comprehensive coverage of all the public content that can be found on the *Participatory Space*. In particular, all the public participation carried out in the platform is available through its integrated API along with documentation on how to use it. The inDICES Decidim instance provides the following URLs dedicated to the API:

- [participate.indices-culture.eu/api](https://participate.indices-culture.eu/api)  
The route where to make requests. Requests need to be done using the POST http verb, otherwise the request will be redirected to the API built-in documentation.
- [participate.indices-culture.eu/api/docs](https://participate.indices-culture.eu/api/docs)  
The built-in documentation for the API. This facilitates anyone to quickly use it and provides a full description of all the resources available.
- [participate.indices-culture.eu/api/graphiql](https://participate.indices-culture.eu/api/graphiql)  
This endpoint serves the application GraphiQL<sup>5</sup>, an in-browser IDE for exploring GraphQL APIs. This interface can be used by anyone to test or play with the data and quickly obtain relevant results.

<sup>3</sup> <https://github.com/decidim/decidim>

<sup>4</sup> <https://graphql.org/>

<sup>5</sup> <https://github.com/graphql/graphiql>

The GraphQL format is a JSON formatted text that is specified in a HTTP query. The response is a JSON object as well. This deliverable does not cover all the specifications of GraphQL because there is nothing particular in the Decidim implementation. For full details about specification, the reader is advised to refer to the official GraphQL site<sup>6</sup>.

The most important characteristic of GraphQL is the way it lets the user request data. Each transaction is defined in a single query that defines the content that needs to be retrieved and only that content. This approach relieves the server of the burden of having to fetch and transmit data that is not needed by the client. It also permits the user to create nested queries in order to fetch all necessary data in a single transaction.

For instance, the version of the inDICES Decidim installation can be checked by using the utility *cURL* in a bash console as shown in Listing 17. Note that Content-Type needs to be specified.

```
curl -sSH "Content-Type: application/json" -d '{"query": "{ decidim { version } }"}' \
https://participate.indices-culture.eu/api/
```

**Listing 17:** Retrieve the version of the current Decidim installation with a *cURL* request.

The query can also be used in GraphiQL ([participate.indices-culture.eu/api/graphiql](https://participate.indices-culture.eu/api/graphiql)), in that case the "query" text can be skipped (see Listing 18). The response in both cases will be formatted as JSON and looks like illustrated in Listing 19.

```
{
  decidim {
    version
  }
}
```

**Listing 18:** Retrieve the version of the current Decidim installation using *GraphiQL*.

```
{
  "data": {
    "decidim": {
      "version": "0.24.1"
    }
  }
}
```

**Listing 19:** Response when requesting the Decidim version.

Note that the participatory space API does not require authentication of any kind. It is available as a public resource for anyone to use it. Only fair usage limits are in place, limiting the maximum connections per minute to a particular client. However, it is also true that it is a read-only resource and can be only used to obtain a structured, programmatically friendly, version of any content publicly browsable on the web.

<sup>6</sup> <https://graphql.org/learn/>

### 3. Participatory Space

Technology has profoundly changed how museums, artists, and people interact with cultural institutions and culture. However, many technological advances in sharing and communications have been created by and for private interests. While the aim of inDICES does not directly address the lack of public space in tech, inDICES inherently responds to this question using Decidim an open source software, generating shareable data and tools, and having at its core creating open access to resources and spaces for collaboration. The inDICES platform is intended to create spaces for interaction, sharing, and engagement across various institutions and professional silos within the CH landscape.

The participatory space is the entry point for anyone participating in the activities of the inDICES project. It is an online platform that allows users to browse, register, post content and interact with content posted by other users in many different ways. The participatory space is based on Decidim, an open-sourced application created originally by the Barcelona city council to carry out democratic participatory processes. At the time of writing this document, it is probably the most important application of its kind in the world with more than 400 dependencies in GitHub<sup>7</sup> (the majority of which are Decidim instances for cities or organisations of different sorts).

An introduction to Decidim has been given in the D4.1 deliverable, in here we will describe the technical aspects, its emphasis on transparency and traceability and the community model that it proposes.

