



Co-funded by  
the European Union

# DE BIAS

## How to use the DE-BIAS tool

Guidelines

Image: [Worldwide Wall](#) - Associazione Culturale GoTellGo, Italy - CC BY



<https://creativecommons.org/licenses/by-sa/4.0/>



## Document information

<b>Current version</b>	1.2	
<b>Contributors</b>	Kristina Rose (Deutsches Filminstitut und Filmmuseum), Jason Liartis (ThinkCode)	
History of changes		
Version	Change(s) applied	Contributors
1.0	First version with documentation for standalone tool and Metis Sandbox integration.	Kristina Rose (DFF)
1.1	Updated chapter on PDF report and Metis Sandbox after revision.	Kristina Rose (DFF)
1.2	Documentation for API added. Formatting and editorial review.	Jason Liartis (ThinkCode) Kerstin Arnold (Archives Portal Europe)

## Target audience

- Cultural heritage professionals, researchers, and general users as users of the standalone version of the DE-BIAS tool.
- Aggregators and other data partners as users of the DE-BIAS tool integrated into the Metis Sandbox.
- Aggregators and cultural heritage institutions, including their software development teams or service providers, interested in integrating the DE-BIAS tool's API endpoint into their own workflows.

## Learning goals

- Understand how to manually input textual metadata for analysis in the standalone tool
- Understand how to prepare data for upload in the standalone tool
- Understand how to use the functions of Named Entity Recognition (NER) and Disambiguation to influence the analysis result appropriately to the data to be checked
- Understand how to generate a bias report when using the Metis Sandbox
- Understand how to read and analyse the outputs generated by the tool
- Apply the findings of the tool to the descriptions of your own cultural heritage collections
- Apply the API endpoint in your own workflows

**Content note: Readers are advised that this document contains distressing words.**



## Table of contents

<b>Document information.....</b>	<b>2</b>
Target audience.....	2
Learning goals.....	2
<b>Standalone version: Data input.....</b>	<b>4</b>
Insert texts.....	4
Language selection.....	4
Named Entity Recognition (NER) and Disambiguation.....	4
Text input.....	5
Upload a file.....	5
How to prepare data for upload.....	5
Language selection.....	6
Named Entity Recognition and Disambiguation.....	6
File upload.....	7
<b>Standalone version: Data output.....</b>	<b>7</b>
When using 'Insert texts'.....	7
When using 'Upload a file'.....	8
Raw data.....	8
Statistical report.....	9
<b>Metis Sandbox integration: Data input.....</b>	<b>11</b>
<b>Metis Sandbox integration: Data output.....</b>	<b>12</b>
<b>Direct use of the API: Data input.....</b>	<b>14</b>
The simplified API call.....	14
Example input.....	14
<b>Direct use of the API: Data output.....</b>	<b>15</b>
Example output.....	15
<b>Annex: Europeana specifications for structured records.....</b>	<b>17</b>
Request.....	17
Response.....	19
Example response.....	20



## Standalone version: Data input

### Insert texts

If you want to analyse only a small amount of data for potential bias, the option of direct data input is your best choice. The tool is available here: <https://debias-tool.ails.ece.ntua.gr/>

### Language selection

The screenshot shows the 'Insert texts' tab selected. Below it, a step indicator '1' is followed by 'Language \*'. A text prompt says 'Select the language of the data you are uploading.' Below this is a dropdown menu with 'German' selected.

You can choose from five languages: **English, Italian, French, German** and **Dutch**. Select the language applicable to the texts you want to check via the drop-down menu. This is a required step to ensure that the tool only considers terms from the underlying DE-BIAS vocabulary in the chosen language during the analysis. In total, the vocabulary contains close to 700 terms in the aforementioned languages alongside their historical context and, where applicable, suggestions for alternative terms.

### Named Entity Recognition (NER) and Disambiguation

The screenshot shows the '2 Operations' section. It contains two toggle switches. The first is for 'NER', which is turned on (blue), with the description 'Filter out detections that are part of named entities'. The second is for 'Disambiguation', which is turned off (grey), with the description 'Use context to decide whether detected terms are contentious'.

These two operations are optional and allow you to enhance the precision of the term matching by the tool, i.e. the number of detected bias terms that are indeed relevant when you intend to 'debias' your data. By default, they will be disabled, but you can easily enable one or both of them by moving the slider to the right (as shown for the operation of Named Entity Recognition (NER) above). If a higher recall instead of a higher precision is favoured, i.e. when you feel the tool filters out too much, these two options can be disabled again. Please note that enabling these operations can increase the time needed for processing the data.



