



Revising Archival Descriptions

Simon P Wilson

Archives Consultant

www.simonpwilson.com

March 2024



This report relates to offensive terminology in archival finding aids. In describing the results of the research that has been undertaken it contains terms and language that might offend.

Table of Contents

Executive Summary	2
Acknowledgements	3
1. Background context	4
1.1 <i>Towards protocols for describing racially offensive language in UK public archives..</i>	4
1.2 <i>Revealing the Lived Experience of Black, Asian and Minority Ethnic people in Wales</i>	6
1.3 <i>Inclusive Terminology Glossary</i>	6
1.4 <i>Legacies of Catalogue Descriptions</i>	7
1.5 University of Leeds testbed project	7
1.6 Project methodology	9
2. Selected sectoral practice.....	11
2.1 The National Archives	11
2.2 Bath, Finding the Words project	12
2.3 Inclusive Cataloguing, The Keep.....	13
2.4 Borthwick Institute for Archives, University of York.....	14
2.5 Derbyshire Record Office	15
3. Project results	17
3.1 Preparing the data.....	17
3.2 Distinguishing between false hits and true hits	18
3.3 Results: Headlines	19
3.4 Results: Overview	20
3.5 Results: True terms	21
3.6 Results: False hits and true hits.....	23
4. Project engagement.....	26
4.1 Project Partners.....	26
4.2 Archives Wales	27
4.3 Axiell Customer Equitable Cataloguing Group	28
4.4 <i>History and Archives in Practice</i> conference.....	28
4.5 Engaging the wider archive profession	29
5. Recommendations	30
5.1 Collaborating with others.....	30
5.2 Welsh language	31
5.3 Paper-based finding aids	32
Conclusion	32

Executive Summary

This research was commissioned by Archives and Records Council Wales funded by the Welsh Government's *Anti-Racist Wales Action Plan*. The project adopted the methodology, using *AntConc* open-source linguistic software and Brotherton 'Full' terms list demonstrated by the University of Leeds Archive Testbed project.

Headline results

Thanks to support from the Archives Hub, **29,099 archive catalogues** from **12 archive services** totalling some **58.4m words** were processed. The *AntConc* software identified **195,000 hits** against the list of offensive terms which were then manually reviewed to see each term in context. A total number of **5425 instances** of terms from **71 distinct words** were identified as being potentially offensive, that is **0.009%** of the complete data set.

The results show the methodology works in taking a large quantity of data and identifying specific instances of terms which some people are likely to find offensive.

Recommendations

Four recommendations have been made:

- **Stakeholders co-curation**; including academic researchers and members from a range of community groups, LGBTQ, people with disabilities, mental health groups
- It is not just the archive sector that is investigating this issue, so **collaboration with museum and other colleagues** will produce stronger results for all participants
- **Welsh language** – the research omitted terms that might be offensive but missed because they were in Welsh. Collaboration with other interested parties should develop a terms list in Welsh, the results of which can feed into multiple projects.
- Review the impact and benefits to be realised by **reducing the number of catalogues that exist in paper format ONLY** – so were excluded from this project (and cannot be discovered via an online catalogue search)

As an immediate outcome of the project we will look at updating our current use of disclaimer in the [named] catalogue based on project recommendations.
[Project partner, based in a University Archive service]

The project has demonstrated the value of collaboration and co-ordination with other archive services across Wales to review the findings from the research. Adopting similar approaches to address the issue of offensive terminology will benefit all service users.

Acknowledgements

Thanks to colleagues from the University of Leeds for undertaking the initial research and for sharing the methodology and their experiences.

My sincere thanks to the project partners who contributed their finding aids for this particular form of scrutiny: Glamorgan Archives (lead) and Anglesey Archives; Bangor University Archives and Special Collections; Cardiff University Archives and Special Collections; Carmarthenshire Archives; Conwy Archives; Glamorgan Archives; Gwent Archives; Gwynedd Archives; National Library of Wales; Powys Archives; Swansea University Richard Burton Archives and West Glamorgan Archives.

I would also like to extend special thanks to Carly Richardson and Ben Crabstick from the Archives Hub whose assistance allowed the scale of the research data to be extended quite significantly.

1. Background context

Archives and Records Council Wales (ARCW) was awarded funding from Welsh Government's Anti-Racist Wales Action Plan Fund. The *Revising Archival Descriptions* project sought to identify biased and offensive language in archival catalogues. By undertaking this work across multiple archive services it was hoped the research will foster discussion and recognition that collaboration serves both archive service and service users interests. In sharing the methodology and results it is hoped that other services will undertake their own review of the language within the finding aids.

Offensive terminology

The Archives and Records Association Inclusive Cataloguing Group is a special interest group formed by the ARA Diversity and Inclusion Allies to provide guidance on inclusive cataloguing. Research identified existing resources and toolkits and the group will continue to increase awareness by highlighting ongoing work to make archival catalogues more inclusive.¹

There are numerous initiatives, some described in more detail below, looking to understand the scale of offensive terminology in our catalogues and finding aids. Some projects like the OCLC project *Reimagine Descriptive Workflows*² have sought to better understand the harm caused by cultural institutions' metadata and collection description practices.

1.1 *Towards protocols for describing racially offensive language in UK public archives*

This article, written in 2019 by Alicia Chilcott, has made a significant contribution to the discussion of the issues and challenges facing the archives sector in the UK.³ It highlights work already in-progress in Australia and North America and the need to decolonise archival practice in terms of the language within our catalogues and the UK archival workforce.

There is no doubt that the language used, for example by one group to describe another, have reflected power dynamics and the mis-representation of marginalized communities. Proposed solutions might include engagement with the communities to create more representative descriptions and an approach towards 'supplementing rather than censoring' the historic language that has been used.

¹ *New approach to providing information on inclusive cataloguing from the Inclusive cataloguing group*, ARA Diversity and Inclusion Allies blog¹ by Victoria Cranna, London School of Hygiene & Tropical Medicine. See <https://www.archives.org.uk/news/new-approach-to-providing-information-on-inclusive-cataloguing-from-the-inclusive-cataloguing-group-ara-diversity-and-inclusion-allies>

² *Reimagine Descriptive Workflows: A Community informed Agenda for Reparative and Inclusive Descriptive Practice* (Apr 2022) featured input from Australia, Canada, New Zealand and the USA.

³ 'Towards protocols for describing racially offensive language in UK public archives' by Alicia Chilcott was published in *Archival Science* (2019), see <https://doi.org/10.1007/s10502-019-09314-y>

As part of the research the author looked into the use of ten words⁴ in the online catalogues of sixteen archival repositories. This highlighted a wide variation in practice from no distinction or recognition, the use of content warnings, the use of quote marks to associate the term with the original document through to the inclusion of further text to provide context. The variations and inconsistencies demonstrate the need for sector-wide discussion on this issue. In acknowledging some services have sought to provide additional guidance relating to language and terminology this is most frequently distinct from the collection(s) themselves.

A 'good, better, best practice model'

In concluding the article the author proposes a model, summarised below, as a starting point for further discussion across the sector:

Good

- Use quotation marks to distinguish terms taken from the records
- Avoid the use of substituting offensive terms with a modern equivalent
- Engage with communities to create appropriate subject indexing
- Feature a content warning on the catalogue homepage detailing why offensive terms appear within the record descriptions
- Apply practices for all future cataloguing activities

Better (in addition to the 'good' recommendations)

- Provide guidance to users about the likelihood of offensive language, how to research particular topics/themes, suggest terms to use in searches and highlight particular collections relevant to the topic/theme

Best (in addition to the 'better' recommendations)

- Engage with communities to support user-generated tagging of catalogue entries on descriptive practices
- Base practice on professional ethics including the principle of representation and radical empathy

Whilst the author's work was focusing on racial aspects, *Revising Archival Descriptions* has adopted a broader consideration of terms that might cause offense.

⁴ Without knowing which 10 words were used in the author's research it has not been possible to extend the findings to the review of catalogues undertaken within this project.

1.2 *Revealing the Lived Experience of Black, Asian and Minority Ethnic people in Wales*

This ARCW project proposes a methodology for enhancing engagement of Black, Asian and Minority Ethnic communities with their collections.⁵ The methodology seeks to both discover, or re-discover, materials that are relevant to the communities and to facilitate access to these resources. The proposed methodology advocates engaging and involving community-led activities to contribute to the curatorial knowledge and to make the collections more accessible.

There is a clear overlap between this research and the *Revising Archival Descriptions* project comes with the former calling for institutions to take “a critical stance against the reproduction of racist narratives and the proactive acknowledgement and amendment of such representation.”⁶ Providing each contributing partner with the results from the processing and review of their finding aids supports the need to identify instances where the language used within the archive catalogue might be serving as a barrier to some users from accessing and using the material.

The project features a list of 303 terms and keywords (including both individuals and organisations) 16 of which it identifies as being Contested, Derogatory or Discriminatory. Some of these terms do feature in the Brotherton Full list but others for example ‘Colour bar’ or ‘People of Colour’ are broader concepts; the terms ‘color’, ‘colored’ and ‘coloured’ all appear on the Brotherton list but ‘colour’ does not.

1.3 *Inclusive Terminology Glossary*

The *Inclusive Terminology Glossary*, seeks to provide guidance for heritage professionals on discriminatory and harmful language relating to race, ethnicity, gender, sexuality, religion and disability. It was created by Carissa Chew as part of an Equalities, Diversity and Inclusion Internship at the National Library of Scotland (September 2020-June 2021).

Creating the Cultural Heritage Terminology Network and publishing the Glossary⁷ online sought to make the work more visible whilst also facilitating contributions from other practitioners. The Glossary can support a range of activities including:

1. **Surveying the problem:** audit catalogues, collections and websites for discriminatory language via (manual or automated) key word searches.