#### 3.1 Decidim Framework

Decidim is not an application per se (as could be WordPress, for instance) but rather a framework to build applications. Decidim is based on Ruby on Rails<sup>8</sup>, a low-level framework for creating rich, REST based applications, for the web using the ruby<sup>9</sup> language. This approach gives flexibility for personalizing Decidim in very different ways while it facilitates keeping the application synchronised along the successive (and rather frequent) versions published by the main Decidim team. The main technical advantages of Decidim are:

- **It is open source** (as free software). This is no minor feature, since being open source is key to avoid being locked with a particular company or product. It also means that (when having the appropriate knowledge) bugs can easily be fixed and new features can be added when needed.
- **It is modular**. To be modular means that features come in “packs”, and it is the webmaster who decides which of these to install on a particular Decidim instance. Decidim comes with some modules by default, some optional but, more importantly, allows anyone to create third party modules to extend functionality.

---

<sup>7</sup> <https://github.com/decidim/decidim/network/dependents>

<sup>8</sup> <https://rubyonrails.org/>

<sup>9</sup> <https://www.ruby-lang.org/>

- **It is “hackable”**. This concept is mainly inherited from the properties of the Ruby language itself and their mother framework, Ruby on Rails. It means that almost everything can be transformed or overridden to match the desired behaviour. The inDICES project takes advantage of this when overriding the proposal’s default view to incorporate the embedded visualisations from webLyzard.
- **Is actively maintained**: There is a common fear when starting a new project based on an existing open source project, which is that it is abandoned. This would make it difficult to maintain the platform and ensure sustainability. Therefore, it is desirable to be sure that the project has “good” sponsors that ensures maintenance and future versions. Luckily, Decidim is well-funded by the Barcelona City Council and has a very active community with contributions from other cities and organisations.
- **GDPR compliance**: One of the foundational aspects of Decidim is the Social Contract<sup>10</sup>, an internal document that anyone using Decidim (or contributing to it) should adhere to in order to be part of the community. Among other things, it enforces the software to be strictly GDPR compliant, granting the user the right to be forgotten or opt in checkboxes for receiving newsletters for instance.

## 3.2 inDICES Decidim Implementation

The inDICES implementation of Decidim uses some official modules as well as several third-party modules provided by the community. In particular:

- **Decidim built-in modules**:<sup>11</sup> Proposals, Meetings, Debates, Surveys, Pages, Blog and the 2 main participatory spaces, Participatory processes and Assemblies.
- **Decidim optional modules**: Consultations (not being used yet) and Conferences (known as “Calendars” in the platform).
- **Decidim Term Customizer**:<sup>12</sup> This external module has been developed by the city of Helsinki and allows admins to customise any text shown in the platform. Even in different ways depending on the specific page that the user is visiting.
- **Decidim Awesome**:<sup>13</sup> This module has been specifically developed by Platoniq in order to boost or transform some functionalities and also provide some extra features that permit administrators to heavily transform behaviour and design of the platform.
- **Decidim Notify**:<sup>14</sup> This is another module developed by Platoniq specifically for its use in live discussion reports. With this module, an assigned user can transcript a conversation and organise it in chapters which is useful to keep an ordered record of what is being said or just the minutes in general.

---

<sup>10</sup> <https://docs.decidim.org/en/understand/social-contract/>

<sup>11</sup> <https://github.com/decidim/decidim>

<sup>12</sup> [https://github.com/mainio/decidim-module-term\\_customizer](https://github.com/mainio/decidim-module-term_customizer)

<sup>13</sup> [https://github.com/Platoniq/decidim-module-decidim\\_awesome](https://github.com/Platoniq/decidim-module-decidim_awesome)

<sup>14</sup> <https://github.com/Platoniq/decidim-module-notify>



Apart from the installed modules, some code has been added in order to improve and adapt the Decidim's default behaviour to the inDICES needs and, in particular, to achieve the desired integration with the webLyzard dashboard. The specific instance of Decidim used to create the inDICES participatory platform has been optimised in two distinct ways:

- **Design and interface texts:** The look and feel of the original Decidim application has been transformed in such a way that matches the design guides provided by the inDICES project guidelines. This has been achieved by overriding most of the CSS rules that come by default in any Decidim implementation. The other part is the adaptation of the interface default texts (known as “copies”), many of these copies are “hard-coded” in certain localized files that are not available for the administrators of a running Decidim. With a combination of programmatic custom locale files and the use of the Term Customizer plugin, some of these texts have been changed to “speak” the language of the whole project.
- **Improvements or changes to original Decidim functionalities:** This is achieved by adding external plugins and overrides of the original Decidim application and the application of the technique called “monkey patching”, which the ruby language is especially suited for. For example, proposals' views are changed in order to integrate webLyzard visualisations.