**Named Entity Recognition (NER)** is the task of finding spans in a text that refer to named entities such as persons, places, organisations etc. Finding named entities in a text is an important step in filtering vocabulary terms detected in the text where these are used in non-contentious ways. For example, the German word “Mohr” is harmful when used to refer to people but it also appears as a last name. NER detects when “Mohr” appears in the text as part of a person’s name and filters it from the annotations.

**Disambiguation** comes into play when a term can hold a contentious meaning in one context while being neutral outside of that context. One such example is the English term “race” that is contested when used for categorising humans based on physical features, but completely unproblematic when used in the context of motor racing and other racing sports. Enabling this operation tells the tool to check with the help of a Large Language Model (LLM) if a detected term is used in a contentious context as described in the vocabulary. ‘Context’ only refers to the textual content directly surrounding the detected term, e.g. the complete title of a cultural heritage object, or its complete description. Digital representations of the described item or other metadata fields, however, are not taken into account when establishing a term’s context.

## Text input

3 Enter text \*

This is a text|

+ Add text

In the last step, you can then enter the text you wish to analyse. You can either type text into the empty field directly or copy/paste text from your source. It is possible to analyse multiple text elements at once by clicking on “Add text” below the last text box and thus creating an additional text box.

## Upload a file

If you want to analyse a bigger amount of records for potential bias at once, the option to upload a file is your best choice. The results of this option will not show in the tool’s interface, but you will receive a report with the analysis results sent to the email address you provided during file submission. The tool is available here: <https://debias-tool.ails.ece.ntua.gr/>

## How to prepare data for upload

When preparing data for upload, each record should be stored in a separate .txt file. The tool will treat each .txt file as one value to be analysed and will associate the annotations for detected bias terms to each value. The output of the tool will accordingly reference the file names. All .txt files to be analysed - even if it is only one - need to be provided to the tool in .zip format. Since the language of the metadata is [selected once per analysis for all records](#), a .zip file should only contain records with metadata in one



language. Prepare separate .zip files if you wish to analyse metadata in different languages.

## Language selection

Insert texts    Upload a file

**1** Language \*

Select the language of the data you are uploading.

German

You can choose from five languages: **English, Italian, French, German** and **Dutch**. Select the language applicable to the texts you want to check via the drop-down menu. This is a required step to ensure that the tool only considers terms from the underlying DE-BIAS vocabulary in the chosen language during the analysis. In total, the vocabulary contains close to 700 terms in the aforementioned languages alongside their historical context and, where applicable, suggestions for alternative terms.

## Named Entity Recognition and Disambiguation

**2** Operations

NER  
Filter out detections that are part of named entities

Disambiguation  
Use context to decide whether detected terms are contentious

These two operations are optional and allow you to enhance the precision of the term matching by the tool, i.e. the number of detected bias terms that are indeed relevant when you intend to 'debias' your data. By default, they will be disabled, but you can easily enable one or both of them by moving the slider to the right (as shown for the operation of Named Entity Recognition (NER) above). If a higher recall instead of a higher precision is favoured, i.e. when you feel the tool filters out too much, these two options can be disabled again. Please note that enabling these operations can increase the time needed for processing the data.

**Named Entity Recognition (NER)** is the task of finding spans in a text that refer to named entities such as persons, places, organisations etc. Finding named entities in a text is an important step in filtering vocabulary terms detected in the text where these are used in non-contentious ways. For example, the



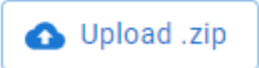
German word “Mohr” is harmful when used to refer to people but it also appears as a last name. NER detects when “Mohr” appears in the text as part of a person’s name and filters it from the annotations.

**Disambiguation** comes into play when a term can hold a contentious meaning in one context while being neutral outside of that context. One such example is the English term “race” that is contested when used for categorising humans based on physical features, but completely unproblematic when used in the context of motor racing and other racing sports. Enabling this operation tells the tool to check with the help of a Large Language Model (LLM) if a detected term is used in a contentious context as described in the vocabulary. ‘Context’ only refers to the textual content directly surrounding the detected term, e.g. the complete title of a cultural heritage object, or its complete description. Digital representations of the described item or other metadata fields, however, are not taken into account when establishing a term’s context.