⁵ See <https://archives.wales/archives-and-records-council-wales/arcw-projects/black-asian-and-minority-ethnic-communities-in-wales/>

⁶ *Survey Methodology* p4, see <https://archives.wales/wp-content/uploads/2022/06/Survey-Methodology.pdf>

⁷ The *Glossary* can be accessed via the Cultural Heritage Terminology Network website, see <https://culturalheritageterminology.co.uk/>

2. **Inclusive description:** selection of non-discriminatory language when writing description or creating resources, as well as the addition of language disclaimers and relevant historical context.
3. **Advisory work:** identification of discriminatory language, which might require the addition of advisory content and creating content advice for finding aids.
4. **Surfacing diversity:** locating materials relating to underrepresented histories, including the addition of information to improve its discoverability.
5. **Inclusive finding aids:** creating guidance on how to find materials relating to underrepresented histories, including recommended search terms.
6. **Challenging controlled vocabularies:** that Cultural heritage professionals use the information within the Glossary to reflect critically on their use of language and controlled vocabularies.
7. **Education:** equip cultural heritage professionals with the knowledge to respond to enquiries relating to terminology and provide advice on search terms to use to discover materials related to underrepresented histories.

1.4 *Legacies of Catalogue Descriptions*

This digital scholarship project, which ran between February 2020 and February 2023, saw the Sussex Humanities Lab, the British Library and Yale University Library use legacy catalogues to support multi-disciplinary research into curatorial practices.⁸ Using two related catalogues, one held by the British Museum the other the Lewis Walpole Library at Yale, they combined linguistic and archival research to investigate and uncover the curatorial voice in the catalogue. This was achieved through the use of a very detailed and highly technical methodology.

Whilst this project and Revising Archival Descriptions project initially appear to be very similar in nature – using linguistic software on archival catalogues – this is actually the end of any similarities.

1.5 *University of Leeds testbed project*

The potential use of *AntConc* to detect offensive words was demonstrated in a University of Leeds Archives testbed project⁹ awarded in April 2021. The project, saw the University Library collaborate with the Leeds Arts and Humanities Research Institute, to look at what they described as ‘problematic language’ and the barriers they represented to engagement.

⁸ See <https://cataloguelegacies.github.io/>

⁹ Archives Testbed grant programme (2019-2022) sought to encourage innovation by awarding grants upto £5000 to services to develop an idea, adopt an alternative approach or develop a new service. Ideas were aligned with 3 key themes of the *Archives Unlocked* policy: digital, resilience and impact. See <https://cdn.nationalarchives.gov.uk/documents/archives/archive-testbed-inspiration-guide.pdf>

Colleagues were also aware of the potential impact that editing the record to remove sensitive language could also be interpreted as “interfering with the integrity of the record and covering up the mistakes of the past.”¹⁰ Their research highlighted the importance on distinguishing between two voices:

Creator the original record that could feature important context, with the decision made to retain the language but to place it within quotation marks to differentiate it from fact or opinion

Curatorial could be subject to biases, the decision made to replace the language with an equally meaningful and accurate term

Transparency was a critical part of the process, so alongside making changes was the need to acknowledge, under Access and Use, that a change had been made to the record and that the legacy description had been preserved and could be accessed by contacting the service. Colleagues at Leeds identified a number of broad areas of offensive language that could potentially be used in legacy archive descriptions including terms that were racist, sexist and misogynist, ableist, homophobic, transphobic, or trans-exclusionary terms that reflected religious discrimination and other terms which are generally offensive.

In testing the methodology the approach was applied to a number of catalogues, with 1.4m tokens (words) processed, resulting in 31 ‘true hits’. This work has prompted the development of rights statements and a review of policies and practices relating to the language and the voices that are present in the collections.

***AntConc* software**

AntConc describes itself as being “a freeware, multiplatform tool for carrying out corpus linguistics research, introducing corpus methods, and doing data-driven language learning.”¹¹ Unless you are experienced in the world of corpus linguistics the potential relevance to archives is not immediately obvious, the potential becomes clearer when we put the functionality into an archives-specific context:

AntConc allows you to search a document (for example a finding aid) for matches against specific terms (for example a list of offensive terms).

¹⁰ Holly Smith, Collections Assistant, Leeds Archive of Vernacular Culture in the University of Leeds Special Collections describing her work in May 2021. See <https://leedsunilibrary.wordpress.com/2021/05/26/sensitive-language-in-archive-description/>

¹¹ *Help file v4.2.4* available from <https://www.laurenceanthony.net/software/antconc/>. It can run on devices running Windows (built on Windows 10), MacOS (built on Mac Catalina) and Linux (built on Linux Mint). It was developed in Python and uses SQLite as the underlying database.

If there were 100 words in the terms list we would effectively be undertaking 100 simultaneous searches of the text. The ability to include an asterisk to serve as a wildcard in your term list increases the breadth and scope of the tool even further.

AntConc allows you to see patterns and relationships between the words, however for our specific needs the key feature is placing the Key-Word-In-Context (KWIC) allowing you to see the words that precede and follow a particular word. We can thus automate the first step in the workflow (see below) to detect the potential existence of a word or phrase and see the specific context to determine whether in this instance the word should be reviewed and an alternative considered in its place.

1.6 Project methodology

Preliminary testing of the *AntConc* software in conjunction with the Brotherton terms list and sample data from the Archives Hub during the scoping phase proved successful. It was decided unnecessary to develop an alternative approach and whilst there was some duplication of terms within the Brotherton terms list refining this list would not change the results. Adopting an existing methodology allowed us to focus on recruiting as many contributing partners to the project as possible.

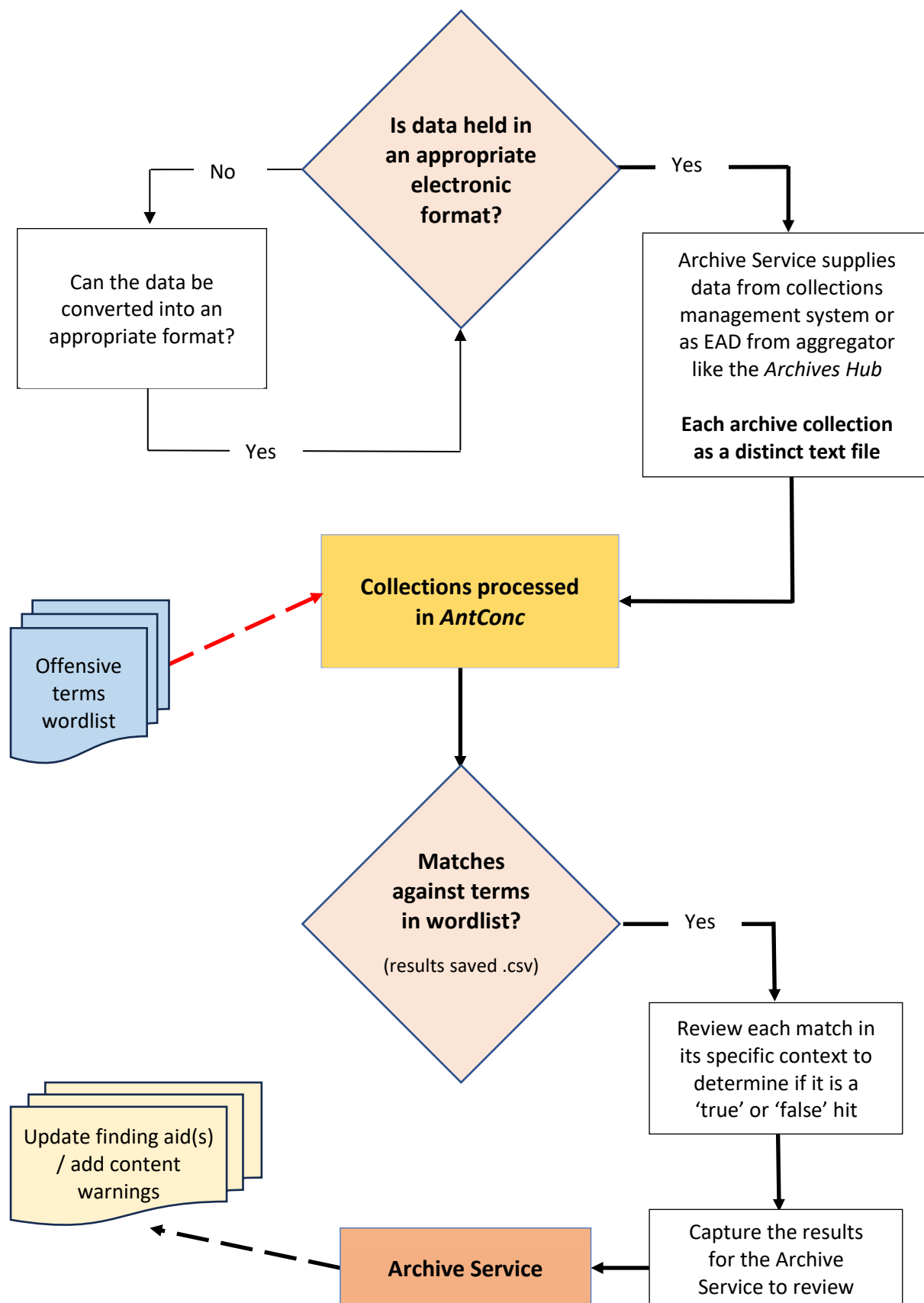
Source data

The majority of data was supplied as published finding aids in XML format prepared for (or downloaded from) the Archives Hub. In describing the methodology at Leeds there is a fair amount of effort spent on preparing the data before processing but the decision was taken when dealing with other service's data that no changes would be made to the data.

