

# OPTED

## Report on Producing a prototype platform for the OPTED project

Susan Banducci, Michele Scotto di Vettimo, Shota Gelovani, Yannis Theocharis & Prasanna Dhamal



## Disclaimer

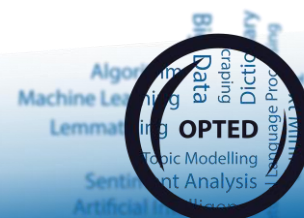
This project has received funding from the European Union's Horizon 2020 Research & Innovation Action under Grant Agreement no. 951832. The document reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained herein.

## Dissemination level

Public

Type

Report



## **OPTED**

Observatory for Political Texts in European Democracies:  
A European research infrastructure

# **Report on Producing a prototype platform for the OPTED project**

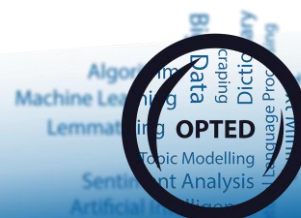
## **Deliverable D9.4**

**Michele Scotto di Vettimo<sup>1</sup>, Susan Banducci<sup>1</sup>, Shota Gelovani<sup>2</sup>, Yannis  
Theocharis<sup>2</sup> & Prasanna Dhamal<sup>1</sup>**

<sup>1</sup> University of Exeter

<sup>2</sup> TU Munich

**Due date:** September 2022



# 1 Executive Summary

## 1.1 Purpose of the production of a prototype platform

The OPTED platform aims to empower interested audiences to easily find the political text data and analysis tools to promote understanding of challenges to democracy. Previous reports WP9 presented our initial work and results on the framework for a curated resource classification system, mapping the principles that will help us organise, classify, and link the appropriate resources including text sources, documentation, software packages, tools and training materials that will feature on the OPTED platform. Additionally, the report of D9.3 covered, among other things, the preferences of potential users about functionalities that a platform dedicated to text analysis should possess, as well as the ways in which it could benefit their work.

In the project proposal, WP9 was tasked to produce a report detailing a prototype platform. In this report, we **present the main features of the prototype platform** as it has been finalised. The prototype strongly builds on the agreed curation workflow and on the feedback of potential users. Though it has been designed around the resources generated by work package 4 (Texts by political organisations), once all work packages collect their final data, the process of harmonisation will start checking (and adjusting if needed) the naming conventions and variables' format so that the prototype can easily be extended to the resources collected by other work packages. Most importantly, the prototype contains a functionality to allow users to upload files that can be stored on the platform, rather than only URLs directing users to other repositories. In this way, the prototype already contains most of the key functionalities that will be needed on the final platform.

Our report proceeds as follows. First, we outline the conceptual and technical design principles. We draw these principles from the conceptual work in D9.1 and D9.2, the taxonomy and the community-wide training needs survey (D9.3). Second, we detail the development of the prototype, briefly clarifying its technical aspects, the role that a shared taxonomy has in the harmonisation of the different resources collected by the work packages, and how the prototype meets the aims set out in the OPTED curation workflow (D9.2). Finally, in Section 4 we summarise the main results and outline the next steps necessary to develop a final version of the OPTED platform.

The main findings of this report are:

- The prototype meets the aims set out in the OPTED curation workflow. Namely, the available functionalities allow: (i) community members to find and integrate into their own research practices political text analysis resources; (ii) it allows community members who create resources to share these resources through depositing on the platform; (iii) it allows community members to link various resources conceptually to produce a better understanding of democracy.
- The development of the prototype has followed the conceptual development from D9.1 and D9.2 that has set out the criteria and process of classifying, (re)appraising and selecting appropriate resources for the platform.
- The use of a shared taxonomy across work packages allows a smooth and harmonised classification of the resources. It also facilitates their linkage and enhances the findability from the user's perspective.
- The prototype has been developed using an SQL database, with backend functionalities developed with C# and frontend ones created using Angular with Bootstrap 5, HTML and CSS.
- The prototype represents an intermediate step in the development of the final OPTED platform. Feedback from testers and responses from participants in the community-wide training needs survey will inform the development of the final platform by spotting features that can be improved and by identifying functionalities or overall aims that have to be prioritised.

## 2 Development of the OPTED platform

The design, development and testing of the prototype is a key step in the creation of the final OPTED platform. The final platform will strongly build on the structure and logic of this prototype, but also on the integration between the prototype and existing tools developed by other work packages for more targeted purposes (e.g. WP3's Meteor and WP7's AmCAT; see details below), and the suggestions we received from peers and testers on the prototype. The platform will have its own website - separated from the existing project website [opted.eu](http://opted.eu) - and will host a wider variety and number of resources.

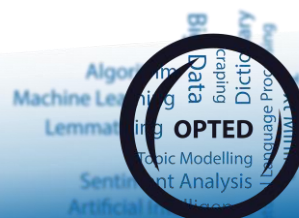
The conceptual framework for the OPTED platform was detailed in D9.1 where we defined the purposes of the platform to be:

- A Research Data Management System, which refers to the structures and processes for selecting, curating, storing, retaining, documenting, and sharing research data. These aspects of the platform are to be developed in collaboration with WP10 (e.g. the data management plan) and implement international standards for RDMS e.g. the DDI – Document, Discover and Interoperable.
- A Repository for different types of resources such as tools, data sources, documentation and training materials. The repository will consist of inventories created by all WPs that have been coded to a common OPTED taxonomy discussed below. WP3 has developed a searchable inventory (Meteor) that overlaps with the prototype we present in this deliverable. Below we discuss how Meteor (D3.2) can operate as a basis for the platform incorporating features of the prototype discussed in this deliverable.
- A Research Platform: “a set of digital resources—including services and content—that enable value-creating interactions between external producers and consumers.” AmCAT, the Amsterdam Content Analysis Toolkit (AmCAT), is an open source infrastructure that makes it easy to do large-scale automatic and manual content analysis (text analysis) and can serve to deliver an integrated research tool. One of the main deliverables of WP7 is (a new version of) the AmCAT infrastructure for document management and text analysis and will allow access to stored document, querying of stored documents and clients for analysis of documents.
- A Recommender System composed of software tools, algorithms and techniques, aimed at providing suggestions that would be of interest to users based on identified characteristics (e.g., area of research, type of analysis, etc.).

