D4.2 Evaluation of the DE-BIAS tool

Date of submission	31 December 2024
Authors	Kristina Rose (DFF)
Reviewers	Adina Ciocoiu, Jochen Vermeulen (EF)
Dissemination level	Public

HISTORY OF CHANGES		
Version	Publication/Submission date	Author(s)
0.1	31 December 2024	Kristina Rose (DFF)

Note: as of January 2025, this deliverable - and the DE-BIAS project - have not been assessed yet by the European Health and Digital Executive Agency (HaDEA).

Table of content

Abbreviations

DFF: DFF - Deutsches Filminstitut & Filmmuseum

MCA: Michael Culture Association

APEF: Archives Portal Europe Foundation

EAF: Europeana Aggregators' Forum

ENA: Europeana Network Association

1. Introduction

The DE-BIAS project aims to promote a more inclusive and respectful approach to the description of digital collections and the telling of stories and histories of minoritised communities. The project has developed an AI-powered tool to automatically detect potentially problematic terms in cultural heritage metadata and provide information about their background. For this, it uses vocabularies that combine offensive language with contextual information and suggestions for alternative, more appropriate terms. These vocabularies focus on three themes: migration and colonial history; gender and sexual identity; and ethnicity and ethno-religious identity.

The DE-BIAS tool is available for users as a stand-alone application on the DE-BIAS platform[1] and as an integration into the Metis Sandbox[2]. In both cases, a web interface enables the user to feed data to the tool and analyse it for biassed terms listed in the DE-BIAS vocabulary[3]. Furthermore, the public API can be integrated in customised data processing workflows, as has happened during the project for the analysis of over 5 million records available on Europeana.eu. All three usage scenarios are explained in detail in the user guidelines made available on the DE-BIAS Knowledge Hub on Europeana Pro.[4]

Part of Work Package 4, "Evaluation and Validation", the usability of the tool was tested in task 4.3 "Evaluation of the usability of the tool by Europeana aggregators". In four online events, users were given the opportunity to test the tool in its integration to the Metis Sandbox and in the stand-alone application. They were asked to provide feedback on its ease of use, about the comprehensibility of the interface and their interest in using the tool after its official release. This document will report on the results and insights gathered during the events and via the feedback forms and it will furthermore explain how this feedback was processed by the project consortium.

This report does not include a validation of the tool's performance in terms of recall and precision, those evaluation results can be found in D4.1 *Report on crowdsourcing campaigns with the CrowdHeritage platform* and D1.7 *Final Technical Progress report*.

---

[1] https://debias-tool.ails.ece.ntua.gr/
[2] https://metis-sandbox.europeana.eu/
[3] https://pro.europeana.eu/page/the-de-bias-vocabulary
The tool uses a machine readable version of the DE-BIAS vocabulary, e.g. its semantic format as DE-BIAS knowledge graph, hosted on EU vocabularies, will be available in January 2025.
[4] https://pro.europeana.eu/files/Europeana_Professional/Projects/debias/DE-BIAS_tool_technical_documentation.pdf

2. Setup of evaluation events

The setup for the four usability evaluation events was the same, however the scope of integration scenarios to be tested was adapted to be most suitable for each event.

2.1 Audience

In general, the events focused on an audience of cultural heritage professionals from the Europeana Initiative or aspiring professionals. As the feedback from Europeana Aggregators was key to this task, they were addressed in a specific event on October 23, 2024 (see 3.1). Furthermore, members of the EuropeanaTech Community and participants of earlier dissemination and capacity building events were contacted for the Europeana Network Association (ENA) event on December 9, 2024 (see 3.3). After all four events had gone through, the feedback of overall 22 participants was collected and processed.

2.2 Material

The material for all events consisted of:
- A presentation slidedeck giving an introduction/recap of the DE-BIAS project, an overview of the DE-BIAS tool and vocabulary and a short overview of the main elements of the live demonstration of the tool.
- A standardised survey collecting the feedback for each component of the tool integration (see Annex I). The survey allows for custom feedback based on the different components tested. E.g. if a participant answers with "yes" to the question "Did you test the data upload functionality of the tool?", they would be shown the questions for this specific feedback. If they answered with "no", they would be guided to the following survey section.
- A document with the guidelines on how to use the different integrations of the DE-BIAS tool. These guidelines were adapted each time the tool's interface or reports were further developed in between the evaluation events. The guidelines can be found on the DE-BIAS Knowledge Hub on Europeana Pro.[5]
- The participants were offered sample data to use either for the input of texts directly via the interface or for the file upload functionality. The data for the former application was given in the form of a table to copy/paste from, the data

---

[5] https://pro.europeana.eu/files/Europeana_Professional/Projects/debias/DE-BIAS_tool_technical_documentation.pdf

for the latter testing case was provided as one .zip file per language containing correctly formatted .txt files.

- For the event conducted during the Europeana Aggregators' Forum (EAF) Autumn meeting, sample data in the form of .zip files containing .xml records in the Europeana Data Model (EDM) format, was made available for testing the Metis Sandbox integration.

2.3 Event outline

All events took place online, only the EAF event had a hybrid setting, as it was part of the bi-annual EAF meetings which are usually held hybridly. The project representatives gave a short project presentation as introduction, followed by a live demonstration of the tool - and, in case of the EAF event, of the Metis Sandbox integration. After a first Q/A session, the participants were given time to test the usability of the tool at their own pace. Furthermore they could ask questions during this stage as well. Before the wrap-up and sharing of the survey link, the feedback received in the breakout sessions, if applicable, was summarised for the whole group and participants were given another chance to ask questions. On the same day as the event, a follow-up mail with all relevant information and links was sent out to the participants and after about one week, the survey was closed.

3. Results

The following chapters illustrate and reflect on the results of each testing event. A general overview of the KPIs and event statistics can be taken from the two tables below.

Key Performance Indicator	Status
10+ accredited aggregators engaged in the validation of the tool's usability	12 accredited aggregators engaged (see section 3.1) (11 of which answered the survey)
80% of aggregators rate the ease of use of the tool as four or higher on a Likert scale of one to five.	Standalone tool: 91% Metis Sandbox integration: 71,5% (see section 3.1)
12+ aggregators apply the tool on their datasets by the end of the project	10 aggregators responded "Yes" or "Maybe" to the question "Would you be interested in running the DE-BIAS tool on your collections yourself?", three aggregators voiced interest depending on specific conditions in free text form responses. *(Note: One aggregator provided two submissions to the survey, one answering "Yes" to this question, and one "Other" with free text comment covered in the previous sentence.)*

Additional statistics

Overall number of testing events	4
Overall number of surveys answered	22

3.1 Europeana Aggregators' Forum event

The tool usability testing session took place on October 23, 2024, as a hybrid workshop during the Europeana Aggregators' Forum Autumn meeting in The Hague and online. Participants were given the standalone tool, the Metis Sandbox integration and the UI/UX component on Europeana.eu for testing. Since the evaluation was conducted as part of the EAF meeting, the exact number of participating aggregators in this specific workshop session was not recorded separately. Eight aggregators replied to the feedback survey. In addition, a separate survey was set up for the Europeana Aggregators within the project consortium that did not reply to this EAF event survey. Four additional aggregator responses were collected via this separate project Aggregator survey and the survey sent out for the Country Managers' Event (see 3.2). It should be noted that one aggregator provided two separate submissions, thus 12 responses overall subsume the responses of 11 Aggregators. All results will be presented in a combined way. Europeana Sounds provided detailed feedback via email, but did not fill out the survey. The complete, anonymised raw data of the feedback shown as charts below, can be consulted in annex II.

The feedback of the following Aggregators is represented in the charts (aggregators represented through project partners are listed in italic):
- *Archives Portal Europe*
- Carare
- Digital Repository of Ireland
- *Dutch Collections for Europe*
- eKultura.hr
- *EUscreen*
- German Digital Library
- Kulturpool
- *MUSEU*
- Slovenian National E-content Aggregator
- *The European Film Gateway*

Chart 1: Ease of use of the standalone tool

Out of the eleven participants that tried the standalone tool, 91% agreed or completely agreed with the statement "The standalone tool was easy to use". The question was asked as a likert scale with 1 equaling complete disagreement and 5 equaling complete agreement.

Legend:
- 3 (Neither agree nor disagree)
- 4 (Agree)
- 5 (Completely agree)

1 (9.1%)
3 (27.3%)
7 (63.6%)

Chart 2: Ease of use of the Metis Sandbox integration

Out of the seven participants that tried the Metis Sandbox integration, 71,5% agreed or completely agreed with the statement "The DE-BIAS tool integration was easy to use". The question was asked as a likert scale with 1 equaling complete disagreement and 5 equaling complete agreement.