Data was successfully processed from a range of sources:

- EAD XML from the Archives Hub
- One service had exported all of its catalogues as a single 600MB XML file as part of a migration to another collections management system
- A handful of catalogues were provided in *MS Word* format, which cannot be read or processed by *AntConc*, but it was possible to include these files by opening the file in *MS Word* and saving it as a Word XML Document
- Data exported from *MODES* system in XML format
- Over 12,000 individual catalogues (1GB in size) as a single data source for processing

Workflow



See 3.1 for details about preparing the data, 3.2 for the review process and 3.3 for the results.

2. Selected sectoral practice

In the same way that the project methodology took inspiration from existing work, so the matter of what to do with the results can also be guided by current practice. In addition to Alicia Chilcott's article 'Towards protocols for describing racially offensive language in UK public archives' (see 1.1 above) other UK-based colleagues have looked to see what is being done to inform their own approach.¹² There is a list of predominantly North American but with some Australian library and archive services (but no UK sites are referenced) that have included a statement relating to bias in the catalogues.¹³

What follows is a consideration of how a small selection of UK archive services have approached the topic of offensive language in the catalogue entries including informing users in this work and the inclusion of content warnings in the finding aid.

2.1 The National Archives

The National Archives has a detailed statement about its view on offensive terminology.¹⁴ The statement clearly details the value in using information from the archive and that used by the record creators as being a key component of the archival record. One consequence of this is the likelihood of users coming across "*..outdated language and representations about matters such as race, sexuality, disability and gender. These can be derogatory, inappropriate, abusive, harmful, and offensive.*"

The statement continues that inclusion is not an endorsement of the language or the content and also highlights that these same terms help researchers identify relevant material. In describing its cataloguing activities it identifies a number of key aspects:

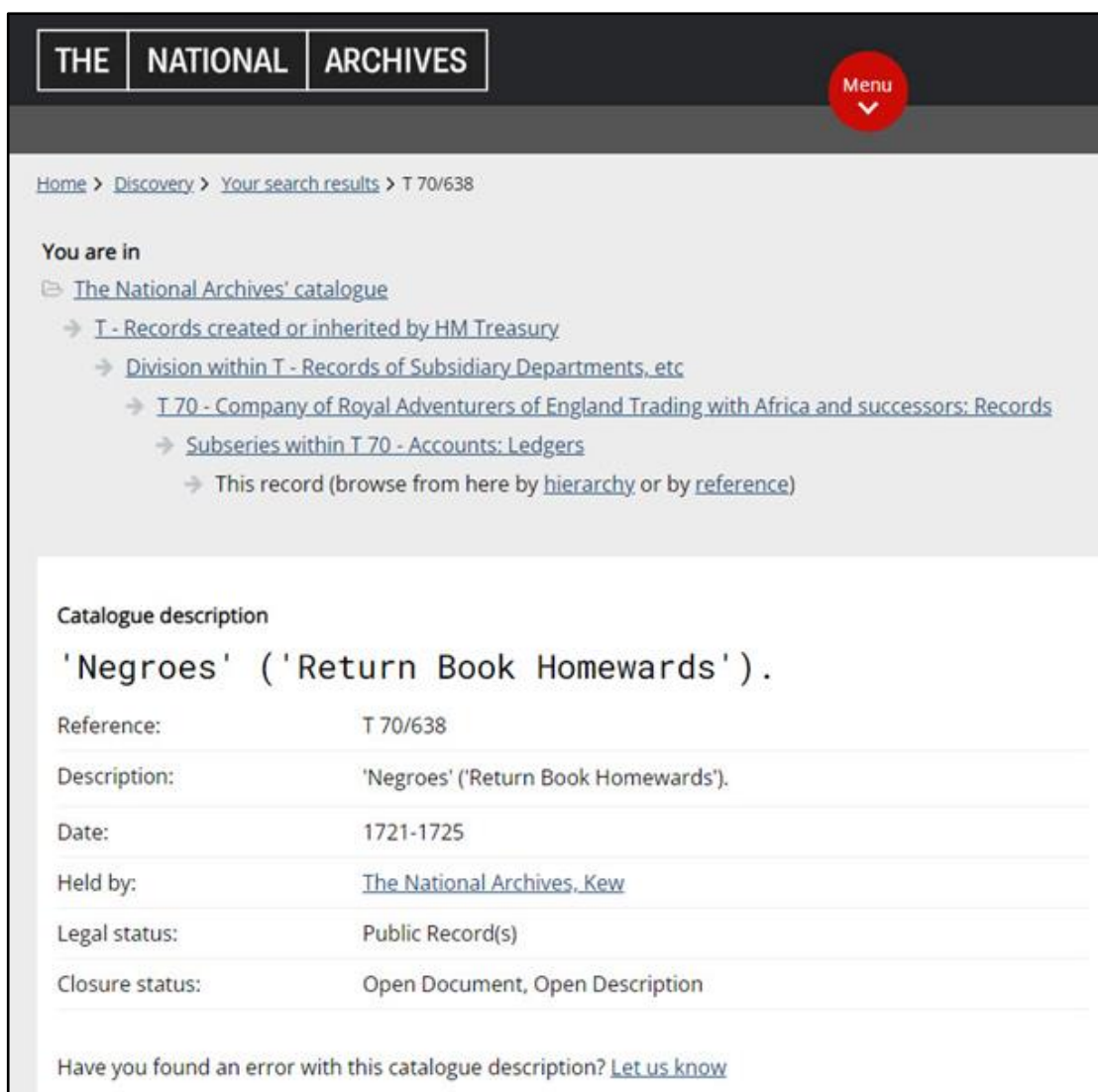
- Language used... should be accurate, inclusive and guided by objectivity
- When using original file titles... any offensive, pejorative or harmful terms are placed in quotation marks to indicate that it is a direct quotation from the record, not an endorsement of these terms
- Staff will use neutral and inclusive language to describe the record accurately. If it is deemed relevant and appropriate to keep offensive or pejorative terminology within a description, cataloguers will clearly indicate that it is a direct quotation from the record.
- Cautionary notes may be added to inform users of descriptions containing offensive or harmful language or representations.

¹² See Jessica Smith 'Rylands Reflects: Content Warnings for Collections and Catalogues' (Nov 2020) <https://rylandscollections.com/2020/11/05/rylands-reflects-content-warnings-for-collections-and-catalogues/> or Lindsay Loebig 'Understanding Harmful Language Statements' (May 2023) <https://aranewprofessionals.wordpress.com/2023/05/21/understanding-harmful-language-statements/>

¹³ See <https://cataloginglab.org/list-of-statements-on-bias-in-library-and-archives-description/>

¹⁴ See <https://cdn.nationalarchives.gov.uk/documents/cataloguing-approach-on-offensive-terminology.pdf>

Within the Frequently Asked Questions there is a reference to offensive or upsetting terminology and the ability to notify the service via its 'Found an error?' link which appears on every page (see screenshot below). However, unless a user has previously seen this note it is unlikely to occur to them that reporting upsetting terminology would fall within the remit of 'reporting an error'.



Screenshot showing a item on the Discovery catalogue including the error prompt

2.2 Bath, Finding the Words project

This project, also funded by The National Archives Testbed Fund, produced guidance for archivists and heritage professionals to consider the implications of derogatory language.¹⁵ The work undertaken in 2021 was initiated by a network of heritage organisations in Bath including colleagues working on The Bath & Colonialism Archive Project¹⁶ and Alicia Chilcott's article (see 1.1) and suggested 'good, better, best'

¹⁵ See <https://www.bath-preservation-trust.org.uk/wp-content/uploads/2022/06/Finding-the-Words-booklet.pdf> The guidance was produced by Lisa Kennedy an independent curator, historian and writer who advocates for the inclusion of wider perspectives within museums and history.

¹⁶ See <https://bathandcolonialism.org/>

approach. The guidance was produced by Lisa Kennedy an independent curator, historian and writer who advocates for the inclusion of wider perspectives within museums and the study of history.

The project took three case studies from the *Bath Chronicle* newspaper - two instances of 'negro' dating from 1761 and 1762 and one instance of 'Africa' from 1763. For each there is a detailed consideration of the term including why it was used and possible approaches to adopt to update the description without losing the context or meaning.

2.3 Inclusive Cataloguing, The Keep

Colleagues at The Keep have written an excellent account about their need to create a content warning for their collections¹⁷ which includes insight from their research into the practices of other services *"We found many that were hard to locate on websites or within catalogues, some were vague and noncommittal ('contains some language some users might find offensive'), some were broad and detached (we are sorry if you are offended but...')."*

The service produced a statement¹⁸ about inclusive cataloguing which acknowledges:

1. That [the service] holds material with direct links to enslavement and British imperialism. The catalogue does not always accurately reflect the reality of these relationships.
2. Our catalogue contains offensive language and outdated terminology that can be a barrier to access. This can be actively degrading, derogatory and harmful. It does not always accurately represent the people, events, and activities it is describing.
3. That there is room for improvement in the way [the service] acquire, catalogue, and make accessible archive collections. We are committed to making these improvements.

