Measuring the Impact of Digital Culture

inDICEs

Deliverable 4.3

Database and Service Layers Architecture V2

D4.3 – Database and Service Layers Architecture V2

Version 1.0
31 October 2022

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**
Project Website:	**www.indices-culture.eu**

Authors:	**Katinka Böhm, Arno Scharl - webLyzard technology (WLT)**
Contributing Partners:	**Nadia Nadesan - Platoniq**
Reviewers:	**Rasa Bocyte - Netherlands Institute of Sound & Vision (NISV)**
	Fiona Mowat - Europeana (EF)

Document History

- 01.09.2022 - Document and ToC created from template (WLT)
- 06.09.2022 - Document and ToC shared with contributing partners
- 18.10.2022 - First version ready for internal review (WLT)
- 19.10.2022 - First version shared with reviewers (WLT)
- 28.10.2022 - Deliverable review (NISV, Europeana)
- 28.10.2022 - Second version including review comments (WLT, Platoniq)
- 31.10.2022 - 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*.

Decidim Platform. A digital platform for citizen participation and community engagement; the underlying software framework which is used for the Open Observatory.

Participatory Space. A collaborative environment for dialogue and experience exchange in the form of surveys, debates and collaborative projects, for experts and members of the cultural and creative sectors. It supports a variety of participation modules in which users can open a *Debate* or create a *Proposal.* For more details on the modules see Section 5 in *D4.1 User scenarios and wireframes report*.

Visual Analytics Dashboard, also: inDICEs dashboard. A 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. The LITE version introduces a simpler data view that focuses on the individual visualisations.

CHI Self Assessment Tool (SAT). An inDICEs online tool for the CH sector that can be adapted to various uses and contexts. The *Self Assessment Tool* created in partnership with Europeana's Enumerate survey enables organisations to receive recommendations to improve their digitisation practices based on the outputs and resources 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 CH digitisation and socio-economic impact and a set of associated indicators.

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

1. Introduction

This deliverable provides the second and final report covering the database and service layers. Together, these layers form the *Open Observatory* including (1) the *Participatory Space,* the collaborative environment developed in *Task 4.2*, (2) the *Self Assessment Tool* (SAT) as well as (3) the *Visual Analytics Dashboard* developed in *Task 4.3*. In addition to presenting technical details of the frontend components' final implementation, D4.3 also outlines the inDICEs API Framework that is used for intra- and inter-platform communication.

The following Section 2 summarises the final System Architecture for data ingestion and data sharing between the inDICEs repositories and the services hosted by webLyzard, and the Decidim Participatory Space application of Platoniq. Section 3 then provides updates on the Participatory Space, highlighting the development of the Self Assessment Tool along with the developments of Platoniq to improve participant experience and the administrator tools. Subsequently, Section 4 introduces the new features of the Visual Analytics Dashboard (VAD), including both a PRO and a LITE version, and shows the integration of the latter into the Participatory Space in order to tie all components of the Open Observatory together. The last section summarises the current status and gives an outlook on further ideas and plans that emerged from the collaborative work in inDICEs.

2. System Architecture

The inDICEs Open Observatory is a participatory and monitoring platform that serves open data and methodological tools, developed and collected by the inDICEs project, for the use by different networks and stakeholders. It integrates the following technical components:

1. the Participatory Space, to communicate and exchange information with other users and cultural experts,

2. the Self Assessment Tool (SAT), a tool that institutions can use to assess their degree of digitization and ascertain the areas in which they can still improve, and

3. the inDICEs Visual Analytics Dashboard that is available as either a LITE version for easy exploration, or as a Professional version with advanced analytics features and configurations.

The final application combines all the components under one access point, where registered users can participate in online community activities and discover methodologies and tools to better assess their target audience and plan strategies and next implementation steps. Its main frontend application is hosted by Platoniq. It is based on Decidim and integrates with the webLyzard components and APIs as described in this section.
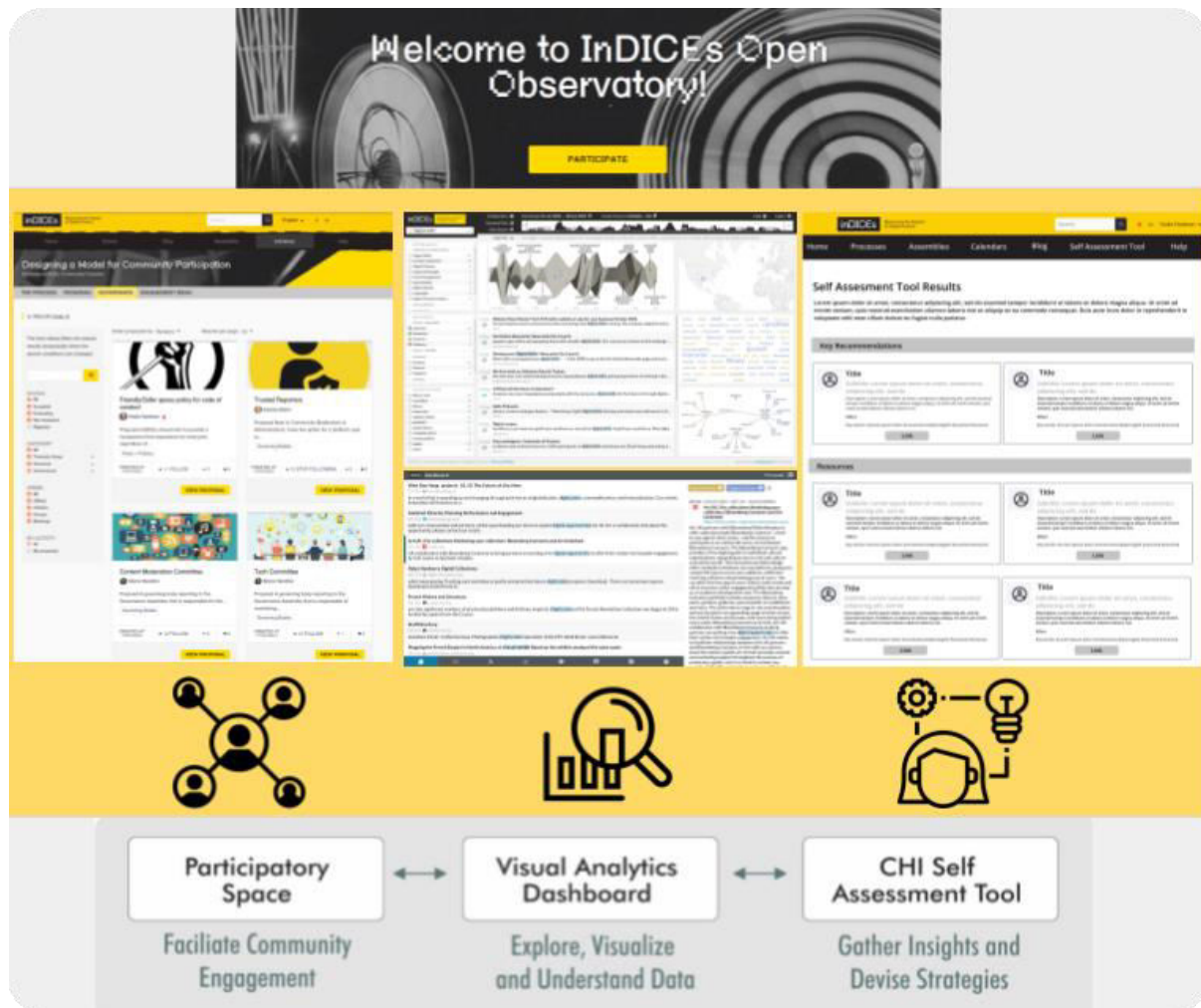
Fig 1: Overview of the three *Open Observatory* components
Icon Image Attribution: Mangsaabguru, Grant Taylor, Phonlaphat Thongsriphong

2.1 Data Infrastructure and Platform Integration

Figure 2 shows the updated visual schematic of the inDICEs system integration. There are two main integrations as information flows from one platform to the other:

(1) The inter-platform communication that assists the Participatory Space user in their proposal creation process with embedded visualisation widgets. In turn, once published on the platform, the content from the created proposal pages is pushed back to the webLyzard API and shapes the content visualisations. Thereby it offers a view of what topics are currently discussed not only on the News and dedicated cultural Websites, but also by others of the platform.

(2) The fully embedded inDICEs Dashboard LITE – a simple yet powerful content exploration tool – which allows the Participatory Space user to search for topics and phrases and visualise different context dimensions from within the Participatory Space. This way they are able to immediately exchange findings and discuss ideas with   other   users.
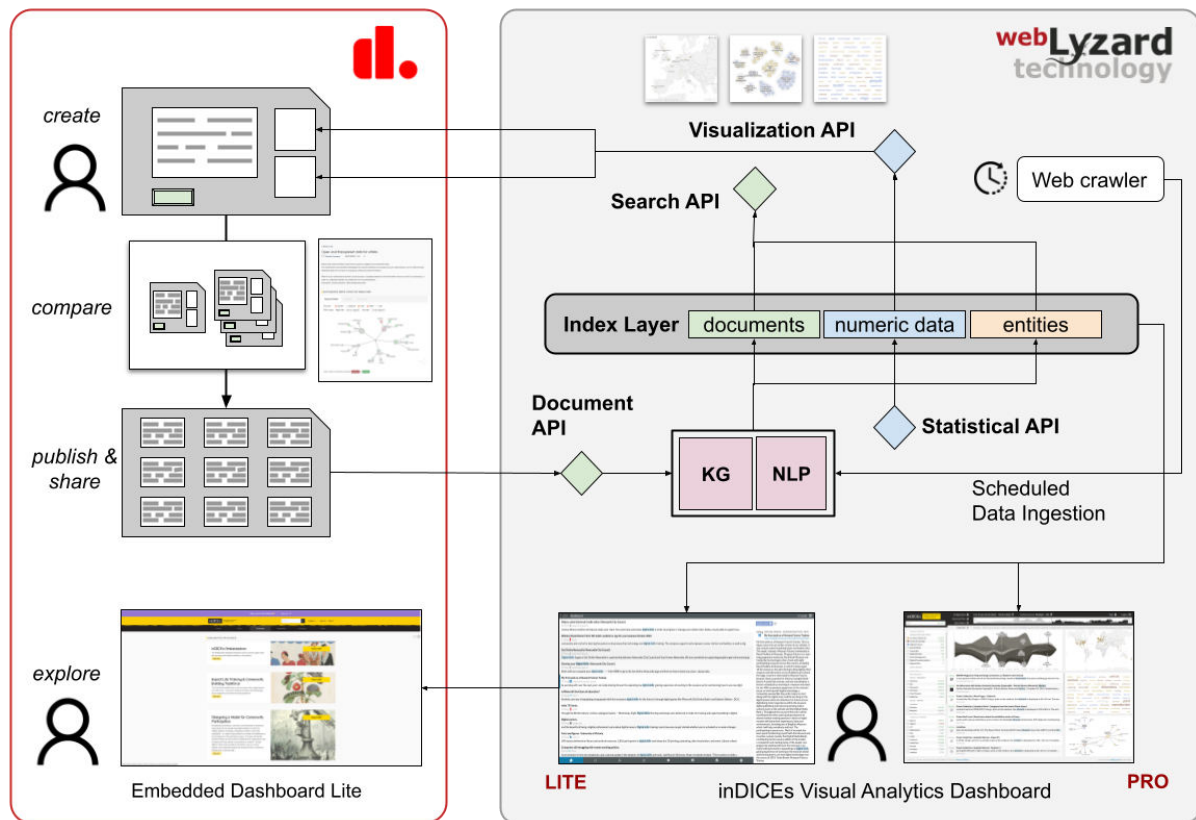
**Fig 2: High-level view of the final system architecture to integrate the
Decidim Participatory Space and the *webLyzard Visual Analytics Platform*.**

Figure 2 illustrates both information flows, showing (1) in the top half of the image and (2) in the bottom half. For (1) visualisation widgets are requested from the Visualisation API and embedded in the proposal page. The seeding keywords that generate the visualisations are based on the proposal title but can also be individually adapted by the proposal creator. In the next step the user can compare their proposal to other relevant ones that were recently created. Once the proposal is published, the content, i.e. proposal title, text and associated metadata are sent to the Document API and thereby ingested by the webLyzard data processing pipeline. After NLP processing (tokenization, POS tagging, dependency parsing) and metadata extraction of keywords, named entities and sentiment, the new document is added to the inDICEs repository and indexed in the Elasticsearch Index Layer. Data for all visualisations is aggregated and visualisations are rendered on demand including any new data that was ingested in the meantime. For (2), a simpler and more approachable version of the inDICEs Visual Analytics Dashboard, the "inDICEs Dashboard LITE" was developed by webLyzard (see Section 4 for a full description) that is available as an embeddable iframe and can be fully integrated into any third party website. The Participatory Space offers a dedicated subpage where users can access and use the embedded dashboard.

The inDICEs Data Repository, initially envisioned as a space to potentially upload and share your own data, was implemented in a reduced fashion in the final version. The Open Observatory hosts a space where users can submit information and link to open datasets that they own or have discovered online. It provides a space to discuss hypotheses around the data with others and make statements on interesting findings and thereby puts emphasis

on sharing, discussing and collaborating data findings. Certain metadata may be provided, such as, who owns the data, how it may be used and when it was collected.

Single Sign-On

To connect both platforms and give users easy access to the inDICEs analytics dashboard, single sign-on authentication was added. It was implemented using the OAuth 2.0 authorisation protocol.

Users that sign up to the Participatory Space can be "officialized" through an additional parameter to also gain access to the full inDICEs analytics dashboard (both PRO and LITE version). Since the Participatory Space of the Open Observatory is completely free for anyone to register, while the analytics dashboard in its full version is only available for selected users within the project scope, this distinction needed to be made to provide a secure workflow for guaranteeing access to the inDICEs analytics dashboard. Which users are officialized can be set by an administrator of the Participatory Space through the Decidim Admin Interface.
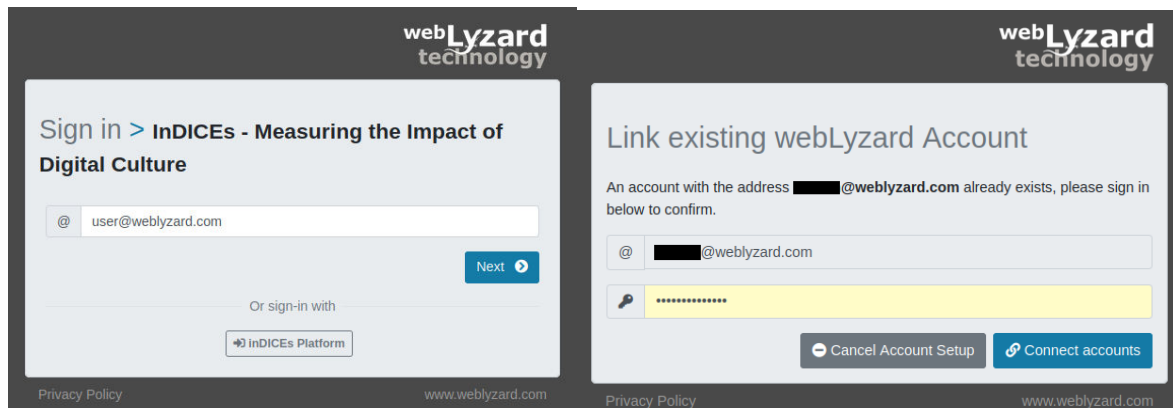

Fig 3: Sign-In of the inDICEs dashboard (left) and prompt to link existing user accounts (right).