### 3.3 Transparency and Traceability

The platform's development and future design is co-designed by a Decidim community that is committed towards ensuring and maximizing transparency and traceability of the various spaces and components to preserve documents, proposals, debates, decisions, or any other mechanism on the participatory platform.

- **Transparency** signifies that all data related to participatory spaces and components are available for download following rigorous standards and formats to share information that take into account accessibility and multi-format downloads.
- **Transparency**, however, does not apply to the sharing or dissemination of personal data or against the interest of the privacy of people and organisations participating in the platform.
- **Traceability** refers to the ability to fully track backward (past) and forward (future) what happened to the proposals, plans, regulations, or any other object of participation or decision recorded on the platform. The platform must at all times show how, why, by whom and with what guarantees an object of a participatory process was dismissed, approved or blocked.

## 4. Visual Analytics Dashboard

The inDICES Visual Analytics Dashboard is an advanced information exploration and retrieval interface that helps to track and analyse emerging stories across sources, regions and languages. It primarily targets researchers and CH professionals who want to conduct in-depth investigation of the data sources and monitor topics of interest to them.

Following a multiple coordinated view approach, the dashboard synchronises various widgets to structure the knowledge repository along multiple metadata dimensions. Different content visualisations, such as the *geographic map*, the *tag cloud*, the *keyword graph* and the *cluster map*, are shown in the right sidebar for the user to explore the underlying knowledge behind the currently active search base along multiple semantic dimensions.

Accessible via all modern Web browsers, an initial version of the inDICES Visual Analytics Dashboard shown in Figure 8 has been launched and made available via authenticated login for project partners at [indices.webyzard.com](http://indices.webyzard.com). All content and metadata items are ingested into a scalable ElasticSearch index that allows for swift queries over millions of documents. The visual tools include the ability to render unstructured textual content together with context information along multiple metadata dimensions, with powerful yet intuitive navigational aids that facilitate the real-time tracking of evolving dialogues. Thereby, inDICES benefits from the storytelling potential of interactive visual tools, including both geospatial and semantic views for browsing community knowledge within a well-structured interface.