In addition to this recognition there is a 'Statement of Intent' detailing activities the Service is committed to undertaking. Those of direct relevance to the *Revising Archival Descriptions* project are:

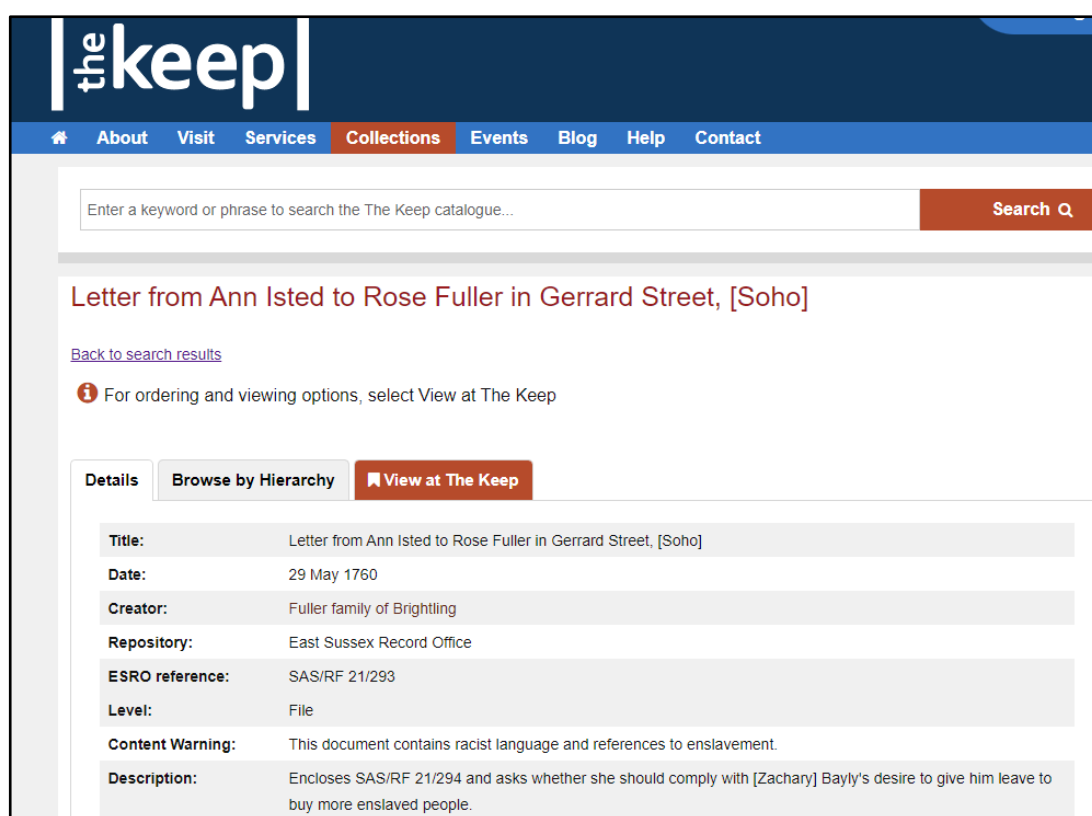
- Updating our cataloguing guidelines and procedures to make our cataloguing more inclusive going forward. Adding a disclaimer to The Keep website to make users aware that offensive language and terminology exists within our collections.
- Add content warning to individual records to alert users to the presence of potentially harmful or offensive language and/or imagery, where it has been identified.

¹⁷ See <https://collectionstrust.org.uk/blog/acknowledgement-and-accountability-creating-a-content-warning-for-the-keep/>

¹⁸ See <https://www.thekeep.info/inclusive-cataloguing/>

- Updating catalogue entries, where appropriate, to clearly indicate where the use of offensive language has been informed directly from the original record.
- Communicating with our professional networks around best practice and how we can do better.
- Communicating with our users and depositors around how we can become more inclusive and accessible.
- We aim to be transparent about this work [changing existing catalogue records] and encourage feedback from our users to help us improve. If you have a concern or query about anything you have seen in our catalogue or in our collections whilst visiting The Keep, please let us know.

As a result of its work content warnings have been added at item level (see below) with a distinct field carrying the following wording: “This document contains racist language and references to enslavement.”



Screenshot showing an item level page with a content warning

2.4 Borthwick Institute for Archives, University of York

The Borthwick website includes a statement about the harmful language which refers to its work “...a programme of ongoing, iterative review of its descriptive practices, revisiting and revising catalogue records where appropriate.”¹⁹ This appears on its collections page with a link on the catalogue search results page:

¹⁹ See <https://www.york.ac.uk/borthwick/holdings/harmful-language-statement/>

Borthwick Institute for Archives

file CFS/40 - Lease and release of manor and advowson of Brafferton

- CFS/40 - Lease and release of manor and advowson of Brafferton, 17 June 1794 - 18 June 1794
- CFS/41 - Chancery enrolment of CFS/40, 18 June 1794
- CFS/42 - Copy of articles of agreement concerning wood sales at Brafferton, 29 March 1774
- CFS/43 - Deeds etc relating to Brafferton rectory, 1682-1857
- CFS/44 - Deeds etc of lands at Pillmoor in possession or occupation of Sir William Payne Galloway, 1853-1895
- CFS/45 - Appletreewick Property, 1820 - c 1821
- CFS/46 - Agreement concerning the division of Pillmoor Common, 8 March 1824
- CFS/47 - Conveyance of land at Brafferton for station depot, 7 April 1896

Records of the Christian Faith Society > Lease and release of manor and advow...

Identity area

Reference code	CFS/40
Title	Lease and release of manor and advowson of Brafferton
Date(s)	17 June 1794 - 18 June 1794 (Creation)
Level of description	file
Extent and medium	1 file

Context area

Name of creator	Christian Faith Society (1691-)
Repository	Borthwick Institute for Archives, University of York

Content and structure area

Scope and content	Lease and release, mayor and commonalty of city of London to the Society for the Conversion of Negro Slaves in the West India Islands, manor and advowson of Brafferton
-------------------	---

Borthwick Institute for Archives
University of York, Heslington, York, YO10 5DD, UK
Tel: 01904 321166 | borthwick-institute@york.ac.uk

Accessibility | Harmful Language Statement
Legal statements | Catalogue feedback
University of York

A catalogue page with links to harmful language and feedback links encircled

Warnings have been added at item level, placed with the Scope and Content field:

Content and structure area

Scope and content

Registration document stating that the previously unregistered enslaved people and property of Mary Julia Butcher have now been registered. It contains details of six of her enslaved people such as their age, sex and ethnicity.

Please note that this collection contains distressing material and language that is offensive or harmful. Please see the Borthwick Institute's harmful language statement (<https://borthcat.york.ac.uk/harmful-language-statement>) for information as to why such language may appear and learn more about work underway at the Borthwick to remediate oppressive language. If you have concerns about language used in this record, please contact us at borthwick-institute@york.ac.uk.

Screenshot showing the warning within the Scope and content field

2.5 Derbyshire Record Office

Derbyshire Record Office has undertaken a review of language in its catalogues and identified 136 entries – most of which related to enslave people on plantations. There were also terms relating to people who are neurodivergent and a very small number relating to swearing in oral histories.

The work has also extended to review and update its approach to indexing and cataloguing. The service's Collection Manual now includes principles guiding the description – for example to including the names of enslaved people, not only the name of the enslaver, and indicating whether the name given is a birth name or a name given by the enslaver. The manual also specifies the terms which should not be used including a preferred alternative. This approach will also be extended to cataloguing records relating to disabilities.

Description of this work on its finding aids²⁰ includes the following statement:

Our catalogues often repeat terminology that was used in the original archive document, but which is now recognised as offensive and oppressive. We're working to improve these catalogue entries to reflect the changes in such language and beliefs. We believe it's important to retain the original language, as it reflects the attitudes of the time, but we'll make clear that this language is offensive and oppressive and add new descriptions that are respectful of all communities.

The content warning appears at the top of the entry (see below)

The screenshot shows the Derbyshire Record Office website. The header includes the logo and navigation links: Home, Search, and Contact Us. The main content area displays a finding aid entry for 'D239 - FitzHerbert family of Tissington - [16th-20th cent]'. Below the entry title, there is a 'Browse this collection' section with a link to view the full catalogue. A 'CONTENT WARNING' section is prominently displayed, stating: 'OFFENSIVE AND OPPRESSIVE LANGUAGE This catalogue and the archives it describes contain highly offensive and oppressive language. Some of the language has been included in the catalogue description to publicly reflect the attitudes of the time the records were written.' Below this, the 'Archive Reference / Library Class No.' is 'D239/M/E/16609'. The 'Title' is 'Letter from Malcom Laing to Frances Perrin' and the 'Date' is '3 May 1761'. The 'Description' section begins with 'Will dispose of the stock and furniture at Retrieve and rent out the land; proposes purchasing 20 young negro men [Black people] for Blue Mountain; a vessel with 18 puncheons from Grange Hill has been captured; also duplicate of D239/M/E/16607.' On the right side, there are 'Useful Links' and a 'Tweets from @DerbyshireDRO' section.