Legend:
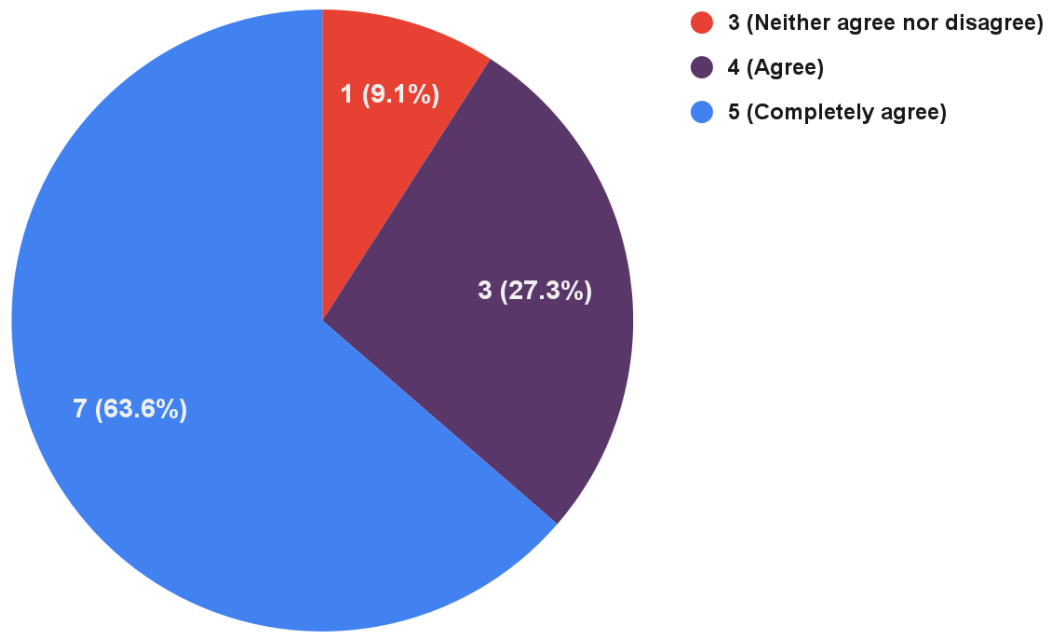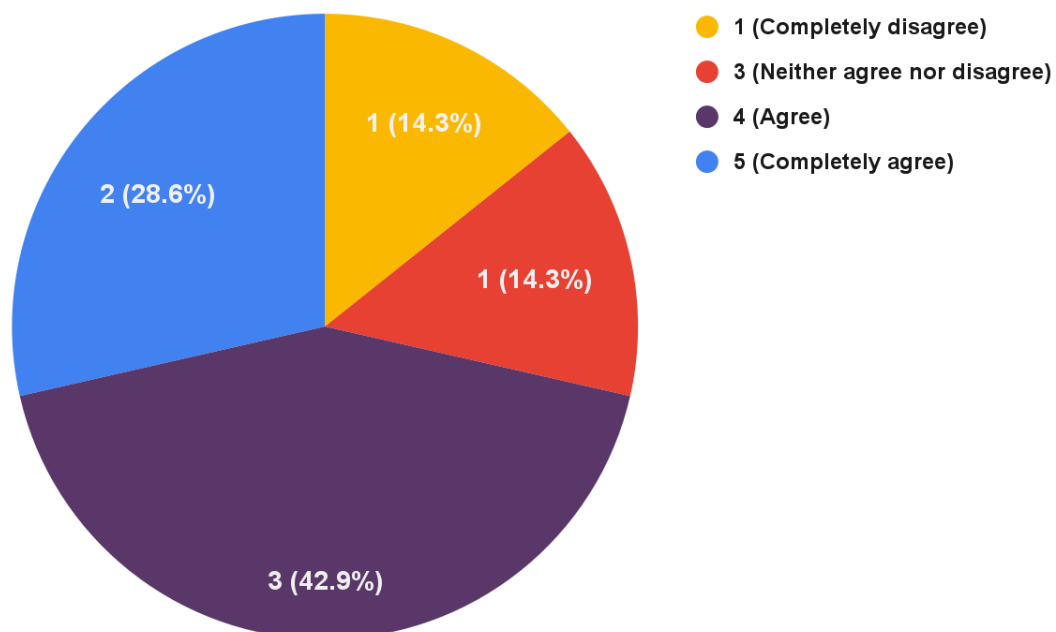- 1 (Completely disagree)
- 3 (Neither agree nor disagree)
- 4 (Agree)
- 5 (Completely agree)

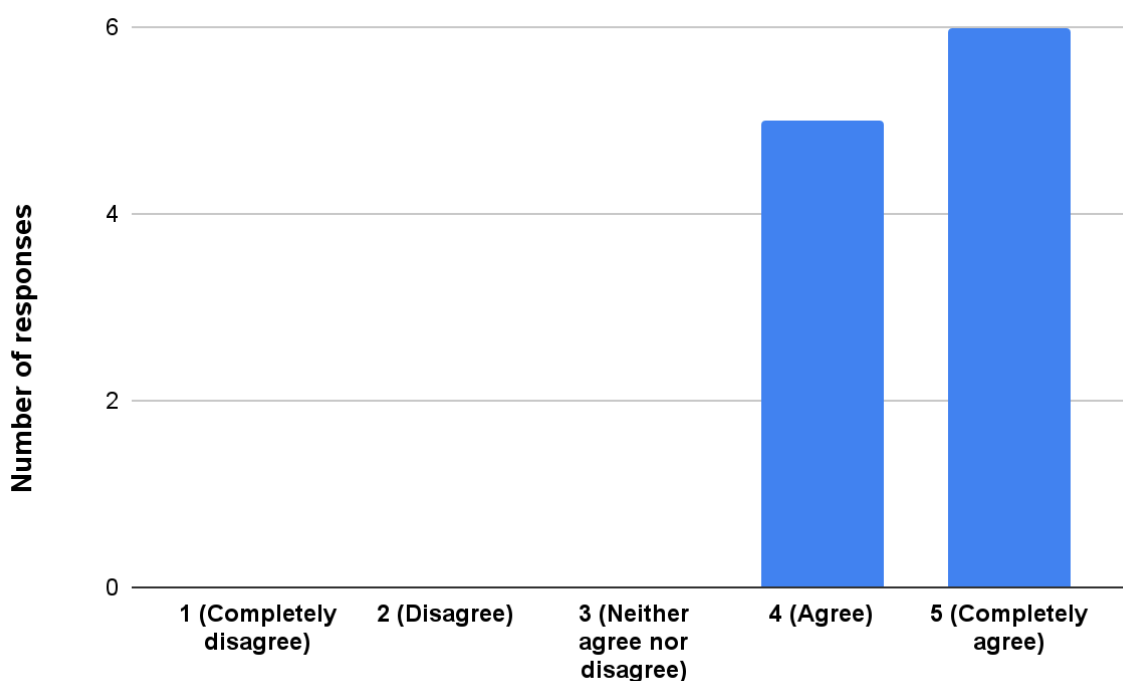1 (14.3%)
1 (14.3%)
2 (28.6%)
3 (42.9%)

The complete disagreement with this statement resulted from one participant not being able to properly upload data to the Metis Sandbox, due to technical complications at the time of testing.
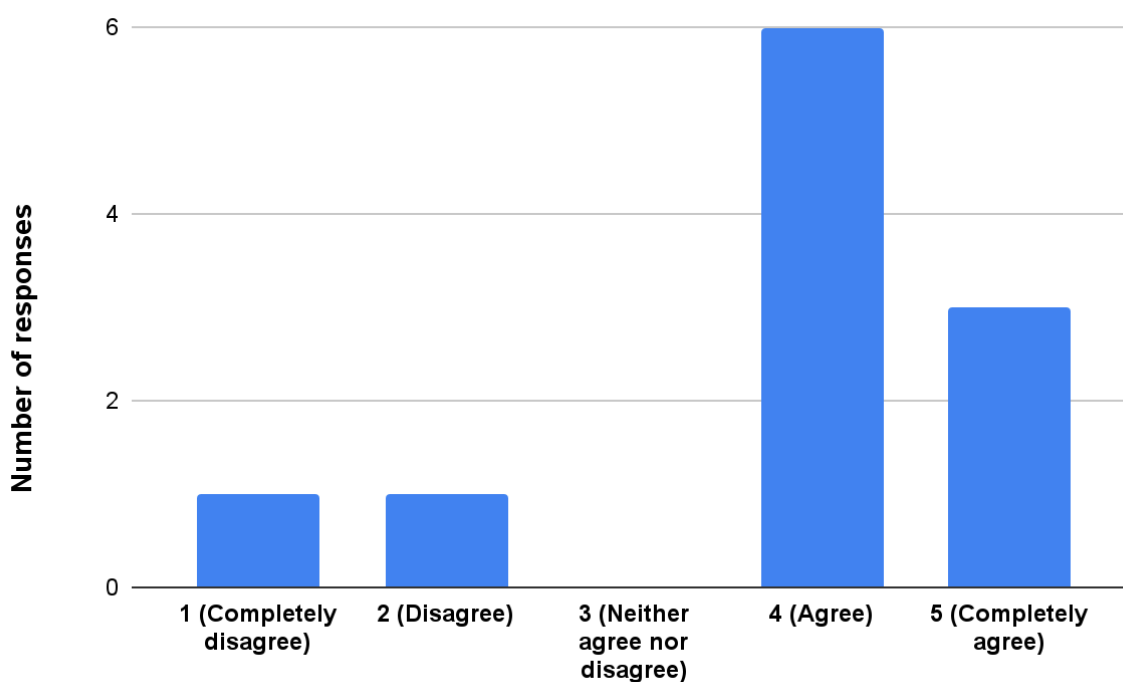
Charts 3 and 4: Comprehensibility of information