As shown in Figure 3, upon accessing indices.weblyzard.com users can "sign-in with inDICEs Platform", which redirects them to the login page of participate.indices-culture.eu. After a successful authentication, the user is redirected to the webLyzard platform and a corresponding user (using metadata passed on from the inDICEs login - first name, last name and email address) is created. If a user account with the same email address (which is not yet linked to an InDICEs login) already exists on the webLyzard platform, the user is prompted for their webLyzard password to link the two accounts. Any officialized user will automatically be given access to the inDICEs dashboard. In the first case, a new user is created in webLyzard's database and granted access if officialized. Any user that is not officialized will have a user account created and linked but not granted permission to access the dashboard.

A general embedded version (see Section 4.2) of the LITE dashboard is available for all users of the Open Observatory and does not need authentication through an individual user account. We foresee that the majority of the inDICEs platform users will find the LITE

version sufficient, with a smaller portion of users (heritage professionals with interest and background in business intelligence and data analytics, researchers) requiring officialization to access the full version. Insights gained there can be shared with others via data exports (pdf reports, csv/xlsx datasets, visualisation images).

2.2 Content Ingestion

Since the first report in M17 on content ingestion, the sources of the inDICEs dashboard and thus the content that is continuously ingested and subsequently processed, stored and indexed, have been extended substantially.

The inDICEs dashboard monitors the following sources:

- Web Sites from the cultural and creative sector provided by project partners and collected in WP1. This includes websites of GLAMs all over Europe as well as sites focused on creative industries, including arts, music, theatre and film.

- Web Sites from GLAMs and cultural institutions that were extracted from Wikidata from the following countries:

  - The Netherlands
  - The United Kingdom
  - Germany
  - France
  - Austria
  - Italy
  - Spain

- News Websites that cover a wide range of countries, including Germany, Austria, U.K., U.S., Canada, Australia, France, the Netherlands, Spain and Italy.

- Other Miscellaneous Web Sites which includes large content publishers from the focus countries (see previous point).

- Twitter and YouTube accounts from the cultural sector provided by project partners, this includes the EuropeanaPro Twitter account.

- Tweets discussing cultural topics that mention any keyword from a set of pre-defined search terms from the cultural and creative sector as defined in WP1.

As mentioned in D4.2, we respect all limitations website owners set for crawling of their content and always try to minimise the load on third-party servers. For social media content, only the official APIs are used and their ToS strictly respected.

Since the inDICEs dashboard is multilingual, the processed content can be in any of the following languages: English, German, Dutch, French, Spanish and Italian.

The inDICEs document repository now grows at a stable rate of approximately 3.2-3.5 million documents per month, of which between 40-70k are continuously gathered from the dedicated sources of the CCS. The initially reported number of 20 million total documents stored per month was primarily due to the Twitter social media content. Since then there have been a few clean-up steps and reconfigurations performed which removed all documents of Retweets and any Tweets with less than 100 followers from the repository. The more regular updates reported in *D1.3-D1.6 Report on Data Gathering*, reflect the

continuous change during the duration of the project. The higher fluctuation from the CCS sources reflects times where websites are publishing more or less content, which for these sources is not as continuous as for News outlets. The additional country-specific GLAM websites that were automatically extracted have been active since September 2022. Numbers of the newly gathered content will be included in *D1.6 Report on data gathering v4,* since it is still too early to report reliable quantities for this content yet.

Update on Web and Social Media Sources

As previously reported, all the Websites and social media accounts from Twitter and Youtube that were provided by WP1 in a joint effort of all project partners were added to the WLT collection of sources. All this content can be searched through, analysed and visualised with the inDICEs dashboard.

Additionally, new Website URLs from GLAMs were semi-automatically extracted from Wikidata. For this the metadata information of the extracted GLAM entities that build the CHI knowledge graph (see Section 2.4) was utilised and the URLs were cleaned up and added to the Web Sites that are crawled for new content on a daily basis.

Table 1 gives an overview of the number of entities extracted and the number of websites that were added to the web crawling configuration per country. The Netherland's pilot is an outlier, since the sources were manually reviewed by NISV before adding them to the sample. In a wide-ranging effort of WLT to drastically increase the Web Sources and in the attempt to limit manual work, a reverse approach was chosen for the other countries; large samples will be initially added and then adjusted and revised on a case-to-case basis should the content not be fitting.

Table 1: New Culture Source Websites per Country.

Country	Culture Entities	Websites
Netherlands (Pilot)	2193 (78 selected)	78
Germany	4209	2328
Austria	4994	425
United Kingdom	2054	1352
Italy	3283	2410
France	1389	895
Spain	2325	1410

Next to increasing the sources of culture related content gathered from GLAM websites, the Italian News sample has also been extended. For this a similar approach – based on Wikidata – was used to extract URLs of Italian newspapers and online news publishers. Any content that is published, including culture and heritage related news, is ingested on a regular basis and stored in the WLT News repositories. Given the representation of Italian partners in the project this sample has also been made available through the inDICEs dashboard. It now consists of content from 73 websites (increased from the previous 22 hand-curated ones) and includes, for example, *RaiNews, Leggo, Avvenire* or *Tiscali.*

The *Open Call for Open Sources*[1] that was launched on the Participatory Space in the second year of the project, also brought a number of new websites and continues to be accessible for contributions from Participatory Space users.

2.3 Contextualisation Services

This subsection summarises the work that has been done to improve the inDICEs Dashboard LITE and PRO with respect to contextualisation and customization of content to the cultural domain. The following new features have been implemented:

1. The set of named entities in the inDICEs knowledge graph has been extended based on insights gained from the NL Pilot case reported in D4.2. For this a new entity type *POIEntity* has been piloted in the KG. By extending the named entities with items from the cultural domain, i.e. museums, galleries, cultural heritage sites etc., more relevant entities are detected in texts and can be surfaced and shown in the dashboard visualisations making it easier to track events and stories that cover cultural points of interest.

2. The collection of Web Site sources where content is gathered from has been significantly enlarged, both with input from stakeholders and project partners, as well as through automatic extraction of information on European CCS organisations from Wikidata. With the larger set of sources the content that is continuously gathered by inDICEs covers a wide range of large European institutions as well as selected, smaller content creators, to better reflect different interest groups and increase content variety.

3. With the tagging feature that was already reported in *D1.5 Report on Data Gathering v3,* sources are assigned multiple matching tags that are used for content filtering for both the APIs and the dashboard. This allows inDICEs dashboard users to quickly narrow down their search to specific content areas, to e.g. view only content published by museums, art websites, or GLAM institutions.

4. The Italian NLP pipeline has been finalised, supporting POS (part-of-speech) tagging, keyword detection, NEL (named-entity linkage) and sentiment analysis for all processed Italian documents, which now also includes a larger set of Italian online News outlets. As a completely new extension, Italian news and stakeholder content can now be viewed and analysed with the inDICEs dashboard, supported by the cross-lingual search functionalities.

5. News documents can be filtered by their IPTC News Media Topic Classification[2]. For inDICEs a new IPTC class model has been in development that not only classifies the first level (17 classes), but also includes classes from the second layer of the "*arts, culture, entertainment and media*" category. This will not only improve the options to filter articles from the News corpus according to their main topic, but make

---

[1] participate.indices-culture.eu/conferences/secondconsultationworkshop/f/66/

[2] cv.iptc.org/newscodes/mediatopic

it easier to quickly filter for content covering the different cultural and creative sectors.

6. Dashboard search filters in the form of bookmarks, have been further developed with feedback from various project partners. Three dedicated online sessions were held in which small groups got a brief introduction to the bookmark configurations and could give input on what information they would like to see captured. These bookmarks will support the visualisations that illustrate the inDICEs perspectives into the data, covering and highlighting areas from social, as well as cultural viewpoints.

7. Traffic statistics for the gathered cultural websites are ingested from *MOZ.com*[3], a Search Engine Optimisation provider. These statistics are visible for an optional sorting criterion to rank content sources by impact. This makes it possible for cultural practitioners and stakeholders to understand which content providers reach larger audiences and compare the impact of coverage.

During the second reporting period the main focus of Platoniq and webLyzard has been the integration of the Visual Analytics Dashboard into the Open Observatory and making it accessible through the Participatory Space Decidim website. For this the "LITE" version of the inDICEs dashboard was developed by partner webLyzard, which is easier to navigate and provides straightforward insights into trends for cultural practitioners and policy makers alike. Built on the first mockup shown in D4.2 the period since the first version of this report has been used to further develop and release a stable live version with new features.

Source Categorization and Source Tagging

To summarise the approach already reported in detail in Section 3.6.2 of *D1.5 Report on data gathering v3*, category tags were introduced to improve the filtering options for the various content sources. Multiple tags – modelled in the WLT LOD namespace – can be automatically assigned to a source website before it is added to the collection of sources. Whenever a document from that source is ingested the respective categorization tags are superimposed on the indexed document. Those tags are then used to do sub-filtering on the various documents coming from cultural sources, to make it possible to gain insights into the different sectors and directly compare their coverage.

A decision was made to stick with the initially proposed 9 categories, since they sufficiently cover most of the cultural domain, especially the GLAM sector as a focus, as well as are comprehensive enough to be able to have enough source websites assigned per category. The nine categories are: *skbtag:gallery, skbtag:film, skbtag:library, skbtag:theater, skbtag:archive, skbtag:music, skbtag:museum, skbtag:site and skbtag:art*.

The new cultural GLAM website sources that WLT extracted from Wikidata (ref. Section 2.2) on a per country basis were also assigned respective categories. For these datasets the algorithm to assign tags was extended to assign tags based on a larger number of additionally extracted metadata attributes, namely based on labels, descriptions, assigned

---

[3] moz.com

Wikidata categories and website URLs for each item, using a predefined set of regular expressions that are matched for each category.

Project partner PIN SCRL, who did some manual assigning of tags to the sources provided by project partners whenever our algorithm could not assign a category, proposed a few additional categories, such as *heritage*, *academia* or *community*. Those would be interesting to include but need some further reflection on exact definitions and scope before they can be applied in a sustainable way. From the current inDICEs sources only a handful would fall into these categories, and agreeing on what constitutes e.g. a community Website, is not straightforward. Additionally it sparks the discussion if those should not be considered dedicated *source segments* since they specify the type of website more than the content that can be expected. Due to this we decided to focus on the 9 categories mentioned and in addition introduce a single new tag skbtag:culture which will be used for the websites that were specified as *heritage site, archaeology* or any other culture-related domain. Introducing individual tags that will only be used for a handful of websites that sporadically produce new content will otherwise result in misleading selectors. Any other content from websites provided by inDICEs partners is also available but not individually tagged.

IPTC Classification for Cultural Content

The inDICEs dashboard makes use of the IPTC classifier that WLT has been developing for a different project. It is a multi-label classification tool that receives a text and predicts labels based on an existing taxonomy. It uses the RoBERTa transformer model that was trained on a large corpus of various News media sources. WLT is currently using it to assign an IPTC NewsCode[4] to all English and German News Documents as part of the NLP pipeline. The IPTC (International Press Telecommunications Council) classification is a global industry standard for the exchange of news data. The base taxonomy level consists of 17 main categories, which includes *Politics*, *Health*, *Science and Technology* etc.

inDICEs has been taken as an incentive to improve the classification for the *Arts, Culture and Entertainment* class (mapped to *Entertainment & Culture* in the Dashboard). The requirements of the project have shown us that it is very wide-ranging and has the inherent problem that the large amount of celebrity and sensationalised news articles that are published are blended with the lower volume of cultural content. WLT is currently working on improving the classification algorithm so that it is also possible to assign more detailed categories for "*Entertainment & Culture*" from the third level of the IPTC taxonomy, especially from the *Arts & Entertainment* and the *Culture* category. The focus there has been on distinguishing content from *Cinema, Dance, Fashion, Literature, Music, Theater, (Visual) Arts, Libraries & Museums* and *Culture.* Figure 4 shows an overview of the IPTC NewsCode taxonomy, with the second and third layer of *Arts, Culture and Entertainment* expanded.

This is still ongoing work with the goal of developing a final model that can be generalised to other domains of interest, e.g. the SDGs, while sticking to the IPTC categories. Following are three example texts extracted from the WLT document repository for the category *Museums.* Sets of texts per category are used to assign labels to the individual clusters in

---

[4] www.iptc.org/std/NewsCodes/treeview/mediatopic/mediatopic-en-GB.html

the new model that builds on *Top2Vec* for initial clustering before predicting the class with the RoBERTa classifier.

"Historic museum celebrating Brazilian independence reopens for country's bicentennial. The Ipiranga Museum in São Paulo, built at the site where independence from Portugal was declared in 1822, has been closed for nearly a decade."

"Philadelphia Museum of Art workers hold one-day strike. The strike by unionised employees of the museum was intended to put pressure on the administration amid protracted contract negotiations."

"The museum has a huge collection of American art going back to colonial days, and lots of modern European paintings and sculptures. It has several Warhols, and works by some of the giants of Abstract Expressionism — de Kooning, Rothko, Frankenthaler and Motherwell. And it has mummies, ancient Chinese porcelain, centuries-old Greek statuary and shining body armor from Germany and Italy. The institute was the first American museum to buy a van Gogh, and it is marking the 100th anniversary of that milestone. It narrowly avoided selling off some of its treasures when Detroit went into bankruptcy in 2013. But the city and the museum have recovered, and Mr. Salort-Pons said the museum's finances have never been stronger."

Bookmarks

During the duration of the project, different approaches were taken to gather input from project partners and potential users of the inDICEs dashboard from the CCS for topics to be transformed into search requests. This section summarises the different approaches and their outputs.