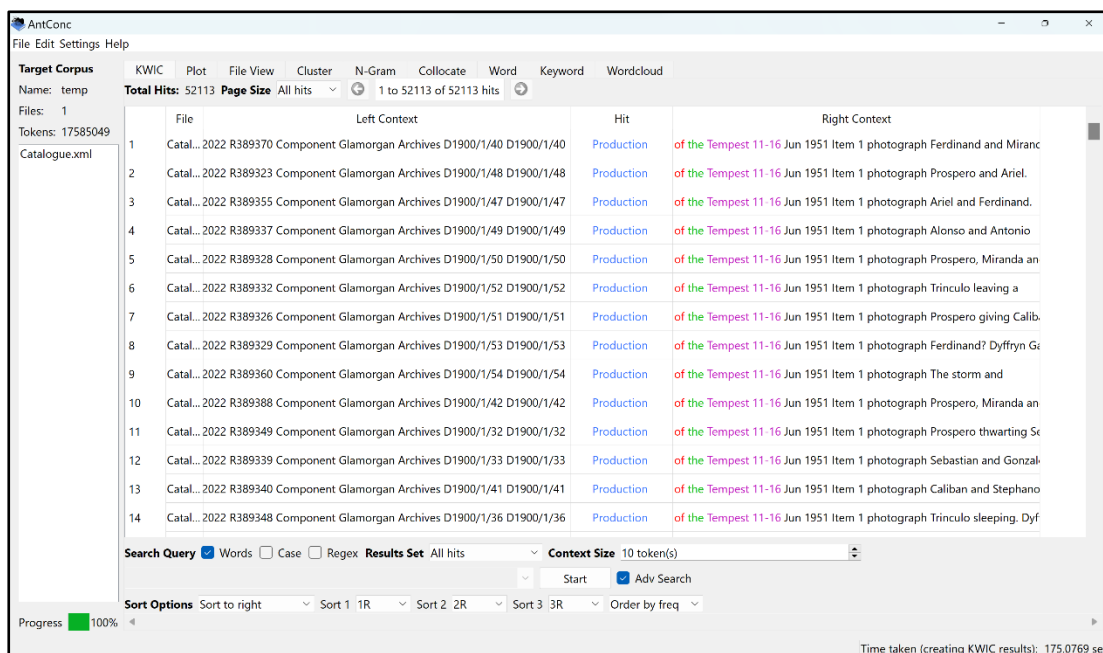
Screenshot showing the warning at the top of the entry

²⁰ See <https://www.derbyshire.gov.uk/leisure/record-office/records/finding-aids/improving-our-finding-aids.aspx>

3. Project results

3.1 Preparing the data

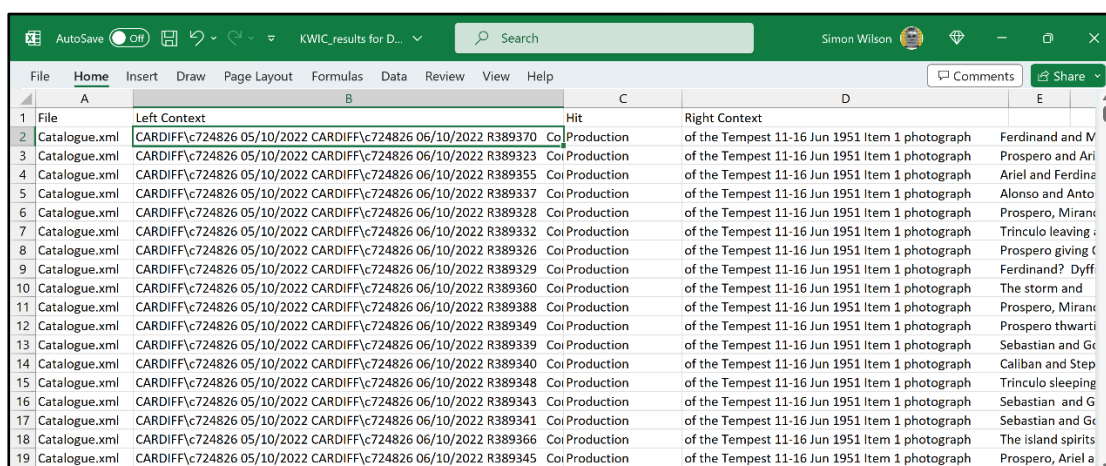
AntConc presents each hit against the terms list and shows the ten words that precede and follow the hit. After running the search against the terms list it is possible to export all of the hits as a csv file which allows us to view the file in *MS Excel*.



Screenshot showing *AntConc* displaying hits (in blue)

The results would be prepared for review by undertaking the following steps:

1. Opening the csv file in *MS Excel* and insert a new column, labelled 'TRUE' with any true terms marked with a Y in this column to allow the data to be filtered for all the 'true' hits
2. Use Ctrl + A to select all of the data and then use Data > Sort to sort the 'hit' column into alphabetical order



Screenshot showing the *AntConc* results in *MS Excel*

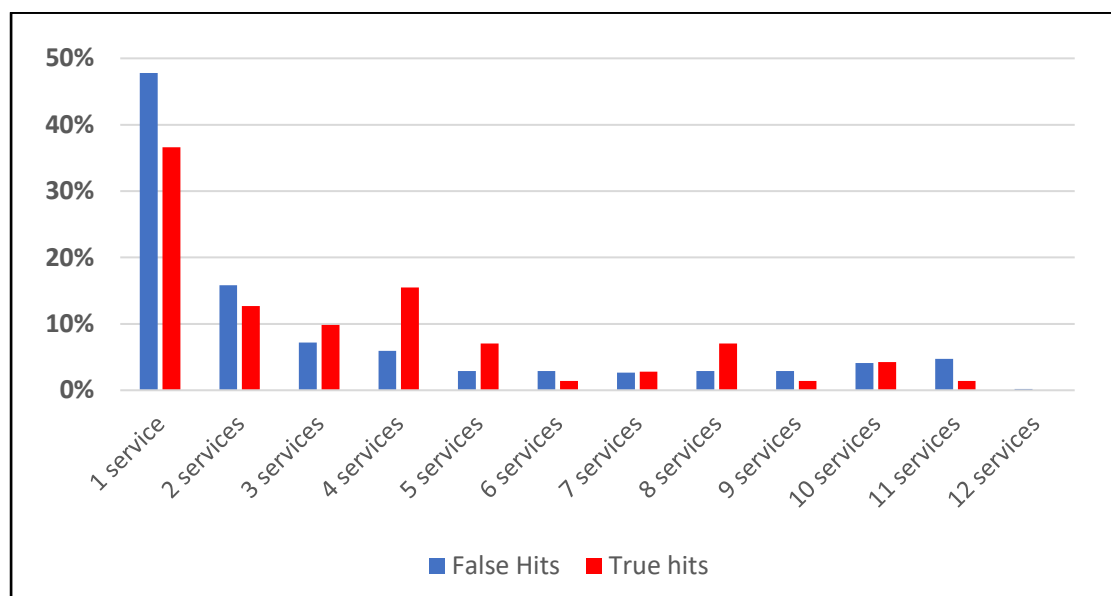
3.2 Distinguishing between false hits and true hits

In describing its methodology, colleagues at Leeds (see 1.5 above) highlighted the likelihood that search results would contain ‘false hits’ especially with adjectives where the same term can be both offensive and innocuous. In recognition of this it advises the need for the search results to be checked manually – referring to *AntConc* as ‘distant reading’ and this second ‘close reading’.²¹

Context is everything, which is why this manual review process considers the specific context in which the word has been used, and in doing so distinguish between:

- A. Terms which are not offensive and appear due to a common core with a term on the list; ‘association’ and ‘assignment’ both match the search for ‘ass*’
- B. Terms which appear in the list but appear in a different context; ‘Cissy’ and ‘Gay’ both feature on the terms list but also appear in finding aids as an individual’s forename
- C. Terms which were used historically, but are no-longer considered appropriate or acceptable; ‘cripple’, ‘dumb’, ‘lunatic’ and ‘mental’ often appear in institutional names
- D. Terms which are offensive, for example ‘coon’, ‘negro’, ‘midget’ and ‘slave’

Terms that fall within the first two scenarios (A & B above) are considered to be ‘false hits’, the latter two (C & D above) are marked as ‘true hits’.



Graph: Frequency of false and true terms by the number of archive services

Looking at the data (see graph above) we can see that a significant proportion of the terms - 47.8% of false terms and 36.6% of true terms - only appear in the results of 1 of the 12 archive services. At the other end of the scale 9% of the false terms appeared

²¹ ‘Corpus searching results’ (Sept 2021) by Dr Vic Clarke and Dr Kevin Matthew Jones, Leeds Arts and Humanities Research Institute included within the Brotherton testbed workshop resource pack.

in the results of 10 services compared with 5.6% of true terms. Three true terms 'lunatic', 'mental' and 'slaves' appear in the catalogues of 10 services with one term 'mentally' being found in 11 of the 12 services.

3.3 Results: Headlines

A significant aspect of the project has been the sheer quantity of data that has been processed from archive catalogues from multiple archive services.

58.4m	Total number of tokens (words) scanned by <i>AntConc</i>
195,000	Total number of hits identified by <i>AntConc</i>
29,099	Total number of catalogues reviewed
5425	Total number of instances of true terms
71	Number of distinct true terms
12	Number of participating archive services contributing data
2.8%	Percentage of hits that were true hits
0.009%	Percentage of tokens (words) that were true hits

Manually reviewing over 58 million words to find potentially offensive terms would not be considered, but the use of specialist software and the existence of a terms list has allowed us to do just that. With a degree of manual intervention the methodology has resulted in the emergence of a list of just over 70 words that should be considered for review.

Even without repeating the *AntConc* processing stage this list (see 3.5 below) can now be used as a starting point by any service looking to review the language in its catalogues and consider the potential for these terms to cause offense to individuals and communities.