To gain insights into whether the information provided in the interface is sufficient for users of the tool, two questions covered the comprehensibility of the standalone tool's interface. While the respondents to the survey widely agreed that the interface is self-explanatory, the specific use and consequences of the two options "NER" (Named Entity Recognition) and "Disambiguation" was less well understood. Looking at this result, we had furthermore to take into consideration that the participants received an introduction session that explained these two options, which future users would not. The mitigation action for this result was to place a prominent link to the guidelines for using the tool on the tool's website and to investigate possibilities to clarify the explanation for NER and Disambiguation in the interface (see chapter 3.1.1).

Chart 3: "The interface was self-explanatory."

Chart 4: "I understood the use of the operations "NER" and "Disambiguation"."

Number of responses (y-axis: 0, 2, 4, 6)

X-axis categories: 1 (Completely disagree), 2 (Disagree), 3 (Neither agree nor disagree), 4 (Agree), 5 (Completely agree)

Charts 5-7: Satisfaction with the display of detection results

When looking at the satisfaction rates of the display of detection results in the different output formats of the DE-BIAS tool integrations, the participants' responses varied. The first chart reflects the display in the web interface of the standalone tool (see screenshot), which was generally regarded as satisfying.

Chart 5: "I am satisfied with the display of detected terms in the web interface."
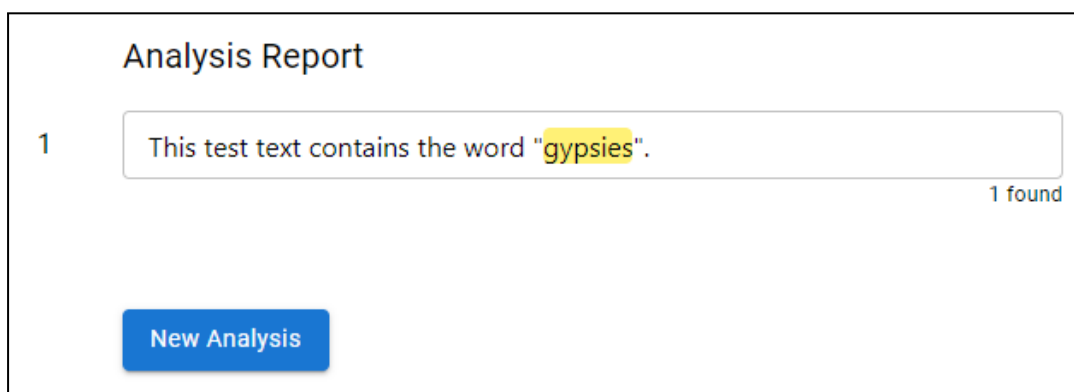
Image 1: Screenshot of the display of a detected term in the standalone tool.
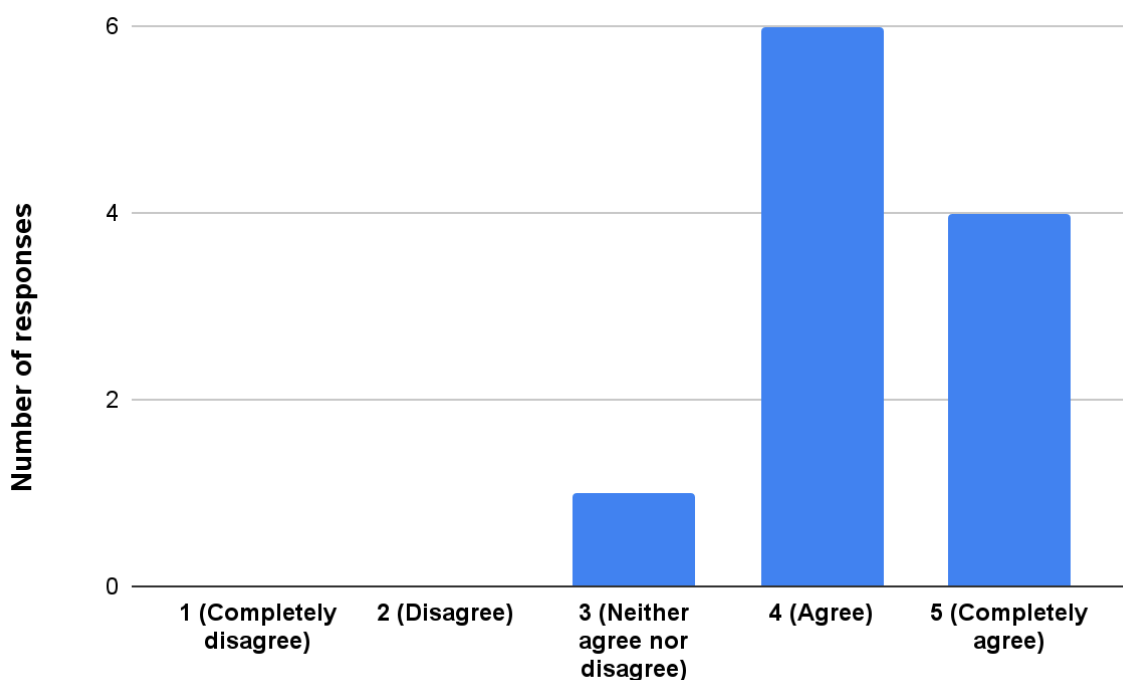
The second chart reflects the satisfaction with the detection results in the PDF report (see Annex III, 5.3.1) sent by the standalone tool when the file upload functionality is used. The satisfaction was lower in this case and participants gave extensive feedback in the form of comments. As a result, the PDF report was revised to address some of the mentioned concerns (see chapter 3.1.1). Please note that one response was excluded from this chart, as this aggregator answered the survey when the new version of the report was already available and thus does not reflect the same data basis.