Three dedicated "Bookmark Sessions" were held via Video Call in March and April 2022 in cooperation of webLyzard, FBK and Platoniq, with invited representatives from different project partner organisations. All sessions, with durations ranging between 1h and 1h 20min, at first had an introduction of the main features of the webLyzard dashboard and then focused on explaining how searches can be configured and bookmarks set up. Then participants could propose their ideas, suggest topics and ask questions.
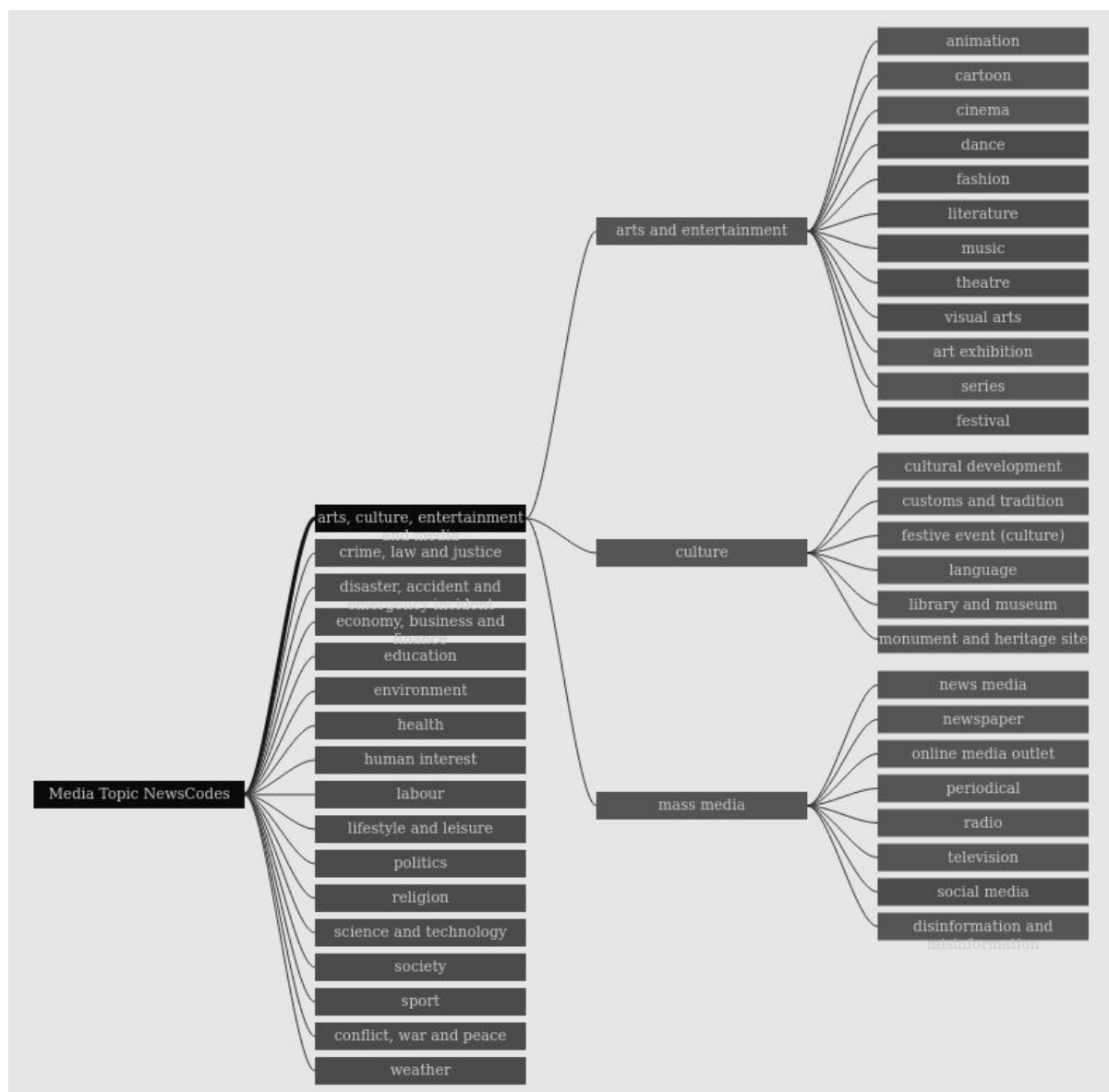
Fig 4: Tree View of the IPTC NewsCode taxonomy showing second and third layer of the *Arts, Culture and Entertainment* class[5].

Table 2: Overview of the key information of the sessions held.

Date	Partners	Topics	Description
15.03. 2022	ICCU	Archaeology	Setting up a bookmark "Archaeology" (general topic) in the full dashboard plus an explanation of the Dashboard LITE and how the same bookmark can be accessed there.
11.04. 2022	NISV, CHT	Lascaux, Decolonising GLAMs	Setting up a bookmark "Lascaux" (specific topic) and exploring the content and different visualisations in the full dashboard version. Creating a more complex bookmark "Decolonising GLAMs", suggested by NISV, and discussing the thought process around it, with input from WLT on how to improve it to get more meaningful results.

---

[5] show.newscodes.org/

11.04.2022	EF	Europeana	Overview of the dashboard LITE and a thorough explanation of the full dashboard and its capabilities and features by WLT. Exploring content from the Europeana website on the dashboard, and some good follow-up questions.

Sessions were video recorded and are available and shared within the project on GoToMeeting and Google Drive.

- *Bookmark Collection Sheet.* A Google Spreadsheet was created and shared on which all partners could propose suggestions for topics and bookmarks that they would like to see as pre-configured settings for the Dashboard LITE integration.

- *Bookmark Voting.* Based on the initial suggestions for the two bookmark categories *Creative Industries* and *Digital Culture* described in D4.2, partners could vote on the respective keywords and make suggestions to improve the associated terms. The bookmarks were then adapted accordingly.

- *WP4 Discussions.* Platoniq, leading WP4, held a few discussion meetings in which bookmarks were talked about and suggested, mainly to agree on the ones that should be available for the public in the Open Observatory.
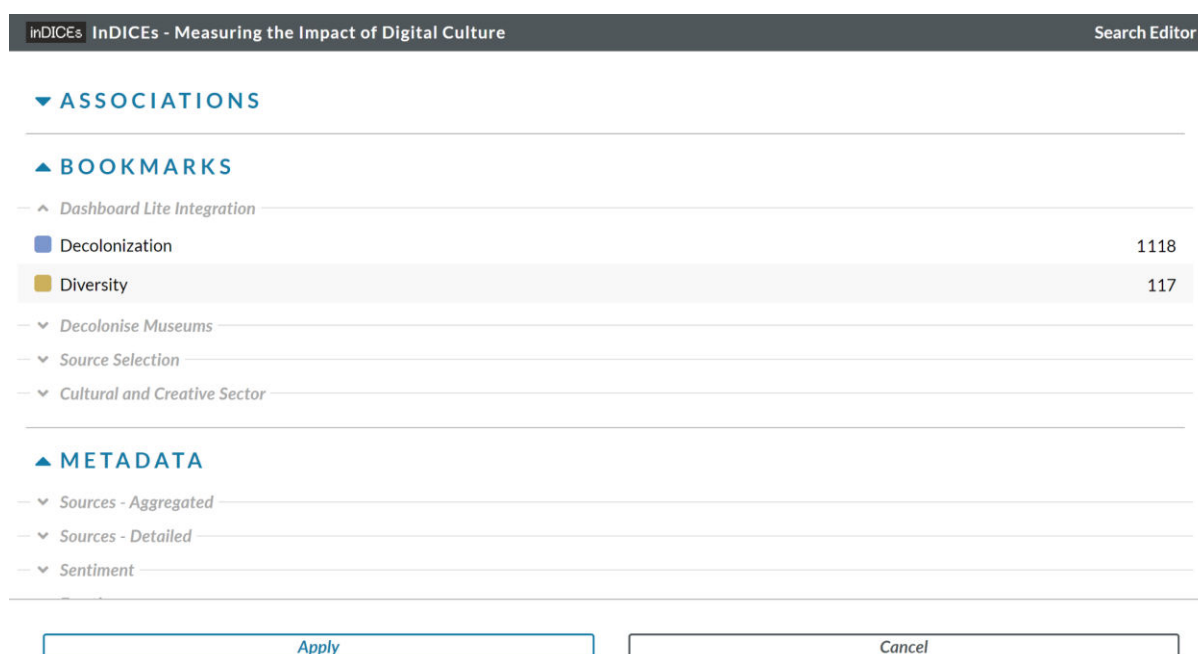
Fig 5: Bookmark Selection in the Configuration Menu of the inDICEs Dashboard LITE.

As a result of the different efforts, a handful of bookmarks will be selected to be directly accessible through the Open Observatory menu as an initial source of inspiration. All other bookmarks that were defined and revised throughout the project will also be available under the "bookmarks" tab in the configuration menu of the Dashboard LITE (ref. Figure 5) and can be selected directly within the embedded application.

2.4 CHI Knowledge Graph

Since the reporting in D4.2, the inDICEs Knowledge Graph has grown further and now contains 3.3 million persons, organisations and locations with metadata from LOD sources, such as *Wikidata*, *GeoNames* and *OpenStreetMap*.

Additional metadata attributes were added and are extracted from Wikidata for each entity: *gender* (for persons) and social media attributes for organisations. For gender we stick to the values Wikidata uses, namely *male*, *female*, *non-binary*, *intersex*, *transgender female*, *transgender male* and *agender*. For social media we added the extraction of *twitter* handles, *instagram* usernames, *facebook* ids, *youtube* channel ids, *tiktok* user usernames, *linkedin* ids and *vimeo* and *pinterest* usernames. Unfortunately the data provided publicly on Wikidata for the social media properties is rather limited, where Twitter account names are by far the most commonly provided attributes. To account for this we have planned an initiative that will focus on extracting information on official social media channels from public websites and include those into the metadata. While this will also be applied for the CHIs added and the inDICEs dashboard will benefit in the long run, it is not feasible to implement this feature as part of the project, but valuable learnings were gained on the distribution of this information on Wikidata.

To further extend the CHI Knowledge Graph with GLAMs and other cultural institutions, insights gained from the Netherlands pilot case were used to further develop the automatic extraction of entities from the GLAM sectors from Wikidata.
The same SPARQL query was revised and used to extract data on CHIs on a per country basis. As a result 5 further spreadsheets were created for Austria, France, Germany, Italy and Spain that contain Wikidata items and their metadata (name, description, country, inception, website, social media channels). The whole process was streamlined and the technical components improved, such that it is easy and an automatic process to create a new "GLAM" Google Spreadsheet for any country. For this the Google Sheets API v4 is used to automatically create, fill and format the spreadsheets. In the future it is planned to put the complete process behind an API so that it is possible to generate, review and directly import new entities into the knowledge graph.

From a technical perspective, all the imported CHI entities are stored in an Apache Jena Fuseki dataset in a dedicated subgraph so that they can be easily retrieved as a batch and separated from other entities. They are then also indexed in Elasticsearch for fast retrieval and querying for the WLT dashboard and API components.
Figure 6 shows an overview of Dutch museum entities in the CHI Knowledge Graph in the SKB Browser Interface, a dedicated frontend to view all data in the SKB. Figure 7 shows a detailed view on the entity for *Amsterdam Museum* and all associated metadata.
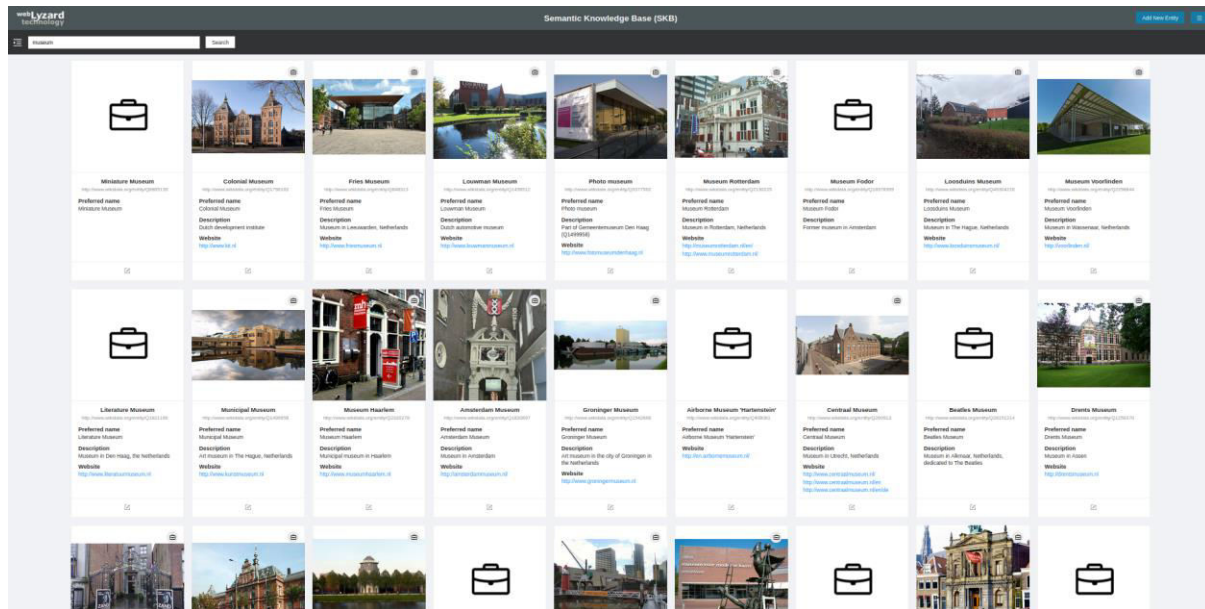
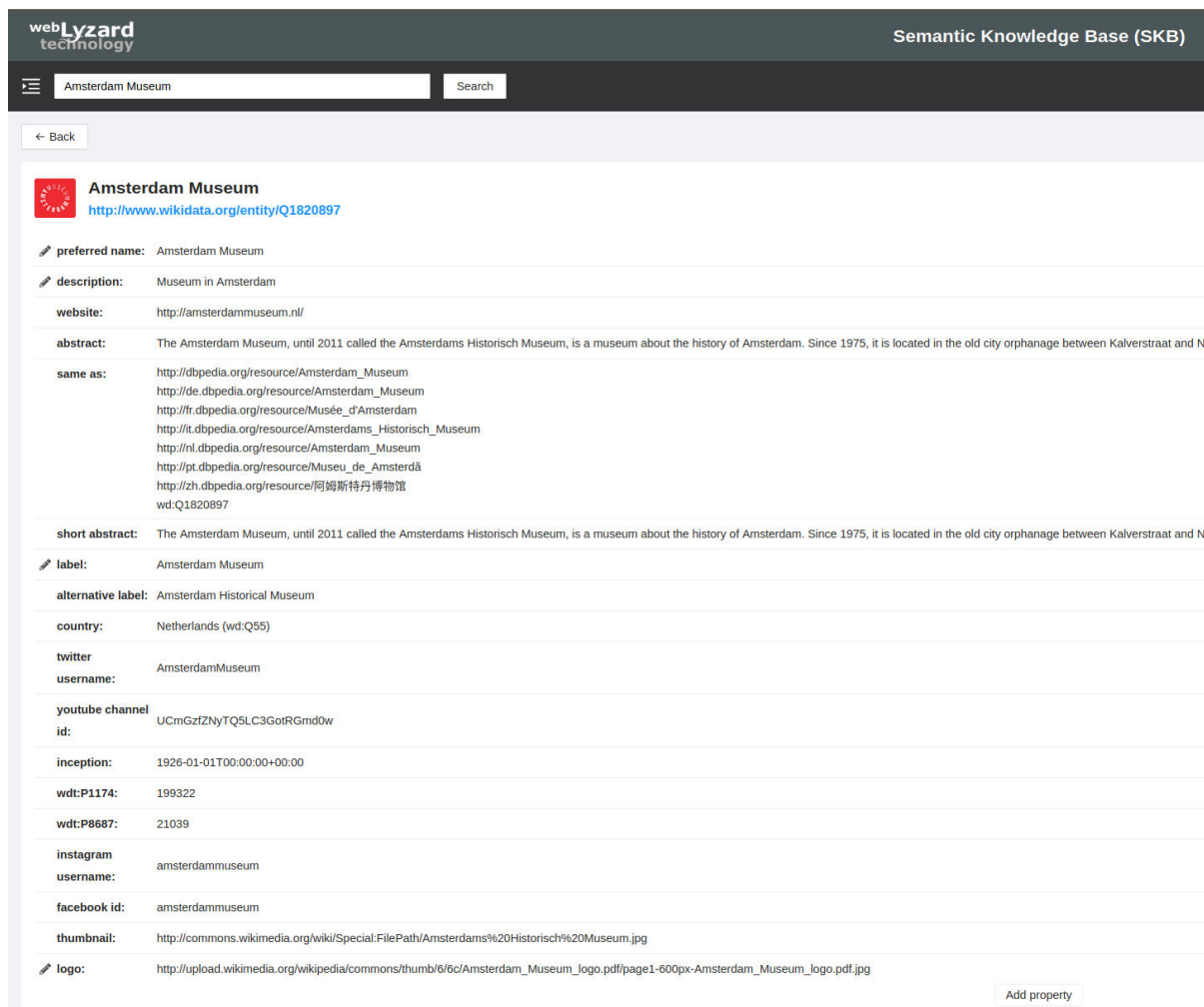**Fig 6: Entity Browser View showing search results for Dutch Museums in the CHI KG**

Fig 7: SKB Browser Entity Card showing the item "Amsterdam Museum".

