



Co-funded by
the European Union



D3.1 User Requirements in the Form of User Stories and Epics

Date of submission	31 May 2023
Author(s)	Sofie Taes, Fred Truyen (KUL) with contributions from Kerstin Arnold (APEF), Marco Rendina (EFHA), Kristina Rose (DFF), Vassilis Tzouvaras (DATOPTRON), Orfeas Menis (ThinkCode), Antoine Issac, Michelle Lewis, Jochen Vermeulen (EF)
Reviewers	Marco Rendina (EFHA), Vassilis Tzouvaras (DATOPTRON)
Dissemination level	PU

HISTORY OF CHANGES		
Version	Publication/Submission date	Author(s)
1.0	31 May 2023	Sofie Taes, Fred Truyen (KUL)

TABLE OF CONTENT

1. Introduction	2
2. Scope of the Deliverable	2
3. The DE-BIAS Tool: Context & Rationale	3
3.1. Objectives	3
3.2. Technological Background	3
3.3. Overall Concept	3
3.4. User- and Workflow-Specific Tool Versions	4
4. User Stories & Epics	4
4.1 Scenario Strand A: Integrated with MINT	5
4.2 Scenario Strand B: Integrated with Europeana's Core Service	7
4.3 Scenario Strand C: Independent Use of the API	10
5. Tool Outline Based on User Requirements and Available Functionalities	12
5.1 Core Elements and Functionalities	12
5.2 AI Components	13
6. Conclusion and Next Steps	14

1. Introduction

The DE-BIAS project aims to promote a more inclusive and respectful approach to describing cultural heritage collections and telling the stories and histories of minority communities. Over the course of two years, the project will develop an artificial intelligence (AI)-based tool to automatically detect potentially harmful and offensive terms in cultural heritage descriptive metadata and provide information about their problematic background.

The DE-BIAS tool will use vocabularies that combine offensive language with contextual information and suggestions for more appropriate terms. AI solutions such as the DE-BIAS tool empower Cultural Heritage Institutions (CHIs) as hubs of valuable knowledge, ideas and artefacts, while making them more sustainable and attuned to what audiences in the new millennium not only want, but expect from their heritage experiences.

2. Scope of the Deliverable

In this deliverable we describe user stories and epics to gain insight into the personas and scenarios that will shape the core functionalities of the DE-BIAS tool. In doing so, we represent the real-life situations, workflows and aspirations of direct stakeholders in the project consortium, primarily aggregators, as they are one of the main target groups for which the tool is intended.

In order to get to a point where the technical components - introduced and described in basic terms in the project's Grant Agreement - can be linked and fine-tuned to the use cases at hand, user requirements will be extrapolated from the stories. These will form the linchpin between what users of the tool would ideally expect as an experience, and the technical elements that will form the backbone of the tool.

What this deliverable does not include is a complete and detailed blueprint of the DE-BIAS tool: translating the stories and their respective requirements into technical specifications is an ongoing project task (T3.1) that runs until M16 (April 2024). Then, in M20 (August 2024), D3.2 will be submitted as a report describing the final version of the functionalities of the DE-BIAS tool from a user and technical perspective, reflecting a starting point that combines readily available information on technical components with the all-important user-driven functionalities.

This deliverable also does not discuss in detail the approach or technical specifications related to the integration of the tool into the Europeana infrastructure (T3.3, 3.4, 3.5 & D1.3), its interoperability with the CrowdHeritage platform (D4.1), or the validation and evaluation process (D4.2) that will be used to ensure that the tool works in accordance with user feedback. This is because the tasks related to these aspects are at an early stage, with specific deliverables planned for later in the project.

Finally, as the present document reflects our investigations and discussions at this particular time and stage of the project, upcoming discussions about integration into the Europeana infrastructure (especially with regard to Metis and/or the Metis Sandbox) and into MINT may

lead to a revision of the user stories, requirements and acceptance criteria. In this case, we will recalibrate the composition and functionality of the tool, while trying to maintain the general principles of the requirements expressed by the stakeholders.