Chart 6: "I am satisfied with the display of detected terms in the report."
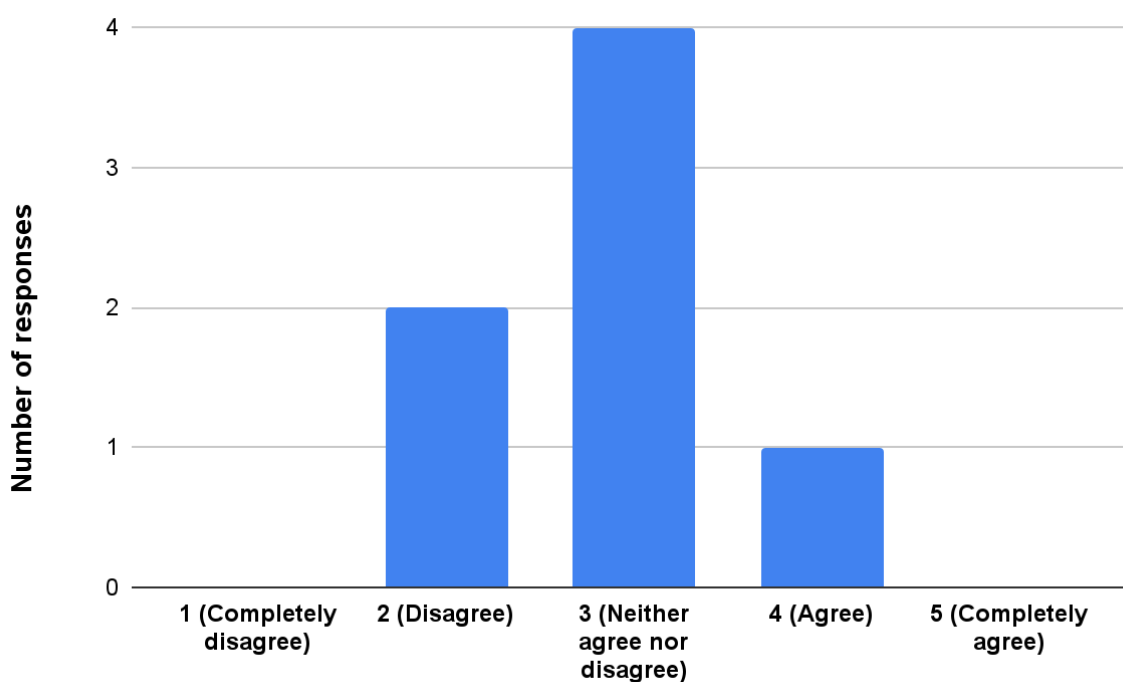
The third chart shows the satisfaction rates with the display of detected terms in the Metis Sandbox integration (see screenshot). The satisfaction rate is quite low, but when considering the free text comments, this can clearly be linked to general technical problems and accessing the integration on the specific day of testing, and less to issues with the set-up of the report as such. Feedback for improvements was addressed and implemented for this tool integration as well (see chapter 3.1.1).
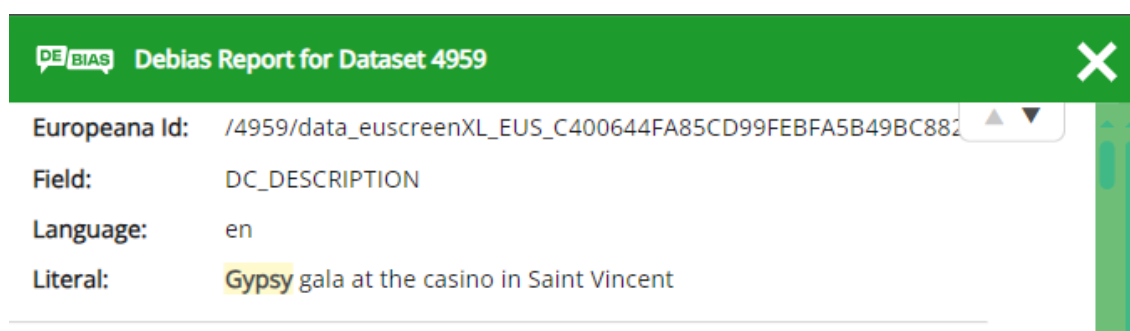
Image 2: Screenshot of the display of detected terms in the DE-BIAS report in the Metis Sandbox.

Chart 7: "I am satisfied with the display of detected terms in the report."
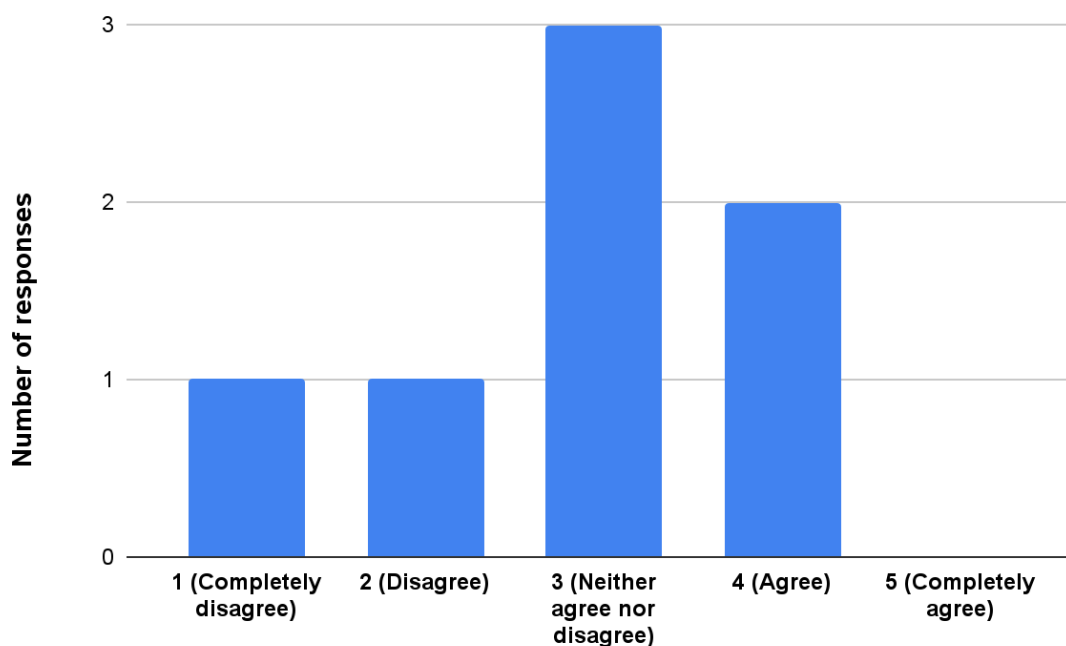
Chart 8: Interest in display of bias detection in collections

All aggregators were consulted about whether they would like to see the annotations of contentious terms, as per the UI/UX design shown below developed by EF, in their own collections as well. 75% of the participants replied with "Yes" or "Maybe". 25% gave detailed comments, explaining that this decision could not be made by them, but by a superior within the Aggregator institution or by the data provider. One respondent declared, they were not able to see the mouse-over in the click dummy provided.

This term is generally used to refer to a member of a travelling or itinerant people, specifically Roma people. The Roma people are divided into different groups. Associated with itinerancy, due to their history of (forced) migration, negative stereotypes of Roma as thieves and vagabonds continue to exist today. For the Roma people the term 'gypsy' is derogatory. Consequently they collectively and officially adopted the term "Roma" in the 1970s.

Read the full explanation provided by DE BIAS

Gypsy 💬 caravans and tents on

Gypsy caravans and tents on Belvedere Marshes , Kent ._x000D_
1936

Image 3: Screenshot of the planned display of detected terms and the mouse-over in Europeana.eu.

"Would you like to have this contextualisation of terms in the records coming from your Aggregator as well?"
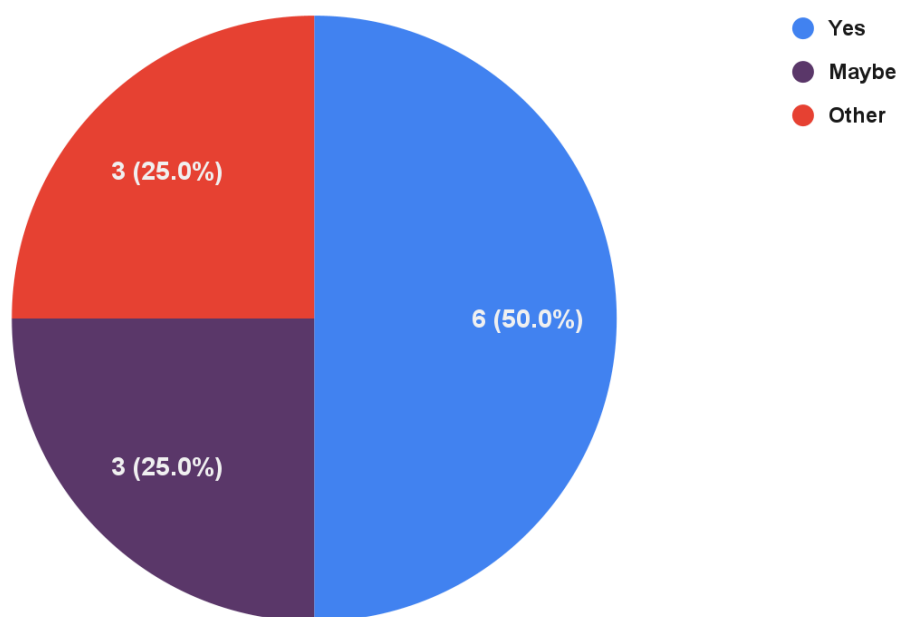
Chart 9: Interest in using the tool in the future