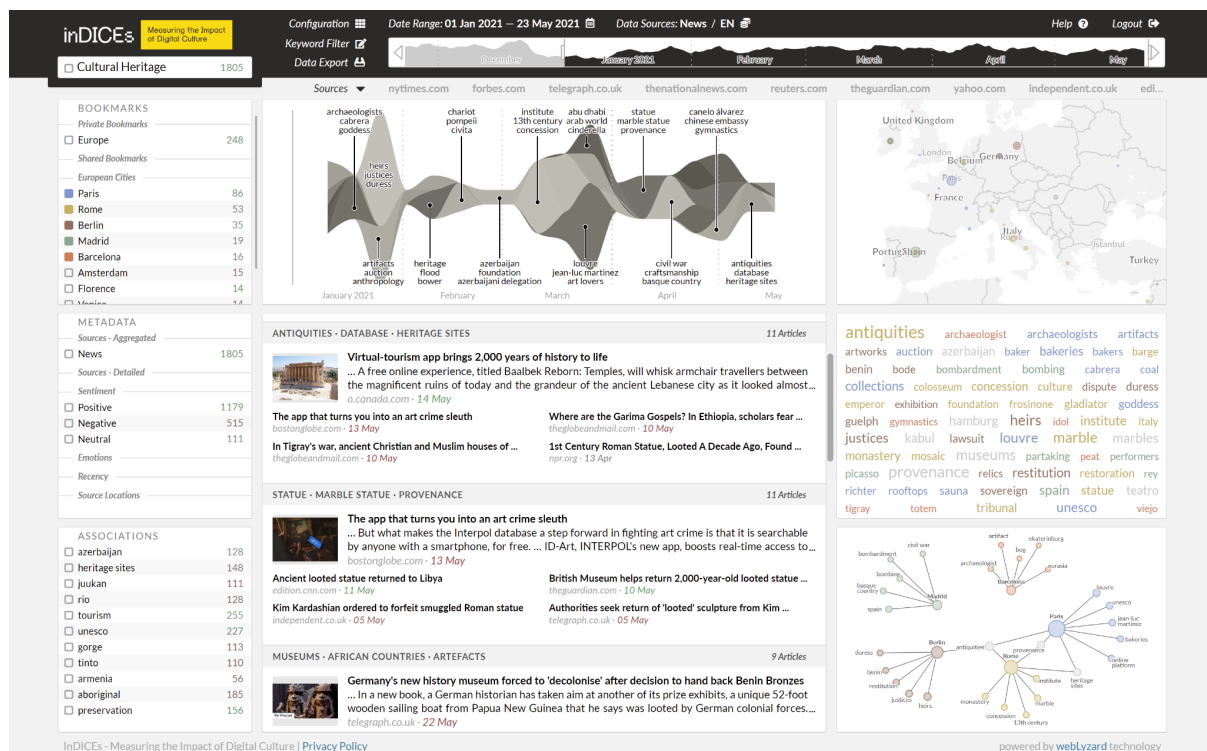


Fig 8: inDICES Visual Analytics Dashboard, showing the results of a search query for articles referencing “cultural heritage” between January and May 2021

These visualisations are also provided as embeddable widgets for the inDICEs Participatory Space accessible through the Visualisation API that is described in the section “Embeddable Widgets” below. The embedding mechanism can be used for various types of integrations: (i) as an individual visualization e.g. during the proposal creation phase, (ii) as a interactive “Dashboard Light” module that allows users to specify simple queries and select the type of visual representation, or (iii) as a predefined combination of multiple visualisations on a dedicated page tailored to a specific user case, curated by project partners.

A dedicated page on the Participatory Platform with the embedded visualisations will target policy makers and CH professionals, for example, to provide an immediate overview of the most relevant trends in the sector (see mockup in Figure 9). The curation of the page will build on the research performed in WP1-WP3, monitoring the impact of digital heritage and relevant activities in the CH sector - e.g. open access, digital transformation, etc. Based on the data displayed in the visualisations, related activities from the Participatory Space could be highlighted on the right side (e.g. proposals that include the same keywords as the latest documents on a selected topic). From here, interested users can also access the fully integrated Visual Analytics Dashboard described above.