To account for the fact that not all cultural institutions need to be organisations, the entity types that WLT extracts were extended with a new type *POIEntity* (POI = Point of Interest), in a preliminary effort. POIEntities are also based on Wikidata, but span a wider context, of which culture is one aspect. The cultural POIEntities include any items that can be derived to be an instance of a subclass of either GLAM (wd:Q1030034), cultural heritage (wd:Q210272) or of cultural institution (wd:Q3152824). Additionally all entities that were *disbanded* (wdt:P576) are not included.

POIEntities are still in development and not fully integrated yet, but Figure 8 shows a preview of some of the results retrieved for Austria. These entities are planned to be fully integrated into the NLP pipeline for named entity linkage and will continuously become available for entity annotations.
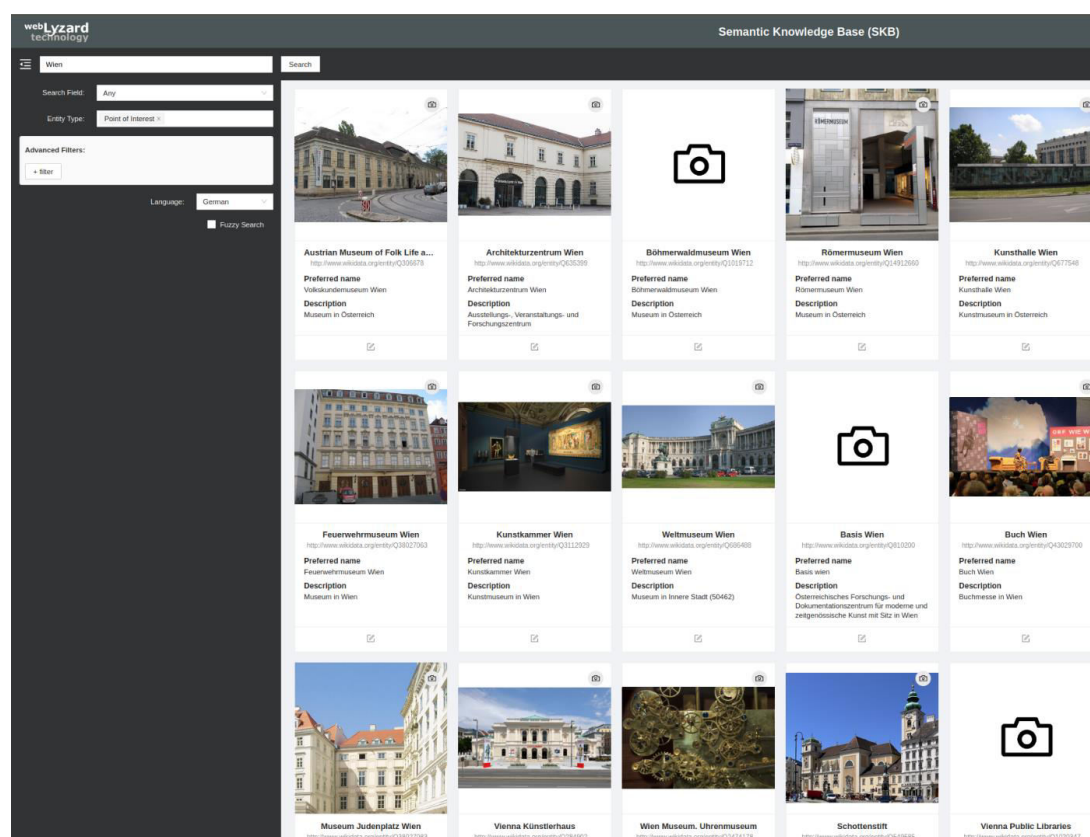
Fig 8: Development View of the SKB Entity Browser showing POIEntities in Vienna.

2.5 inDICEs API Framework

The overall *inDICEs API Framework* was already reported in D4.2 (see Figure 9). After authentication through a JSON Web Token, different webLyzard APIs can be accessed to push (POST) or retrieve (GET) documents, statistical data or entities from the webLyzard servers. Through the Visualisation API, different embeddable visualisation widgets can be retrieved. A new addition in this reporting period is the availability of the inDICEs *Dashboard LITE* as an iframe embed with configurable URL parameter settings (see Section 4.2).
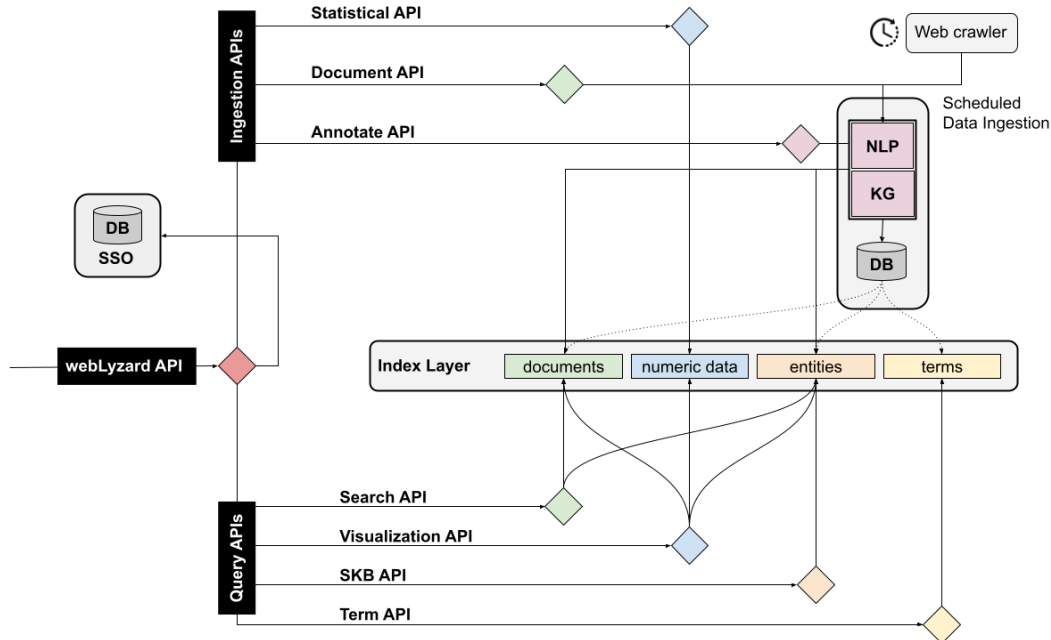
Fig 9: Overview of the webLyzard API framework (described in detail in D4.2)

The Document API was improved and now also accepts NDJSON (Newline Delimited JSON) requests for uploading multiple documents to the inDICEs repository where the documents generated from the user content on the Participatory Space get stored, at once. It can be used to stream multiple documents at once that then get processed one record at a time. Thereby each line needs to be a valid JSON value and lines are separated by a "\n" newline delimiter. Listing 1 and Listing 2 show a minimal example of an NDJSON accepted by the Document API and the respective cURL request.

```
{"uri":"https://participate.indices-culture.eu/…", "title": "Post One", "meta_data":
{"published_date": "2022-07-01T19:00:00Z","user_name": "User X",...}, "content":...}
{"uri":"https://participate.indices-culture.eu/…", "title": "Post Two", "meta_data":
{"published_date": "2022-07-01T19:10:00Z","user_name": "User X",...}, "content":...}
…
```

Listing 1: A sample NDJSON document containing two user posts from the Decidim platform.

```
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 2: Push (POST) a stream of JSON documents to the inDICEs repository.

Decidim API Specification

Decidim provides an API that follows the GraphQL.org specification. GraphQL is a technology developed originally by the company Facebook in 2012 and liberated with an open source licence 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. For the full API specification, please refer to Deliverable D4.2.

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. Users are also able to participate and use the platform to create online spaces for their own communities. One example would the Europeana Impact Lite Taskforce that hosts their meetings and has even created a space for participatory budgeting.

3.1 inDICEs Decidim Implementation

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, it is probably the most important application of its kind globally with more than 400 dependencies in GitHub[6] (the majority of which are Decidim instances for cities or different types of organisations). An introduction to Decidim has been given in the D4.1 deliverable, with additional technical details and the underlying community model reported in D4.2.

3.2 The Self Assessment Tool

Through collaboration between all partners and in consideration of the outputs of the project and the scope of Decidim, the tool has gone through several iterations. Ultimately the tool – in collaboration with Europeana – has been created to utilise the *Europeana Enumerate Survey*[7] and to provide feedback using the knowledge and resources produced by inDICEs. This shift has been to create a synthesis of successful elements of the project and an existing audience of participants from the Europeana community. The Self Assessment Tool goes beyond the Enumerate integration. It adapts the Decidim survey component to enable administrators to give feedback for any survey created on the platform. The replicability of the survey opens up possibilities for creating an interactive experience and for creating unique self assessments from different aspects of the project - e.g., WP1 impact areas or WP3 insights from policy recommendations; see *SAT 'Lite' in Europeana* later in this section.

The Design Process
The initial concept for the Self Assessment Tool was developed by WP3 and reported in *D3.2 Guidelines for CHIs Digital Transformation*.

---

[6] github.com/decidim/decidim/network/dependents
[7] pro.europeana.eu/page/enumerate

It summarises as follows:
- Survey - a questionnaire about the typology of the organisation along with questions to assess their digital transformation
- Self-Assessment Report - a series of visualisations that communicate how CHIs compare in terms of the topics covered in the survey
- Guidance and Knowledge - strategic guidelines as well as targeted, domain-specific resources to support digital transformation processes
- Participatory Space - embedded within the participatory space to encourage community engagement and development among CHIs

Within this scope, WP3 developed a questionnaire and resources from the consortium for the guidelines. The knowledge was sourced from the partners by WP4 and WP3. To validate the concept a simplified paper prototype was created, along with an exercise where partners took on the role of future users or participants to evaluate their satisfaction with the tool. It was piloted in October 2021.
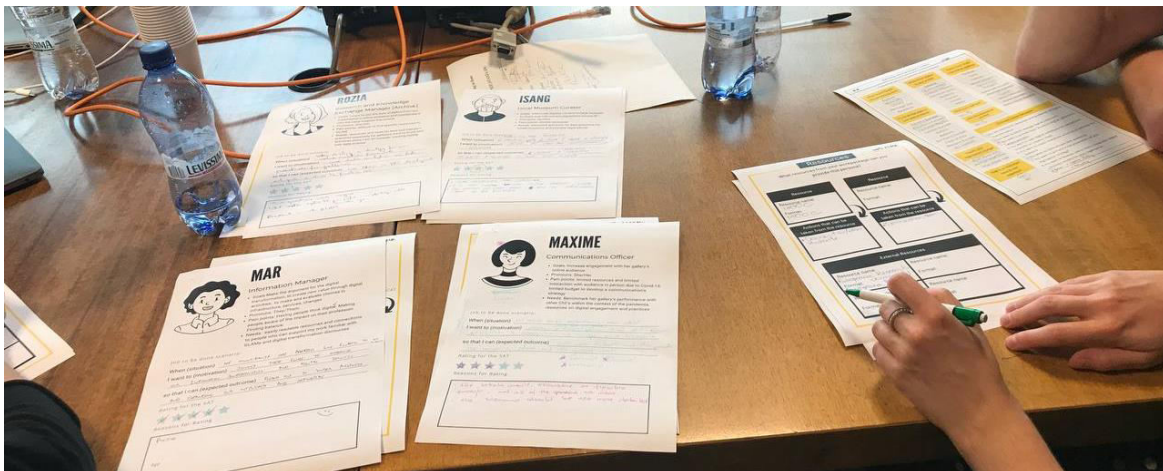
Fig 10: **Partners participating in the workshop to validate the concept of knowledge and guidelines to be produced by the SAT at the Rome PMB in October 2021.**

The paper prototype was created within the scope of the Decidim platform and the adaptability of the various components such as the survey component. Of the four facets of the Self Assessment Tool, the visualisation aspect had not at this point been further developed without an existing data baseline. Still, there was a consensus on the progress made and the success of a flexible framework that could synthesise the outputs of the project with the questionnaire produced by WP3.

Creating Feedback or an 'Assessment' for each Survey

Like the templates, the feature to add feedback to a survey has been added to the admin dashboard. Adding it to the dashboard created the simplest pathway for the feature within Decidim's systems architecture.

The Survey Component on Decidim

The survey component generally allows administrators to create questionnaires that both registered and unregistered users of the platform can answer. The survey component allows the administrator to set the following scope around each survey:

- Administrators can limit or set a time period
- They can have the following answer types: short text (one line), long text (multiple lines), single or multiple options, sorting through a list, matrix questions, conditional questions etc.
- They can also define the minimum and maximum characters of each answer
- They can define the maximum number of options a participant can select

Exporting Answers
Administrators can export respondents' answers as: CSV, JSON, XLSX, or PDF.

Templates
Administrators can create survey templates through the administrator dashboard. This allows them to then use the template multiple times across different Decidim modules, such as processes, assemblies, or conferences. Administrators can adapt the template for each unique use. The capacity to create templates was developed due to foreseeable future needs of the Enumerate survey on Decidim. The update has been made available to the Decidim community.