Of all eleven Aggregators 83,3 % answered "Yes" or "Maybe" to the question "Would you be interested in running the DE-BIAS tool on your collections yourself?". Two aggregators chose the option other and elaborated in free text form. Both raised their interest in the tool as well, but with the constraints of not being the person in charge of this workflow decision or the tool needing to cover more languages in order to be applied to the whole range of datasets. Furthermore, the additional feedback received via mail by one aggregator favoured the use of the tool in the future.
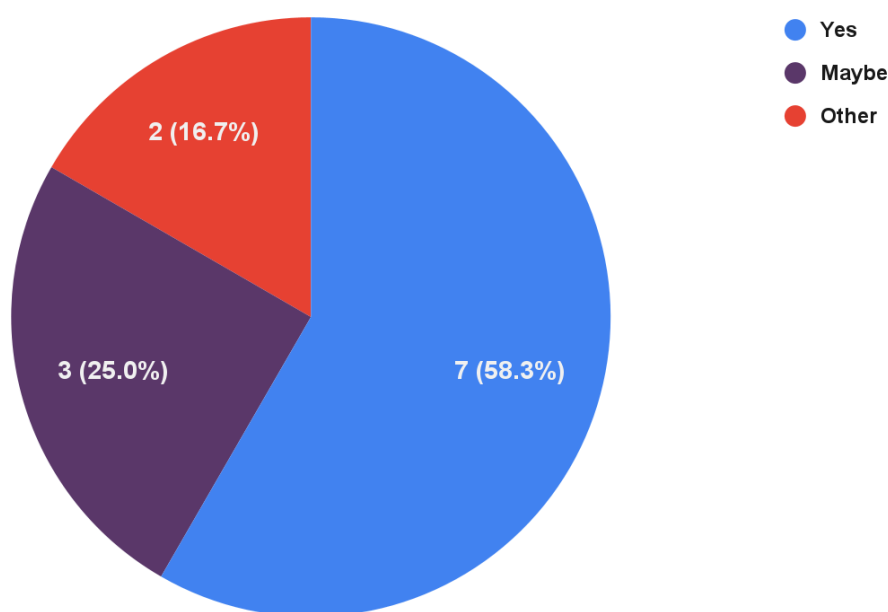
3.1.1 Processing of feedback

Since the results from the answers based on the likert scale already showed a high level of positive evaluation and acceptance, the processing of the feedback was conducted especially on the basis of free text answers that contained suggestions and/or negative feedback. An overview, grouped by component of the tool was created together with the feedback documented directly during the event and received via mail. The following table reflects this feedback together with the mitigation action taken in accordance with ThinkCode, Datoptron and Europeana Foundation. Feedback that did not refer to the usability of the tool, but rather pointed to the DE-BIAS vocabulary or general concepts of the project are not included here. The original feedback given by the test users in the survey can be consulted in annex II.

Component	Feedback (grouped)	Mitigation action	Outcome
Standalone tool - Disambiguation functionality	For Dutch terms disambiguation wasn't working as expected. As an example, colours ('zwart' and 'wit') referring to clothing were given. It was expected that these terms would not be flagged when enabling disambiguation. Instead the terms were still flagged.	The action was twofold: a) the editorial team of the Dutch vocabulary re-assured that all terms needing to be disambiguated would have the respective marker (debias-o:isAmbiguous = "true"). b) ThinkCode refined the disambiguation functionality for the Dutch components of the tool.	The tool correctly excludes colours referring to clothing from Dutch texts. This was already implemented by the time of the ENA evaluation event.
	The functionalities of NER and Disambiguation were only clear because of the oral introduction at the workshop, they should be better explained in the tool.	Revise the explanation of the two options in the tool's interface.	Revised wording was added to the interface to allow for a better understanding of the purpose of both functions .
Standalone tool - results view in the web interface	A collation of detected terms in a list in addition to the highlights was suggested, especially with regard to very long input texts, to improve overview of detected terms.	No action following up on this issue in order to not overcomplicate the user interface. The option to discuss this further was kept open if this suggestion would come up again in future responses.	n/a

Component	Feedback (grouped)	Mitigation action	Outcome
	Evaluators were missing further information and context and reasons for the detections, e.g. link to the source vocabulary, background information on the term.	The tool's response cannot include reasoning, but vocabulary details can be added once term URIs are dereferenceable. As with UI/UX integration in Europeana.eu and CrowdHeritage, contentious issue descriptions can appear as mouse-over text.	The display of further information about detected terms in the web interface is planned, but dependent on the DE-BIAS vocabulary publication in the EU Vocabularies environment, scheduled for January 2025.
Standalone tool - File upload	UTF-8 requirements might exclude some languages from the outset.	At the moment, no concrete plans for further languages to add to the tool are fixed. But this concern will be considered when applicable.	n/a
	Users reported error messages after uploading files.	The action was twofold: a) Consider adding error feedback or a pre-check in the uploaded file to improve user experience. b) Check the tool for potential problems with file handling.	Bug fixing for file handling issues was applied.
	It was suggested to allow for the upload of .csv files next to .txt files.	Since .csv files require a distinct file-handling setup (e.g., mapping	n/a

Component	Feedback (grouped)	Mitigation action	Outcome
		tabular values to specific literals), this suggestion is beyond the project's scope but will be considered for future extensions.	
	General issues with preparing the .txt files in the correct format were flagged.	Give clear instructions on how a .txt file needs to be prepared in the tool's guidelines available on the DE-BIAS Knowledge Hub.	A section on .txt file preparation is added to the guidelines for the DE-BIAS tool.
Standalone tool - Data input	It was suggested to allow for pointing to online catalogues or websites via URL as a means for data input.	This request may already be addressed for online catalogues via custom API integration. As this requires programming resources from cultural heritage institutions (CHIs), the option should be included in the tool's guidelines on the DE-BIAS Knowledge Hub.	Information on how to integrate the DE-BIAS API into a CHI's own technical infrastructure is available in the DE-BIAS tool usage guidelines.
Standalone tool - .pdf report	The PDF report lacks clarity on what users should take away from it. Users felt unguided regarding	Include clearer information about what users can expect from the report, along with links to relevant	Introduction text in the report rephrased (see annex III).

Component	Feedback (grouped)	Mitigation action	Outcome
	potential follow-up actions they could take.	documentation and publications for further guidance.	
	Improvement of usability if detected terms were linked to the relevant vocabulary term, rather than just giving the term identifier.	Working links and labels instead of URIs can be added, once the URIs of the detected terms are dereferenceable.	This request will be resolved, once the DE-BIAS vocabulary is published in the EU Vocabularies environment, scheduled for January 2025.
	The language of the introduction is quite difficult to understand.	Rephrase the text parts of the .pdf report.	The text parts of the .pdf report were re-phrased (see annex III).
	Formatting issues where text overlaps.	Tilt values on the x axis of charts to avoid overlap.	Consider changing the layout of the axis
	It is not easily comprehensible which terms were found in which record.	A simplified, tabular overview can be added to the .pdf report.	A table listing the file name and the detected terms was added directly to the .pdf report.
Standalone tool - .json report	The content and intended use of the JSON report are unclear, and users reported difficulties handling the .json files.	Provide clearer instructions on what users can expect from the .json file, emphasizing that it is intended for computer processing.	An explanation for the .json file usage and contents was added to the email text and to the .pdf report.
	Suggestion to include a .csv output of the tool next to the .json	A .csv output of the contents from the .json file is too	A table listing the file name and the detected terms was

Component	Feedback (grouped)	Mitigation action	Outcome
	file that contains the same information.	complex, because of the nested structure of the output. A simplified, tabular overview can be added to the .pdf report.	added directly to the .pdf report.
Metis Sandbox integration	General remarks on error messages and dataset submission.	Mostly not applicable as users will be using a more performant and robust production environment (as opposed to the testing environment used during the evaluation event). Further bug fixing for remaining issues as needed.	Users can upload their data as per usual in the Sandbox without errors.
Metis Sandbox integration - data input	It is not clear to the users whether the NER and disambiguation functionalities are enabled or not.	Decision whether to enable these functions and whether to give the option for the user to disable them. Add note accordingly in the Metis Sandbox documentation and/or the info text pop-up available in the application.	The NER and disambiguation functionality are enabled by default without the option for the user to disable them. This information is included in the usage guidelines for the tool.
Metis Sandbox integration - data output	Suggestion to make the report downloadable.	Investigate whether the report can be made	The report is downloadable in .csv format.

Component	Feedback (grouped)	Mitigation action	Outcome
		downloadable in a user-friendly format (see issues with .json report above).	
	It is unintuitive to click on "view report" after it was generated instead of it opening up directly after completion.	Design more intuitive UX for signalling that processing has completed and that a report is available.	Users will be notified of the availability of the report, or presented with the report, in a more intuitive and unobtrusive way in a future version of the Sandbox.
UI/UX component on Europeana.eu	It would be useful to know when and why a term detail page was last updated.	The date of last modification can be retrieved from the Knowledge Graph. The addition to the detail page will be kept in the backlog for further adjustments of the detail page. Documentation about the reasoning behind changes are not recorded. (Only the nature of the change can be viewed in the editing history in VocBench.)	n/a
General	Availability of bias detection for more languages.	This feedback could only be resolved in dedicated	n/a

Component	Feedback (grouped)	Mitigation action	Outcome
		follow-up projects that build on the workflows for vocabulary creation and tool development set up in the DE-BIAS project.	