### 2.1 Aims, Objectives & Design Principles

In D9.1 and D9.2 we set out a framework for the OPTED platform (D9.1) and a workflow to curate the resources (D9.2). These conceptual models were then supplemented by a community-wide survey (D9.3). While these earlier reports provide detailed accounts of the aims, objectives and principles, we provide a brief summary here of the main findings of D9.1 and D9.2 that underpin the development of the prototype:

- The platform acts as a research management system, repository and recommender system that supports research using political texts to better understand the challenges facing democracy. We are particularly interested in designing a platform that promotes the resources of OPTED among those currently under-represented such as those seeking to develop computational skills and women.
- The resources that are part of the platform are managed through the curation workflow proposed in D9.2.
- Our objectives are to automate processes where possible, avoid duplicating efforts where protocols and systems exist elsewhere to result in a platform that is flexible.
- These systems and protocols include F.A.I.R, the European Code of Research Integrity and DDI – Document, Discover and Interoperate (<https://ddialliance.org/>).
- In Table 2.1 of D9.2 we set out the criteria for appraisal and high-level classification of resources for the platform.
- The taxonomy (see 3.1 below) provides a lower level system of classification and linkage across the resources and will support linkages at the conceptual level for the study of democracy.



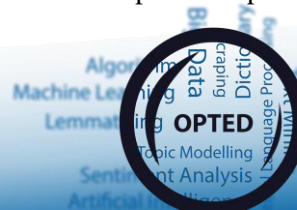
## 2.2. User Views and Community Building

As part of a community-wide training needs survey (D9.3), we asked our respondents how a platform for sharing text analysis resources could benefit them, and what kind of features they would like to see implemented on it. More specifically, we listed different aspects and functionalities which could be included in the platform, and asked respondents which ones should be prioritised and also allowed them to comment on possible additions so as to gather relevant external suggestions. Table 2.1 shows the percentages of respondents (out of 198 participants giving a valid response) highlighting each aspect as a priority. It emerges that the three aspects that potential users want to see prioritised and that they would personally make use of are: an open repository for different types of data sources and relevant documentation, an open repository for different types of training materials, and an open repository for different types of tools, software and packages. On the contrary, the aspects that are not seen as priorities by users concern: a platform that can host discussions about text analysis resources, a platform that can be used to “re-appraise” existing resources, highlight and solve issues (e.g., bugs) and track updates, and a platform that can be used to work collaboratively on the discovery, creation and sharing of text analysis resources (e.g., codes or data).

**Table 2.1** WHICH ASPECTS DO YOU THINK DESERVE TO BE PRIORITISED TO BEST SATISFY THE USERS' COMMUNITY NEEDS, INCLUDING YOUR OWN?

|  | <i>Not a priority</i> | <i>Definitely a priority</i> | <i>Definitely a priority and I would make use of it</i> |
|--|-----------------------|------------------------------|---|
| An open repository for different types of <b>data sources and relevant documentation</b>   | 6.6                   | 30.3                         | 63.1  |
| An open repository for different types of <b>training materials</b>  | 6.1                   | 29.8                         | 62.1  |
| An open repository for different types of <b>tools, software and packages</b>  | 11.6                  | 21.7                         | 67.7  |
| A platform where users can find <b>recommendations for relevant resources</b> for their research   | 14.6                  | 37.4                         | 48  |
| A platform users can <b>contribute to with their own resources</b> (e.g., new databases, software or packages)                                     | 18.2                  | 45                           | 36.8  |
| A platform with tools and standards for the <b>validation</b> of computational methods   | 14.2                  | 33.8                         | 51  |
| A platform that can be used to work <b>collaboratively on the discovery, creation and sharing of text analysis resources</b> (e.g., codes or data) | 19.7                  | 32.3                         | 48  |
| A platform that can <b>host discussions</b> about text analysis resources  | 30.3                  | 41.4                         | 27.3  |
| A platform for where users can <b>access training materials</b> (e.g., videos, slides, Shiny apps)   | 11.6                  | 37.8                         | 50.6  |
| A platform that can be used to <b>“re-appraise” existing resources</b> , highlight and solve issues (e.g., bugs) and track updates                 | 24.2                  | 47.5                         | 28.3  |

Additionally, we also asked an open-ended question in order to receive more precise inputs on the features





that would benefit the user community the most. Among the most reported answers there is user friendliness, accessibility, a clear structure and good search functionality, and the possibility for users to leave comments (or “up and down votes”) under each resource. The more substantive comments related to the content of the platform, instead, touched upon various aspects such as: the breadth of the resources available, availability of tutorials and training materials, and a multilingual focus.

We also looked at whether priorities vary across different types of potential users, namely men versus women or novices versus more expert users. Table A1 and A2 in the Appendix report the share of responses to the 10 items for these different categories of respondents. Generally, the four groups identify a similar set of priorities, which of course is similar to those highlighted by the full sample. The only key, but not surprising, difference is that expert users list among the top three priorities also a platform with tools and standards for the validation of computational methods. On the other hand, novice users and women prioritise a repository for training material as well as an open repository for tools and software. Also, novices do not see as a priority a platform where users can contribute with their own resources or that can be used to find tools and standards for the validation of computational methods. The platform in terms of providing an infrastructure for the political text community should reflect these differences in priorities (e.g. to prioritise training) but also reflect that the community had a number of shared priorities for a platform.

### ***Recommendations and next steps:***

Based on the above results we recommend that the platform focus on the following areas:

- The platform should reflect the priorities identified in the community-wide survey: repository for training materials, tools, software (e.g. provide an inventory of these resources) but also, given the priority of women and novel users, provide access to the material through the sharing of videos, slides and Shiny apps. This recommendation is consistent with the early conceptual framework for the OPTED platform.
- We should also ensure the platform is “inclusive by design” -- i.e. accessible to all users including those with impairments and those with diverse abilities.
- The OPTED project should not invest in developing services on the platform that are aimed at hosting discussions about political text analysis, a collaborative workspace or a platform invested in providing in “fixing bugs” in existing programmes. On this last point, we need to distinguish between this type of “reappraisal” and the reappraisal of tools that we propose as part of the OPTED curation workflow.