3. The DE-BIAS Tool: Context & Rationale

3.1. Objectives

The goal of the tool is to parse cultural heritage metadata records and identify terms that are considered contentious, based on the vocabulary and associated information to be defined in the project. The data management module will store and handle input and output metadata, as well as reports generated as a result of the bias detection process.

3.2. Technological Background

In recent years, AI has been widely used to automatically detect bias and offensive language. Many different technologies and techniques have been applied, including machine learning, deep learning, meta-learning, and knowledge graphs for bias detection and debiasing in various domains. There is active research with various applications with remarkable results in the field of NLP regarding bias and debiasing, with studies dealing with all forms of bias such as racial, social, and gender.¹

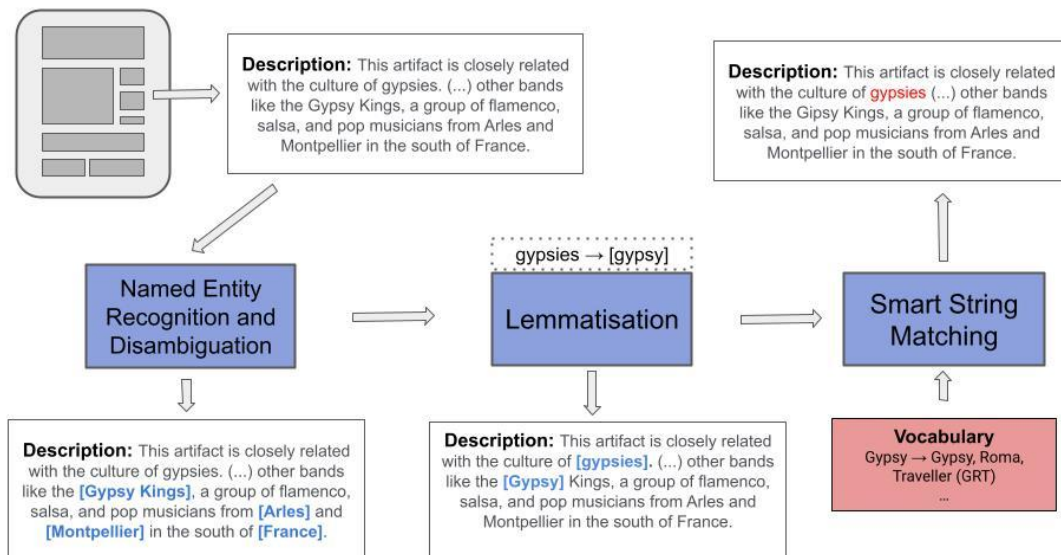
In addition, knowledge graphs have emerged as a curated and transparent component of AI systems that can produce effective results either through reasoning algorithms or in collaboration with other techniques such as lemmatizers and intelligent string matching, but they can also work with deep learning models that mitigate the disadvantage of the complexity of such systems and improve the quality of the result.

3.3. Overall Concept

In order to best serve the intentions and setup of this project, and to fit in with its goals and the workflows of its main stakeholders (aggregators and CHIs), we will combine techniques from the fields of Natural Language Processing (NLP) and Knowledge Graphs to develop an AI-based DE-BIAS tool that meets the specific needs of both the cultural heritage sector and

¹ See e.g. Zhao, Jieyu, and Kai-Wei Chang. "LOGAN: Local group bias detection by clustering." arXiv preprint arXiv:2010.02867 (2020); Chen, Jiawei, et al. "AutoDebias: Learning to debias for recommendation." Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021; Wei, Feng, et al. "Adaptive Alleviation for Popularity Bias in Recommender Systems with Knowledge Graph." Security and Communication Networks 2022 (2022); Nikolov, Andriy, and Mathieu d'Aquin. "Uncovering Semantic Bias in Neural Network Models Using a Knowledge Graph." Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020; Dev, Sunipa, et al. "What do Bias Measures Measure?." arXiv preprint arXiv:2108.03362 (2021); Babaeianjelodar, Marzieh, et al. "Quantifying gender bias in different corpora." Companion Proceedings of the Web Conference 2020; Brate, Ryan et al. / Capturing Contentiousness: Constructing the Contentious Terms in Context Corpus. K-CAP 2021 - Proceedings of the 11th Knowledge Capture Conference. Association for Computing Machinery, Inc, 2021. pp. 17-24; Hogan, Aidan, et al. "Knowledge graphs." Synthesis Lectures on Data, Semantics, and Knowledge 12.2 (2021): 1-257.

the communities involved in the project, ultimately improving the user experience on the Europeana website.