3.2 Country Managers event

The APEF Country Managers are representatives of mostly national level archival institutions or national level archival aggregators in the network of Archives Portal Europe. They meet twice a year, usually once in person, once online. The testing session for the DE-BIAS tool took place during their Autumn/Winter meeting 2024 on November 25 and 26. It was held as part of a repeated session to allow for a broader participation as not everyone was able to attend the first iteration of the event. In total, 11 Country Managers attended the first testing session, while the second testing session was attended by 5 additional participants. As the DE-BIAs project had been introduced to the Country Managers already during previous meetings, the session focussed on a demonstration of the standalone version of the tool, followed by some time for individual tests and exploring. As the Country Managers don't use the Metis Sandbox in their usual workflows in Archives Portal Europe, this integration was not part of the evaluation.

3.2.1 Processing of feedback

Since only APEF answered to the survey and this data was compiled into the Aggregator responses listed in chapter 3.1, the following table focuses on the feedback received directly during the event. Only feedback not already raised in the EAF event is addressed in the table below.

| Component	Feedback (grouped)	Mitigation action	Outcome
Standalone tool - Data input	Issues with German (in both, the "Insert text" and "Upload a	This may have been caused by server issues or	n/a

	file", options using German created an error message on the first day of testing)	technical work on that same day and is not indicative of a general problem, as it could not be reproduced.	
Standalone tool - .pdf report	The email sent by the tool ended up in a spam folder.	This behaviour can't be impacted by the tool, but a general reminder to check the spam folder as well can be added.	Note to users to check the spam folder is added to the screen following the data upload.

3.3 Europeana Network Association event

The event was organised by DFF and NISV and took place on December 9, 2024. Participants from earlier capacity building and disseminations events, who left their contact details for future events, were invited. Additionally, the invitation was sent out to the mailing list of the EuropeanaTech Community of the ENA. This community was chosen to raise awareness for the tool with a tech-focused audience and because of the technical focus of the usability testing events. The audience consisted of cultural heritage professionals, researchers and educators. During the event, only the standalone tool integration and the UI/UX component on Europeana.eu were tested by the users, as the Metis Sandbox integration requires previous knowledge about the platform and Europeana's integration workflows which could not be expected of the users. Furthermore, ENA members are not necessarily the target audience of the Metis Sandbox. 11 test users participated in the event and nine of them provided feedback via the survey. The complete, anonymised raw data of the feedback shown as charts below, can be consulted in annex II.

Chart 1: Ease of use of the standalone tool

All respondents agreed or completely agreed that the tool was easy to use.

**4 (Agree)**
5 (Completely agree)

2 (22.2%)

7 (77.8%)
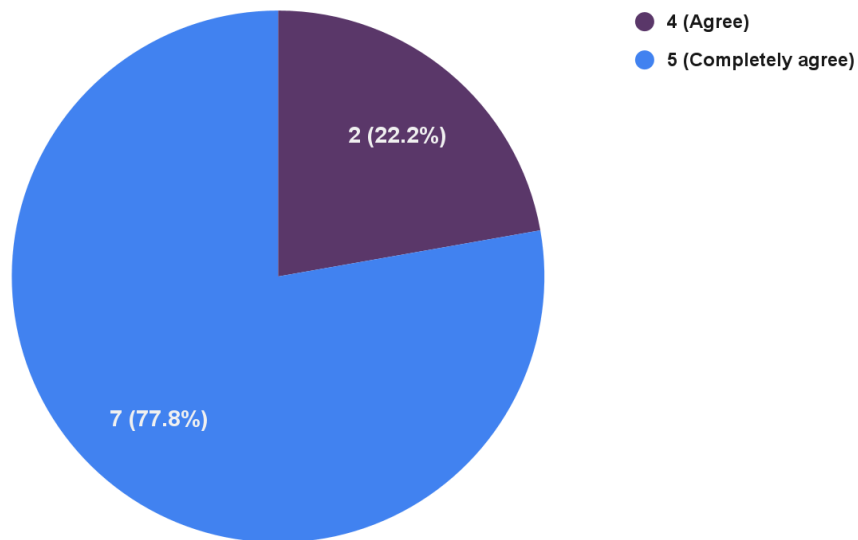
## Chart 2 and 3: Comprehensibility of information

Similarly to the Europeana Aggregators, the respondents showed a high rate of agreement, that the interface was self-explanatory, but when asking specifically about the two options NER and Disambiguation, the responses indicate that further explanation is needed.

Chart 2: "The interface was self-explanatory."
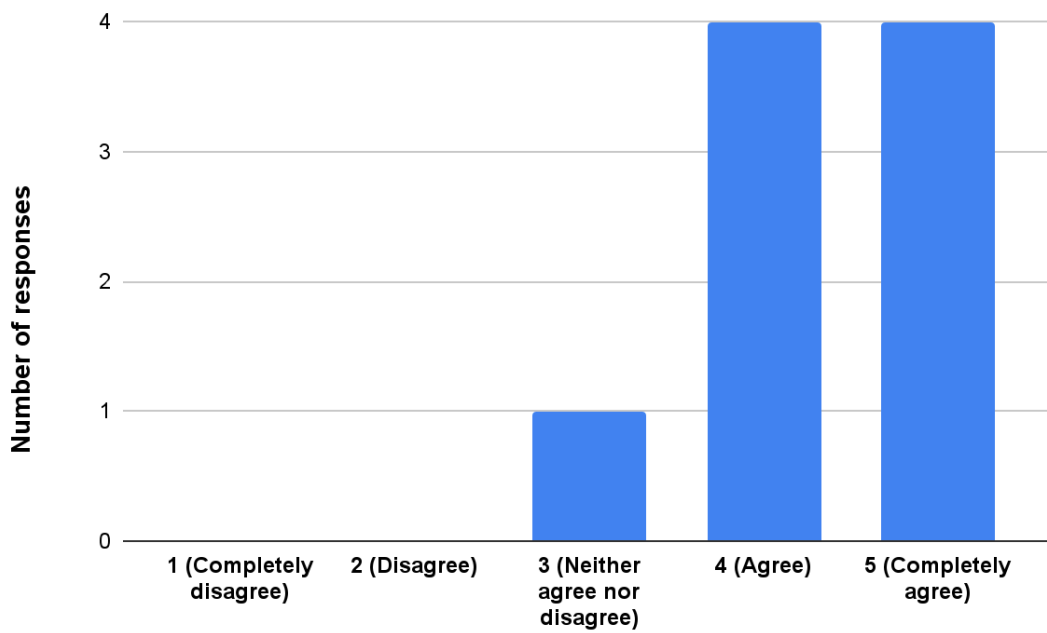
Chart 3: "I understood the use of the operations "NER" and "Disambiguation"
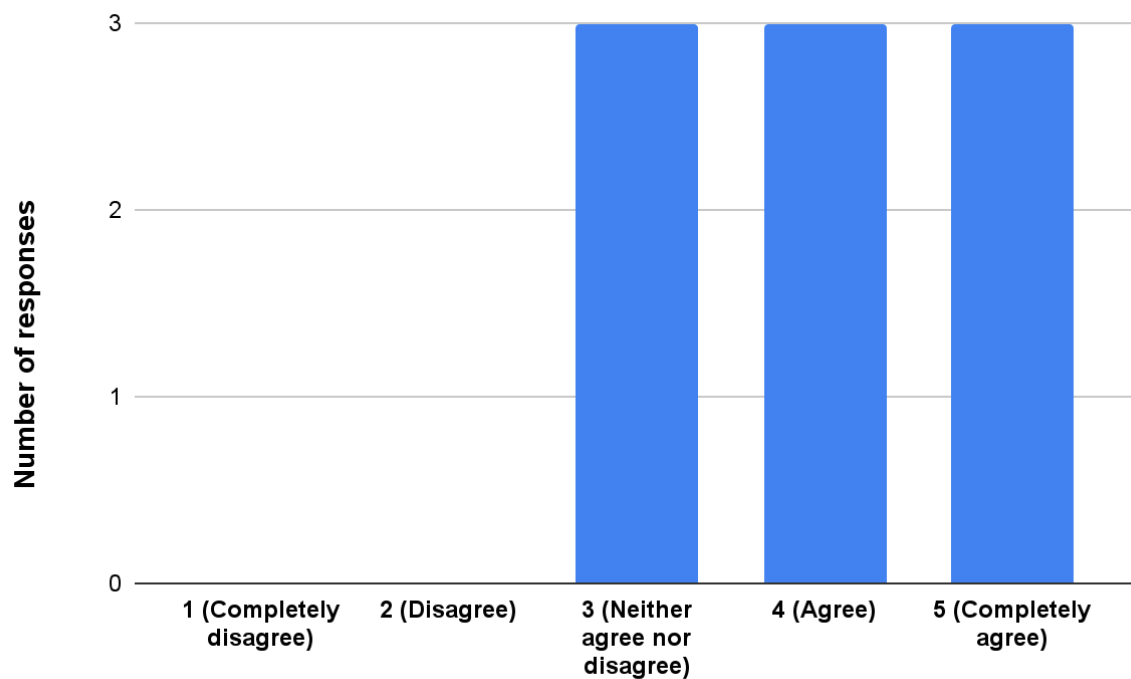
Charts 4 and 5: Satisfaction with the display of detection results

The following two charts illustrate the level of satisfaction with how the detection results are displayed in the web interface of the standalone tool and in the PDF report. While the detected terms in the web interface were displayed in the same manner as during the EAF event in late October (see chapter 3.1), the PDF report had been revised in the meantime.