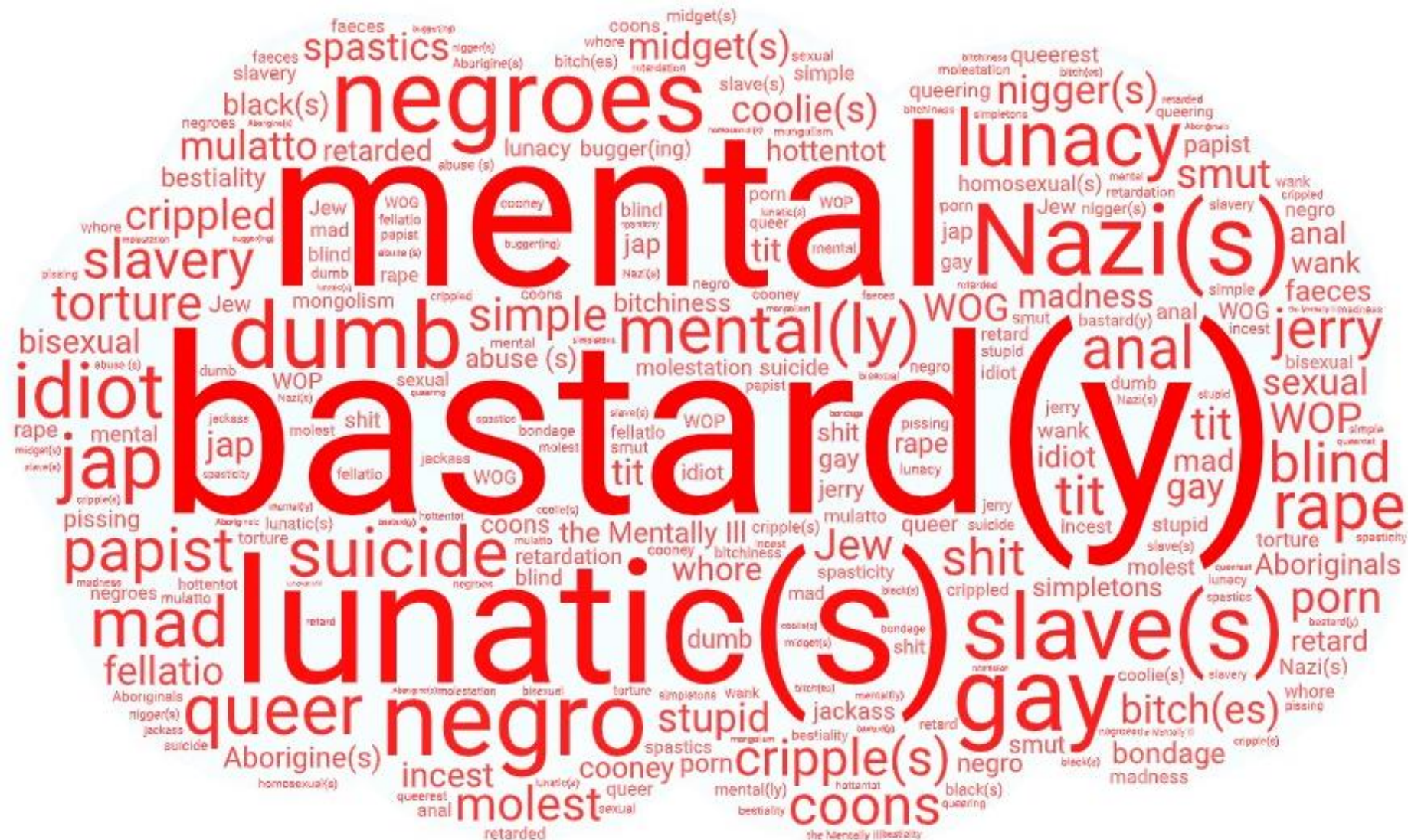
3.4 Results: Overview

Service	Number of Tokens	% of all tokens	False Hits	False Terms	True Hits	True Terms	True hits as % of all hits
Anglesey Archives	8,722,867	14.9%	24,176	332	2184	27	8.3%
Bangor University Archives and Special Collections	2,630,415	4.5%	8584	360	543	32	5.9%
Cardiff University Archives and Special Collections	846,678	1.4%	5078	237	37	12	0.7%
Carmarthenshire Archives	2,275,047	3.9%	7153	235	88	17	1.2%
Conwy Archives	2,904,625	5.0%	11,347	295	237	20	2.0%
Glamorgan Archives	17,585,049	30.1%	50,804	558	1105	29	2.1%
Gwent Archives	459,286	0.8%	2043	118	222	11	9.8%
Gwynedd Archives	956,120	1.6%	2817	186	44	9	1.5%
National Library of Wales	14,635,874	25.1%	41,446	618	486	44	1.2%
Powys	1,899	0.0%	2	2	0	0	0.0%
Swansea University Richard Burton Archives	1,267,560	2.2%	7635	206	182	18	2.3%
West Glamorgan	6,134,685	10.5%	28,558	368	297	25	1.0%
Total			58,420,105	189,643	1099	5425	71

Notes:

- 1 *AntConc* counts the number of tokens (words) in the dataset, most of the data was in the form of catalogues in EAD format which means this figure will include the XML encoding that provides the structure of the finding aid and the field labels.
- 2 The total number of terms is the number of distinct terms across the entire data set and not a sum of the terms for each archive service.

3.5 Results: True terms



Word cloud showing all of the true terms in proportion to the number of instances across the entire dataset

Term	Total number of instances of the term	Number of services featuring this term
bastard(y)	2097	8
lunatic(s)	730	10
mental	627	13
negroes	251	3
slave(s)	206	12
Nazi(s)	184	9
negro	184	8
dumb	175	7
mental(ly)	154	12
lunacy	147	9
gay	75	8
suicide	72	8
cripple(s)	69	8
queer	52	3
slavery	46	5
papist	42	5
idiot	20	4
nigger(s)	20	4
midget(s)	17	8
bitch(es)	14	4
mad	14	4
the Mentally Ill	12	5
coolie(s)	12	4
torture	11	5
molest	11	4
simple	11	3
spastics	11	3
coons	10	4
crippled	9	5
Aboriginals	9	4
madness	9	1
Aborigine(s)	8	2
mulatto	7	2
rape	6	4
jap	6	3
jerry	6	3
stupid	6	3
simpletons	5	4
Fellatio	5	1
bugger(ing)	4	4
anal	4	2
sexual	4	2
black(s)	4	2
bondage	4	2
WOP	4	2

Term	Total number of instances of the term	Number of services featuring this term
bisexual	4	1
smut	4	1
blind	3	3
homosexual(s)	3	2
shit	3	2
bestiality	3	1
Hottentot	3	1
Jew	3	1
porn	3	1
cooney	2	2
molestation	2	1
retarded	2	1
whore	2	1
WOG	2	1
bitchiness	1	1
faeces	1	1
incest	1	1
jackass	1	1
mongolism	1	1
pissing	1	1
queerest	1	1
queering	1	1
retard	1	1
retardation	1	1
spasticity	1	1
wank	1	1

Table: Number of instances and the number of services featuring a true term

3.6 Results: False hits and true hits

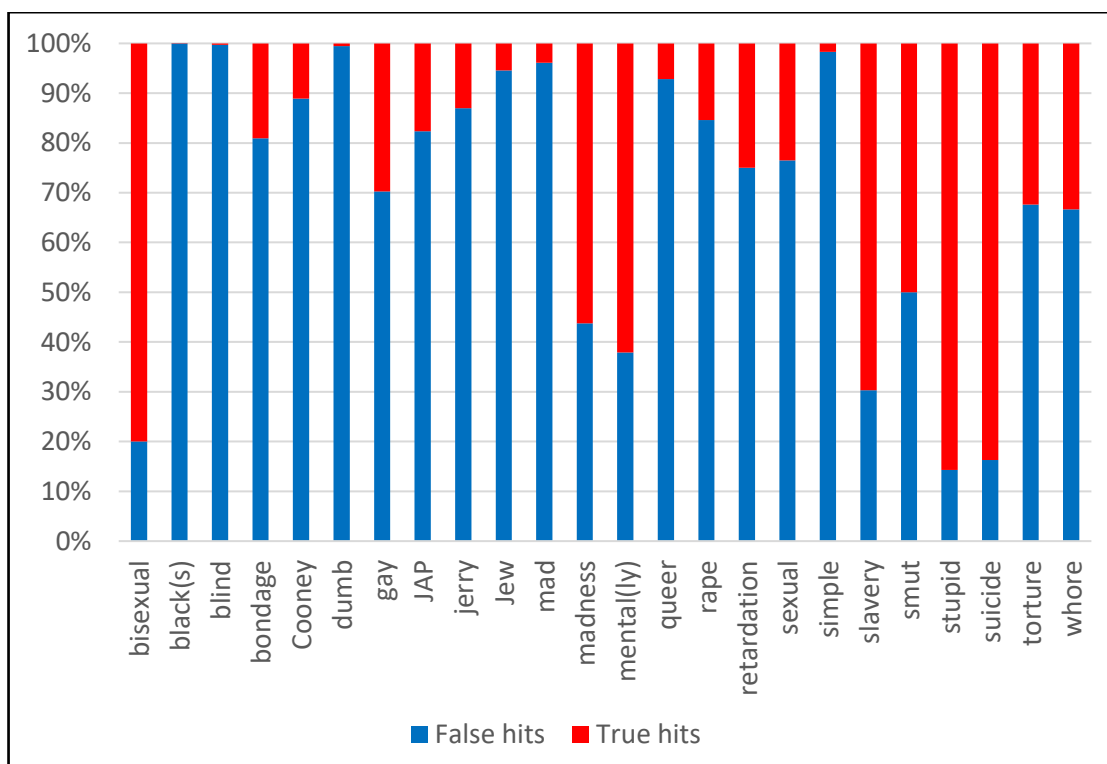
The manual review process (see 3.2) which considers the specific context in which the term appears means that it is possible for a word to be, at different times, a false hit and a true hit. In total 24 terms appear in the dataset as both a false and a true hit:

Term	Total number of instances of the term as a TRUE hit	Total number of instances of the term as a FALSE hit
bisexual	4	1
black(s)	4	14,902
blind	3	1020
bondage	4	17
Cooney	2	16
dumb	1	175
gay	75	177
JAP	6	28
jerry	6	40

Term	Total number of instances of the term as a TRUE hit	Total number of instances of the term as a FALSE hit
Jew	3	52
mad	14	344
madness	9	7
mental(ly)	154	94
queer	4	52
rape	6	33
retardation	1	3
sexual	4	13
simple	11	642
slavery	46	20
smut	4	4
stupid	6	1
suicide	72	14
torture	11	23
whore	2	4