## File upload

**3** Upload .zip file \*

The .zip file should contain text files (.txt) with the data you want to analyze. The text files should be encoded in UTF-8.



At this point, the .zip file containing the records in .txt format needs to be uploaded. Clicking the ‘Upload .zip’ button will open your local file explorer and allow you to select the .zip file you prepared (see [How to prepare data for upload](#) for details). Once the upload is complete, the name of your .zip file will show next to the button and a waste bin icon will become available to its right. Clicking the waste bin icon, you can remove your upload again, if necessary. Should your .zip file contain any files in a different format (e.g. .csv, .xml, .json etc.), the tool will display an error message.

## Standalone version: Data output

### When using ‘Insert texts’

When a contentious term is detected by the tool, it will be highlighted in the original text in the results display. When you click on the highlighted term, a pop-up window opens where you can access the link to the relevant term in the DE-BIAS vocabulary with all its contextual information. (Please note that at the time of writing this manual, these URIs are still example links and do not have their final domain yet.)

Having access to the full vocabulary entry will provide you with more information to understand why a term has been flagged by the tool in a specific context. Where applicable, this will also include suggestions for alternative phrasings or terms, which you can then consider using for describing - or searching - cultural heritage objects moving forward.



### Analysis Report

1 This test text contains the word "gypsies". 1 found

New Analysis

## When using 'Upload a file'

When uploading a bulk of records via .zip file upload, a twofold report with the analysis results is sent to you via email.

### Raw data

On the one hand, you will receive a .json file that lists the text value (*literal*) and the annotations (*tags*), including their position (*start*, *end*, *length*), per analysed .txt file (*id*). It also includes the link to the vocabulary entry (*uri*), where you can find more context information. *(Please note that at the time of writing this manual, these URIs are still example links and do not have their final domain yet.)*

```

7:
  language: "de"
  literal: "Indianerkapelle, Markgraf Christian von Bran
289 x 527 mm. Dresden: SLUB Mscr.Dresd.J.18\r
gnd/4006568-6\r\nDatenpartner\r\nDeutsche Fot
\r\nErstellungsdatum\r\n1990\r\nHerkunft\r\nS
k.de/item/4NA4VY56ATII67YP2RLNT5WON2SZIVGC\r\
ver\u00f6ffentlicht\r\n2020-12-16T09:50:12.330Z\r\
  tags:
    0:
      uri: "http://www.example.org/debias#t 38 de"
      start: 0
      end: 15
      length: 15
    1:
      uri: "http://www.example.org/debias#t 38 de"
      start: 56
    
```

You can use this .json file for further processing, if necessary with support from your software development team or service provider. With the information about the detected term's position within the text, you can e.g. consider highlighting the term in the user interface of your own systems. And with the information from the full vocabulary entry, you can e.g. create annotations to be added to the highlighted terms.