3.4. User- and Workflow-Specific Tool Versions

The DE-BIAS project will investigate the integration of the tool with the Metis suite as part of the Data Space infrastructure to detect biased expressions attached to digital datasets at the publication stage, as well as the integration of the tool with the MINT aggregation platform, where content providers who have their data uploaded can apply the DE-BIAS tool.

Furthermore, the project will provide all the necessary documentation to enable the integration of the tool by other external platforms through an API - e.g. the respective infrastructures of aggregators or CHIs - in order to analyse datasets for bias prior to any data publication steps. The implementation of this scenario will then be up to each individual aggregator or CHI wishing to use the DE-BIAS tool within their own infrastructure.

4. User Stories & Epics

In the following grid, we explore personas and use cases relevant to the proposed DE-BIAS tool. We focus on four main stakeholder groups:

- **The operator of the Europeana Core Service** and the Data Space in which the tool will be integrated and the results of its analysis will reach the end user.
- **Aggregators**
 - as users of the Metis suite to prepare aggregated collections for publication on the website, and as users of Metis itself to process and publish their data,
 - as operators of their MINT aggregation infrastructures and workflows for processing and publishing digital datasets,

- as operators of their own data infrastructures and workflows for processing and publishing digital records.
- **CHIs**
 - as providers of said collections,
 - as operators of their MINT aggregation infrastructures and workflows for processing and publishing digital records,
 - as operators of their own data infrastructures and workflows for processing and publishing digital records.
- **End-users** who visit the Europeana website and explore or re-use the digital collections offered there.

For each user group and for the integrated use in the different contexts (Metis suite, MINT, user's own infrastructure), we define possible scenarios that could translate into specific tool requirements. We envisage the use of the tool in the context of CHIs and aggregators preparing datasets for publication either using MINT (Scenario Strand A: Integrated with MINT), in the context of their existing infrastructures and workflows (Scenario Strand C: Independent use of the API), or in the core Europeana aggregation workflow when CHIs and aggregators specifically prepare the publication of their datasets via Europeana (Scenario Strand B: Integrated with the Europeana Infrastructure).

4.1 Scenario Strand A: Integrated with MINT

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
As an aggregator ² I want to check whether the metadata I intend to publish contains harmful wording. The metadata comes from different providers I represent as an aggregator. This means that I don't have control at the source, so I need a tool that allows me to do a quick check while working on the datasets in MINT. It is an extra service for the providers if I can warn them about harmful data.	As an aggregator I will process the datasets from the different providers and do the necessary conversions to EDM. This is done by using the MINT tool, which allows me to do an intermediate mapping to the aggregator uniform format (e.g. LIDO) and a conversion to EDM.	The aggregator can use the DE-BIAS tool as an integrated step in the existing workflow within MINT.
	As an aggregator I want to work with the DE-BIAS tool on the files uploaded by the content providers in MINT rather than having to do this step in another tool.	

² It is important to take on board that the aggregator is an in-between between the content provider and the final publishing platform. It is not the role of the aggregator to change the source data, but (possibly) to normalise the data by mapping the information to the defined metadata standard and to enrich the data for increased discoverability and interoperability.