## **3 Design and functionalities of the prototype**

The development of the prototype platform has been informed by two main goals, both set out in D9.2. Firstly, D9.2 set out the aim of creating a workflow for selecting and classifying the resources, which should allow the platform to fulfil the following objectives: (i) community members to find, compare and use (i.e., understand) the resources; (ii) members of the community, tool developers, data producers, producers of text corpora, and trainers can contribute their resources to the platform; and (iii) provide a set of standards for the political text community in the use of these resources. Secondly, previous deliverables have emphasised the need for presenting all the resources in a harmonised and systematised way, possibly putting all of them in a relationship between each other. This aim is achieved through the use of a common taxonomy, which will be exploited by the prototype platform as well as by the final platform.

Concretely, the current prototype has been developed using an SQL database, with backend functionalities developed with C# and frontend ones created using Angular with Bootstrap 5, HTML and CSS (the source code is available on GitHub at this [link](#)). Alongside the development of this prototype, WP3 developed [meteor.opted.eu](http://meteor.opted.eu), a searchable “database of news sources, organisations, datasets, corpora, tools” that has many of the features proposed for the prototype and OPTED platform (Balluff et al. 2022). Both the current prototype and Meteor will be used as starting points for the development of the final platform, which will be the result

of a cooperation between WP3 and WP9.

The resources displayed on and used to test the prototype are the ones related to political parties collected by WP4 (Texts by political organisations). Drawing on the taxonomy (3.1), re-coding has been conducted so as to harmonise variable formats and names across work packages, so that this prototype can easily include other resources as well.

### 3.1 OPTED taxonomy and linkage of the resources

The harmonisation of the resources collected by the various OPTED work packages will be conducted through the use of a shared taxonomy. This taxonomy has been developed with the purpose of identifying key variables that are common and can be collected for all (or most) types of resources shared on the platform and to harmonise their names and format across work packages. In this way, the various work packages contribute their resources to the prototype/platform without the need of adapting the latter to work package-specific conventions.

The provisional taxonomy can be accessed at this [link](#). A number of variables have been identified and are being harmonised across work packages. To minimise the additional work required to the work packages which have already created inventories or datasets, the harmonisation consists of a renaming and re-coding of existing variables, so as to meet the shared conventions and formats, conducted by work package 9. The most important ones are: type of the resource, a short description, name, author, countries covered, language of the resource, the years covered and the date of the last update, quality of the resource, information related to the access and license, URLs or DOIs, the concept being covered by the resource (e.g., party positions, issue salience), and a unique identifier. Additional variables that are captured are intended to allow researchers to link resources at the conceptual level to aid the analysis of dimensions of democracy. These variables include, for instance, those identified by WP8 (D8.2 “Inventory of Concepts and Measurements”) such as policy preferences, party priorities, agenda power, lobby goals and so on.

#### ***Recommendations and next steps:***

In order to implement the taxonomy we recommend the following:

- Working with WP8 to finalise the taxonomy and incorporate substantive conceptual categories that are core to understanding democracy.
- Integrate the taxonomy with the searchable database (e.g., the query form and the database format).

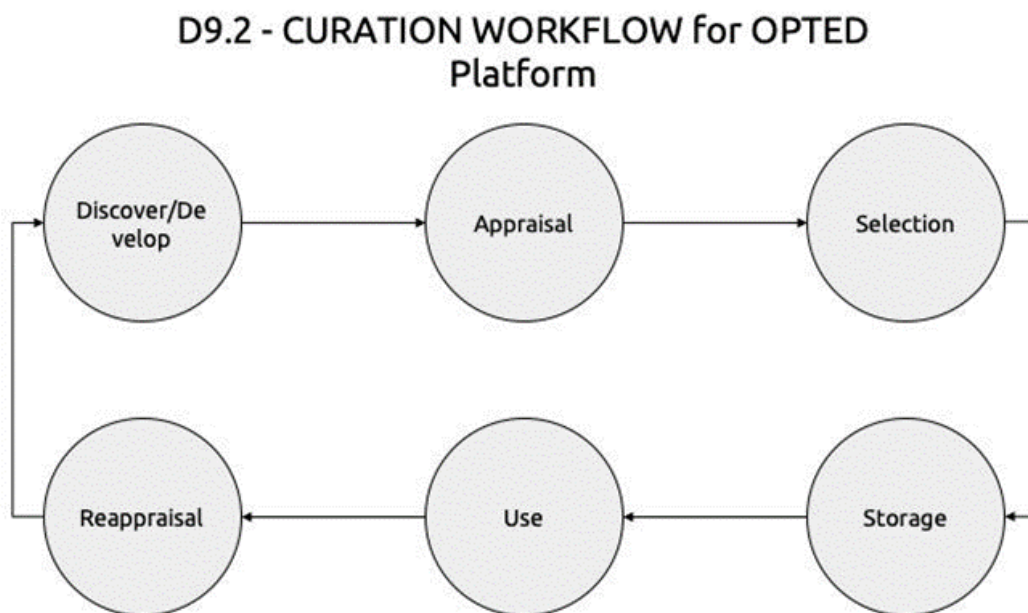


### 3.2 OPTED Curation Workflow

The OPTED curation workflow contains six stages. We replicate the workflow proposed in D9.2 below to illustrate the stages. Below we detail the stages of the workflow and how they have been or will be implemented in the prototype.

Figure 3.1

Provisional OPTED WP9 Workflow: Curation of Political Text Resources



#### 3.2.1 Discover and Develop

There are three main avenues for this stage: development within OPTED, discovery through inventories and deposit by members of the community. There are several necessary steps here to make the platform functional:

- Harmonise existing inventories according to the taxonomy and create a searchable database.
- Collate tools and resources developed by OPTED network.
- Create a mechanism to allow the deposit of resources (upon registration) by interested users.

Various OPTED work packages are generating inventories collecting existing resources or are generating new tools and resources which would feature on the final OPTED platform. However, in the prototype we add functionalities so that new resources can also be easily contributed to the platform by interested users upon registration (and login). These functionalities inform different stages of the workflow:

- i) Contributors are required to fill a submission form containing most of the key metadata. This step is specifically about aiding the discovery of new resources.
- ii) Users with administrator rights have to review and approve (or reject) the resource, with the possibility of sending comments and requesting amendments to the contributors before approving the resource (Selection stage, section 3.2.2)
- iii) Various validation channels have been put in place to make sure the information provided is correct (Appraisal stage, section 3.2.2).

