

Text Anonymisation Helper Tool: User Guide



06/04/22

Copyright © 2022 University of Essex. Created by UK Data Archive, UK Data Service.

Version No. 02.00



Table of contents

Introduction	3
Copyright and disclaimer	3
Anonymisation.....	3
The Text Anonymisation Helper Tool.....	4
Precautions to consider when using this tool	4
Installation.....	5
Step 1. Show developer ribbon	5
Step 2. Install the Macros	7
Step 3. Install the Anonymisation Helper tab	9
Testing	10

Introduction

The Text Anonymisation Helper Tool can be used as an aid to identify and remove disclosive information from the qualitative data files you make available for re-use. Where informed consent for raw data sharing was not obtained, pseudonyms and/or vague descriptors should be used to replace any problematic identifying information.

To date it has only been used with Microsoft Word versions (2010, 2013 and 2016) and may not work with other versions.

Copyright and disclaimer

Copyright of the content of this macro and associated documentation files belongs to the University of Essex.

Re-distribution and re-use of the Text Anonymisation Helper Tool is permitted on the condition that the University of Essex is appropriately credited.

The Text Anonymisation Helper Tool has not been developed or tested to a commercial standard and therefore is made available strictly on the basis that you accept it as is, and that you are solely responsible for any use made of it. The University of Essex does not give any warranties, express or implied, including without limitation, as to the accuracy of the Text Anonymisation Helper Tool and disclaims any liability to you or any third party anywhere in the world for any injury, damage, direct or indirect loss, consequential or economic loss or any other loss suffered as a result of the use of or reliance upon the Text Anonymisation Helper Tool to the maximum extent permitted by law.

Anonymisation

When personal and/or sensitive data have been collected, information should not simply be removed from the text during the anonymisation process. This could result in loss of data accuracy or distortion of data. Instead, remaining within the context of the data, suitable replacements should be found to replace real names, place and company names or full addresses. Even when the respondent's name, company name and full address are removed, very detailed information on socio-demographics such as employment, educational institutions qualifications gained, occupations of other family members could compromise confidentiality. Care must be taken when anonymising to avoid removing too much information and making the data unusable. However, care must be taken when anonymising to avoid removing too much information and making the data unusable. An anonymisation log should be created to

keep track of all changes made during anonymisation. Given that the anonymisation log contain personal data, this should be destroyed alongside the raw data according to the information provided to participants in the Participant Information Sheet.

The Text Anonymisation Helper Tool

The Text Anonymisation Helper Tool makes use of Microsoft Word macros. Functions include:

1. Finding all numbers and words starting with capital letters.
2. Resetting highlighting of numbers and words starting with capital letters.

The Text Anonymisation Helper Tool does not anonymise data or make changes to data in any other way except formatting. It does not identify names, organisations, dates or place names as some text analytics tools do. The Text Anonymisation Helper Tool only highlights numbers and words which start with capital letters to make the anonymisation process easier. Words which start with capital letters will not only be the first words of sentences, but also any capitalised names, companies, addresses, educational institutions, countries and other identifiable information. In the same way, numbers can represent page numbers, counts, but also age, date of birth and address. All this information could potentially be disclosive and must be reviewed during the anonymisation process.

The Text Anonymisation Helper Tool does not change or remove any data. After all numbers and words with capital letters have been highlighted, it is the responsibility of the reader to read the document and make decisions on each highlighted word. Any information that can result in breaches of confidentiality must be anonymised. After checks and replacements have been made, all highlighting produced by the Text Anonymisation Helper Tool can be removed.

Precautions to consider when using this tool

It is important to note that Text Anonymisation Helper Tool does not search for names, places, dates or organisations. A set of dictionaries could be implemented to search for words that are in those dictionaries to identify them as names or dates. However, this is highly impractical for qualitative data anonymisation. While there are comprehensive dictionaries for English names, countries, UK, USA or international organisations, dictionaries in other languages are very limited, if they exist at all. Even if data are collected in English, respondents' answers might involve disclosive information in foreign languages. Also, even if names and places are recorded in the dictionary, they can be misspelled or purposefully entered wrongly to preserve original pronunciation. Information that does not match dictionaries correctly would not be picked up by such tools, but would still be potentially disclosive. Therefore, the Text

Anonymisation Helper Tool highlights all numbers and capitalised words to enable the reader to make judgements on preserving confidentiality as they read through.