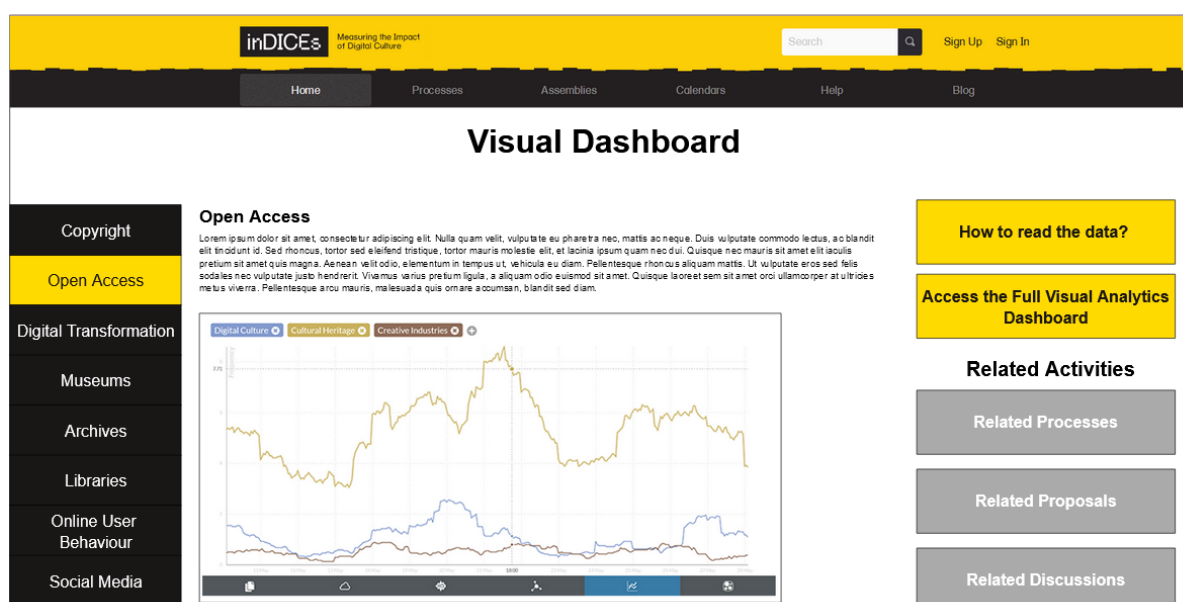


Fig 9: Mockup of the embedded “Dashboard Light” component for the inDICEs Open Observatory

Data literacy is essential for users to interpret the data, understand what is included and excluded from the visualisations and draw actionable conclusions. Hence, inDICEs will offer examples and practical guidance on reading the data.

## 4.1 Dashboard Structure and Configuration

The overall setup and configuration of the dashboard is managed via various header menus at the top of the dashboard that include Configuration, Data Export (which provides access to the PDF generator) and advanced filter options via the Keyword Filter. The Date Range as the interval for analysis can be selected using the interactive timeline or drop-down menu.

Different Data Sources that include news media, social media and the Decidim participatory platform and content language can be changed in the respective menu.

Figure 8 shows a screenshot of the latest dashboard version. The input box in the upper left corner of the portal header represents the currently active global filter. It can be used for full-text search and allows users to explore content and find documents based on their personal search queries. A “\*” indicates that the filter is turned off, “<advanced>” that a complex query is being used.

To trigger a search, one or more checkboxes in the left sidebar can be selected. It is possible to combine checkboxes from different categories including *bookmarks*, *associations* or *metadata attributes*. A preview of the results matching the global filter over a longer time span is provided in the line chart at the top. The info bar below either shows context-aware help texts, or a list of the most frequent sources or named entity references to persons, organisations or locations.

The left sidebar includes the *Bookmarks* section, offering separate categories for private and shared bookmarks. The latter category contains three predefined topics based on the inDICEs fields of interest “cultural heritage”, “creative industries” and “digital culture” as pilot domain definitions. On hover, a gear symbol appears for accessing the overlay menu to view, configure, rename or delete bookmarks, or to define email alerts. Below, the *Metadata* section classifies the coverage by *source*, *sentiment*, *recency*, *emotion* and *source location*. Users can click on the check boxes to select those bookmarks or metadata attributes that should be included in the set of search results. The main content area contains the trend chart to explore longitudinal trends and the story view that displays groups of related document clusters, similar to aggregators such as Google News. For each story, the top three keywords and a lead article including a thumbnail are shown, followed by related articles or postings from other sources.

Complementing this brief summary of the dashboard’s core features, a more extensive and continuously updated description that would be beyond the scope of this deliverable is maintained at [www.weblyzard.com/dashboard](http://www.weblyzard.com/dashboard).

For inDICEs, the data sources that are used in the Web crawling process have been extended to include web sites of museums, libraries, archives, contemporary art institutions and web resources and blogs of cultural heritage practitioners.

## 4.2 Embeddable Widgets

The *inDICEs Visual Analytics Dashboard* offers a feature-rich and customisable solution for visual analytics and semantic search. To support use cases that require a more granular approach (instead of the fully integrated dashboard), the webLyZard Visualisation API enables the integration of distinct portal components into third-party Web applications. For integration with the *Decidim* participatory space, <iframe> tags are used to embed these components. This approach ensures widespread compatibility across platforms.

Visualisations from the *inDICES Dashboard* can be integrated as stand-alone widgets into other HTML-based applications. For the integration with the *Participatory Space*, we have selected the visualisations *Keyword Graph*, *Geographic Map* and *Tag Cloud* as widgets to be displayed as (i) part of the creation process and as (ii) discussion impulse on the Decidim platform. The motivation for this choice is that we wish to support the comment creator in their ideation process during the comment authoring through provisioning of the most relevant context from the surrounding discussion and existing proposals on the platform, and alternatively from current online coverage in the media.