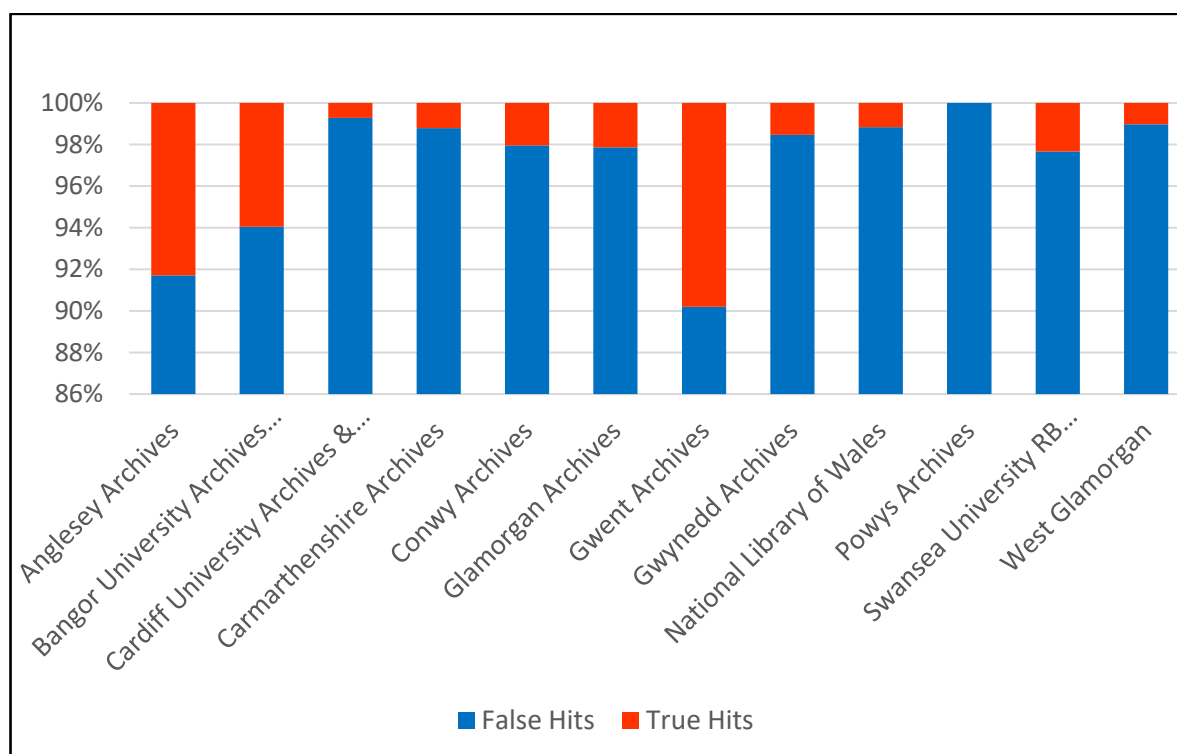
Table: All of the terms which appear as both true and false terms across the dataset

Without reviewing each instance or each term, some of the 'false' contexts include a novel entitled 'Come Whore, Come Conscience'; an individual named Field Marshal Smut and instances of 'Gay' and 'Jerry' as an individual's forename. Representing this data visually (see below) shows the variation between the words and how some are predominantly one of the other though not exclusively.



Graph: Terms which appear as both true and false terms across the dataset as a %

The table detailing each participating service (see 3.4 above) shows the variations in terms of the number of catalogues processed, the number of false hits and the number of true hits. Converting this data to % allows a degree of comparison, note the graph below shows the detail for 86-100% to see the variations between services. If the bar showed 1-100% the true hits would be barely discernible.



Graph: showing False hits and True hits as a percentage for each service

4. Project engagement

There was a strong desire to engage the wider profession and involve and engage as many other archivists and archive services as possible.

4.1 Project Partners

There were four partner meetings with colleagues from participating archive services during the project. Each meeting featured a progress update and then discussion about future activities including a review of the methodology and terminology lists; the collections to be featured in the project and content warnings. The final meeting previewed the results from the analysis and the report recommendations.

The results from the processing was shared with each project partner with the hope that it would encourage discussion about the words that had been found and possible next steps. Some of the services deliberately suggested for inclusion catalogues which they thought might be more likely to include offensive terms – for example collections relating to plantations worked by enslaved peoples.

A few of the responses from participating services are given below:

We now have the tools, and greater confidence, to revise the descriptions to ensure that they are more sensitive and inclusive. The first stage is highlighting the catalogues that need to be worked on, which this project has started.

[Project partner, based in a Local Government Archive service]

We were not surprised that the majority of our true hits related to mental illness and disability terms as we were aware of those in [named] collection... for the last couple of years we have discussed such terms as part of an archive teaching session

[Project partner, based in a University Archive service]

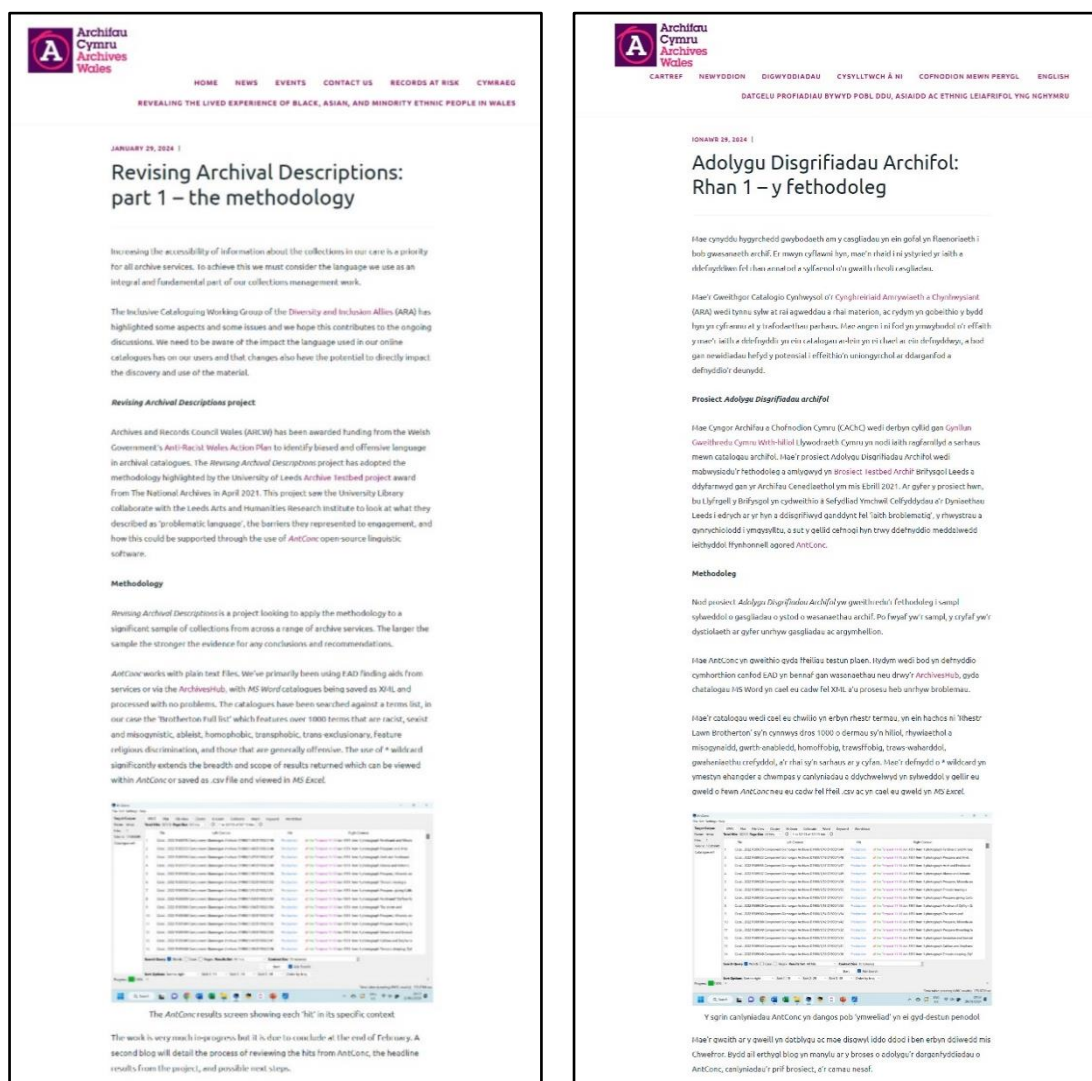
[we are] already reviewing catalogues following upload into a new collections management system and learning from this project will help staff to assess the terminology as part of the general reviewing work.

[Project partner, based in a Local Government Archive service]

One archive service was unsure whether to participate in the project because “we've checked our descriptions at [repository] and we can't find anything obvious we would need to flag up”. A selection of catalogues were chosen and analysed and the results shared with the archive service who replied “This is great and I've shared this will colleagues at Cardiff Uni. Fascinating to see this”.

4.2 Archives Wales

It was decided to write two blogs about the project on the Archives Wales website²²; the first describing the methodology and the second focussing on sharing the results.