As the Text Anonymisation Helper Tool highlights words starting with capital letters only, any non-capitalised names, places, companies or other potentially disclosive information will be missed.

Once disclosive information is found, it is important to search for the same information without capitalisation to find known disclosive words that were not capitalised. It is important to note that *Reset all Numms and Caps* function removes all highlighting from the document. Any text should be saved before using Text Anonymisation Helper Tool.

Previous file version should be used if any undesired changes are made when using the Text Anonymisation Helper Tool. When using the Text Anonymisation Helper Tool and making changes to the data, care should be taken to avoid mistakes which could occur if the automated search and replace function is used. For example, the name Jan should be anonymised. 'Replace all' option when using Microsoft Word search and replace function is not suitable as it can result in undesired changes in data. This can happen even when Match case option is chosen. If it was decided to change Jan to Mike to preserve anonymity, all instances of 'Jan' will be changed into Mike. This could introduce mistakes such as 'Mikeuary' instead of 'January'. If such mistakes occur, it does not only result in misspelled or unrecognisable words, but also the anonymisation could be reverted, especially if such mistakes happen more than once, for example, a word 'Mikeitor' is found instead of 'Janitor'. Also, the order of replacements matters if the automated option *Replace all* is used. In this case, if Jan was anonymised before Jane, all instances of Jane would be changed to Mikee. As an alternative, use the *Replace* option instead of *Replace all* to review each word individually before making changes.