```
Usage

Templates will be available in their own section in the admin

Installation

Add this line to your application's Gemfile:

gem 'decidim-templates'

And then execute:

bundle
bundle exec rails decidim_templates:install:migrations
bundle exec rails db:migrate
```

Fig 11: A snapshot of the instructions to update a Decidim instance with the template module, available on GitHub.

Adaptation of the Self Assessment Tool

Currently, instead of questionnaires that do not provide immediate feedback, the Self Assessment Tool will provide respondents with recommendations set by the administrator and allow them to download their own answers. These improvements add value to the participant experience because they are given immediate feedback and can access their own answers. In the case of the Enumerate survey which is lengthy and requires multiple team members from a CHI organisation to answer the survey, being able to access this data

after answering the survey is an important aspect to the organisation's own data collection and preservation.

Figure 12 shows an initial sketch, developed by WP3, of the response that participants should receive when completing a Self Assessment.
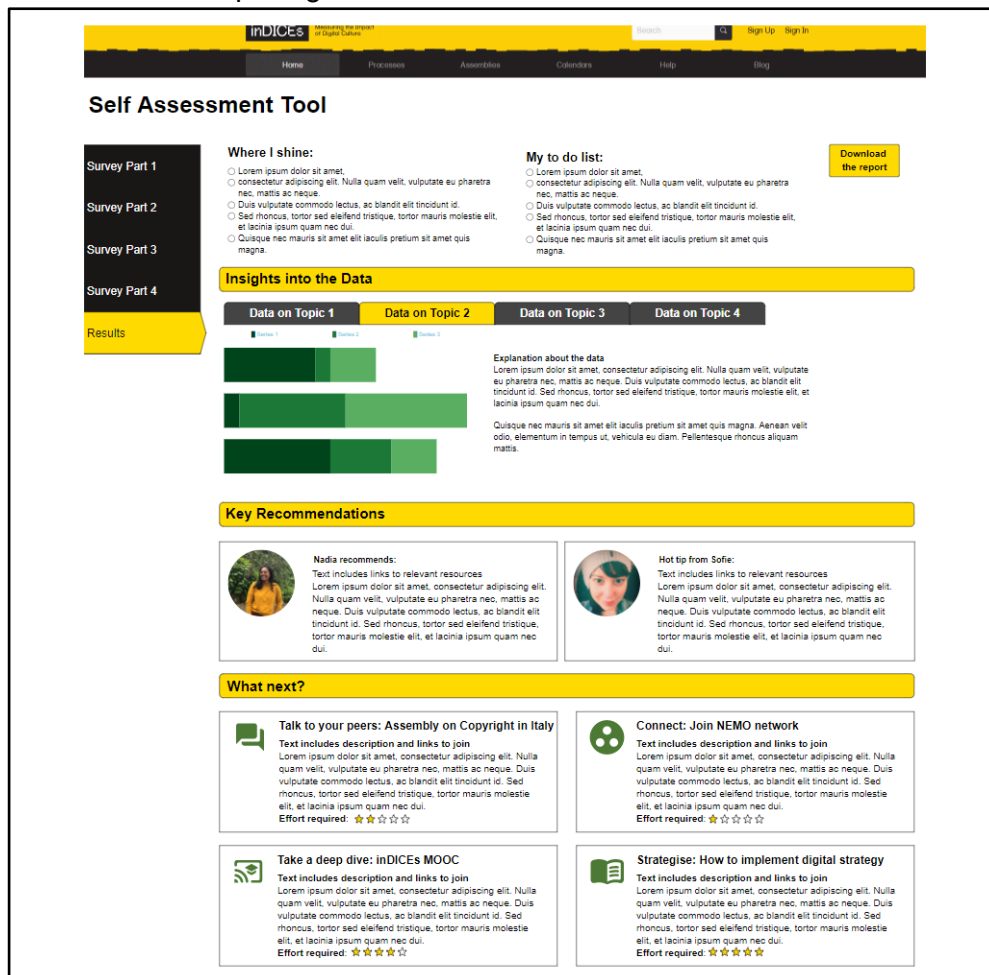
Fig 12: An initial mockup of the Self Assessment Tool response page.

The following iterations of this design removed the following sections:

- *Where I shine*
- *My to do list*
- The visualisation for *Insights into the Data*

The reason for eliminating these areas was that there was not enough data and no framework to approximate where participants were doing well, or what actions could be taken based on the developed questionnaire. To make the synthesis possible and create flexible feedback that can be adapted to the questions and the resources and outputs produced by the InDICEs project, the iteration depicted in Figure 13 was developed.
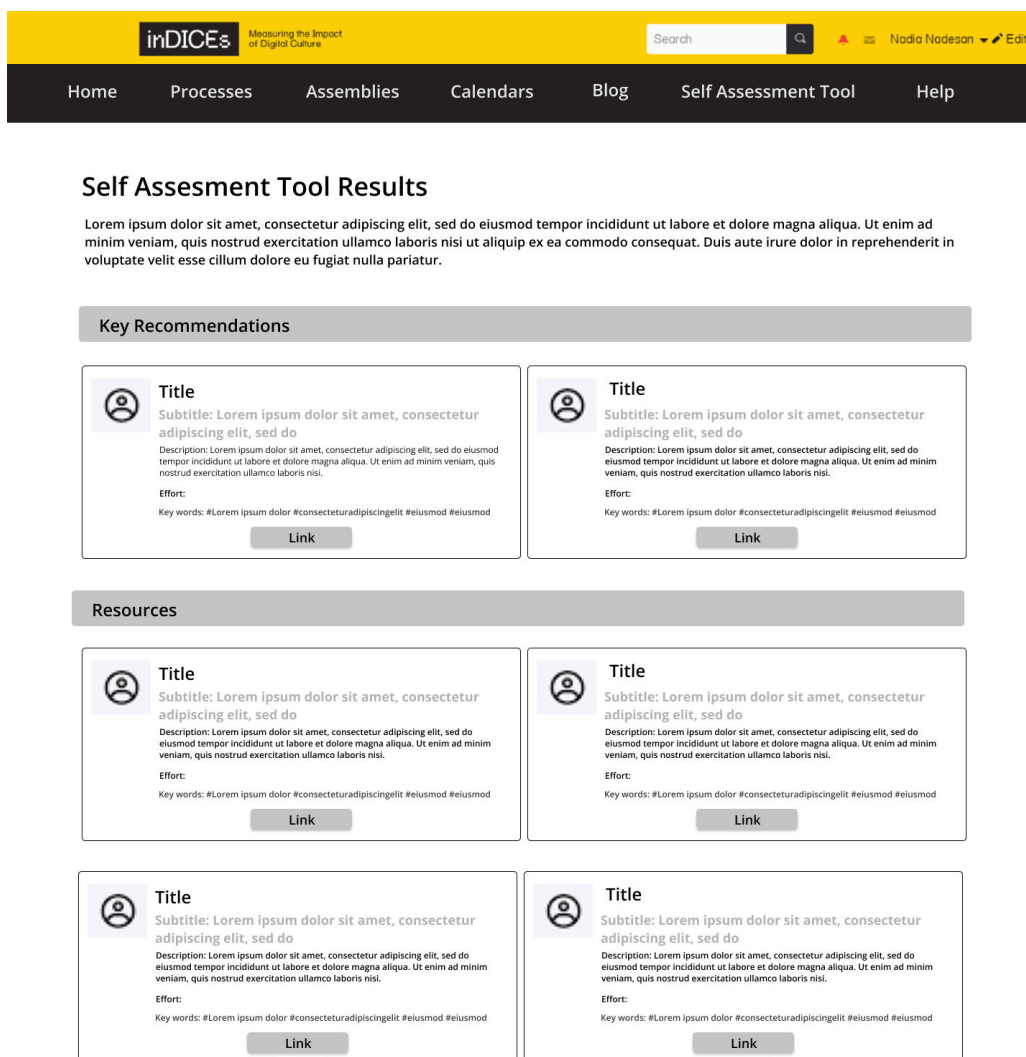
Fig 13: Second iteration of the Self Assessment Tool response page.

Administrators can create multiple recommendations or resources based on the answers of the respondents. They can do so by embedding keywords into the answers within the survey. When a participant selects a particular question with a keyword that is associated with a recommendation they will receive that recommendation. This is a highly flexible system that can facilitate both simple and complex self assessment depending on the administrator's needs. Recommendations are prioritised. The administrator decides the 'weight' or importance, i.e the administrator can assign a weight of 1 to a recommendation and it will appear at the top, a weight of 2 and it will appear second. The remaining recommendations will appear under the section "Resources".

The elements in each recommendation are as follows:

- Thumbnail image
- Title
- Subtitle
- Description
- Keywords associated with the recommendation

- Effort, or an approximation of difficulty which can be expressed with words or emojis.

For the inDICEs SAT a number of strong arm emojis was used to indicate the required effort. As previously mentioned, participants can download their survey responses as a CSV file.