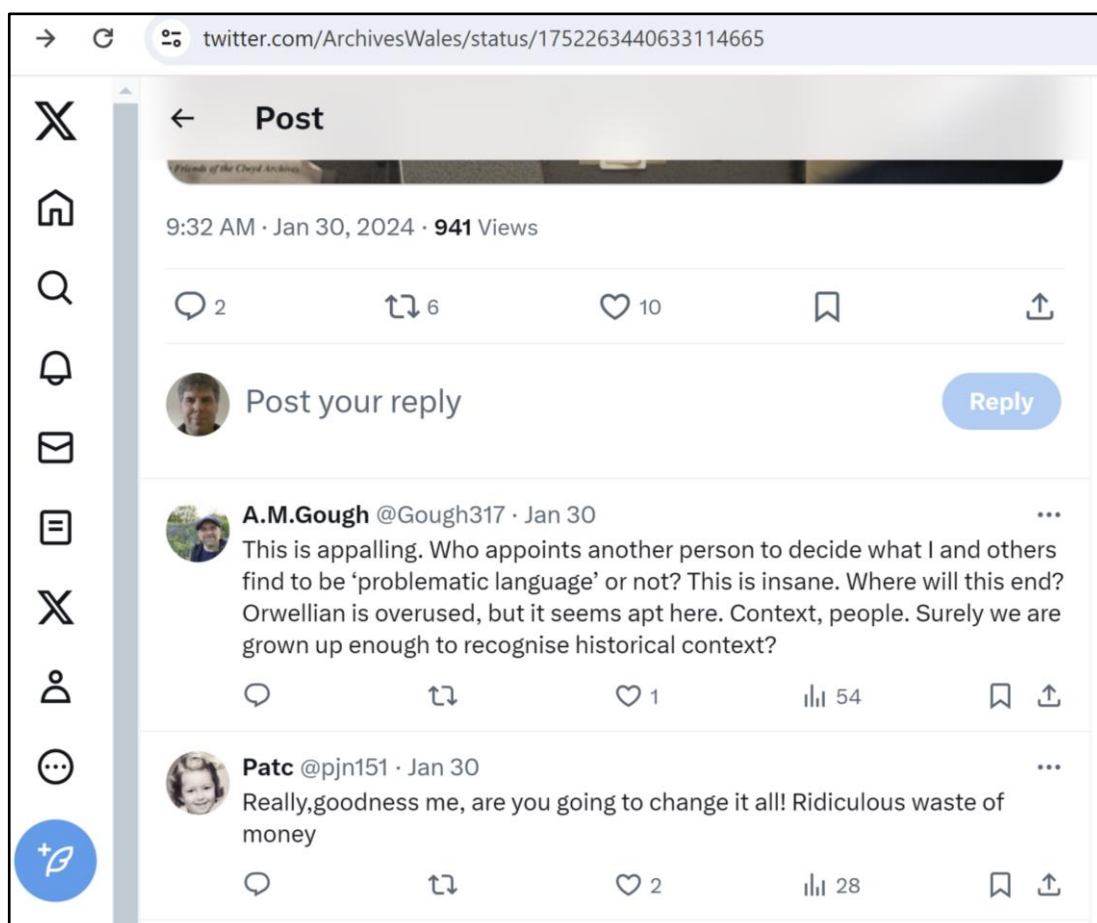


The first blog on the Archives Wales website in English (left) and Welsh (right)

Two comments were made in response to the first blog (see below), these misinterpret the research as determining which terms offend. The research was never intended to create or imply such a definitive list, the focus was always on the potential for a term to potentially cause offense.

What emerged during the review stage was the issue of professional archival bias. We are familiar with the use of terms in institutional names and in documents like bastardy bonds; but we must exclude this knowledge and focus on the potential for a term to cause offense. In the review stage the question was never 'Do I find this offensive?' but rather 'Might **somebody** find this offensive?'

²² See <https://archives.wales/>



Screenshot showing the two comments made in response to the first blog

4.3 Axiell Customer Equitable Cataloguing Group

This informal group is for users of Axiell’s collections management system to discuss equitable cataloguing work of all kinds - this might be decolonisation, repatriation, content warnings, terminology, co-production and collaborative interpretation, or related topics. The focus is on how the collections management systems can help, or hinder, this work, and to share documentation and systems work with each other.

Although many of the partner archive services were using *Axiell* software the use of *AntConc* is a vendor-neutral solution. Presenting the project methodology to the group at the meeting (26 January 2024) was an ideal opportunity to engage with colleagues about the topic. There was considerable interest in the methodology and the project results will be shared with the group at its next meeting (22 April 2024) in the hope it will encourage more services to review their collections in this way.

4.4 *History and Archives in Practice* conference

A proposal was submitted to the ‘History and Archives in Practice 2024 Conference’ as the project seemed to align very closely with the conference theme of legacies and was exploring the “intersection between history, archives, collections, and research and reflect on shared practice across and between disciplines”.

The proposal was accepted and a presentation about the methodology and results (so far) was given in Cardiff on 6 March 2024. The 20 minute session included an aspect of audience engagement by asking them whether the terms “slave”, “bastard” and “negro” were always offensive. The unscientific consensus of those present was generally that “it depends” with most feeling that the term “negro” was always offensive.

As archivists we bring with us a professional bias – we are familiar with the use of terms in institutional names and in documents like bastardy bonds; but we must exclude this knowledge and focus on the potential for a term to cause offense.

4.5 Engaging the wider archive profession

The timing of the project allowed the publication of a brief piece about the project that will appear in the May/June issue of the *ARC Magazine*. The article, along with a blog promoting the methodology and the results produced for the ARA Diversity Allies, are designed to highlight the project methodology and the headline results with a pointer to this report on the ARCW website.

5. Recommendations

The project sought to utilise the methodology highlighted by the University of Leeds and how it might support discussion across the Welsh archive sector. The intention was not to see **IF** offensive terminology was to be found amongst the finding aids but to identify terms that might offend a user and to share this with the archive service so that they could review the terms and consider possible next steps.

Thanks to project partners and the further assistance of colleagues at the Archives Hub the methodology was extended to an unprecedented number of archive catalogues (as the headline results report see 3.3 - over 29,000) containing over 58 million words. Whilst acknowledging that this includes the XML mark-up for each catalogue again it represents a significant quantity of material. To put this into another context this well exceeds the full printed edition of the *Encyclopaedia Britannica* which is believed to have contained about 40 million words.²³

As an immediate outcome of the project we will look at updating our current use of disclaimer in the [named] catalogue based on project recommendations.

[Project partner, based in a University Archive service]

Guidance and information from the project, including advice and templates on content warnings, will be shared and discussed with staff and added to our Collections Manual for use by current and future staff.

[Project partner, based in a Local Government Archive service]

At the outset of the research there was a belief that the type of collection might make a difference to the likelihood of it containing offensive terminology, but it quickly became clear that the effort to do this far exceeded the likely narrative the results would show. The following four recommendations are based on the methodology, the results and the wider reflections for the archive sector.

5.1 Collaborating with others

By undertaking research across multiple services it allowed the dialogue to focus on the findings and the results rather than a sense of justifying or explaining why a particular word appeared within the catalogue. Individual services were given the results from their finding aids, but in the partner group meetings the discussion was initially on the results and then quickly re-focussed onto possible next steps.

Some archive services are clearly addressing the issue of offensive terminology in their finding aids. Whilst each individual service needs to decide what they want to do, working together is not only more efficient as we are basically talking about the same

²³ See https://en.wikipedia.org/wiki/Encyclopaedia_Britannica

issue even if there is variation in specific words. Although collaboration will by necessity slow-down a response to this research it could also serve as a huge opportunity to demonstrate to the user community the importance given to this issue.

The focus for this report has been on the terms appearing within published finding aids produced by archive services. The archive sector is not alone in looking at this issue, the benefits of collaboration would be further magnified if colleagues across the heritage sector worked together.

1. It is not just the archive sector that is investigating this issue, so collaboration with museum and other colleagues will produce stronger results for all participants

A strong message that comes from research and current sectoral practice is the need to engage and involve communities in its very broadest sense – not just geographical but also racial, LGBTQ+, people with disabilities and mental health issues – groups to discuss the language that currently appears and to allow them to identify a preferred approach to the issue.

2. Co-curation with a range of stakeholders, including academic researchers and members from a range of community groups, LGBTQ, people with disabilities, mental health

5.2 Welsh language

AntConc uses Unicode allowing it to support all European and Asian languages. Only English terms appear within the Brotherton word list, so any offensive Welsh words amongst the finding aids would only appear in the results if the spelling was similar to its English equivalent.

3. Welsh language – the terms list was in English, collaboration with other interested parties to develop a terms list in Welsh would benefit and support multiple projects

The development of a Welsh list of offensive terms would support multiple initiatives. This would not need to be a translation of all 1093 terms into Welsh. Many of the terms from this list are contemporary slurs and most unlikely to appear in a finding aid. A combination of the Brotherton list and the results from this research could be used to create a heritage orientated list of terms in both English and Welsh. Consulting

with Welsh language experts and with colleagues from Welsh heritage organisations investigating offensive terminology would identify Welsh terms for inclusion.

In terms of implementation keeping the Welsh terms distinct, and not integrated into a single alphabetical list. This would allow the Welsh terms to be applied to the catalogues where the Brotherton list has already been applied.

5.3 Paper-based finding aids

Some archive services were unable to include some of their collections for consideration because *AntConc* can only process catalogues held in text format. Those held in PDF could not be processed nor could the considerable number of catalogues that currently only exist in paper form.

Participation in the project highlighted descriptions that need to be reviewed and provided us with the tools to undertake similar reviews in future. This is very important to us because the majority of our catalogues are in paper format and could not be analysed.

[Project partner, based in a Local Government Archive service]

4. Review the impact and benefits to be realised by reducing the number of catalogues that exist in paper format ONLY – so were excluded from this project (and cannot be discovered via an online catalogue search)

Project partners were asked to estimate the percentage of catalogues they held which only existed in paper format, with responses varying from 10% to something closer to 90%. Combining the conversion of these finding aids into a digital format with a review to identify offensive terminology will benefit the archive service and significantly enhance the discoverability and use of the collections by users.

Conclusion

The project has demonstrated the methodology works at scale and the value of collaboration with other archive services across Wales to review the research results. Adopting similar approaches to address the issue of offensive terminology will benefit all service users.

Findings could also be used as a talking point/ideas for future sessions re language use in our collections relating to mental health/disability/social care etc. ...the University's Library Leadership Team are discussing similar disclaimers for the library website and will recommend the project report.

[Project partner, based in a University Archive service]