Chart 4: "I am satisfied with the display of detected terms in the web interface."


Image 3: Screenshot of the display of a detected term in the standalone tool.
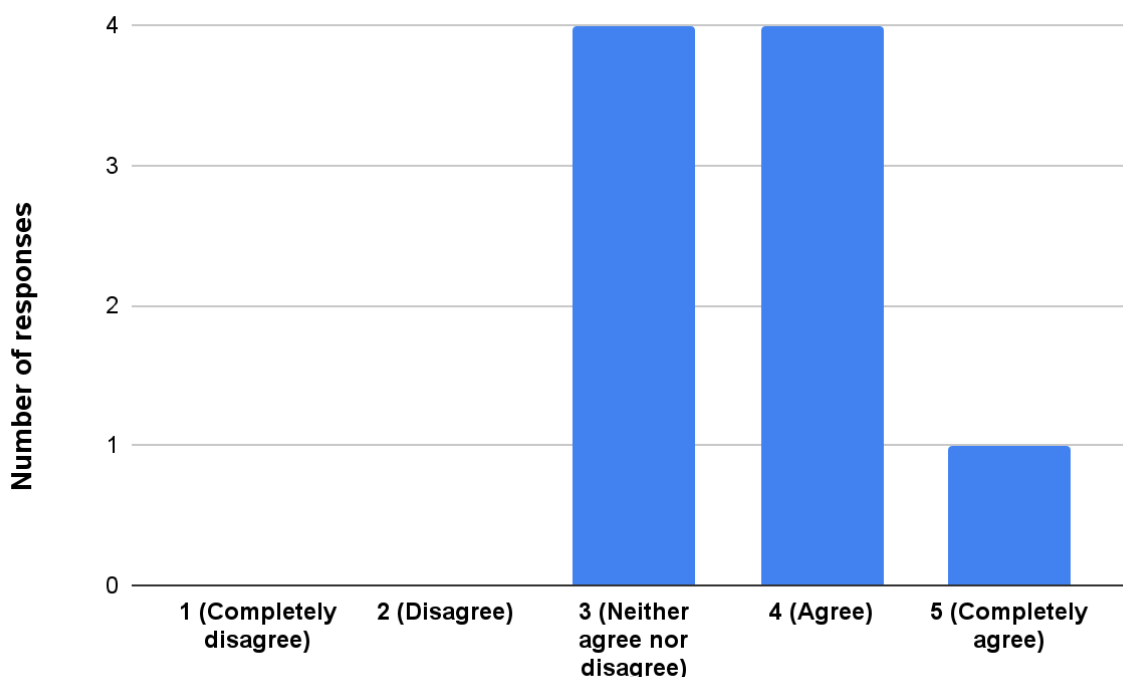
The participants of the ENA showed a lower level of agreement with the statement than the EAF, but none disagreed with the statement. The feedback given in the form of

comments largely overlapped with the feedback from Europeana Aggregators, so concerns raised by both sides can be covered with the same mitigation actions.

Chart 5: "I am satisfied with the display of detected terms in the report."
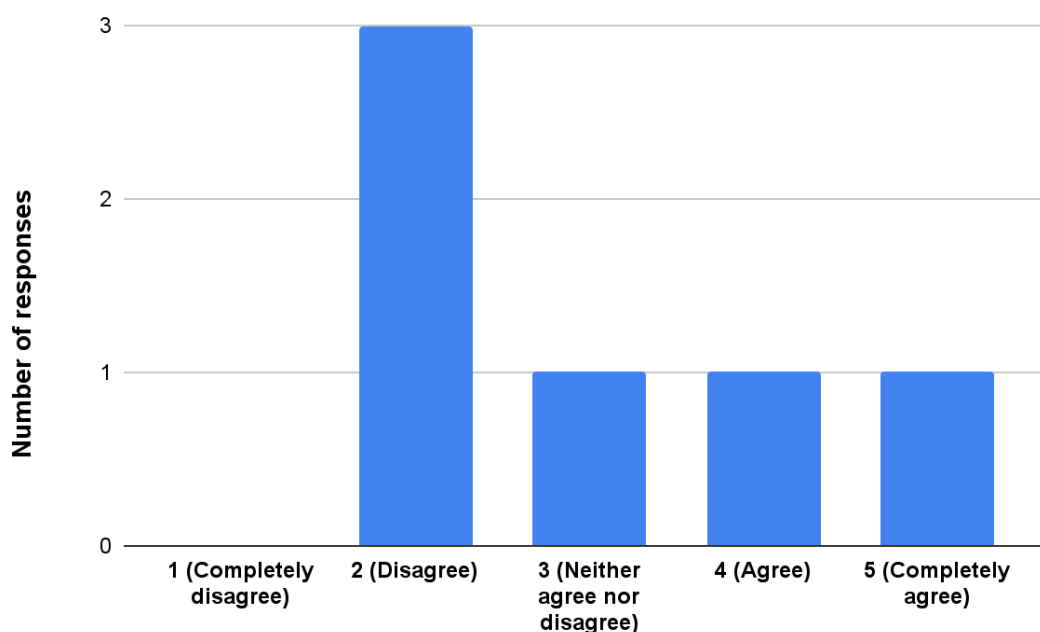

The satisfaction with the display of detected terms in the PDF report did not improve significantly among the ENA respondent as compared to the EAF results, even though parts of the PDF had been revised. However, one participant completely agreed with the statement, which could not be achieved in the previous testing session.

The results can be explained by considering that the change request most often received - displaying the labels of detected terms instead of the URIs (see table in chapter 3.1.1), could not be implemented yet by the time of the ENA event. Satisfaction with the PDF report is expected to be higher, once this display of information is revised.

Chart 6: Interest in using the tool in the future

Of all nine respondents, 89% said they would be interested in using the tool in the future. These answers show that the scenarios of application for the DE-BIAS tool transcend the scope of collection descriptions in CHI catalogues, and additionally professionals from the educational domain and researchers see benefits for their work.
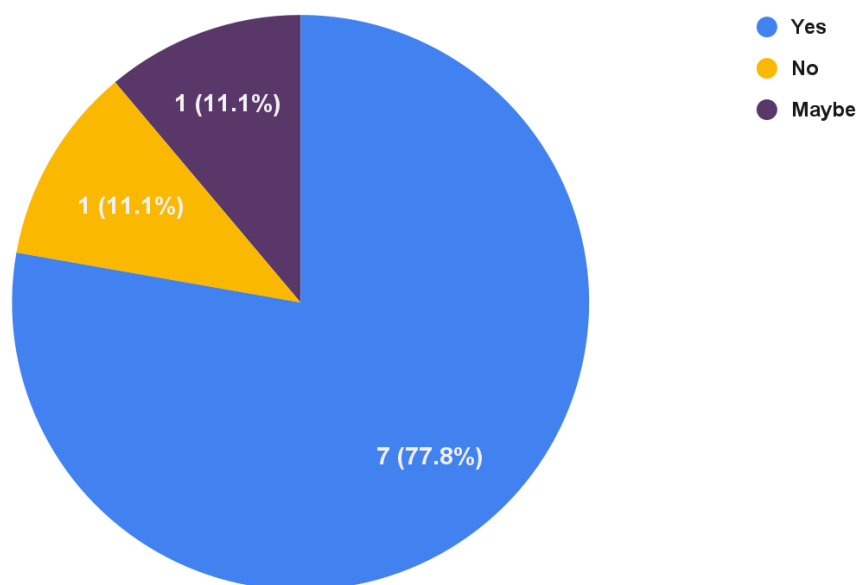
3.3.1 Processing of feedback

Similarly to the EAF testing event, the feedback from the free text comment fields is grouped and mitigation actions discussed with the technical developers. Only feedback not already raised in the EAF event is addressed in the table below.