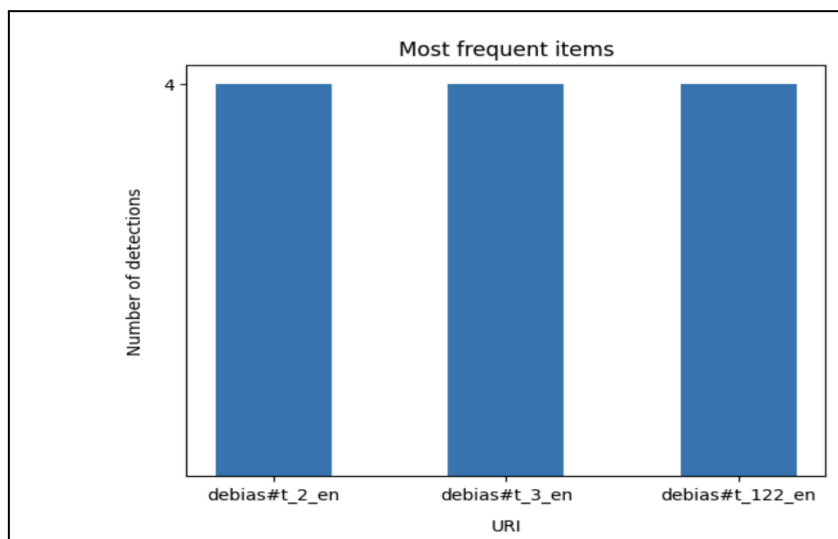




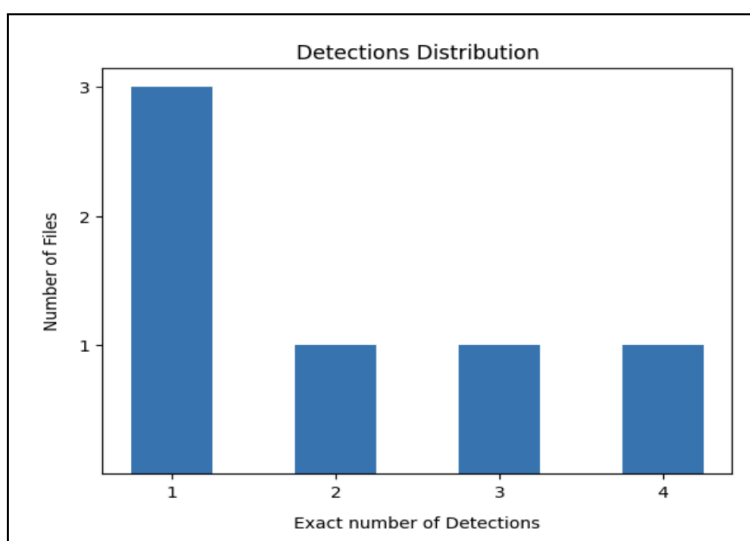
## Statistical report

This report (in PDF format) contains general information about the number of files processed, the metadata language, and the following statistics:

- Total number of annotations, i.e. instances of detected bias.
- The absolute number of files processed containing at least 1 detected term.
- The 10 terms or issues from the vocabulary that have been detected most frequently.



- An overview grouping the analysed files by the number of annotations per file (the x-axis shows the number of detections per file and the y-axis indicates the number of files with the respective number of annotations)



- A table that lists the name of the analysed file(s) and the vocabulary URI of the term(s) that was or were detected. *(Please note that at the time of writing this manual, these URIs are still example links and do not have their final domain yet.)*



This table shows the bias detections per file:

<b>File Name</b>	<b>Detected terms</b>
13766472057481376647205748.txt	<a href="http://www.example.org/debias#t_31_de">http://www.example.org/debias#t_31_de</a>
item_7LFNOOJRO32JZ56PDCZ4RBCFKIGBRX2M.txt	<a href="http://www.example.org/debias#t_25_de">http://www.example.org/debias#t_25_de</a> <a href="http://www.example.org/debias#t_25_de">http://www.example.org/debias#t_25_de</a>



## Metis Sandbox integration: Data input

If you want to analyse a bigger amount of records for potential bias as part of the already established process for preparing data for submission to Europeana.eu, the option to initiate a bias report in the Metis Sandbox is your best choice. The integration of the DE-BIAS tool in the Metis Sandbox is currently available in a test environment: <https://metis-sandbox.acceptance.eanadev.org/>.

The data input for analysing records for bias via the Metis Sandbox works just as the standard data submission to the Sandbox. You can find detailed instructions on how to prepare and upload records to the Sandbox here: [Metis Sandbox User Guide - Europeana Knowledge Base](#).

Once you have submitted your dataset, please wait until the whole processing is finished which is indicated by a green dot next to the *Publish* process. Additionally, the link *view published records* will be visible at that point in the upper right corner of the user interface.

Although having the records visible in the Sandbox version of the Europeana.eu frontend takes a couple of minutes, you can now click on *run report* to start the bias analysis.

The screenshot shows the DEBIAS tool interface for a dataset named 'DEBIAS\_FR\_Test'. The dataset is associated with 'France / French' and was submitted on '17/12/2024, 10:15:36'. The number of records is 5151. There are two main buttons: 'view published records' and 'run report' (with a DEBIAS icon). A progress bar shows the status of various processing steps, with a green circle indicating completion. Below the progress bar is a 'Track Processing' table.

Step	Progress
Harvest (file)	10 / 10
Validation (external)	10 / 10
Transformation	10 / 10
Validation (internal)	10 / 10
Normalisation	10 / 10
Enrichment	10 / 10
Media Processing	10 / 10
Publish	10 / 10

Additional buttons visible in the interface include 'view details (6)' and 'view details (6)'.