### Template-Based Widgets

All visualisations requested via the Visualisation API initiate a real-time search query on the webLyZard metadata repository and a subsequent server-side rendering of the visualisation by webLyZard's visualisation engine. The resulting visualisation is side-loaded into the external application via HTML embed. This setup requires a search query to be specified a priori which will be used for restricting the document space on which to render the visualisation. Valid filters for this search query are date ranges and term matches, as well as content and language filters. For the first step of integration with the Participatory Space, we have provided template-based widgets on the three topics *digital-culture*, *cultural-heritage* and *creative-industries* that are defined and also available as bookmarks in the inDICES dashboard. Listing 20 shows an example of the query template for the topic *cultural heritage*. Listing 21 provides an example of how to retrieve a Keyword Graph embedding for a specified date range for English news, miscellaneous Web and Decidim for the topic.

```

"query": {
  "bool": {
    "should": [
      {
        "text": {
          "phrase": "cultural heritage"
        }
      }
    ]
  },
  "beginDate": "2021-01-01",
  "endDate": "2021-04-30"
}

```

**Listing 20:** A configuration template for the WLT Visualisation API, describing a topic definition to use as a document basis for the requested visualisation.

```

https://api.indices.weblyzard.com/embed/ayBVrdTaUoabsUQ6fPp9zkxK8WrTV2Fg/keywordgraph/to
pic=cultural-heritage/date=2021-01-01,2021-04-30/source=decidim,news,web/language=en

```

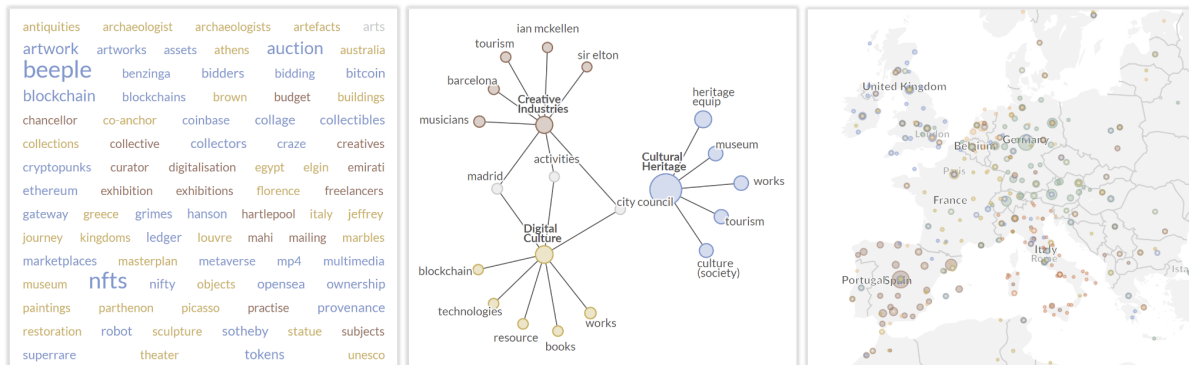
**Listing 21:** Example of a Keyword Graph embedding for a predefined inDICES topic with additional date, content source and content language restrictions.

### Thread-Based Widgets

inDICES extended the Visualisation API of webLyZard by new functions to support thread-based document representations via embeddable widgets. This allows an external application such as the *Open Observatory* platform to provide their users with visual context

on a single discussion, or existing contributions, e.g. proposals of an assembly on the inDICES Participatory Space, and thereby gives more precise insights into the topics of highest relevance during content creation.

As an argument we provide the visualisation engine with the seed URL of the discussion thread in URL-encoded form. The Visualisation Engine queries for all those documents that reference the given seed URL to be their container document.



**Fig 10: Visualisation Examples - Tag Cloud, Keyword Graph and Geographic Map**

```
https://api.indices.weblyzard.com/embed/WuvFyPa0e4yqIi4NTYrguGbTEI/tagcloud/url/thread/<url-encoded-url>
https://api.indices.weblyzard.com/embed/WuvFyPa0e4yqIi4NTYrguGbTEI/keyword/url/thread/<url-encoded-url>
https://api.indices.weblyzard.com/embed/WuvFyPa0e4yqIi4NTYrguGbTEI/geomap/url/thread/<url-encoded-url>
```

**Listings 22-24:** Thread-based embedding, single discussion rooted at *url-encoded-url* (tag cloud, keyword graph, geographic map).