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
	As an aggregator, I want to check these files for bias and get a report that gives me the details I need to take action. I can then share this report with the provider to avoid bias at the source of the publishing process.	The aggregator is able to flag biased wording downstream to the content provider via the report provided by the DE-BIAS tool and surfaced in MINT.
	As an aggregator I want to be able to enrich the EDM data (as an output of MINT) with the detected bias information if the provider asks me to do so.	The aggregator is able to flag biased content upstream to Europeana (by enriching MINT's EDM output with the detected bias information).
As a CHI, I want to be able to check my datasets for possible bias when processing them in MINT, so that I can take action to correct/flag the content. Bias is not only a concern for me when uploading to Europeana. It could be very useful to have this information to take corrective action in our own systems, which often publish directly to external audiences, on our own website or at a national aggregation level.	As a CHI, I first need to select and export the data from my own content management system, in a format that can be used for processing in MINT.	The CHI can use the DE-BIAS tool as an integrated step in the existing workflow within MINT.
	As a CHI, before sending my data upstream, I want to make use of the DE-BIAS tool to detect bias in the metadata within the existing workflows in MINT.	The CHI receives a report with statistical information on the bias status of their datasets so that they can take action at source if they wish.
	As a CHI, I want to use the statistical analysis of bias in my datasets from the tool's output to inform my institution of the extent of the problem so that a management/strategic decision can be made. It's possible that we may need to take action at the source level in our own databases, but perhaps we can limit our efforts to direct intervention based on the exported datasets.	The CHI has the opportunity to address biased wording in the exported datasets and to provide annotations to suggest replacements before sending their data upstream.

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
	For me as a CHI, direct intervention could mean adding markers/corrections at the level of the exported datasets that I would send upstream for publication.	The CHI has the ability to identify cases of bias in exported datasets and intervene before they are sent upstream.

4.2 Scenario Strand B: Integrated with Europeana's Core Service

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
As an aggregator not using MINT for the data transformation process, I need to be able to assess any potential issues that might arise from the use of derogatory terms or problematic language in the metadata of records in the collections I represent. Before submitting a dataset to Europeana, I would therefore like to check it for any potential bias so that I can flag this up to the CHI providing the data. If, on the basis of this feedback, the CHI wishes to proceed with the submission of the dataset to Europeana, I, as the aggregator, need to be able to request the inclusion of the detected bias information, which is carried out by the Europeana Foundation's Data	As an aggregator, I want to run an automatic detection check for biased wording when processing a dataset in the Metis Sandbox.	The aggregator can choose bias detection as an additional step in the data processing performed in the Metis Sandbox. ³
	As an aggregator, I want to be able to share the statistical overview of the detected bias from the Metis Sandbox interface downstream with the CHI providing the data, so that they can decide on the next step, and upwards with the Europeana Foundation DPS team, should the CHI decide that they wish to proceed.	The aggregator receives a summary report of the bias check via the Metis Sandbox interface. This allows them to 1) indicate to the CHI (collection provider) that the dataset may benefit from a review; 2) confirm that no issues were found and no further follow-up is required.

³ It should be noted that the Metis Sandbox can currently only process datasets of up to 1,000 records. For datasets with more than 1,000 records, repeated processing via the Metis Sandbox would be necessary, e.g. using a different step size with each repetition to increase the total number of records from the dataset reviewed for bias eventually, though this method will always leave gaps. To reach complete coverage for a bigger dataset, this would need to be broken down in packages of 1,000 records each, which could then be processed separately via the Metis Sandbox.