```ruby
module Decidim
 module Admin
   class CreateIndicesSatFeedback < Rectify::Command
     def initialize(form, sat_set)
       @form = form
       @sat_set = sat_set
     end
     def call
       return broadcast(:invalid) unless form.valid?
       begin
         create_sat_set!
       rescue StandardError => e
         return broadcast(:invalid, e.message)
       end
       broadcast(:ok, sat_feedback)
     end
     private
     attr_reader :form, :sat_set, :sat_feedback
     def create_sat_set!
       @sat_feedback = Indices::SatFeedback.create!(
         sat_set: sat_set,
         title: form.title,
         description: form.description,
         hashtags: form.hashtags
       )
     end
   end
 end
end
```

Fig 14: A snapshot of the code to develop the Self Assessment Tool feedback system from Github (app/commands/decidim/admin/create_indices_sat_feedback.rb).
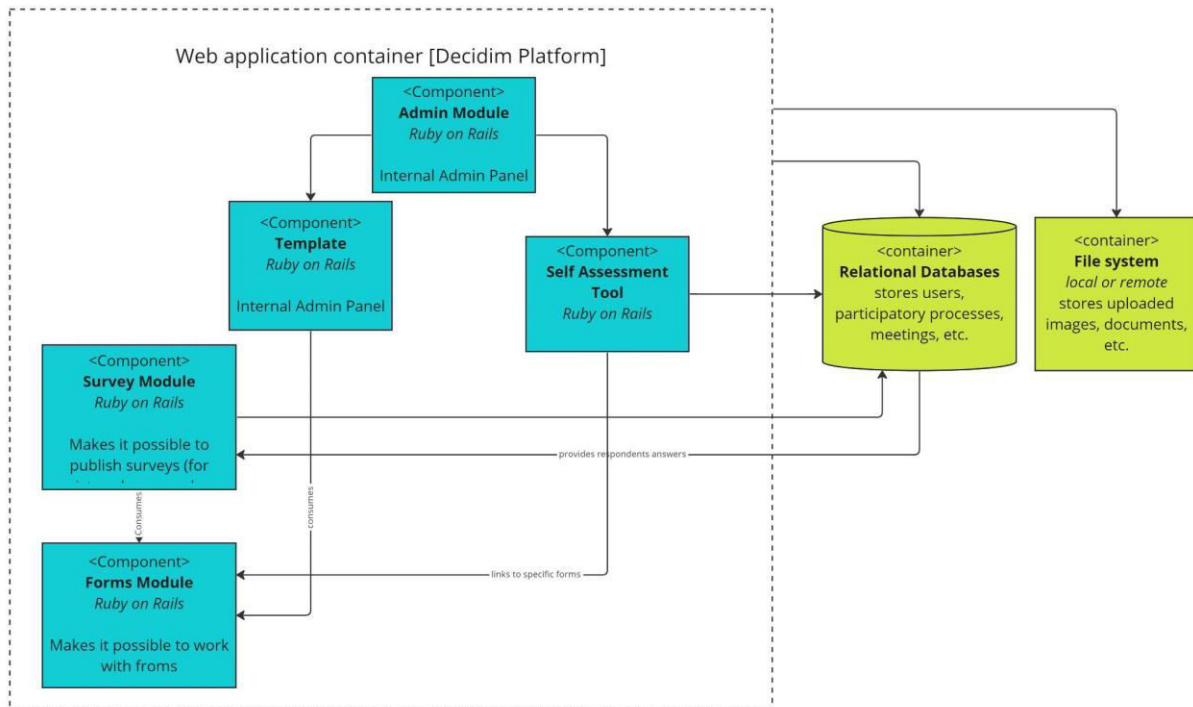
Fig 15: A simplified image of the system architecture of the SAT and Templates showing the relationship between the components and the new features.

Integrating the Enumerate Survey

As mentioned previously an facet of the Self Assessment that also needed attention was baseline data along with an audience motivated to use the tool to produce more data in the future to come. To resolve this gap a strategic integration was made in collaboration with WP3 to integrate the Europeana Enumerate Survey into the Self Assessment Tool methodology. Since many of the questions from the SAT questionnaire bore striking similarity to the Enumerate survey and that the Enumerate survey had a built in audience, communications strategy, and baseline data the shift was a strategic move towards a Self Assessment Tool with a greater chance of success and more aligned with the original conception.

Initially, the survey was incredibly long with 50+ questions depending on the typology of the institution. The questions were divided into sections. To make the survey more amenable WP4 improved the UI of the survey component. Previously each section of the survey served as a divider between questions. Platoniq added a title and description when creating new sections in the survey. However, the length of the survey was not amenable to the platform. After the summer of 2022 the sections were then transformed into 'mini' surveys as a strategic design developed by Platoniq and Europeana. To display the surveys in a single space, the surveys were hyperlinked in a single page component that could be stylised by the administrator and have additional supporting text.
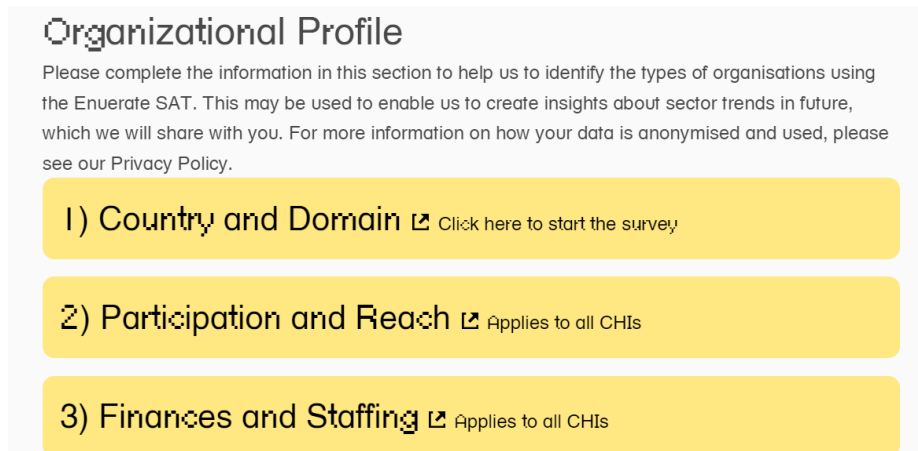
Fig 16: The Enumerate SAT with the first series of questions divided into sections and displayed on the page component.

SAT 'Lite' at the Europeana Conference

To apply the concept of the SAT's flexibility and adaptability to make other resources and outputs available from the InDICEs project the SAT 'Lite' was ideated and executed. The SAT 'Lite' unlike the Enumerate Survey had fewer sections and different questions to make it 'lighter' for participants and utilised the insights from Deliverable 3.3 that visually summarised the CHIs DSM readiness Assessment. Methodology. The SAT Lite was deployed in a workshop at the Europeana Conference in September 2022.

Fig 17: SAT Lite implemented at the Europeana Conference in September 2022.

Participants were able to receive recommendations based on their answers to lighter surveys based on the feedback module.

Fig 18: Feedback provided for SAT Lite based on answers from Platoniq's team.

Fig 19: First section with only 5 questions and a more visual approach.

Finally, in a small experiment to try out new methods for visualising the results a page was added to show a sample of the data based on participant answers for the SAT Lite.
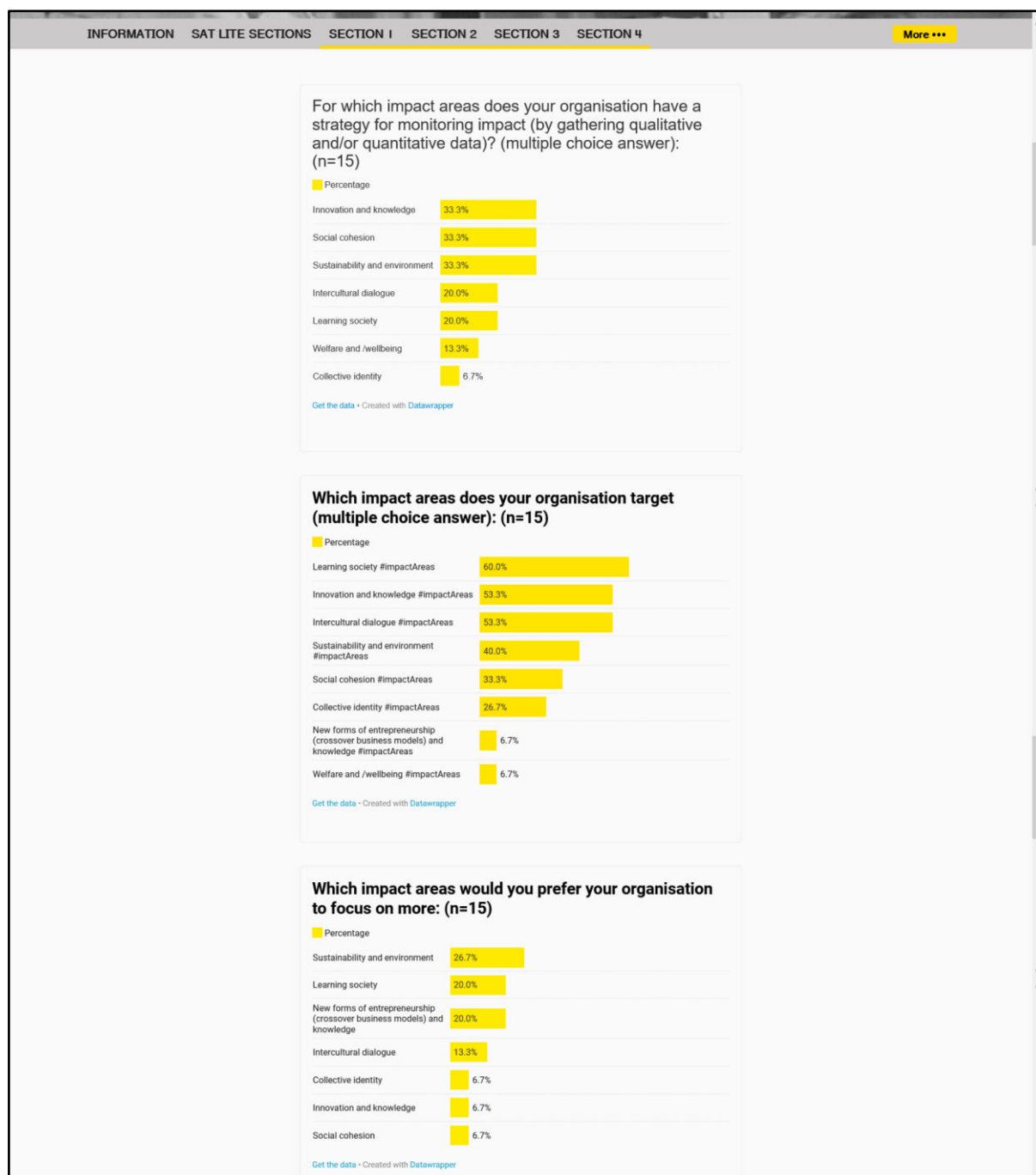
Fig 20: Visualisation produced from participant answers of the Europeana Workshop.

The aim is to continue developing better practices around visualising the data produced by the SAT and further add value for the participants to the Open Observatory.

3.3 User Interface Improvements

The Decidim interface has received criticism over the years because of the limitations around making the interface more user friendly and especially the landing page where administrators had little control over customising the look or content that was highlighted. In the next months a Decidim sprint has been planned for December 2022 with preliminary

planning sessions in November 2022. This sprint will improve the information architecture so that visitors and regular participants of the platform can more easily access tools, resources and active participatory spaces on the platform. The following sections highlight the digital tools on the platform that will also support creating a more streamlined experience on the platform that have been developed by Platoniq over the past two years.

Landing Page Content Blocks

Integral to making information more accessible on the platform is having an effective and easily legible landing page. Platoniq developed tools so that administrators have greater control and flexibility to highlight content on the landing page. The landing page consists of different content blocks. Administrators choose which blocks are 'Active' or seen on the landing page.
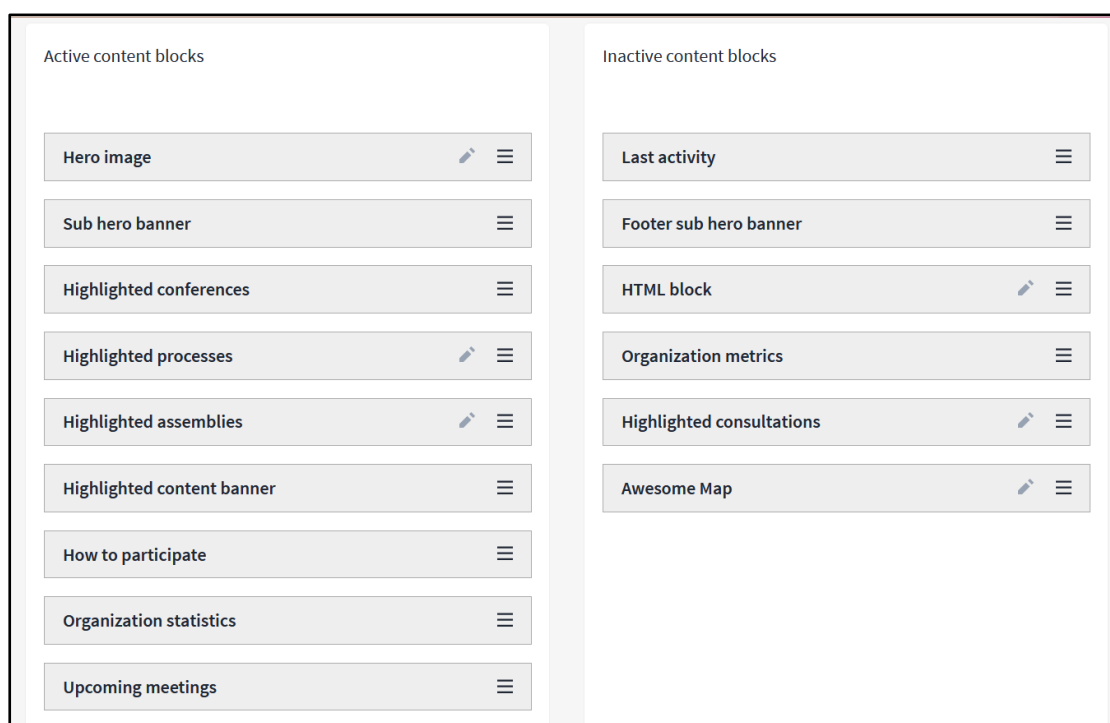
Fig 21: Traditional landing page options for instances of Decidim.

The improvements Platoniq developed are the following:
- A calendar widget for the homepage that displays active events
- Alternative "Hero" style blocks, with images and text positioned in several ways
  - A set of 3 horizontal blocks each with a title, description, and button for a hyperlink
  - A vertical stack of blocks each with a title, description, and button for a hyperlink
  - A block of 5 tiles each tile containing a title, description, and button for a hyperlink
- Highlighted consultations
- Latest blog posts (with the ability to choose which blog posts from a specific participatory space)

● Upcoming meetings

Fig 22: A sample of a landing using the alternative landing options developed by Platoniq on the administrator dashboard.

The updates are available on Github for any organisation or individual to employ on their instance of Decidim.
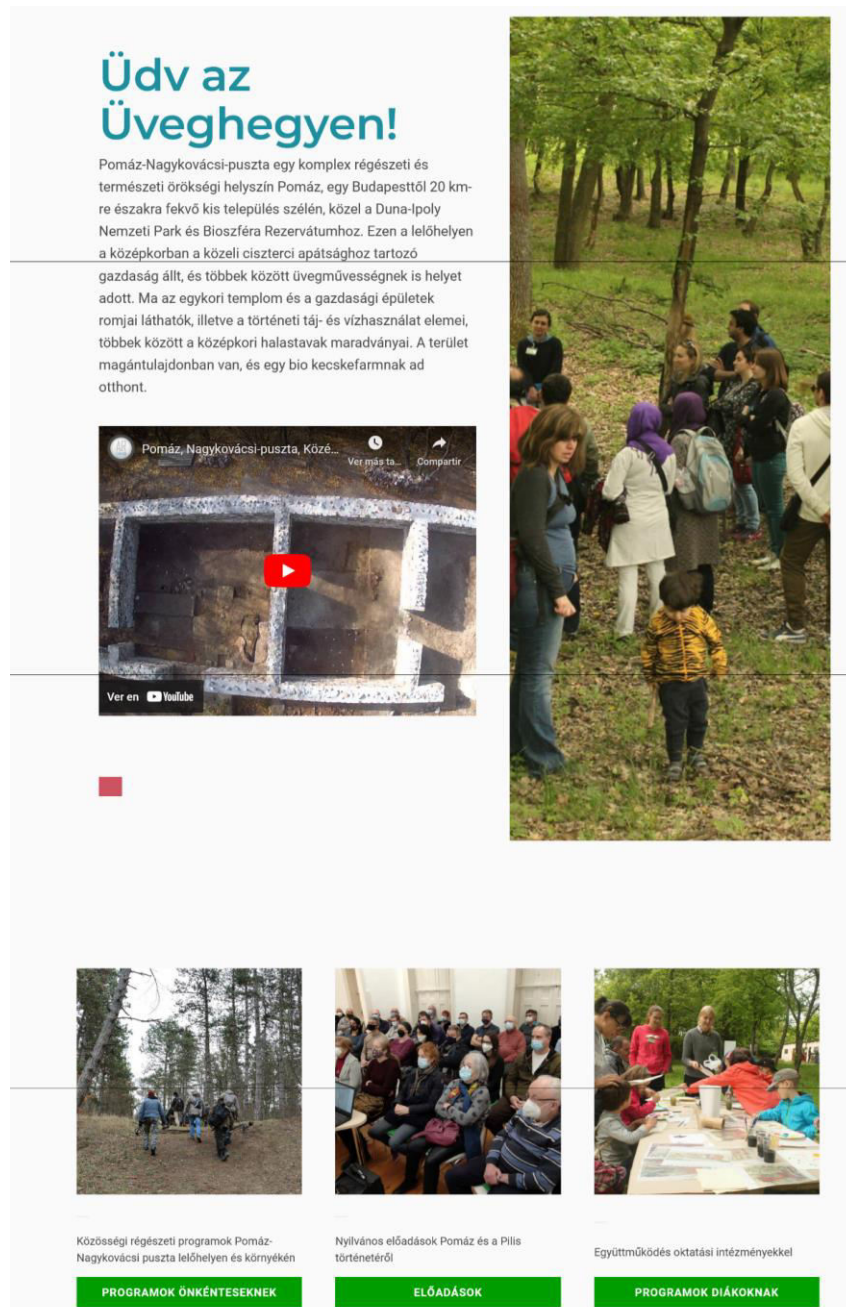
Fig 23: Use of the alternative landing blocks created by Platoniq for a Decidim dedicated to a heritage site in Pómaz, Hungary.

Decidim Awesome

Over the past two years Platoniq has taken in some of the constructive criticism to updating the platform and created a series of tweaks for administrators within the Github repository 'Decidim Awesome' that is available to the public for public use.

The InDICEs platform is currently updated with the Awesome repository which allows administrators to do the following:

- **Image support for the Quill editor** - this allows administrators to drag and drop images into the Quill editor which is the application used to create pages or add descriptions across the platform.
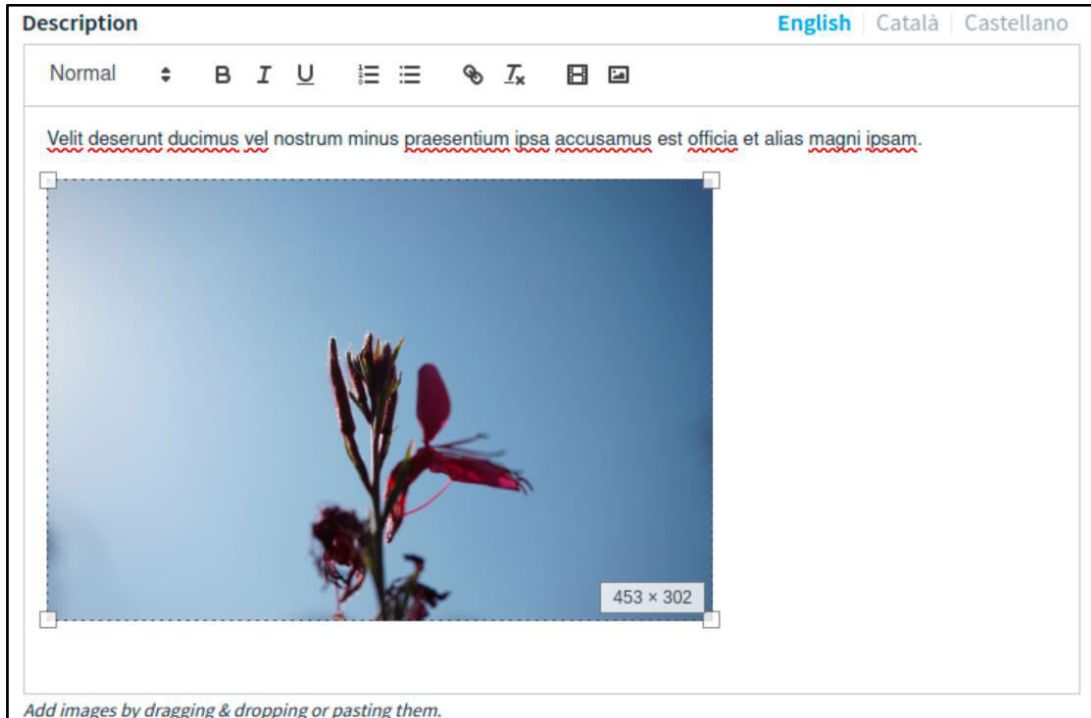
Fig 24: Images in the Quill editor which is used to edit pages along with descriptions for various spaces or modules and components.

- **Auto-save for surveys and forms** - With this feature admins can activate (globally or scoped) an auto-save feature for any form in Decidim. It works purely on the client side by using LocalStorage capabilities of the browser. Data is stored every time any field changes and retrieved automatically if the same user with the same browser returns to it in the future. This feature is currently still being improved upon and is going through a bug fix phase for the next release of the Enumerate SAT.
- **Images in proposals** - You can allow users to upload images in them by drag & drop over the text area.
- **Fullscreen Iframe component** - Another simple component that can be used to embed an Iframe with any external content in it that fills all the viewport. This feature has been used to integrate google docs within the InDICEs platform to support tools already in use by the InDICEs consortium.

Fig 25: Example of integration of google docs within the InDICEs platform.

- **Custom CSS applied only according scoped restrictions** - With this feature you can create directly in the admin a CSS snipped that is only applied globally, in a particular assembly or even a single proposal. This has been applied across the platform especially to hide unnecessary elements built into the platform. One example would be for online events in which a venue tab is built into each event. The custom CSS allows administrators to hide the venue tab or venue map when it is not applicable.
- **Menu Tweaks** allow administrators to hide, modify or add items in the Decidim's main menu. They can also change the order, establish some conditions (like showing only for logged users) or open in a new window. Previously to edit the menu required a backend tweak but now it is accessible within the administrator dashboard. It is the latest tweak and most helpful when reorienting and customising the platform to the specific needs of the InDICEs project.
- **Assign admins to specific scopes** and prevent them modify anything else - Administrators can convert any user on the platform (that is not currently an admin) to a limited subset of participatory spaces or event components. Just add users to a box and scope them to some constraints. These users will see the "Edit" button

36

everywhere they have permissions. Any access to unallowed zones will redirect the user to the admin index page.

● **Custom fields for proposals** - Now admins can substitute the body of a proposal with a set of form fields. Edition is made with a Drag & Drop interface in the admin and can (and should) be scoped to apply only to certain proposal components. The custom fields proposal has been used in multiple workshops and participatory processes with the consortium partners on the platform most notable the Hypothesis Assembly. Technically, the content is stored in the database as an XML document compatible with normal HTML (it uses the DL/DT/DD elements).

Fig 26: Example of custom fields for proposals.

These edits support customising each Decidim instance to flexibly adapt to InDICEs project. The Decidim community at large has also been working on UX updates that will be released in February 2023. This has been commissioned by Decidim Barcelona. The advantage of the InDICEs Open Observatory is that it will also be able to receive the benefits of the UX updates as a result of being a community participant.

4. InDICEs Visual Analytics Dashboard

The *Visual Analytics Dashboard* is now available in an extended and improved PRO Version (indices.weblyzard.com/pro) and a new LITE Version (indices.weblyzard.com/lite) version. Figure 27 compares the two versions, based on a search query on *"digital skills"* using English Web Sites from the cultural sample (GLAMs, Arts, Film/Theatre and Music) as the content source.

Fig 27: inDICEs Dashboard PRO (top) and LITE (bottom), showing the results of a search query for articles mentioning "digital skills" from Jan-Oct 2022, colour-coded by GLAM sector

Both the dashboard PRO and the Dashboard LITE access the same content repositories and have the same configurations. The main difference is the depth at which search queries can be configured and the separated view of the individual visualisation components in the LITE version that is more tailored towards users that want to get a fast overview of a pre-configured topic or want to do simple searches. While the PRO dashboard version has a synchronised view where all different visualisations are displayed and updated simultaneously in the LITE version the focus can be kept on one aspect at a time.