Installation

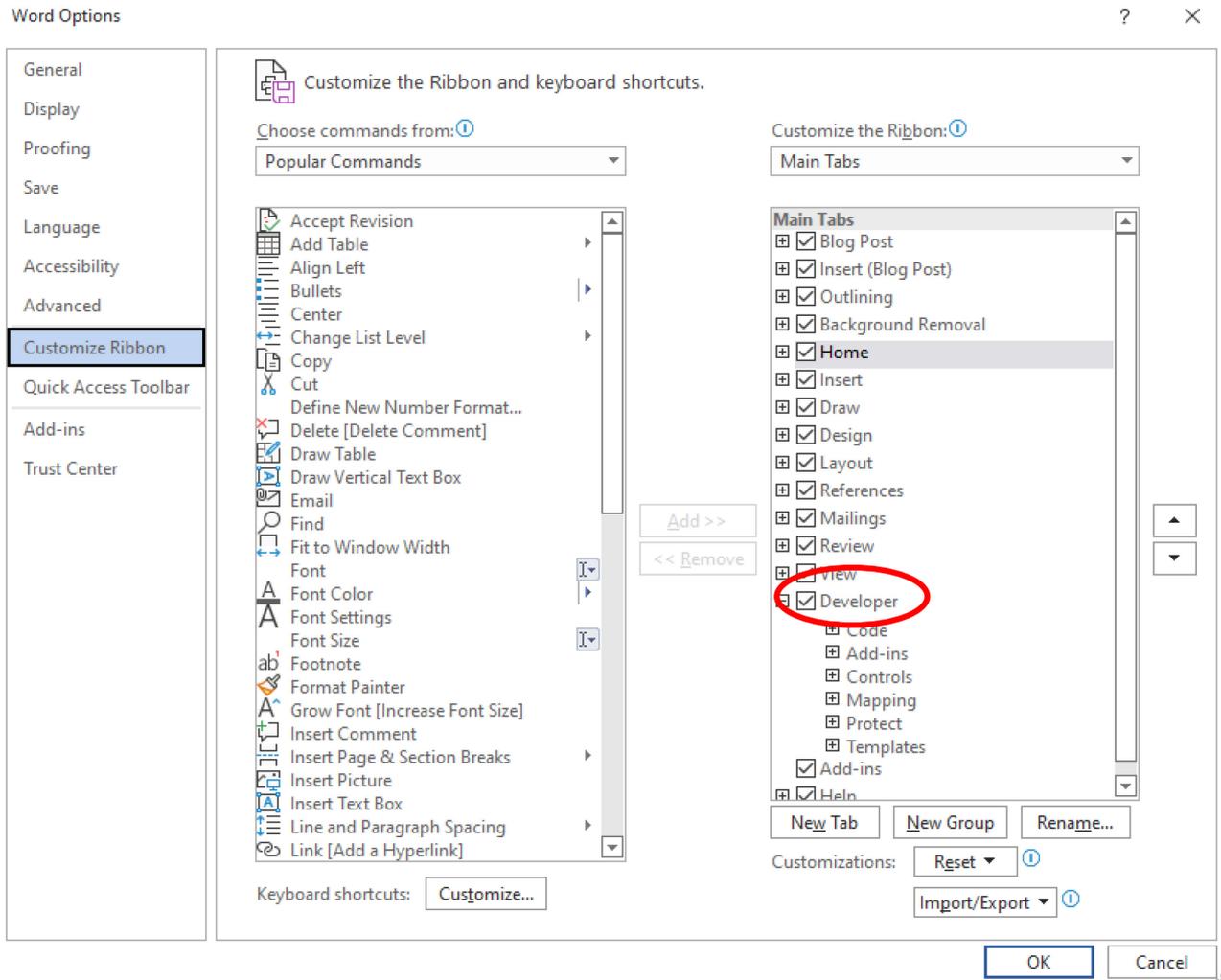
Step 1. Show developer ribbon

The Developer Ribbon is not shown by default. This step covers the instructions on how to make it visible.

When Microsoft Word editor is opened, click on File -> Options to open a Word Options window (Figure 1.1). On the Options window click on Customize Ribbon. Make sure the Developer Ribbon is ticked. When all done, click OK to save changes. There should be a new Developer Ribbon at the top of the MS Word window.

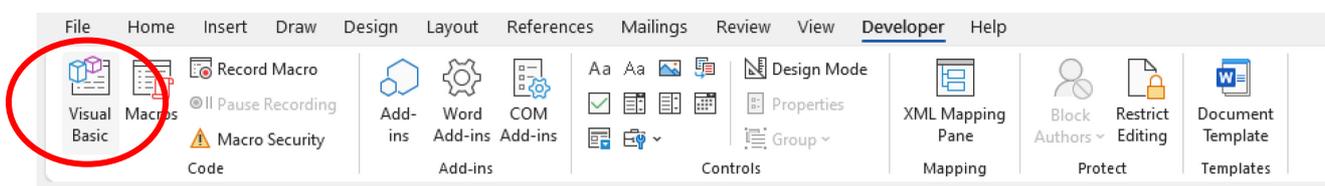
The image shows a software interface with a dark blue sidebar on the left and a white main content area on the right. The sidebar contains a navigation menu with the following items: Home (selected), New, Open, Info, Save, Save As, Print, Share, Export, Transform, Close, Account, Feedback, and Options (circled in red). The main content area displays a greeting 'Good afternoon' and a 'New' section with two options: 'Blank document' and 'Numbers coloring book'. Below this is a search bar and three tabs: 'Recent', 'Pinned', and 'Shared with Me'. The 'Recent' tab is active, showing a list of documents with a 'Name' header. The documents listed are: Collaborative Research Edits Documents, Anonymisation Documents, Plan to share Documents, depositing_data_in_the_curated_collection Documents, zendtoinstructions_depositor Documents (highlighted), Projects_SR Documents, Documentation File_SR Documents, Data File_SR Documents, Main Switchboard_SR Documents, and Data Collection Inventory_SR Documents.

When Microsoft Word editor is opened, click on File -> Options to open a Word Options window (Figure 1.1). On the Options window click on Customize Ribbon (Figure 1.2). Make sure the Developer Ribbon is ticked (Figure 1.2). When all done, click OK to save changes. There should be a new Developer Ribbon at the top of the MS Word window.