Figure 3.2 shows part of the submission form that users have to fill in order to contribute resources to the platform. In general, the form covers key variables used for the classification and linkage of resources as well as fields used by OPTED administrators to assess their quality (see Appraisal below). In terms of key variables, apart from those shown in the figure, users are also asked to provide a name and short description of the new resource, a working link (if it exists), and information about the years, entities, countries and languages covered. Not all this information is compulsory, so resources without, for example, an indication of the languages used can still be uploaded, but administrators and contributors will make sure that all the relevant metadata are updated at the (re-)appraisal stage.

Figure 3.2 Submission form to be filled to contribute new resources to the platform

ADD RESOURCES

Source Type Select Type

Sub Type Enter Sub Type

DOI Enter DOI

General Concept Enter General Concept

Programming Language Enter Programming Text

Next →

Figure 3.3 Section where contributors can attach files to the newly contributed resources

Manage Resources

Manage Users

Add User

Upload Resources

Select a file: Choose File No file chosen

Upload

### **Recommendations and next steps:**

To implement 3.2.1 (Discovery), we recommend the following steps:

- To integrate current D9.4 prototype with Meteor, so as to avoid duplications. This means to keep Meteor’s functionalities and features whilst also adding those developed in the prototype which are not currently implemented in Meteor.
- Update the submission form to include all types of resources and adapt the fields to the needs of these various types.
- Update the submission form to cover all criteria as detailed in Table 3.1 (below).

### 3.2.2 Classification and Appraisal

At this stage the resource has to be classified according to the predefined schema introduced in Section 2.1 and then assessed so as to ensure that it complies with the agreed standards, is relevant for the community we aim to serve, and it is appropriately annotated so as to be easily discovered by users. Resources identified by the OPTED network have been already classified and appraised by network members.

The criteria to be included in the submission page and in classification of the existing resources include, at the highest level, the ones listed in D9.2 and reported in the table below and concern the standards for research integrity and the respect of FAIR principles.

| <b>Table 3.1</b> |   | <b>CRITERIA FOR THE APPRAISAL AND CLASSIFICATION OF RESOURCES FOR THE PLATFORM</b> |           |
|------------------|---|--|-----------|
| Category         | Criteria                                      | Classification   | Appraisal |
| FAIR             | Interoperable, open, discoverable             |  | X         |
| Domain Relevance | Text as data workflow                         | X  |           |
|                  | Usable with political texts as defined in WPs |  | X         |
| Reproducibility  | Completeness                                  | X  |           |
|                  | Transparent                                   |  | X         |
| Usability        | Level   | X  |           |
|                  | Quality [functioning URL – working link].     |  | X         |

Once a resource has been submitted (3.2.1), it is not immediately visible in the platform. Rather, it is put on hold awaiting a decision (whether to accept or reject it) by the OPTED administrators. In particular, Figure 3.4 shows what administrators see when a new resource has been submitted for approval.



Figure 3.4

Section where administrators can view and manage newly submitted resources

|                      | Source Name              | Description                        | Source Link                         | Country Name                        | Language                            |
|----------------------|--------------------------|------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| <a href="#">View</a> | wikidata                 | Query to get links to party web... | https://query.wikidata.org/#SEL...  | Germany                             |                                     |
| <a href="#">View</a> | AUTNES                   | Data set containing content an...  | https://data.oussda.at/dataset.x... | Austria                             | de                                  |
| <a href="#">View</a> | Political Party Database | collection of party statutes       | https://www.politicalpartydb.or...  | Austria, Belgium, Croatia, Cypru... | en, de, es, fr, nl, pt, hr, el, cs, |
| <a href="#">View</a> | AUTNES                   | Data set including a content an... | https://data.oussda.at/dataset.x... | Austria                             | de                                  |
| <a href="#">View</a> | Friedrich Ebert Stiftung | Database of press releases fro...  | https://library.fes.de/pressemit... | Estonia                             | et                                  |
| <a href="#">View</a> | AUTNES                   | Data set containing content an...  | https://data.oussda.at/dataset.x... | Austria                             | de                                  |

The figure shows a list of seven resources submitted to the platform and awaiting administrators' approval. Some key information is displayed in various columns (administrators can decide which information they want to be displayed there) and, by clicking on "view", administrators can access all the metadata submitted along with the resource.

Figure 3.5

Details of a newly submitted resource with decision buttons and comment section for administrators.