Please note that differently to the standalone tool, the functionalities [Named Entity Recognition \(NER\)](#) and [Disambiguation](#) are enabled by default, when using the bias detection tool in the Sandbox. This means, that e.g. proper names of persons and geographical names are filtered out from the detection results, and that terms having contentious and non-contentious meanings are flagged only for instances, where a contentious use is assumed.

Furthermore, the language of your metadata will be detected automatically based on the values of the `xml:lang` attribute in the metadata elements `dc:title`, `dcterms:alternative`, `dc:description`, `dc:subject`, `dc:type`



and *skos:prefLabel* in linked contextual classes, if included. This means that the DE-BIAS tool integrated in the Sandbox can also analyse multilingual datasets and you do not have to split your data per language like in the '[Upload a file](#)' option of the standalone tool. Please note, that only the following languages can be analysed by the tool: de, en, fr, it, nl.

## Metis Sandbox integration: Data output

Once the compilation of the bias report is finished, click on *view report* in the upper right corner to access it.

DEBIAS\_FR\_Test ✓

France / French

17/12/2024, 10:15:36

5151

[view published records](#) ↗

[view report](#) 

A pop-up window will open and display a list of all detected instances of bias including a record identifier (*Europeana Id*), the name of the respective metadata element (*Field*), the language of the detected term (*Language*) and part of the text where the term was found (*Literal*).

The detected terms are highlighted in yellow and linked with the URI of the term in the DE-BIAS vocabulary. (*Please note that at the time of writing this manual, these URIs are still example links and do not have their final domain yet.*)

 Dataset 5151 - Debias Report (10) ✕

  **Europeana Id:** [/5151/proxy\\_provider\\_794\\_ark\\_\\_12148\\_bpt6k3339004k](/5151/proxy_provider_794_ark__12148_bpt6k3339004k) ↗

**Field:** [dc:title](#)

**Language:** fr (French)

**Literal:** [La Chinetoque](#) ↗ et le gigolo / J. Numed

---

**Europeana Id:** [/5151/proxy\\_provider\\_794\\_ark\\_\\_12148\\_bpt6k57735787](/5151/proxy_provider_794_ark__12148_bpt6k57735787) ↗

**Field:** dc:title

**Language:** fr (French)

**Literal:** Voyage au pays des [Esquimaux](#) ↗...

Close



Having access to the full vocabulary entry will provide you with more information to understand why a term has been flagged by the tool in a specific context. Where applicable, this will also include suggestions for alternative phrasings or terms, which you can then consider using for describing - or searching - cultural heritage objects moving forward.

Furthermore, the *Europeana Id* will give you access to the preview of the object with all its metadata and its digital representation as it will appear on Europeana.eu once officially uploaded. While this won't show the DE-BIAS highlight itself, having access to this full display can provide you with more information related to the object in question that might help in deciding whether to act upon the bias detection. For example, there might be cases where the tool correctly flags a term based on the information available in the metadata elements included in the analysis, but where the digital object or other metadata fields indicate that the term is actually used in a non-contentious way.

The bias report will remain available to you in the Metis Sandbox as long as your dataset is available. Please note that, due to being a test environment, the Sandbox is cleared regularly and datasets will be removed after a certain amount of time. To ensure that you still have access to the analysis results from the bias detection, you can also download the full report via the little download icon on the top left of the pop-up including the bias report.

The report will be downloaded in .csv format, which you can open in Excel, Numbers or OpenDocument spreadsheets locally or share as an upload to Google Sheets. Next to the information available in the report online, this download also includes information about the position (*start, end, length*) of the detected terms within the metadata element where they were found.

Having the analysis report available in this downloadable format, allows you to run any statistics you might be interested in or filter detected terms based on specific criteria, e.g. if you would want to only look at bias detected in titles and descriptions. Similar to the [raw data report](#) received when using the 'Upload a file' option in the standalone tool, you can use the information in this downloaded report as a basis for further processing, e.g. when considering to highlight or annotate terms.