Listings 22-24 give three examples for a Tag Cloud, Keyword Graph and Geographic Map, respectively displayed in Figure 10, as embedding for a specific discussion thread from the *Decidim* Participatory Space.

### Similarity-Based Widgets

On top of the thread-based widgets, we are also providing document similarity as a measure to determine the scope of the context to be used by the Visualisation API. We see the application of document similarity instead of an explicit query or a fixed context thread to provide ambient search features as useful in two inDICES scenarios:

1. Since discussion threads found on collaboration platforms are often restricted in depth, the resulting visualisations would be sparsely populated as well. For this situation, using the similarity-based widget as a fallback allows to still provide ample context even for small discussions.
2. Through similarity search the document space can potentially be extended to cover similar proposals and associated discussions currently conducted on the Participatory Space platform.

### Similarity Requests via Text Snippets

It is possible to request a visualisation widget based on a text snippet or a set of keywords. The contextual scope is then based on a similarity measure calculated on the documents in the content repository. The top 500 most similar documents from the requested context space, i.e. sources and source languages, are used to build the visualisation. Listing 18 shows an example of a Tag Cloud for document keywords similar to the text “*covid-19 vaccination progress in Europe*” for English news content in May 2021. Other visualisations can be generated analogue to Listing 25 where the *topic* selection is replaced with *similar*to.

```
https://api.indices.weblyzard.com/embed/ayBVrdTaUoabsUQ6fPp9zkxK8WrTV2Fg/tagcloud/similar?to=covid-19%20vaccination%20progress%20in%20europe/date=2021-05-01,2021-05-31/source=news/language=en
```

**Listing 25:** Similarity-based Tag Cloud embedding of the top 500 most similar documents to the text “covid-19 vaccination progress in europe” .

### Similarity Requests via URLs

The request for HTML embeds of the Tag Cloud based on document rather than thread-containment is shown in Listing 26. Other visualisations can be generated and embedded similar to Listings 22-24, where the **thread/** part of the path must be replaced with **similar/**, respectively.

```
https://api.indices.weblyzard.com/embed/WuvFyPa0e4yqIi4NTYrguGbTEI/tagcloud/url/similar/<url-encoded-url>
```

**Listing 26:** Similarity-based Tag Cloud embedding of the most similar documents to the seed document with URL *url-encoded-url* (up to 100 documents).

## 5. Summary and Outlook

Deliverable D4.2 summarises the intermediate state of the participatory space platform and the inDICES dashboard, and reports on the individual components of the inDICES API framework. In the first reporting period the development focused on the initial set-up of the individual components and implementation of first integration pilots.

Work in the second reporting period will continue with the integration between the two platforms by finalizing the Single Sign-On functionality to make it possible for the Open Observatory user to directly switch between content discovery and participation, using interactive user controls and thereby interlinking both platforms. The embeddable components will be further developed and extended to improve the contextualisation by offering list views of related documents or metadata resources, such as entities and contributing sources.

To maximise the value of the extracted metadata and the CHI knowledge graph, we plan to develop additional dashboard components - *extended tooltips* to show metadata attributes on CH organisations and their Web sites, for example, or *threaded dialogues* to contextualise discussion threads. The latter will provide insight into the evolution of individual proposals made by users of the inDICES Participatory Space. In terms of impact assessment, traffic statistics for the gathered cultural organisations will be ingested from *MOZ.com*<sup>15</sup>, a Search Engine Optimisation provider. These statistics will serve as an optional sorting criterion to rank matching organisations by impact.

We will also assess which portions of the annotated metadata attributes would represent useful additions to the *Wikidata.org*<sup>16</sup> public knowledge graph. Enriching this graph with inDICES metadata would help to increase visibility of the cultural sector and provide valuable data resources for researchers as well as the linked open data community.

---

<sup>15</sup> <https://moz.com/>

<sup>16</sup> <https://www.wikidata.org/>