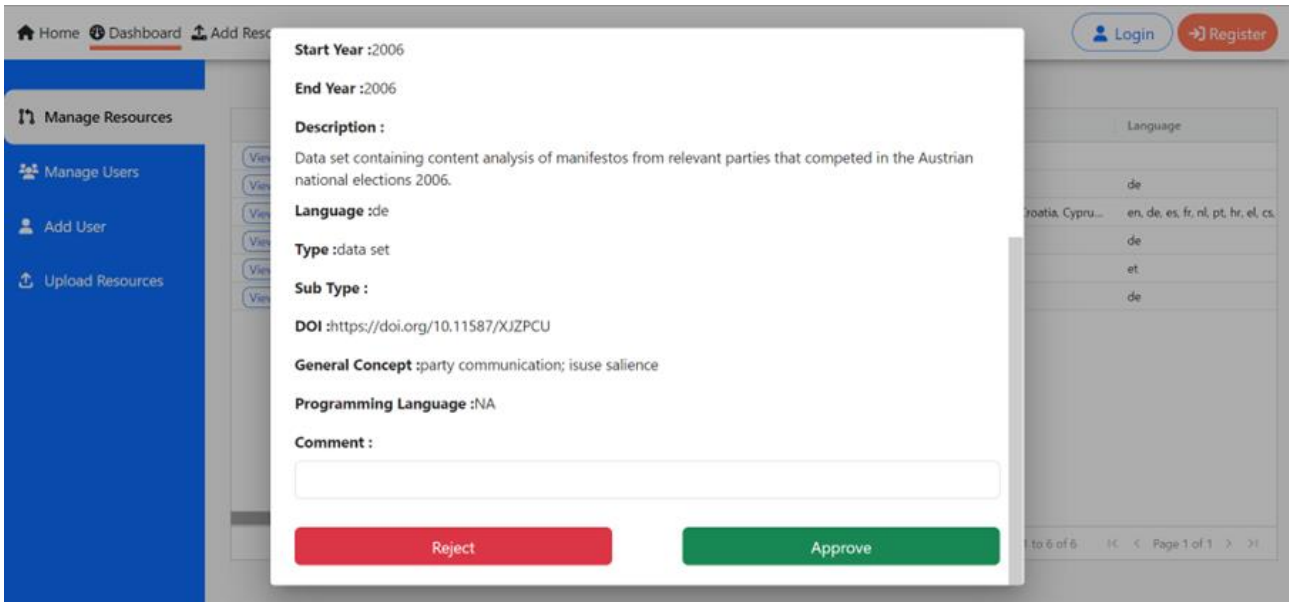


Figure 3.5 shows the detail of a specific resource once the “view” button is clicked. Along with the details added by contributors, users with administrator rights have here to make a decision as to whether to approve or reject the resource, with the possibility of sending comments to the contributors asking for amendments or corrections in the information provided. This effectively leads us to the selection stage of the workflow.

### ***Recommendations and next steps:***

In consideration of the Meteor resources, the following consideration and steps can be taken:

- Ensure Meteor can classify tools according to the criteria listed in Table 3.1. This will need to be operational for both the submission of new resources and the classification of existing inventories that is achieved with the taxonomy.

### 3.2.3 Selection

As we indicated in D9.2, selection means that the resource is approved for the inclusion on the platform and key metadata (e.g., a DOI) and functionalities (e.g., instrument to track download) are added to the resource. At the moment, the selection stage occurs simultaneously with the appraisal stage. Once the administrators are shown the newly contributed resource with all its metadata (Figure 3.5), they will be able to make three types of decisions: accept the resource as it is, request amendments or corrections to the contributor, or reject the resource. In all cases they are allowed to easily communicate with the contributors through a dedicated “Comment” section.

At the moment, we ask contributors to fill a DOI field (if the resource has a DOI already) but do not provide a DOI directly. As per the information offered on the [website](#) of the DOI system, DOIs are provided only by an approved Registration Agency (RA). Hence, to provide DOIs either OPTED applies for becoming one of such RAs or contacts one whose services best meet our needs.

Additional functionalities like the download tracker can be added, but they have also to take into account the fact that a resource can be shared on the platform through an URL, but the download of the actual files is made from that link and not from the OPTED platform.

In terms of validation of the information provided by the users, three different strategies have been put in



place to aid the administrators during the selection stage. First, when possible (e.g., for URLs or DOIs), automatic forms of validation will be preferred. In this way administrators (or even the users before they finalise the submission) can be alerted that a field has been filled with an invalid value (e.g., a non-existing DOI). Secondly, as shown in the screenshots above, users will be asked to self-assess their resource with regard to key inclusion criteria and their responses will help administrators in deciding on the selection of the resource. Finally, human-checking by the administrators will deal with the remaining fields which cannot be easily validated automatically.

### 3.2.4 Storage

As mentioned with regard to the download tracker functionalities, some resources available on the prototype and on the final platform will not be actually stored on it, in which case the contributor will provide a working URL where the resource can be accessed. However, the prototype offers the possibility of adding files (e.g., PDFs, datasets or other forms of multimedia materials like videos) at the moment of the submission of the resource (Figure 3.3). These files will be then hosted on the platform, and we will make sure that contributors of resources can count on the OPTED platform for getting credit for their role as “generator” of resources. For instance, the platform will also provide predefined and standard ways for referencing each resource. This should ensure that contributors are adequately credited when their resources are being used.

In light to avoid a duplication of efforts within the OPTED network, further developments of the storage capacity of the platform have to take into account the fact that for some types of data (e.g., documents with document-level metadata and annotations), other tools have already been developed that could better accommodate the needs of the final platform. For instance, particularly useful for the OPTED infrastructure is the ability for “non-consumptive” research – computational analysis is performed on texts in their stored location. This would allow the platform to ease the requirements and copyright-related implications for the storage of some resources, while also offering users a valuable tool for conducting quantitative text analysis. However, neither the prototype nor Meteor represent a valid solution for this task.

Therefore, with regard to the storage of the resources, we propose to integrate the final platform with the [AmCAT](#) tool developed by work package 7. AmCAT is a non-SQL database and can complement what can be achieved in terms of storage and access to the resources through the OPTED platform. In particular, AmCAT, as developed for OPTED, can provide the following :

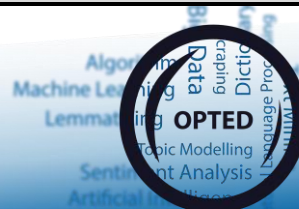
- A backend based on elasticsearch that provides API access to stored documents;
- A pre-made query front-end that allows management, exploration, and querying of documents stored in a backend;
- Components for custom front-ends/clients/analyses in R, Python and React.

We see the integration of AmCAT in the OPTED platform also as a way to make the content of the latter more accessible to users that are not (yet) very experienced, but that nevertheless have an interest in exploring OPTED resources without the need of actually accessing and processing the raw data. Similarly, in view of offering training materials to prospective users, the possibility of searching through textual material not stored on the platform (for instance by looking for words occurring in a text) can facilitate the access to academic articles, documentation or other potentially relevant textual material without restricting the search filters to the metadata attached to the resource whilst also being compliant with copyright rules.

#### ***Recommendations and next steps:***

To implement 3.2.4 (Storage), we recommend the following steps:

- To integrate the current D9.4 prototype with Meteor. At the moment, the prototype can accommodate different types of resources in terms of formats (e.g., databases, texts or even multimedia material), whereas most resources on Meteor come in the form of URLs or databases. The integration should allow Meteor to host other types of resources as well.