5139\_debias\_report

dataset-id	creation-date	detection_recordid	detection_europeanald	detection_sourceField	Detection_language	Detection_literal	tags_start	tags_end	tags_length	detection_value	Detection_tags_uri
5139	2024-12-11T00	242505	/5139/item_2022362_Rc	DC_TITLE	en	?Negroes at Schc	1	8	7	debias:t_165_en	
		242507	/5139/item_08602_IL500	DC_DESCRIPTION	en	3 - the rag giraffe	29	34	5	debias:t_76_en	
		242508	/5139/item_2048015_Ath	DC_TITLE	en	Sturdza, the hatr	95	100	5	debias:t_99_en	
		242501	/5139/item_08602_IL500	DC_DESCRIPTION	en	2 - the visiting tee	61	72	11	debias:t_171_en	
		242501	/5139/item_08602_IL500	DC_DESCRIPTION	en	3 - then the team	45	56	11	debias:t_171_en	
		242501	/5139/item_08602_IL500	DC_DESCRIPTION	en	Rome: Undersecr	84	92	8	debias:t_128_en	
		242507	/5139/item_08602_IL500	DC_DESCRIPTION	it	3 - la giraffa di pe	35	39	4	debias:t_18_it	
		242509	/5139/item_2064121_Mu	DC_DESCRIPTION	de	Dieses Blatt gehö Auf diesem Blatt Beschriftung: Plat	105	110	5	debias:t_153_de	
		242504	/5139/item_440_item_HX	DC_TITLE	de	Zigeunerin, einer	0	10	10	debias:t_32_de	
		242503	/5139/item_2048221_eur	DC_TITLE	en	Women of the Bo	18	23	5	debias:t_208_en	
		242509	/5139/item_2064121_Mu	DC_SUBJECT_REFER	en	Prostitution	0	12	12	debias:t_180_en	
		242500	/5139/item_08602_IL500	DC_DESCRIPTION	en	15 - ella crosses f	57	62	5	debias:t_165_en	
		242506	/5139/item_2048221_eur	DC_TITLE	en	Costume design f	0	7	7	debias:t_53_en	
								19	25	6	debias:t_212_en
		242504	/5139/item_440_item_HX	DC_TITLE	en	Gypsy woman, tr	0	5	5	debias:t_99_en	
		242500	/5139/item_08602_IL500	DC_DESCRIPTION	it	15 - ella attravers	60	65	5	debias:t_7_it	



## Direct use of the API: Data input

The DE-BIAS API is available at <https://debias-api.ails.ece.ntua.gr/>. It implements two methods, a simplified POST request at <https://debias-api.ails.ece.ntua.gr/simple> and a POST request at the base URL that follows the [Europeana specifications for structured records](#).

You can access and use the API using any program or programming language capable of sending POST requests and receiving JSON objects, no other prerequisites are needed. You might need the support from your software development team or service provider to make full use of this option.

### The simplified API call

For the simplified API call (<https://debias-api.ails.ece.ntua.gr/simple>) you will need to send a JSON object as part of the request body containing the fields **"language"** and **"values"**.

- The **"language"** field must have an ISO 639-1 language code as a value such as "en" for English or "de" for German. The languages currently supported are Dutch (nl), English (en), French (fr), German (de) and Italian (it).
- The **"values"** field must be a list of character strings compatible with the JSON format. The character strings are the texts that the tool will process, e.g. the title and/or description of a cultural heritage object.
- There are two additional optional arguments, **"useNER"** and **"useLLM"**, which enable/disable contextual disambiguation using Named Entity Recognition and Large Language Models, respectively. By default, these are set to "true", i.e. NER and disambiguation will always be applied. For further information about these two options, see the chapter [Named Entity Recognition and Disambiguation](#) above or refer to the report on the functionalities of the DE-BIAS tool<sup>1</sup>.

### Example input

A user desires to detect terms from the DE-BIAS vocabulary for descriptions of these two items found on Europeana.eu:

- [https://www.europeana.eu/en/item/2051943/data\\_euscreenXL\\_EUS\\_AFBECA25E7F843DE842CC3D6ECDDCECE](https://www.europeana.eu/en/item/2051943/data_euscreenXL_EUS_AFBECA25E7F843DE842CC3D6ECDDCECE) with the description *"Negro-parade". The students from Aurehoej upper secondary school in Gentofte celebrate their last day in school by throwing a parade. They have painted themselves black and dressed up to look like an afro-american parade in New Orleans.* and
- [https://www.europeana.eu/en/item/322/Museu\\_ProvidedCHO\\_Mus\\_es\\_Royaux\\_d\\_Art\\_et\\_d\\_Histoire\\_86669](https://www.europeana.eu/en/item/322/Museu_ProvidedCHO_Mus_es_Royaux_d_Art_et_d_Histoire_86669) with the description *Sales contract of a slave for 10 golden shekels, dated to the 4th year of Gimil-Sin.*