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
Publishing Services (DPS) team.	As an aggregator, I want to be able to share record level information about detected bias from the Metis Sandbox interface downstream to the CHI providing the data, so that they can more easily take action.	Through the Metis Sandbox interface, the aggregator receives record-level details of the bias information detected, which can be used in further communication with the CHI, which provides the data for next steps.
As a CHI, I need to be able to assess any potential issues that may arise from the use of derogatory terms or problematic language in the metadata of records in my collections. Before submitting a dataset to Europeana, I would therefore want to check it for potential bias so that I can either take action at source or ask for the inclusion of the detected bias information during the enrichment process conducted by Europeana Foundation's Data Publishing Services (DPS) team, should I decide to proceed with publication.	As a CHI I want to run an automatic detection check for biased wording when processing a dataset in the Metis Sandbox.	The CHI can select bias detection as an additional step in the data processing performed in the Metis sandbox.
	As a CHI, I want to get an overview of the detected bias from the Metis Sandbox interface, which will allow me to decide how to proceed - either to ask the Europeana Foundation's DPS team to include information about detected bias during enrichment, or to confirm that no problems have been detected and that publication will not be hindered by biased language expressions.	The CHI receives an overview report for the bias check via the Metis Sandbox interface.
	As a CHI, I want to be able to see details of detected bias at record level within the Metis Sandbox interface, so that I can take action more easily.	The CHI receives record level details of detected bias information via the Metis Sandbox interface.
As a user of the Europeana website, I want to discover valuable cultural heritage collections in a safe environment where respectful language is paramount.	As a user of the Europeana website, I need to be made aware of any language issues that may arise while using the site, so that I can make an informed decision about whether or not to consult the record.	The user is alerted to any language-related issues that may be present and can choose to continue reading or explore another item.

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
	As a user of the Europeana website, I need to be able to learn more about the context of any language-related issues that are flagged, so that I can gain insight into the historical, ethical, societal, cultural, ... issues at their root.	The user can access the detected bias information via the UI of the Europeana website, based on the DE-BIAS thesaurus, which provides context on the harmful nature and origin of the bias in the item.
	I only want to be informed about flagged terms in the specific language I have chosen to view the record (i.e. I do not want to be "warned" about flagged terms in the English description when I only see the Spanish one provided by CHI).	Visitors of Europeana.eu benefit from an enhanced multilingual approach, in which only language-relevant cases of biased expressions are flagged and contextualised.
As Europeana, operator of the Common Data Space for Cultural Heritage, I want users to be able to enjoy digital collections without getting exposed to un-contextualized instances of harmful language expressions.	As Europeana, I would like to use the DE-BIAS tool via Metis to review large datasets from legacy CHI collections. This would allow me to 1) indicate to the aggregator and/or CHI that the dataset could benefit from a review; 2) confirm that no problems have been found and no further follow-up is required.	Europeana receives an overview report for the bias check via the Metis interface.
	As Europeana, I would like to use the DE-BIAS tool in Metis to check CHI collections in the pipeline for publication. This will allow me to 1) indicate to the aggregator and/or CHI (data partner) that the dataset may benefit from review; 2) consider the sample test successful without the need for follow-up action.	Europeana receives an overview report for the bias check via the Metis interface.

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
	As Europeana, I want to include information on bias detected during enrichment in the EDM, provided that the aggregator or CHI submitting the data has asked me to do so. This would apply equally to legacy data and newly submitted datasets.	Europeana can add detected bias information as part of the enrichment process in Metis, which annotates the EDM data with alternative term(s), an explanation and other contextual information based on the DE-BIAS vocabulary.

4.3 Scenario Strand C: Independent Use of the API

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
As an aggregator or CHI running its own data infrastructure, I want to use the DE-BIAS tool in conjunction with my own data processing workflows to check my collection datasets for biased data. Based on the results of the tool, I would like to be able to develop mitigation solutions and possibly enrich my data with the detected bias information in a second step.	As a user of the open API I need to check my collection data on bias, in a way that fits our internal workflows and facilitates corrective action. This includes the ability to work with our own data formats as much as possible.	The user can upload data in standardised formats such as XML or JSON, according to the data models used in their institution. As a result of using the tool, and based on the statistics it generates, the user can be confident that the vocabularies used and the way in which tagging is performed conform to accepted standards.
	As an open API user with our own IT systems and workflows, I need easy access to a tool that fits into my existing digital infrastructure and workflows.	The user can easily use the DE-BIAS tool as open source software, accessible via a freely available API, to which data sets can be uploaded in their own data formats.
	As a potential user of the open API, I would like to find documentation on Github or another freely available platform.	Potential users interested in the tool can find all relevant documentation (from technical aspects of installing the tool, to using the tool, to the actual code) on Github.