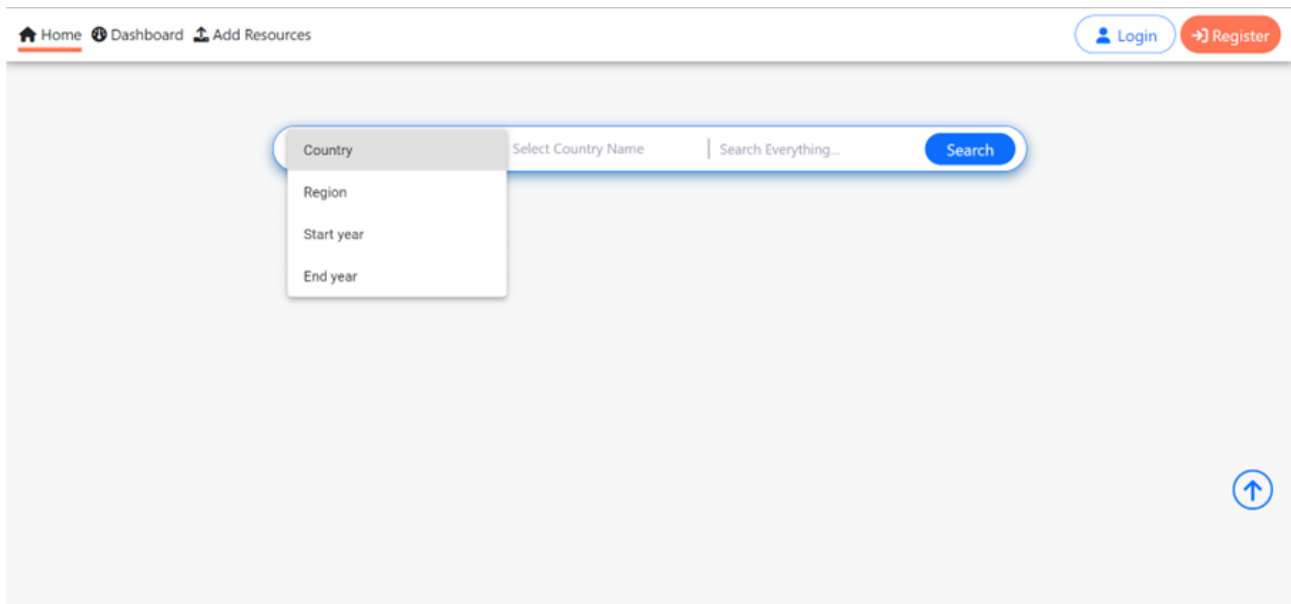
- Integrate AmCAT as a tool for the non-consumptive method for querying, analysing and annotating texts not stored on the OPTED platform. This is also relevant for the discussion on use below (3.2.5).

### 3.2.5 Use

This stage of the workflow aims at engaging users with the resources registered on the system. First, this requires the creation of a searchable database with an easy-to-use interface. In the home page users (whether they are registered or not) can search for resources on the prototype by entering some key words or by entering specific values of key variables in dropdown menus or filtering fields. At the moment, the search bar offers the possibility to add a string of words that will be searched among the names and description of the resources as well as to select filters over key resource metadata. Figure 3.6 shows how users can select among four key variables from a dropdown menu so as to refine their search.

Figure 3.6

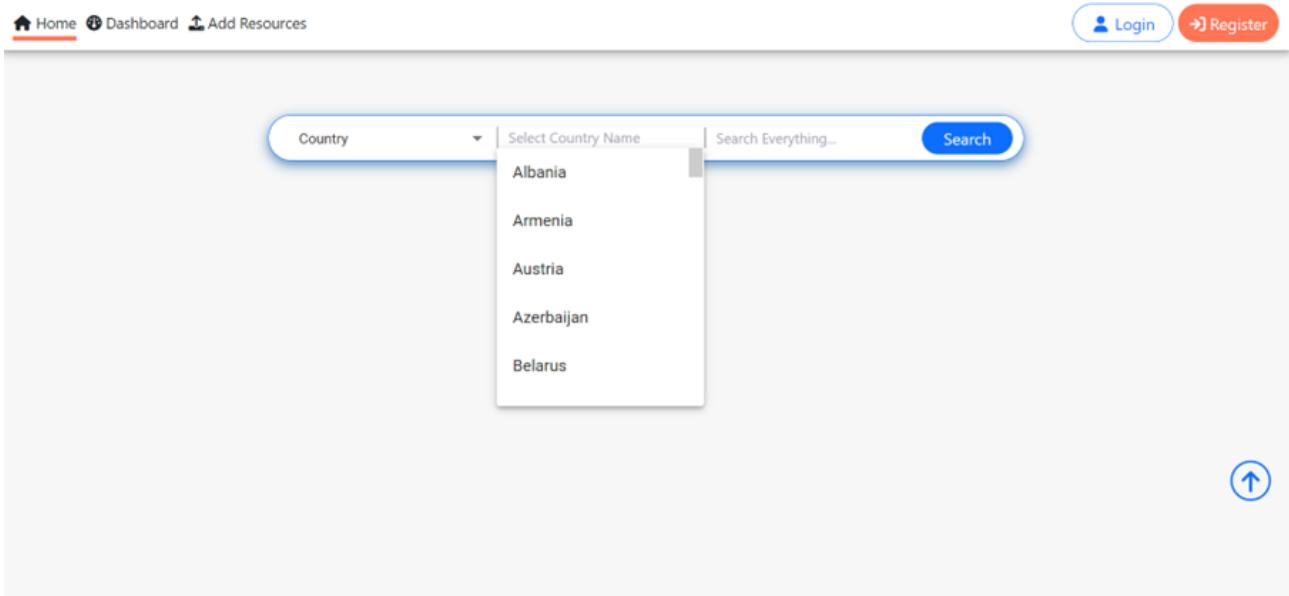
Home page with search functionality for browsing OPTED resources, with four filtering variables



Once a filtering variable has been selected, users can then select from a dropdown menu the relevant value (or values) they are interested in (Figure 3.7), so as to restrict the search results they get only to those resources matching the value of the filtering variables in their metadata.