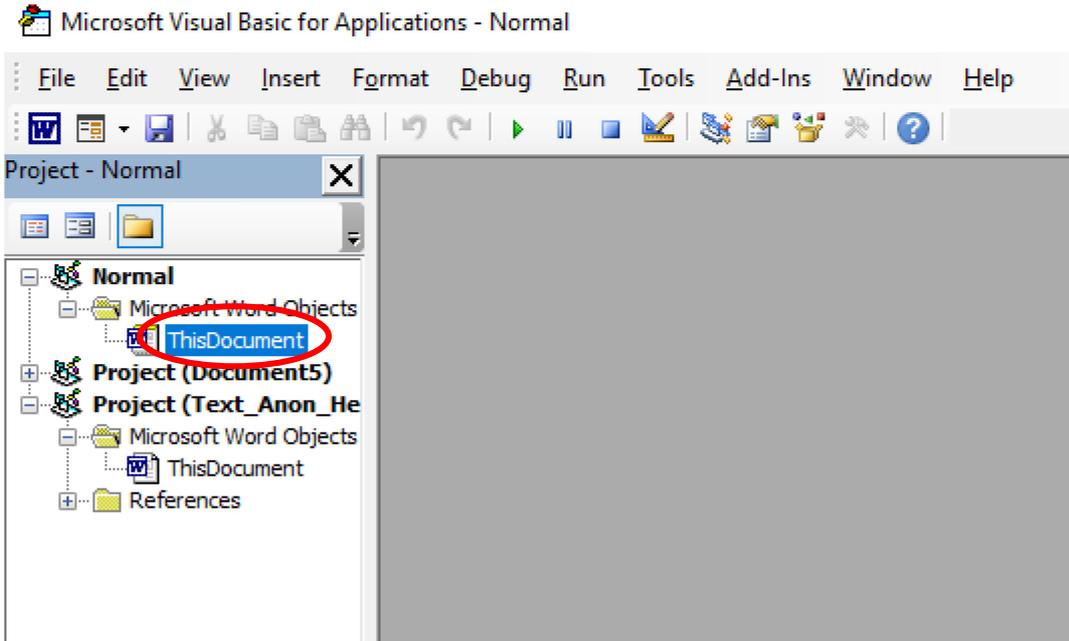


Step 2. Install the Macros

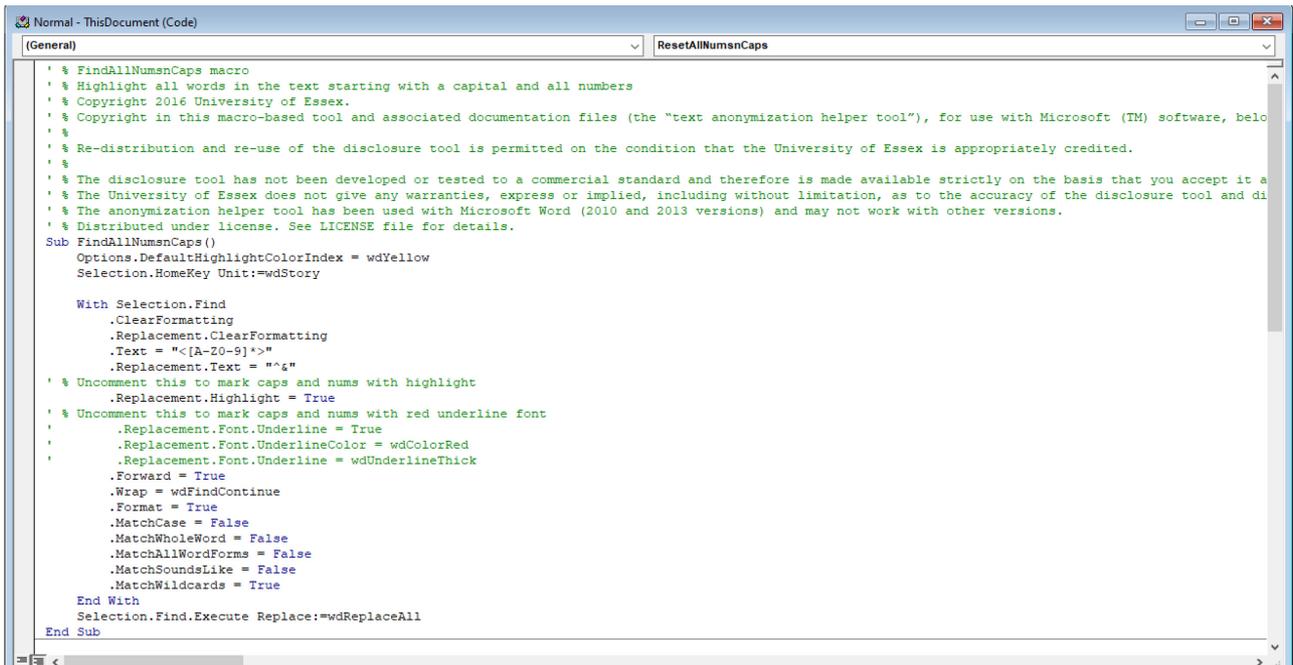
From the Developer ribbon, choose Visual Basic



This will open a new page. Ensure you are editing the 'Normal' document, as shown on the left-hand side under project. Double click 'ThisDocument'. This will open the text editor.



Copy the text from the AnonymisationHelperMacros.txt file into the macro text editor and close this window.