EPIC The user and their main tasks/scope of interest	USER STORY The user's tasks broken into actions	ACCEPTANCE CRITERIA "The user is able to..."
	<p>As a user of the open API, I would need to be sure that the online software does not keep copies of the uploaded data. This is because some of the collections we want to test may not be published in the end, as they may contain quite sensitive data (these are the collections we would use this tool for in the first place).</p>	<p>The user has the option to actively delete the submitted files from the service once processing is complete.</p>
	<p>As a user of the open API, I need an easy to use assessment tool that gives me good statistics on how biased my collections are, so that I can estimate the scale of the problem and how much resources we should/can spend on it.</p>	<p>The user receives statistics on the extent of biased language in the data set provided. Based on this feedback, the user can decide whether or not to take further action.</p>
	<p>As a user of the open API, I need a response that flags the detected bias information, at least in textual format, along with the record identifier, so that we can write a script to make the necessary corrections/additions/flags in our own database.</p>	<p>The user is able to act on the information obtained from the tool's analysis by self-mitigating and remediating instances of bias in their institution's database.</p>
	<p>As a user of the open API, I want alternative output formats for the detected bias information (e.g. predefined XML, JSON or CSV outputs) so that I can use what works best with our existing data formats and workflows.</p>	<p>The user can use the feedback from the tool to address cases of bias in their dataset with no disruption of or added efforts/skills/infrastructures to the ones already in place.</p>

5. Tool Outline Based on User Requirements and Available Functionalities

5.1 Core Elements and Functionalities

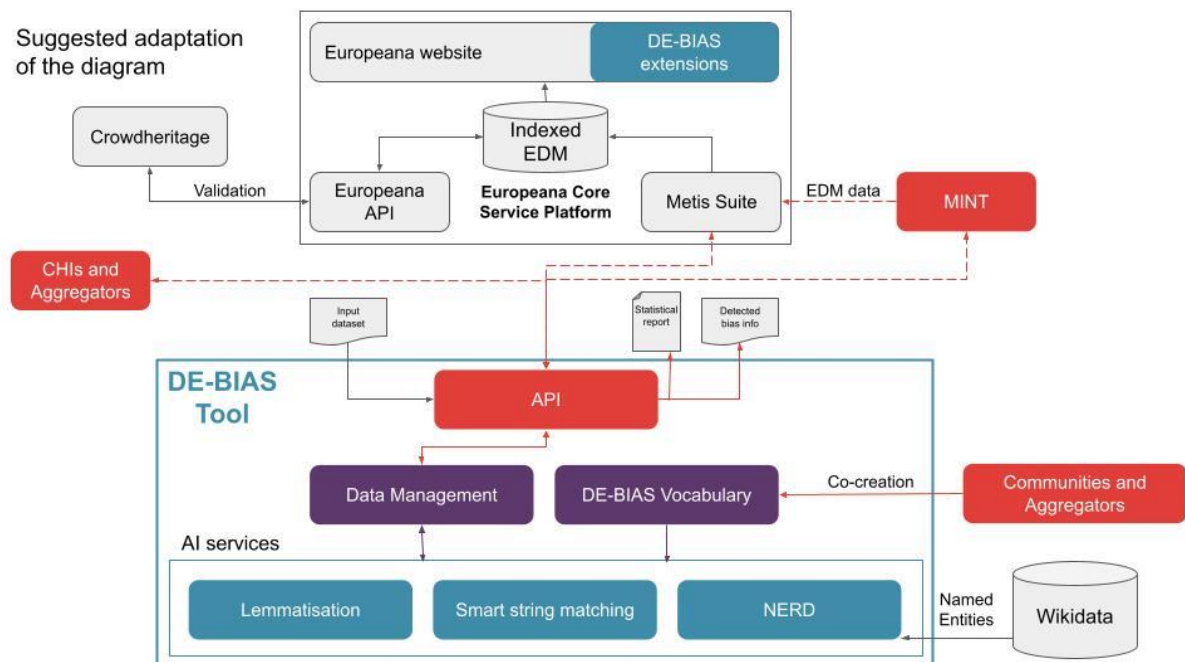
The DE-BIAS tool will consist of three parts, which will be described in more detail in the following sections:

- A suite of AI services (also referred to as the core functionality of the tool)
- A data management module to store and handle input and output metadata, as well as reports generated as a result of the bias detection process
- An API for connecting to external input and output platforms.