Figure 3.7

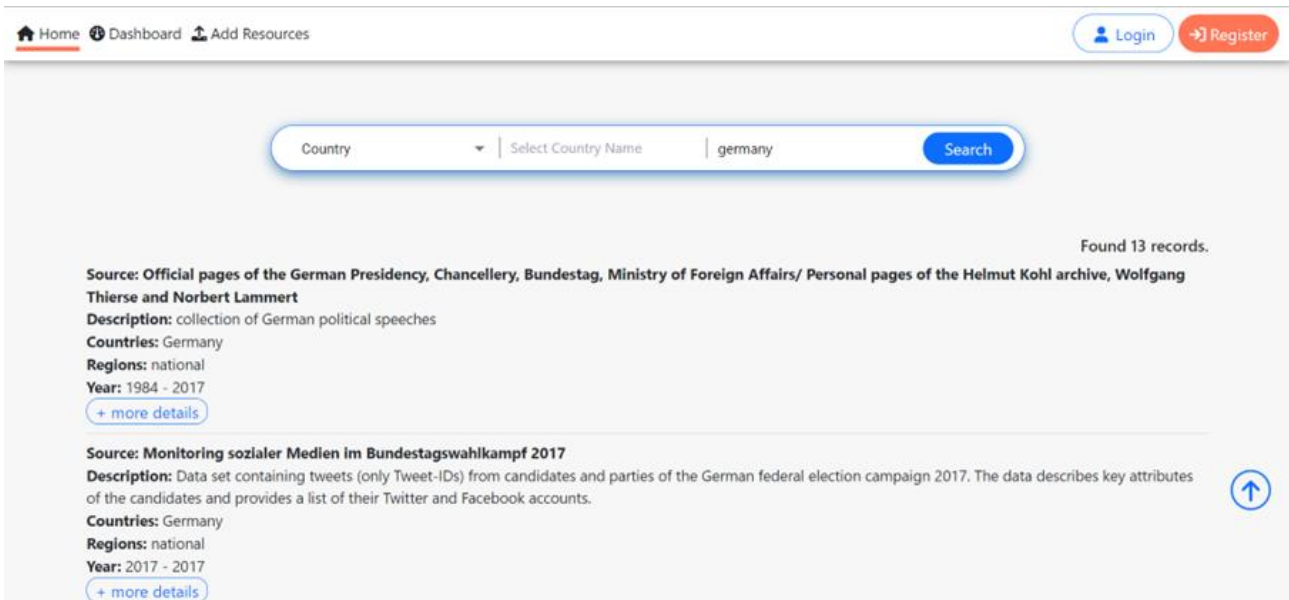
Example of dropdown menu for the selection of search filters



Finally, Figure 3.8 shows how the search results are displayed. Users see a list of results matching their search criteria, and can easily access key information such as the name of the source, a brief description of its content, information about the country (or countries) and year(s) covered. Furthermore, by clicking on the “+more details” button, users can also access the URL and DOI information of the resource as well as to other variables like the entities covered and the general concept (along the lines of those defined by WP8) being captured by the resource.

Figure 3.8

Example of the visualisation of search results



*Recommendations and next steps:*

To implement 3.2.5 (Use), we recommend the following steps:

- Integrate the prototype with Meteor and add more filtering fields in the main search bar. Ideally, allow fields to change according to the type of resources selected (e.g., if users search for tools, it might be more relevant to filter for programming language than for countries).
- When possible, the platform should leverage on existing APIs of resources stored elsewhere to facilitate the search and quick assessment of relevant resources.

### 3.2.6 Reappraisal

Once resources have been made available for users, the curation workflow contemplates a reappraisal stage. Reappraisal involves different considerations that might lead the administrators of the platform to conclude that the resource is outdated or, in more generic terms, not useful anymore. In general, the reappraisal step should conclude with a decision on whether to keep the resource or dispose of it. However, if needed this step could also have an intermediate outcome where curators ask resource contributors to address some problems related to the resource (e.g., update it or resolve bugs) before making the final decision on whether the resource is fit for re-use or it should be disposed of.

At the moment, no particular features have been implemented so as to deal specifically with this stage. Registered users are differentiated between those with administrator rights and those without. For what concerns the reappraisal stage, few questions remain open that should lead to consequential choices in terms of platform design:

- 1) Granted that the contributors of a resource should be able to update/edit what they have submitted, who is in charge for the reappraisal task: only administrators? All registered users with administrators having to approve suggested changes?
- 2) Timing of the reappraisal: Will there be a “scheduled” reappraisal (performed by administrators) at regular intervals after the first publication of the resource (for instance aided by a short checklist of elements that have to be reappraised)?

#### ***Recommendations and next steps:***

To implement 3.2.6 (Reappraisal), we recommend the following steps:

- Addition of functionalities aimed at facilitating a more automated reappraisal process (e.g. user star ratings and/or rankings).
- Add the possibility for users to submit reviews and post comments on the resource.

## 4 Summary and Next Steps

This report presented the main features of the prototype platform as it is being currently finalised. The prototype strongly builds on the agreed curation workflow and on the feedback of potential users, and already contains most of the key functionalities that will be needed on the final platform. In particular, the prototype meets the aims set out in the OPTED curation workflow. Namely, the available functionalities allow: (i) network members to classify, (re)appraise and select appropriate resources; (ii) to store resources on the prototype itself; and (iii) users to search and easily find relevant materials.

The prototype represents, therefore, an intermediate step in the development of the final OPTED platform. Feedback from testers and responses from participants in the community-wide training needs survey will inform the development of the final platform by spotting features that can be improved and by identifying functionalities or overall aims that have to be prioritised. We will trial the developed prototype and seek feedback from potential users at the Second Annual Exeter/NCRM Spring Computational Communications

School to be held April 2023.

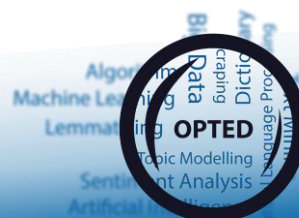
The key next steps leading to D9.5 – the production of a final platform due in 10 months – entail: (1) improving the prototype functionalities that have to be kept for the final platform and understanding which other key functionalities have to be implemented and how; (2) the implementation of any necessary change and improvement in terms of back-end infrastructure; and (3) getting user feedback on the platform during the development stage..

More specifically, with regard to (1), there are some additional functionalities that have been agreed upon but not yet been implemented in the prototype. We have detailed these functionalities in the concluding paragraphs of each section. These represent the less controversial points. However, in D9.2 we promised different features that, at the current stage, it is not clear whether we could include in the final platform as promised. For instance, becoming a DOI provider might entail unreasonable additional costs for the OPTED platform and, therefore, alternative ways of getting credit to contributors have to be devised. Similarly, a download-tracking function might not be relevant for all resources (e.g., the ones not physically stored on the platform). Hence, a different way of tracking the traffic can be implemented so as to fit better with all OPTED resources. Working with WP1 we will also need to develop a plan to disseminate and encourage use of the platform.