Component	Feedback (grouped)	Mitigation action	Outcome
Standalone tool - Data input	It would be great if the tools could also offer the option to modify the text.	Investigate whether returning to the previously inserted texts is possible.	Decision to keep this suggested change in the backlog of possible extensions to the user interface if interest is voiced repeatedly by users.
Standalone tool - Results view in	Option to save the results from the web	Adding the option to download the	The guidelines document is

Component	Feedback (grouped)	Mitigation action	Outcome
the web interface	interface.	report from the standalone tool has many implications and it's not easy to implement. The file upload functionality offers a downloadable report. Investigate whether this can be explained more clearly to the user.	adapted to make users aware of the report with bias detections available when using the file upload functionality.
Standalone tool - .pdf report	It would be useful to add to the table the highlighted passages with yellow highlights similar to the web version.	Investigate whether this addition to the table is possible.	Decision to keep this suggested change in the backlog of possible extensions to the PDF report if interest is voiced repeatedly by users.
UI/UX integration on Europeana.eu	The text in the pop-up is too long.	The length of the description texts is limited to 90 words as per the editorial guidelines. This length was proposed by the Europeana Foundation.	n/a
	The colour of the icon might be better in one of the tool colours (blue or red) to be more eye-catching.	Using a colourful icon would draw unwanted attention to the contentious term. The grey icon does not highlight the term but does invite the user to	n/a

Component	Feedback (grouped)	Mitigation action	Outcome
		access more context about it.	

3.4 ESACH event

The last tool usability testing event took place on December 11, 2024 and was organised by MCA and European Students' Association for Cultural Heritage (ESACH)[6] conducted together with DFF. The audience was composed of students and young professionals of the cultural heritage domain, as this is also the members background of ESACH. Out of the five participants, one person replied to the survey, showing a general satisfaction with the DE-BIAS tool and indicating to be interested in using it in the future. As they did not give feedback in free text form, only the replies with standardised likert scale questions (1 (Completely disagree) - 5 (Completely agree)) are listed in the table below. The complete anonymised survey response can be accessed in Annex II.

Standalone tool	
The standalone tool was easy to use.	4
The interface was self-explanatory.	4
I understood the use of the operations "NER" and "Disambiguation".	3
Standalone tool web interface	
I am satisfied with the detection results.	4
I am satisfied with the display of detected terms in the web interface.	4
Standalone tool file upload	
I am satisfied with the detection results.	4
I am satisfied with the explanation of the analysis results in the report sent to me by e-mail.	4
Europeana UI/UX design for bias detections	
Do you find the information in the pop-up useful?	4

---

[6] https://www.esach.org/

3.5. List of suggestions to be kept for future extensions of the tool

The feedback received from test users received during the testing sessions included potential future developments for the DE-BIAS tool that are out of scope during the project duration, but shall be recorded here for reference:

- **Include more languages**: This suggestion ranked among the most commonly offered feedback. The inclusion of more languages would require a complete repetition of the vocabulary creation and the tool development workflow, as each language requires the integration of specific components. The DE-BIAS project focused on English, German, Dutch, French and Italian, but the insights gained during the project could be adapted in follow-up projects, if the opportunity arises.

- **.csv export of detections**: similarly to the download functionality of detected terms in the Metis Sandbox integration, the stand-alone tool could provide a .csv report of detected terms for download. It would need to be further evaluated, if this is feasible for the data upload functionality and/or the data input via the web interface.

- **OpenRefine integration**: this feedback refers to the possibility to integrate the DE-BIAS tool into the functionalities of the data clean-up and transformation tool OpenRefine[7], similarly to its reconciliation functionality, to provide an additional means of feeding data to the DE-BIAS API. This option would require dedicated investigation, if such an integration would be possible with the current setup of the API and the DE-BIAS vocabulary (for the display of results).

- **Modify inserted texts**: This change refers to the feedback of one test user who suggested the ability of the stand-alone tool to go back to the start page of the web interface after analysis and still show the text that was inserted instead of showing blank text boxes again.

---

[7] https://openrefine.org/

4. Conclusion

The evaluation with test users has shown that the bias detection tool is easy to use and sparked interest among the users for applying it on their data in the future.

Combined results	
"The stand-alone tool was easy to use."	95,3% agreed or agreed completely
"Would you be interested in running the DE-BIAS tool on your collections yourself?"[8]	86,4% replied with "Yes" or "Maybe"
"How satisfied are you with this event?"	95,2% replied with "satisfied" or "very satisfied"

During the evaluation events, the project team encountered a committed and vocal audience, including participants of previous dissemination and capacity building events. This illustrates that the DE-BIAS project and tool meet the CHIs' needs for support in examining cultural heritage collections' metadata for contentious language. Critique was mainly targeted at the display of detection results, and the project partners incorporated the suggestions, if possible, to allow the users an intuitive interpretation of the tool's output. Furthermore, all explanatory texts in the interface and the report were revised for maximum explicitness. Outstanding adjustments to the user interface and PDF report are dependent on the publication of the DE-BIAS knowledge graph in January 2025 and will be considered for potential future improvements.

---

[8] For the ESACH event, this question was rephrased to "Would you be interested in using the DE-BIAS tool in the future?", as the participants were not expected to work with cultural heritage collections directly.

5. Annex

5.1 Annex I - Feedback questionnaire

Thank you again for your participation in the DE-BIAS Evaluation Event on xxx 2024. To keep improving the usability of the DE-BIAS tool we would highly appreciate it, if you could take a few minutes to reply to this survey. We will ask you to provide the name of the Europeana Aggregator you represent, but otherwise collect and analyse your feedback anonymously together with other feedback received. In case you provide us an email address in order to be kept in the loop about future developments, we will use this information separately from your other answers. We will only use your contact information in the context of the DE-BIAS project and we will not share your details with anyone else.

1. Name of Aggregator
2. Did you try out the stand-alone version of the DE-BIAS tool? ([Debias Tool (ntua.gr)](ntua.gr))
   a. Yes
   b. No

Usability of the stand-alone tool

3. To what extent do you agree with the following statements? Rank for each statement: 1 - Completely disagree; 2 - Somewhat disagree; 3 - Neutral; 4 - Somewhat agree; 5 - Completely agree:

   a. The stand-alone tool was easy to use.
   b. The interface was self-explanatory.
   c. I understood the use of the operations "NER" and "Disambiguation".

4. Did you use the option to insert text directly to check metadata for biassed language?
   a. Yes
   b. No

Stand-alone tool: web interface satisfaction

5. To what extent do you agree with the following statements? Rank for each statement: 1 - Completely disagree; 2 - Somewhat disagree; 3 - Neutral; 4 - Somewhat agree; 5 - Completely agree:

   a. I am satisfied with the detection results.
   b. I am satisfied with the display of detected terms in the web interface.

If you wish to elaborate on your previous answers, please provide your feedback here.

Stand-alone tool: file upload

6. Did you use the file upload functionality of the stand-alone tool to check metadata for biassed language?
   a. Yes
   b. No

Stand-alone tool: file upload satisfaction

7. To what extent do you agree with the following statements? Rank for each statement: 1 - Completely disagree; 2 - Somewhat disagree; 3 - Neutral; 4 - Somewhat agree; 5 - Completely agree:

   a. I am satisfied with the detection results.
   b. I am satisfied with the explanation of the analysis results in the report sent to me by e-mail.

If you wish to elaborate on your previous answers, please provide your feedback here.

Metis Sandbox

8. Did you try out the DE-BIAS tool integrated into the Metis Sandbox?
   a. Yes
   b. No

Usability of the integration in the Metis Sandbox

9. To what extent do you agree with the following statements? Rank for each statement: 1 - Completely disagree; 2 - Somewhat disagree; 3 - Neutral; 4 - Somewhat agree; 5 - Completely agree:

   a. The DE-BIAS tool integration was easy to use.
   b. I am satisfied with the detection results.
   c. I am satisfied with the display of detected terms in the report.

If you wish to elaborate on your previous answers, please provide your feedback here.

Feedback on UI/UX integration on the Europeana Website

10. Do you find the information in the pop-up useful? (1 - Not useful at all to 5 - Very useful)
11. In your opinion, was there anything missing on the detail page?
12. Would you like to have this contextualisation of terms in the records coming from your Aggregator as well?
    a. Yes
    b. No
    c. Maybe
    d. Other

If you wish to elaborate on your previous answers, please provide your feedback here.

Future action

13. Would you be interested in running the DE-BIAS tool on your collections yourself?
14. Will you take any other action or make any changes as a result of this event? If yes, please tell us more.

General Feedback

15. How satisfied are you with this event? (1 - Very dissatisfied to 5 - very satisfied)

5.2 Annex II - Raw survey data

EAF Aggregators' Forum event, project partner aggregators and APEF event

📄 Annex II - EAF Aggregators.pdf

Europeana Network Association event

📄 Annex II - Europeana Network Association.pdf

ESACH event

📄 Annex II - ESACH event.pdf

5.3 Annex III - Comparison .pdf reports before and after T4.3

5.3.1 First iteration of the standalone tool (available until Dec 5, 2024)

📄 Annex III - 20241104_bias_report_testing-Before.pdf

5.3.2 Second iteration of the standalone tool (available from Dec 6, 2024)

📄 Annex III - 20241206_bias_report_testing_After.pdf