The Dashboard LITE has been designed with novice users in mind who have no previous experience with using the tool, to make it easily accessible for CHI professionals. Basic instructions are provided on the platform to guide the users, alongside an instructional video tutorial that is available via YouTube at youtu.be/oMcRSpfuVSA. The PRO version is more powerful and requires an introductory workshop to be able to utilise its full potential. It allows users to configure their own search setups and save them as personalised bookmarks – something that is not possible with the LITE version that is purely for viewing and analysing pre-set configurations. Though it has access to any bookmarks previously configured by that user account on the PRO version. Once search configurations have been set up and are satisfactory, all results and the newest updated content can be viewed with the LITE version.

4.1 Dashboard LITE Structure and Configuration

As detailed in the Introduction in Section 7.2 of D4.5: *Train the Trainers Methodology,* the Dashboard LITE contains a series of tabs for different ways to present and visualise search results. It is easy to compare different search terms or select existing search configurations through the colour-coded pills in the top right menu. Similar to the full dashboard version, the analysed time range and content sources can be changed as desired. Also the associations and additional metadata breakdown selections, such as sentiment and emotions, that are accessible on the left side of the full dashboard are preserved in the LITE version. Table 3 shows an overview of the eight different content visualisations with a brief description. Further details can be found in the overview and step-by-step guide provided in D4.5.

Table 3: Overview of the different Dashboard LITE visualisations showing icon, name, a short description and the scope of the visualisation.

	Content Area	View a list of matching documents, sorted by recency, with a document preview.	documents
	Tag Cloud	Visualise colour-coded associated keywords.	keywords
	Keyword Graph	Visualise associated keywords in a hierarchical graph and discover the strongest semantic associations within the search results.	keywords
	Trend Chart	Show frequency of search result occurrences over a timeframe.	documents/ keywords
	Story Graph	View a streamgraph to show the emergence and evolution of stories in the search results.	documents/ stories
	Story View	Show emerging stories, represented by a lead	documents/

		article and a number of related documents	stories
	Cluster Map	Visualise documents by their semantic similarity, using clustering algorithms and methods to form groups of similar documents.	documents/ keywords
	Geographic Map	Show a map with the regional distribution of search results.	documents

The right sidebar menu has been redesigned to make it clearer how to add search terms or access the search editor to apply metadata filters and pre-configured bookmarks (Figure 28). The metadata that can be selected to visualise the respective context dimension, includes polar sentiment, emotion categories, sources and source locations, language and recency.
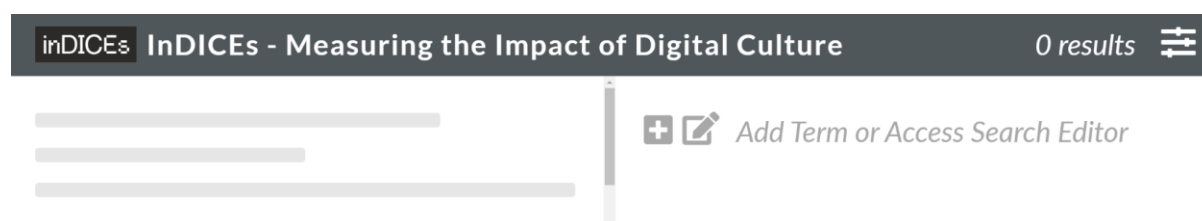

inDICEs InDICEs - Measuring the Impact of Digital Culture   *0 results*

 Add Term or Access Search Editor

Fig 28: Top and Sidebar Menu of the Dashboard LITE without any search term added.

Additionally, the right sidebar of the Dashboard LITE previews documents that match the search, which next to the document title and text includes document metadata, such as the date and original website URL where the content was published. It also shows the text snippet where the underlying search terms occurred and highlights any respective keywords.

One of the newest features is the option to further restrict the search results by interacting with the different visualisations. Clicking on any of the displayed keywords in e.g. the tag cloud or keyword graph adds an additional filter that restricts the search to only documents that also contain the selected keyword. Similarly each story, represented by three keywords, can be selected to apply a filter and preview documents that belong to that story. Selecting an area on the geomap automatically applies a filter that restricts the results to anything associated with the region around the selected area. In the advanced settings certain words or phrases can be excluded from the search, which gives the user even more options to adjust the desired results. The list of new features includes:

● Overall faster response times and loading
● Document preview in the right sidebar
● Activation of sub-filters by interacting with the visualisations
● New indicator in the top right corner when a content filter is active
● Importing of existing bookmarks and metadata topics as (global) content filters
● Selecting bookmarks and visualisations on startup via URL parameters
● Improved search match highlighting
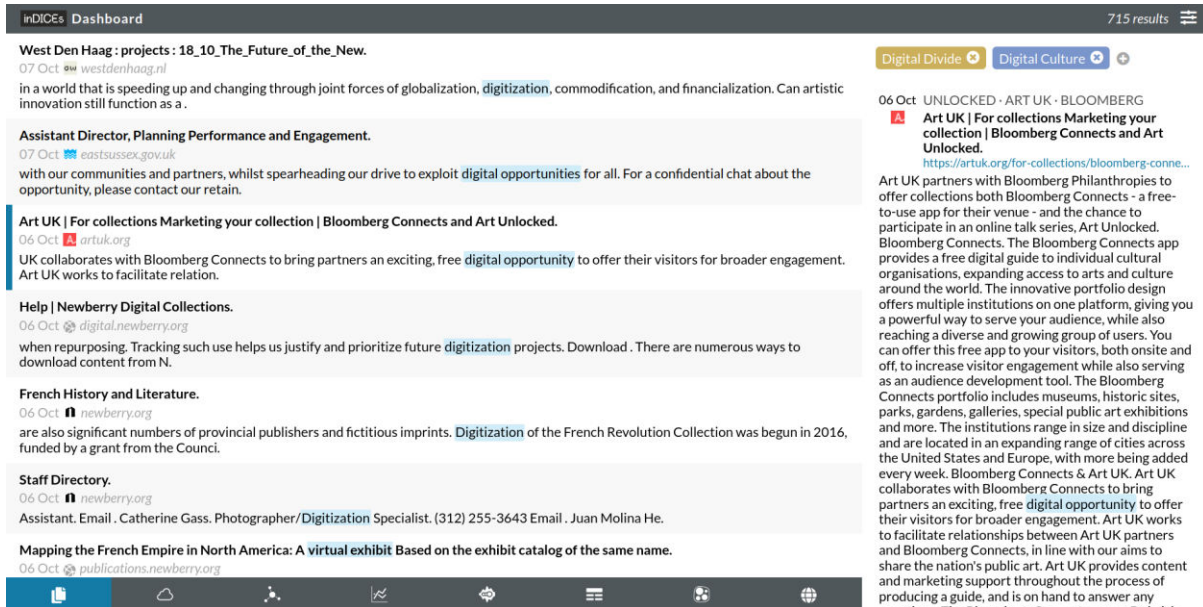
Fig 29: Content Area of the inDICEs Dashboard LITE
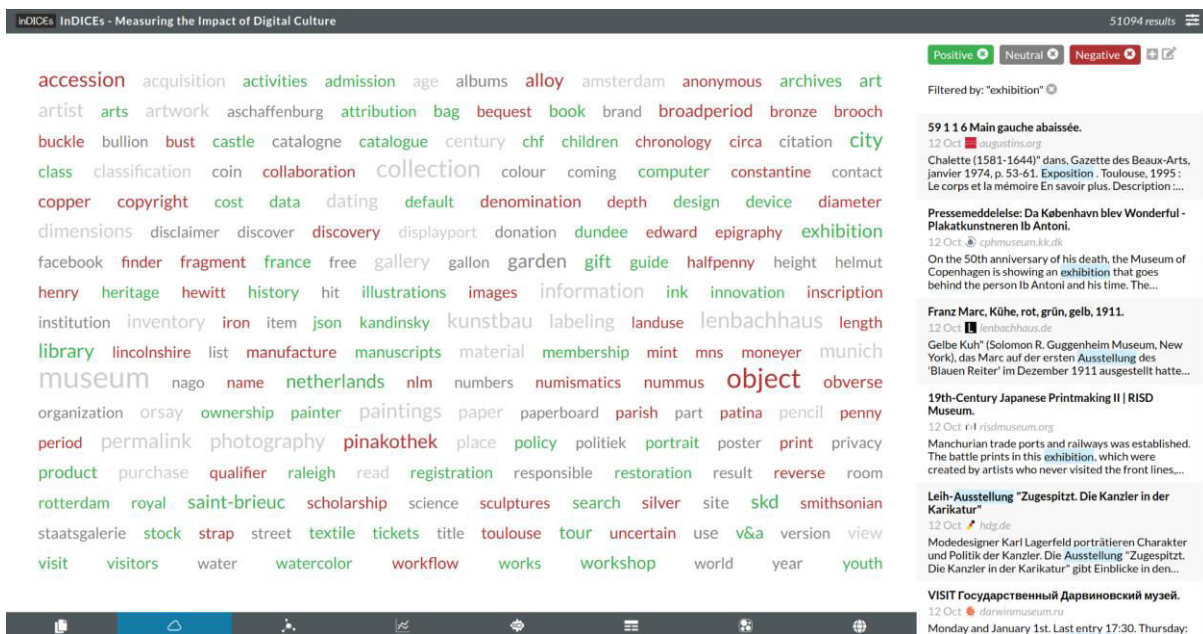
Fig 30: Colour-coded Tag Cloud showing polar sentiment of keywords occuring in websites from sources in the "Museums" sample filtered by the keyword "exhibition".

The default selection of the inDICEs Dashboard LITE contains all documents processed from websites of GLAM institutions as well as all websites that were assigned to the "art" and "film/theatre" categories. Additional sources, either from News categories or social media need to be explicitly included by the user via the configuration menu that offers settings such as time range, content sources and additional filters.

Figure 30 shows the tag cloud, colour-coded by polar sentiment for content coming from all websites from the "Museums" source sample. As visible in the right sidebar an additional filter for the keyword "exhibition" was applied (by clicking the term in the tag cloud). The

results in the right sidebar highlight the term as it was detected in the different documents in multiple languages.

4.2 Dashboard LITE in the Open Observatory

Initially the "Dashboard LITE" as the integration of the dashboard into the Open Observatory was envisioned as the visual display of individual widgets, connected by explanations and selection tabs. During the development this was replaced by the idea of integrating the completely remade "Mobile/LITE" version of the webLyzard Visual Analytics Dashboard as an iframe, since it already offers the desired options to switch between visualisations and inherently integrates additional search functionalities and features. Figures 31 and 32 show mockups of the evolution of the Dashboard LITE integration in the Open Observatory.
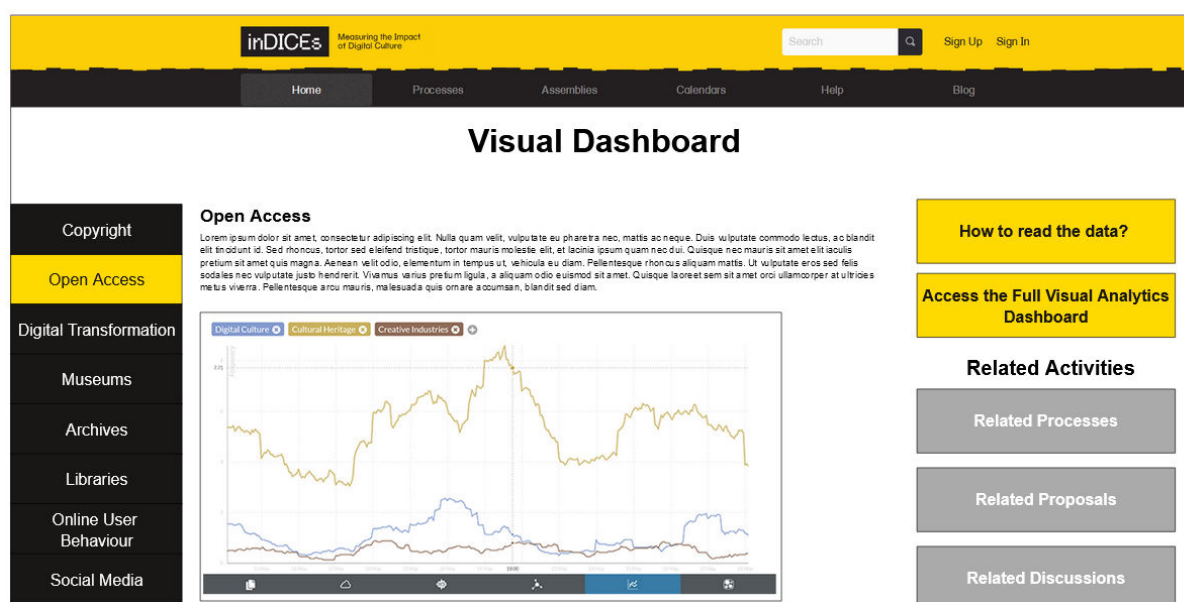
Fig 31: First mockup of the embedded "Dashboard LITE" component for the inDICEs Open Observatory as reported in D4.2

Decolonizing GLAMs. A comparison between ingested content from the GLAM sector on *Decolonization* and on *Diversity*.

The following regular expressions were used to model the *Decolonization* topic: decoloni(s|z)ation, decoloni(s|z)e, colonialism, decolonial(ity)?, post-colonial, reparations?, restitution, loot(ed|ing), imperialism, colony, raci(sm|al|st). Additionally, content was focused on GLAMs and any documents with mentions of "Ukraine" were excluded to remove the articles related to the current conflict.

The following regular expressions were used to model the *Diversity* topic: (cultural|gender|sexual) (diversity|sensitivity), monoculture, diverse cultural perspectives?, culturally diverse, inter(-)?cultural. The Diversity topic additionally includes all results that contain both words diversity and inclusion and focus on GLAMs, culture or heritage.
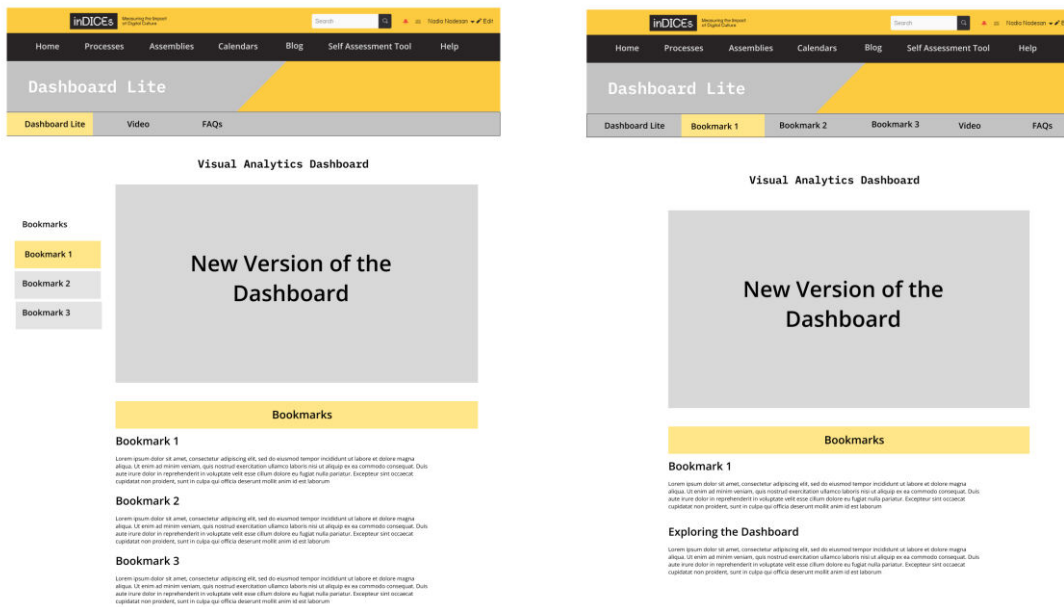
Fig 32: Mockups of the embedded "Dashboard LITE" component showing bookmark selections in a sidebar menu (left) or in the top menu (right).
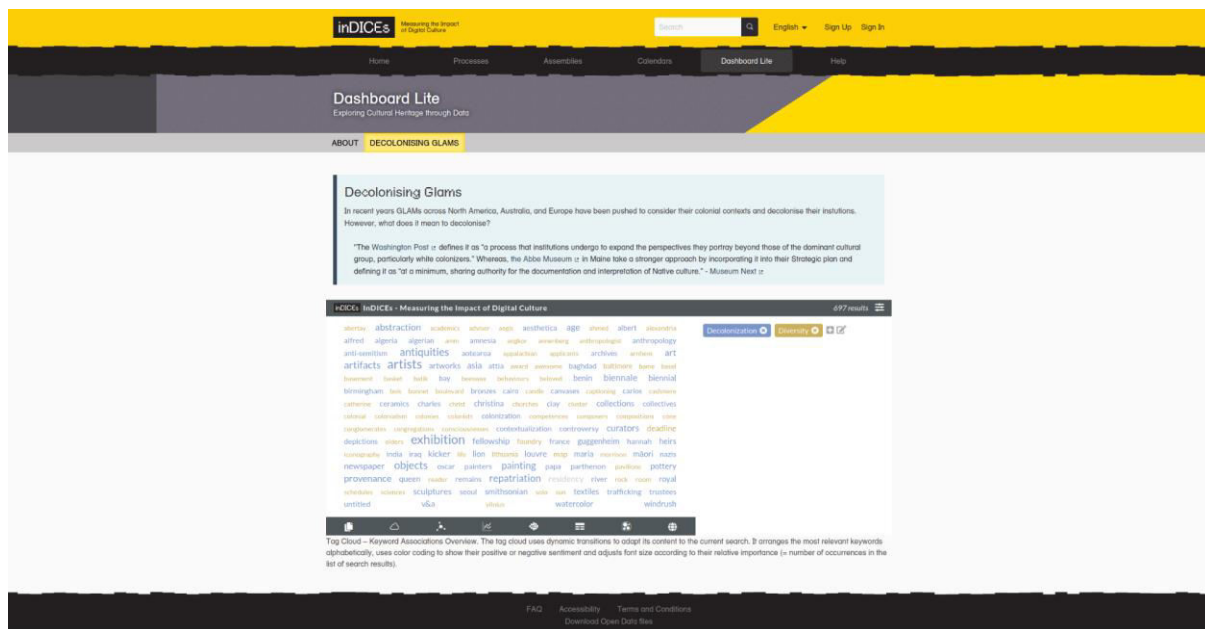
Fig 33: "Dashboard LITE" Integration in the Participatory Space of the Open Observatory.

Figure 33 shows the latest version of the embedded Dashboard LITE with the bookmark selection in the top menu. The predefined selections are planned to include the following bookmarks:

Digital Culture. The *Digital Culture* bookmark was further improved and project partners could vote on phrases they found fitting, and suggest new terms to add. The final bookmark

focuses on any type of virtual, augmented, or interactive tours, digitised collections and assets, and digital participation and community.

Creative Industries. This bookmark was also refined, with the help from project partners, from the initial version reported in D4.2 and includes terms related to the creative sector, i.e. arts, dance, music etc.

GLAMs. Another bookmark will be set up that can be used to filter documents that mention any of the following: glams, archives?, museums?, library(y|ies), galler(y|ies). This can then be combined with other relevant topics, such as bias, gender etc. and also used with the News or Miscellaneous corpus.

Impact Areas

The eight impact areas determined by WP1 and presented in *D1.7 Guidelines for the best practices regarding the maximisation of the impact of digitisation of cultural heritage* could be represented as bookmarks to be displayed in the Dashboard LITE.

- Innovation and Knowledge
- Welfare and Well-Being
- Sustainability and Environment
- Social Cohesion
- Entrepreneurship
- Learning Society
- Collective Identity
- Soft Power

Discussions on whether to include this in the final version are still ongoing. Here the proposed collection of phrases for the first Impact Area *Innovation and Knowledge*: innovation, co-creati(on|ng), capacity building, co(-)?operative learning, collaboration (initiative)?, knowledge (acquisition|creation)?, creative (production|collaboration), participation, business development, entertainment platform.

The bookmarks that can be selected in the top menu will be further finalised in the last months of the project based on the feedback collected by WP4 through the work of project partners in WP1, WP2 and WP3 as described in Section 2.3. Among the topics discussed was also *Copyright*, which is more challenging to put into the correct context. The bookmark configuration is set up in such a way that it is straightforward for WLT to modify bookmarks and create new ones. If permission is given, any bookmark can then be directly accessed in the embedded iframe using its defined name in the passed request parameters.

As an entry point, the latest implementation features an "about" section, explaining the Dashboard LITE along with a 10-min video tutorial created by WLT for inDICEs, that demonstrates the dashboard's main capabilities. The embedded inDICEs Dashboard LITE will target policy makers and CH professionals to provide an immediate overview of the most relevant trends in the sector. As it is integrated directly into the Participatory Space, as an integral part of the Open Observatory, users can directly start discussions or point others to their findings on the same website.

Dashboard LITE Iframe

The Dashboard LITE is fully embedded as an iframe. Listing 3-5 show examples of different iframe configurations.

Additional URL parameters allow for flexible initial configurations. As the dashboard is interactive the user can adapt and make any changes to the initial configuration once the iframe embedding has loaded. The following URL parameters are supported:

- Sources /source=xxx,yyy,zzz
- Pills (pre-selected search configurations) /pills=xxx,yyy,zzz
- Date Range /date=yyyy-mm-dd,yyyy-mm-dd
- Language /language=xx,yy,zz

Listing 4 shows a configuration that initially includes the four different GLAM sources and visualises two of the predefined bookmarks (*creative industries* and *cultural heritage*). Content of all 6 languages is included.

Listing 5 shows how to point to a different initial visualisation using the hash mark (**#**) separator and an URI fragment that points to the desired visualisation. After the initial configuration is loaded the user can still switch normally between the different visualisations using the tabs at the bottom of the dashboard.