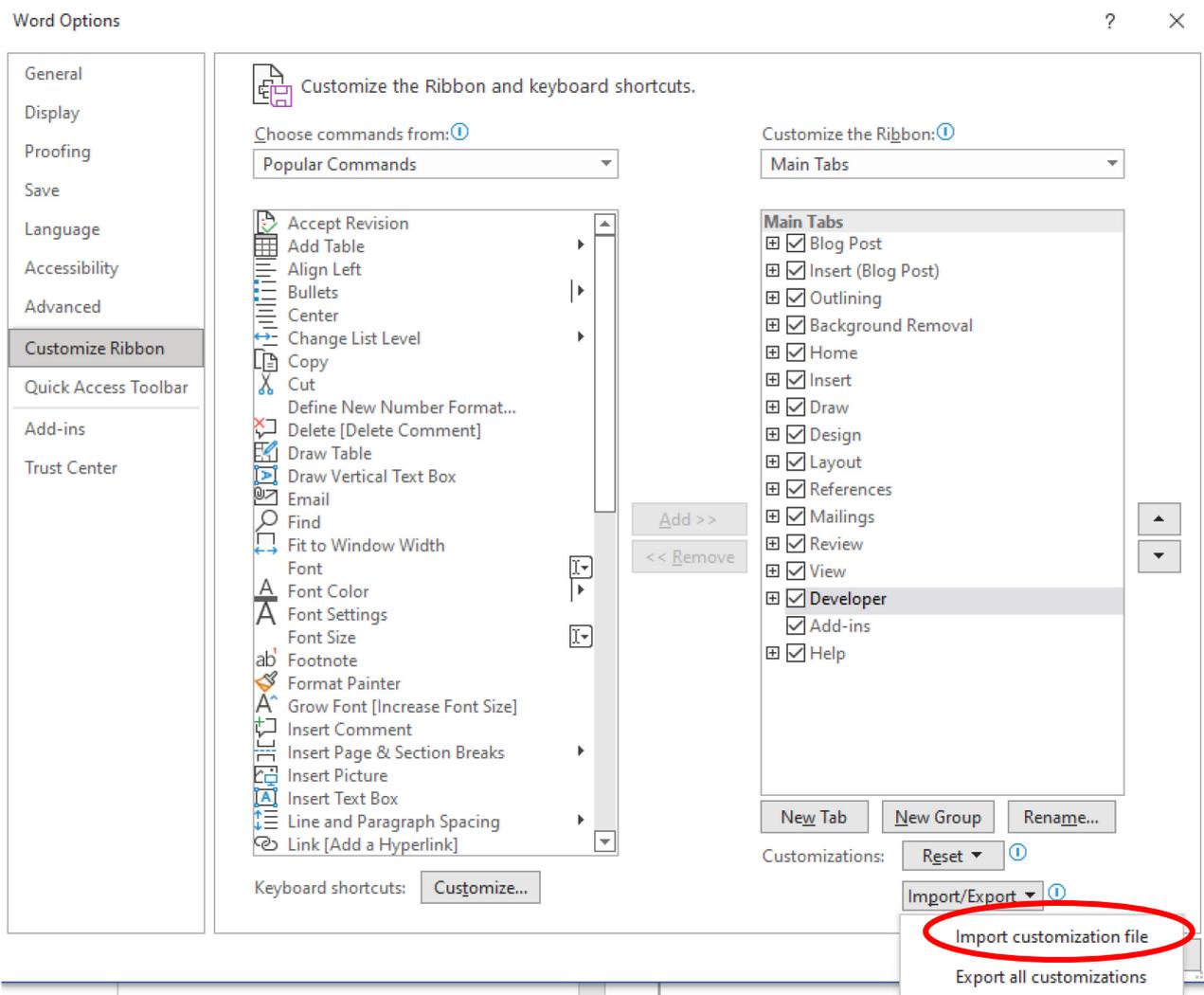


Click Ctrl+S to save the file and close Microsoft Visual Basic for Applications window. You may get an alert saying, "The following cannot be saved in a macro-free document", in which

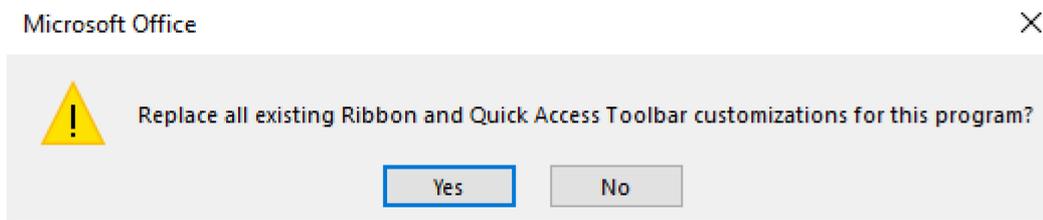
case, select 'No' and change the 'Save as type' to Word Macro-Enabled document and press save.