---

1



The simplified API call should have the following body:

```
{
  "language": "en",
  "values" :
  [
    "\"Negro-parade\". The students from Aurehoej upper secondary school in Gentofte celebrate their last day in school by throwing a parade. They have painted themselves black and dressed up to look like an afro-american parade in New Orleans.",
    "Sales contract of a slave for 10 golden shekels, dated to the 4th year of Gimil-Sin."
  ]
}
```

The “values” field is a list containing the two texts to be analysed as specified above. Notice the escaped quotation marks to ensure compatibility with the JSON format.

## Direct use of the API: Data output

The response body of the simplified API call is a JSON object containing two fields, “**metadata**” and “**results**”.

- The **metadata** field contains some simple metadata about the API call.
- The **results** field contains a list of JSON objects with terms detected from the DE-BIAS vocabulary. Each JSON object of the “results” field corresponds to one of the texts in the “values” field of the API call. They are returned in the same order. Each JSON object of the “results” field contains the fields “**language**”, “**literal**” and “**tags**”.
  - The “**language**” and “**literal**” fields contain the information from the API call with the language and the value of the input text.
  - The “**tags**” field contains the URIs of the detected terms and their location within the text. Specifically, it is a list of JSON objects, each corresponding to one detected term. The JSON objects contain the fields “**uri**”, “**start**”, “**end**” and “**length**”.
    - The “**uri**” field contains the URI of the detected term as specified by the DE-BIAS vocabulary.
    - The “**start**” and “**end**” fields specify the positions of the starting and ending characters of the text span containing the term, and the “**length**” field specifies the text span length.

### Example output

Considering the example for the API call from the previous chapter, the API response is the following:

```
{
  "metadata": {
    "annotator": "de-bias",
    "thesaurus": null,
    "date": "2024-12-11T17:18:01"
  },
  "results": [
    {
```



```
    "language": "en",
    "literal": "\"Negro-parade\". The students from Aurehoej upper secondary school in
Gentofte celebrate their last day in school by throwing a parade. They have painted themselves
black and dressed up to look like an afro-american parade in New Orleans.",
    "tags": [
      {
        "uri": "debias:t_165_en",
        "start": 1,
        "end": 6,
        "length": 5
      },
      {
        "uri": "debias:t_6_en",
        "start": 201,
        "end": 214,
        "length": 13
      }
    ]
  },
  {
    "language": "en",
    "literal": "Sales contract of a slave for 10 golden shekels, dated to the 4th year of
Gimil-Sin.",
    "tags": [
      {
        "uri": "debias:t_198_en",
        "start": 20,
        "end": 25,
        "length": 5
      }
    ]
  }
]
```

Besides the metadata and the repetitions of the analysed texts, the response contains the terms detected in the "tags" field of each item of the "results" field. The first item in the "results" field corresponds to the first text of the API call and its "tags" field signifies that the DE-BIAS URIs "debias:t\_165\_en" (corresponds to the literal "Negro") and "debias:t\_6\_en" (corresponds to the literal "Afro-American") were detected. The location of the terms within the text is also specified. The second item in the "results" field corresponds to the second text of the API call and its "tags" field signifies that the DE-BIAS URI "debias:t\_198\_en" (corresponds to the literal "Slave") was detected. The location of the term within the text is also specified.

Further information about the API calls can be found at <https://debias-api.ails.ece.ntua.gr/docs>.





## Annex: Europeana specifications for structured records

This version of the API call and response are tailored for the integration of the DE-BIAS API with Europeana.eu and contain further fields in the call and response for a more structured output.

### Request

```
HTTP: POST
```

```
{
  "@context": [
    <CONTEXT_URI>,
    { "@base": "http://data.europeana.eu/item/" }
  ],
  "type": "Request",
  "params": {
    "limitPerPredicate": <LIMIT>, \\ optional, can be omitted
    "language": <LANGUAGE>,
    "provenance": <PROVENANCE>
  },
  "totalItems": <TOTAL>,
  "items" : [
    {
      "id": <RECORD_ID>,
      "<FIELD_NAME>": [ <FIELD_VALUE>, ... ]
      ...
    },
    ...
  ]
}
```

The fields in this call body ask for the following input:

- <CONTEXT\_URI>, standing for the URI of the JSON-LD context
- <LIMIT>, meaning the number of mentions to be returned per vocabulary URI, per record, per metadata field, i.e. here the Europeana Data Model (EDM) element (optional, can be omitted to get all occurrences of all detected terms)



- <LANGUAGE>, giving the language of the metadata
- <PROVENANCE>, being a boolean value to inform the tool to return or not the provenance information associated with the annotation, meaning “created” and “creator” fields and any other additional provenance information such as confidence levels
- <TOTAL>, indicating the total number of items that are part of the request (optional)
- <RECORD\_ID>, providing the local identifier of the record, in relation to the @base, or otherwise, the complete URI of the record within - here - the data.europeana.eu namespace.
- <FIELD\_NAME>, standing for the namespace prefixed name of the EDM element (e.g. “dc:title”)
- <FIELD\_VALUE>, meaning the literal value that was associated with the field; more than one value may be supplied as an array

```
{
  "@context": [
    <CONTEXT_URI>,
    { "@base": "http://data.europeana.eu/item/" }
  ],
  "type": "Request",
  "params": {
    "limitPerPredicate": 1,
    "language": "en",
    "provenance": false
  },
  "totalItems": 2,
  "items" : [
    {
      "id": "12345/XPTO",
      "dc:title": [ "a sample title", "a second sample title" ],
      "dc:description": [ "a sample description", "a second sample description" ]
      ...
    },
    {
      "id": "12345/XPTO_2",
      "dc:title": [ "another sample title", "another second sample title" ],
      "dc:description": [ "another sample description", "another second sample
description" ]
      ...
    },
  ],
}
```



```
...
]
}
```

## Response

```
{
  "@context": [
    "http://www.w3.org/ns/anno.jsonld",
    "https://www.europeana.eu/schemas/context/edm.jsonld",
    "@base": "http://data.europeana.eu/item/",
  ],
  "type": "AnnotationPage",
  "partOf": {
    "type": "AnnotationCollection",
    "modified": <DATE_TIME>
  },
  "items" : [
    <ANNOTATION>,
    ...
  ]
}
```

Where <ANNOTATION> is indicated, the following information will be provided:

```
{
  "id": "http://example.org/anno29",
  "type": "Annotation",
  "motivation": "highlighting",
  "body": <VOCABULARY_URI>,
  "target": [
    {
      "source": <RECORD_ID>,
      "selector": {
        "type" : "RDFStatementSelector",
        "hasPredicate": "<FIELD_NAME>",
        "refinedBy" : {
          "type" : "TextQuoteSelector",
```



```
    "exact": {
      "@value": <BIASED_TERM>,
      "@language": <LANGUAGE>
    },
    "prefix": <PREFIX>
    "Suffix": <SUFFIX>
  }
}
},
{
  "source": ...,
  "selector": {
    "type" : "RDFStatementSelector",
    "predicate": ...,
    "refinedBy" : {
      "type" : "TextQuoteSelector",
      "exact": {
        "@value": ...,
        "@language": ...
      },
      "prefix": ...,
      "suffix": ...
    }
  },
  ...
}
...
}
```

The "prefix" and "suffix" should also be stated unless the detected term itself is at the start or at the end of the analysed text fragment. The minimum character length of both prefix and suffix (separately) should be 50 characters. If the 50th character happens to be in the middle of a word, then the complete word should be included respecting this way the 50 characters minimum.

### Example response

```
{
  "id": "http://example.org/anno29",
  "type": "Annotation",
```



```
"motivation": "highlighting",
"body": "http://debias.example.org/term/xpto",
"target": [
  {
    "source": "http://data.europeana.eu/item/some_cho",
    "selector": {
      "type" : "RDFStatementSelector",
      "predicate": "dc:title",
      "refinedBy" : {
        "type" : "TextQuoteSelector",
        "exact": {
          "@value": "bias term",
          "@language": "en"
        },
        "prefix": "same "
      }
    }
  },
  {
    "source": "http://data.europeana.eu/item/some_cho",
    "selector": {
      "type" : "RDFStatementSelector",
      "predicate": "dc:description",
      "refinedBy" : {
        "type" : "TextQuoteSelector",
        "exact": {
          "@value": "bias term",
          "@language": "en"
        },
        "prefix": "with a ",
        "suffix": ". Also"
      }
    }
  },
  ...
]
...
}
```