As regards to (2), it will be important to assess whether the existing backend infrastructure can be used as it is for the final platform or, rather, some changes have to be made so as to accommodate a bigger and more diverse set of resources. The choice of using a SQL-based database proved appropriate and flexible for the development of a working prototype. However, as the number, the diversity and the overall size of resources will increase, we have to consider the possibility that the current approach will represent a limit to the further and desired development of the OPTED platform and, if that is the case, to think about more appropriate backend structures. Additionally, this approach can be developed so as to complement, rather than replace, what can currently be achieved using a non-SQL database as AmCAT and, eventually has to be integrated with Meteor and AmCAT as outlined above.

## References

- Balluff, P., Lind, F., Boomgaarden, H., and Waldherr, A. (2022). Mapping the European media landscape - Meteor, a curated and community-coded inventory of news resources. *European Journal of Communication*, 1-14. <https://doi.org/10.1177/02673231221112006>





# Appendix

Table A1

WHICH ASPECTS DO YOU THINK DESERVE TO BE PRIORITISED TO BEST SATISFY THE USERS' COMMUNITY NEEDS, INCLUDING YOUR OWN? PERCENTAGES BY GENDER OF THE RESPONDENT

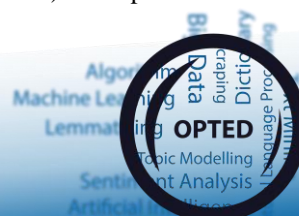


|  | <i>Not a priority</i> | <i>Definitely a priority</i> | <i>Definitely a priority and I would make use of it</i> |
|--|-----------------------|------------------------------|---|
| <b>Gender:<sup>1</sup></b>   | <i>Men- Women</i>     | <i>Men- Womens</i>           | <i>Men- Women</i>                                       |
| An open repository for different types of <b>data sources and relevant documentation</b>   | 9.7 - 2.4             | 32.8 - 25.9                  | 57.5 - 71.6   |
| An open repository for different types of <b>training materials</b>  | 12.3 - 2.4            | 30.9 - 25.9                  | 56.6 - 71.6   |
| An open repository for different types of <b>tools, software and packages</b>  | 16.6 - 4.9            | 23.6 - 18.5                  | 59.6 - 76.5   |
| A platform where users can find <b>recommendations for relevant resources</b> for their research   | 18.5 - 9.8            | 35.3 - 39.5                  | 46.1 - 50.1   |
| A platform users can <b>contribute to with their own resources</b> (e.g., new databases, software or packages)                                     | 19.4 - 17.2           | 41.5 - 50.1                  | 38.9 - 32   |
| A platform with tools and standards for the <b>validation</b> of computational methods   | 16.9 - 12.5           | 36.2 - 31.2                  | 46.9 - 56.2   |
| A platform that can be used to work <b>collaboratively on the discovery, creation and sharing of text analysis resources</b> (e.g., codes or data) | 24.7 - 13.5           | 30.9 - 34.5                  | 44.2 - 51.8   |
| A platform that can <b>host discussions</b> about text analysis resources  | 35.3 - 26.2           | 37.1 - 46.2                  | 27.4 - 27.5   |
| A platform for where users can <b>access training materials</b> (e.g., videos, slides, Shiny apps)   | 15.1 - 7.5            | 37.2 - 38.7                  | 47.8 - 53.7   |
| A platform that can be used to <b>“re-appraise” existing resources</b> , highlight and solve issues (e.g., bugs) and track updates                 | 28.3 - 19.8           | 43.3 - 53.1                  | 28.3 - 27.2   |

**Table A2**

**WHICH ASPECTS DO YOU THINK DESERVE TO BE PRIORITISED TO BEST SATISFY THE USERS' COMMUNITY NEEDS, INCLUDING YOUR OWN? PERCENTAGES BY FAMILIARITY WITH COMPUTATIONAL TEXT ANALYSIS**

<sup>1</sup> Only respondents identifying as either males (N=117) or females (N=82) are reported. Other respondents have been excluded from the table given the very low number (N=4).



|  | <i>Not a priority</i> | <i>Definitely a priority</i> | <i>Definitely a priority and I would make use of it</i> |
|--|-----------------------|------------------------------|---|
| <b><i>Familiarity with computational text analysis:<sup>2</sup></i></b>  | <i>Novice- Expert</i> | <i>Novice-Expert</i>         | <i>Novice-Expert</i>                                    |
| An open repository for different types of <b>data sources and relevant documentation</b>   | 7.4 - 5.7             | 27.6 - 32.7                  | 64.8 - 61.5   |
| An open repository for different types of <b>training materials</b>  | 5.3 - 10.5            | 26.5 - 32.7                  | 68.1 - 56.7   |
| An open repository for different types of <b>tools, software and packages</b>  | 3.2 - 19.1            | 24.4 - 19.1                  | 72.3 - 61.9   |
| A platform where users can find <b>recommendations for relevant resources</b> for their research   | 9.6 - 19.2            | 39.4 - 35.5                  | 51.1 - 45.2   |
| A platform users can <b>contribute to with their own resources</b> (e.g., new databases, software or packages)                                     | 20.1 - 16.3           | 52.1 - 38.4                  | 27.6 - 45.1   |
| A platform with tools and standards for the <b>validation</b> of computational methods   | 20.4 - 9.6            | 45.2 - 24.1                  | 34.4 - 66.3   |
| A platform that can be used to work <b>collaboratively on the discovery, creation and sharing of text analysis resources</b> (e.g., codes or data) | 19.1 - 20.2           | 35.1 - 29.8                  | 45.7 - 50   |
| A platform that can <b>host discussions</b> about text analysis resources  | 30.1 - 31.7           | 40.8 - 42.3                  | 29.1 - 25.9   |
| A platform for where users can <b>access training materials</b> (e.g., videos, slides, Shiny apps)   | 7.4 - 15.5            | 35.1 - 40.7                  | 57.4 - 43.6   |
| A platform that can be used to <b>“re-appraise” existing resources</b> , highlight and solve issues (e.g., bugs) and track updates                 | 19.1 - 28.8           | 56.3 - 39.4                  | 24.4 - 31.7   |

<sup>2</sup> Novices and experts are identified using the replies to question q2\_3. Regular users of computational text analysis are coded as experts (N=125) and respondents giving any other answer option as novices (N=114).