While the AI services and the data management module will operate in the same way in any user scenario, the datasets submitted as input via the API and the way the output from the API is used and integrated into further processing steps will depend on the existing and expected workflows in the specific application and implementation contexts.

For input, the tool will generally support XML, JSON and their archived forms (e.g. tar), although the specifics will again depend on the application and implementation context. As output, the tool will provide (1) statistical overviews of the biased terms identified in the datasets and (2) annotations to the metadata containing information about the detected biases.

The tool will be connected to the CrowdHeritage platform via the Europeana APIs to allow human validation of the results of automatic bias detection on metadata records. The software of the DE-BIAS tool will be open source and will be maintained as a freely available online service for at least two years after the end of the project.



5.2 AI Components

The core of the DE-BIAS tool consists of three internal AI components, each of which is in use in a different phase:

PHASE 1

In the first phase, lemmatisation is applied to the metadata records for the properties and languages considered by the project.

PHASE 2

In the second phase, smart string matching is applied to the DE-BIAS vocabulary in order to automatically match literals present in the descriptive metadata fields with vocabulary terms. The DE-BIAS vocabulary will be transformed into a machine-readable format using the SKOS format and will also be lemmatised to facilitate string matching.

PHASE 3

In the third phase, named entities will be extracted and disambiguated using Wikidata. These entities can then be used to revisit the biased matches identified in the second phase to exclude terms that occur as part of named entities. This last step can provide contextual insight into individual terms, as terms may not be considered biased if they occur as part of standard phrases.

These three phases will be served by the following AI components, using existing mature technologies:

Lemmatiser. Lemmatisers are an algorithmic technique for finding the lemma of a word that is a root. Existing lemmatisers for the supported languages (Dutch, English, French, German, Italian) are used, such as Stanford NLP. Lemmatisers assemble the inflected parts of a word so that they can be recognised as a single element, called the lemma of the word or its vocabulary form. In simple terms, they connect text with similar meanings to a single word.

Smart string matching. Based on the terms contained in the DE-BIAS vocabulary, smart string matching detects biased terms in cultural heritage metadata records. It uses the output of lemmatisation to produce accurate results.

Named Entity Recognition and Disambiguation. Named Entity Recognition and Disambiguation (NERD) modules assign a unique identity to entities (such as famous people, places or companies) mentioned in the text. Named Entity Recognition (NER) identifies the occurrence of a named entity in a given text, but does not identify which specific entity it is, while an additional step is applied to disambiguate the term in Wikidata. Existing NERD tools are used, such as spacy40, stanza41 (formerly called stanfordNLP).

6. Conclusion and Next Steps

With these initial user epics and tool requirements contributed, vetted, discussed and approved by the consortium - in particular the aggregation partners - we will review the tool components available to us through the technical partners and initiate a matching process to pair acceptance criteria with functionalities.

Using a multi-stage approach, we will first focus on functionalities that are already available or can be adapted to the specific use case of the DE-BIAS project. By examining the feasibility of the proposed scenarios in relation to the functionalities of the tool components, we can then investigate how to achieve the most favourable set of user epics expressed in the tables above, and assess the range of use cases that the DE-BIAS tool will ultimately be able to serve. The desired outcome of this exercise is to ensure that a well-functioning basic version of the tool (as specified in the Grant Agreement) can be developed and made available for testing and refinement.

As part of this process, a number of issues that emerged from our initial user story discussions will continue to be the focus of our investigations. These include 1) the exact scope and applicability of the report that the tool will produce, 2) an exploration of how flexible the tool can be and what input/output formats it supports, 3) the granularity of the ideal user experience (on a full dataset, record selection or individual record level?), and 4) the alignment of/differences in the functionality of the tool as integrated with MINT, with the Metis suite or with an aggregator's or a CHI's own data infrastructure.