Step 3. Install the Anonymisation Helper tab

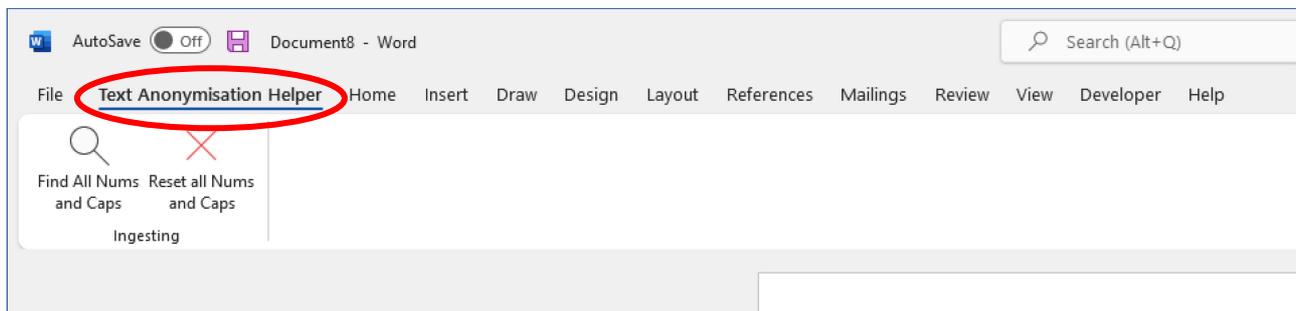
Go to File -> Options -> Customise Ribbon (as in Step 1). Under Import/Export (right bottom of the screen) select Import customisation file.



Find the AnonymisationHelperTab.xml file in your file directory. If you cannot find it in the folders, type the name of the file 'AnonymisationHelperTab.xml' and click open. You may get the message following message. Click yes.



A new tab should appear in your list of tabs in your Word Document. Click OK/Back to see the ribbon in your Word environment.



Testing

Once you have completed the installation steps above, copy and paste the contents of the TestData.txt file (from the src/test directory of the downloaded bundle) into a new Word document.

To use the Text Anonymisation Helper Tool, click '*Find All Nums and Caps*'. This will highlight numbers and words with capital letters in any text found in the word document, making it easier to search for addresses, names, or years. Anonymisation should be done carefully and only where needed to prevent unnecessary data loss. Precautions mentioned above should be taken when anonymising data.

File **Text Anonymisation Helper** Home Insert Draw Design Layout References Mailings Review View Developer Help

 Find All Numbs and Caps
 Reset all Numbs and Caps
 Ingesting

I: So the wartime Secretaries of State, Lord Lloyd, Moyne, Cranborne, Aby and then Oliver Stanley, their work was this work as you describe rather than looking at political and economic development?

R: Yes, it had to be. And of course the period when we had Lord Lloyd was a bit of a revelation. He just had to lift his finger and what he said went. There was no use having any argument: if you got into an argument with some other Department you said "Well, I'm very sorry but that's how Lord Lloyd wants it". The whole opposition just collapsed. It was wonderful.

I: Why did he have this tremendous impact?

R: I don't know. He always had this kind of forceful and ruthless personality.

I: He was also close to Churchill?

R: Yes. Ah, now of course one thing one's got to take into account during that wartime period is the American pressure for the bases in the West Indies. I think they drove us a very hard bargain over that, frankly, but I think it was inevitable in return for the vitally needed destroyers. They settled down with much less disruption than you might think and most of them were given up after the war - they were officially 99 year leases.

I: Did that lead to many disagreements?

R: You mean inside the Office?

To remove the highlighting, press 'Reset all Numbs and Caps' to revert the text back to the original.

www.ukdataservice.ac.uk

datasharing@ukdataservice.ac.uk

+44 (0) 1206 872143

We are supported by the Universities of Essex, Manchester, Southampton, Edinburgh, University College London and Jisc. We are funded by UKRI through the Economic and Social Research Council.