```
https://api.indices.weblyzard.com/embed/ayBVrdTaUoabsUQ6fPp9zkxK8WrTV2Fg/lite
```

Listing 3: Basic URL to the inDICEs Dashboard LITE embedding without any additional configuration.

```
https://api.indices.weblyzard.com/embed/ayBVrdTaUoabsUQ6fPp9zkxK8WrTV2Fg/lite/source=gall
eries,libraries,archives,museums/pills=creative-industries,cultural-heritage/date=2022-
10-01,2022-10-31/language=en,de,fr,es,it,nl
```

Listing 4: inDICEs Dashboard LITE embedding URL with additional configuration parameters.

```
https://api.indices.weblyzard.com/embed/ayBVrdTaUoabsUQ6fPp9zkxK8WrTV2Fg/lite/source=news
/pills=copyright/date=2022-06-01,2022-09-01/language=en#story-graph
```

Listing 5: inDICEs Dashboard LITE embedding URL with initial focus on the Story Graph.

4.3 Embeddable Widgets

As reported in D4.2, selected visualisations of the Dashboard LITE are also provided as embeddable, stand-alone iframes and have been utilised by the inDICEs project to give some additional context to proposals and user contributions on the Participatory Space. These iframes are provided by the webLyzard Visualisation API and can be integrated in any third-party Web application.

There are three types of widgets: (1) Template-based widgets, (2) Thread-based widgets and (3) Similarity-Based Widgets. The last finds usage in inDICEs as additional contextualization for the content generation user interfaces in the Participatory Space on the Decidim platform. To accomplish this the title of the proposal is utilised as seeding

keywords. Through on-the-fly clustering similar documents are then retrieved and the visualisations generated based on the document results.

The initial plans to also include thread-based widgets to visualise longer discussion threads were ultimately not implemented because the user generated content did not fit the requirements, since a critical-mass of comments needs to exist in a single discussion thread – as often found in more traditional forums – for this feature to bring meaningful results.

The widgets themselves have been improved to be faster and more stable. With the continuously growing content in the inDICEs document repositories, the visualisations always show insights about current topics, visualising key terms and geo-spacial distribution around the discussed subject. The GeoMap widget has been improved to be more stable and have more responsive and fluent zoom and dragging functionality. Additionally, the displayed geo locations were adapted to not only include one main location per similar document, but to visualise all aggregated occurring geo entities.

Figure 34 shows examples of the latest state of the similarity-based embedded visualisations using "museum" and "digital" as seed terms, aggregating News and Web content in all supported languages between June and October 2022. The colour-coding is by polar sentiment (green-positive, red-negative, grey-neutral).
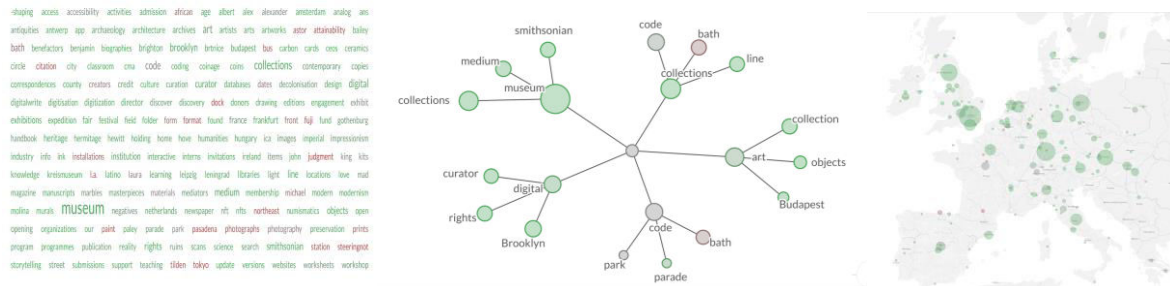

Fig 34: Embeddable Visualisation Widgets for a "similar-to" search.

5. Summary and Outlook

Deliverable D4.3 presented the final development state of the *inDICEs Open Observatory*, consisting of the *Participatory Platform,* the *Self Assessment Tool (SAT)* and PRO and LITE versions of the *inDICEs Visual Analytics Dashboard.* In the second reporting period, development efforts focused on platform integration including the finalisation of the Single Sign-On feature, the development of the inDICEs Dashboard LITE, and implementing the Self-Assessment Tool (to be reported in more detail in D4.6).

All components will continue to be accessible beyond the end of the project to help CHI practitioners and policy makers to interconnect, inform themselves and better understand the drivers of emerging trends. The tools developed for the Open Observatory will be available within the Decidim open source community, except the Dashboard which will remain accessible but represents proprietary code of partner WLT.

As part of an active community, the InDICEs project is not just a contributor but also benefits from updates and features developed by other community participants. In February 2023, Decidim will produce an updated, more accessible user interface that the InDICEs project will be able to adopt. In addition, in terms of the cultural heritage networks and community, during the past Europeana conference in September 2022, the exhibition of the InDICEs project was able to show the replicability of tools that will be at the disposal of CHI communities, for example the *Europeana Impact Task Force*. As future administrators of the platform, several members of Europeana are autonomous administrators of the platform using it for spaces such as task forces and core practices such as the Enumerate survey.

In the following, we briefly discuss some learnings and ideas that emerged from inDICEs, which we wish to further explore in potential future projects.

Sharing of Content & Insights. In regard to dashboard features, the post-project exploitation efforts of WLT will further develop the concept used for the URL parameter configurations of the Dashboard LITE. Extending the existing framework will make it straightforward to share a full dashboard setup, and to jointly view it together with other users via a single link. This would facilitate the sharing of search results and interesting findings and generate value in communication support - currently, only PDF reports generated based on search queries can be downloaded and shared, which can make it difficult to recreate the initial search configuration once an interesting result was found.

Semantic Data Integration. It would be valuable to link statistical data with knowledge graph entities in a sustainable manner. Ideally, data points would not be stored as additional metadata in the knowledge graph, like it was done for the visitor numbers of CHIs provided by Wikidata, but to automatically link knowledge graph entities to statistical datasets that can be continuously updated with the latest data points. WLT already supports the creation of new entities and the ingestion of statistical data along a timeline. A link between those would open new possibilities and allow the development of additional overlays for the various visualisations to show statistical data and various peaks and fluctuations. This idea grew from indicators discussed throughout the project, e.g. the number of digitised collections per cultural institute, or the number of social media followers and visitors for a museum. Unfortunately, statistical datasets with a critical mass of data points across a given time range are hard to find and expensive to gather.

Platform Evolution. In terms of the entire InDICEs platform and its post-project exploitation, the aspiration is that it will be continuously updated by adding new features, continuously improving the user experience and incorporating updates from the community as well as other R&D projects. We hope that groups such as the *Impact Task Force* will not only make active use of the provided tools, but also help us spread the word in order to create and maintain sustainable online communities around the developed technologies.