Ezy De-Identifier Tool

National university of singapore | Tahir Foundation Building, 12 Science Drive 2, Singapore

User Guide – Software Version V1R1

ephsmk@nus.edu.sg

2016

Contents

[Ezy De-Identifier Tool Software Requirements 2](#_Toc440014771)

[Prepare a new PC to use the Ezy De-Identifier Tool (cloud version) with from R Studio 3](#_Toc440014772)

[Prepare a new PC to use the Ezy De-Identifier Tool (downloaded version) with from R Studio 4](#_Toc440014773)

[Starting the Ezy De-Identifier Tool with from R Studio (downloaded version) 5](#_Toc440014774)

[Using the Ezy De-Identifier Tool for use with PortableEnv 7](#_Toc440014775)

[Protecting Data using the Ezy De-Identifier Tool 8](#_Toc440014776)

[NRIC Column Check 10](#_Toc440014777)

[Random Password 10](#_Toc440014778)

[Product Mapping File 10](#_Toc440014779)

[Produce Validation Report 11](#_Toc440014780)

[Shutting Down the Ezy De-Identifier Tool 12](#_Toc440014781)

[Ezy De-Identifier Tool End User License Agreement 13](#_Toc440014782)

# Ezy De-Identifier Tool Software Requirements

To use the Ezy De-Identifier Tool software requires the following minimum specifications

* PC running Microsoft Windows 7,8 or 10 1\*
* Intel Core 2 Duo or better 2\*
* 2GB System Memory 2\*
* Free hard disk space equivalent to twice the data being de-Identified
* One free USB Port if using the PortableEnv Feature
* Default web browser set to Microsoft Internet Explorer 11 or Google Chrome

Note 1\*: This software is not supported or tested on earlier versions, such as Windows XP.

Note 2\*: To improve performance when using larger data sets a faster processor and more memory is recommended.

Prepare a new PC to use the Ezy De-Identifier Tool with R Studio from GitHub

For first time use on a new PC

1. Install R and R Studio on your PC.

<https://cran.r-project.org/bin/windows/base/>

<https://www.rstudio.com/products/rstudio/download/>

1. Install Pandoc and Miktex  
   <https://github.com/jgm/pandoc/releases/download/1.15.2/pandoc-1.15.2-windows.msi> <http://mirrors.ctan.org/systems/win32/miktex/setup/basic-miktex-2.9.5823.exe>
2. Open RStudio and run the following command  
   source("https://raw.githubusercontent.com/dasasmk/EzyDeident/master/run\_from\_github.R")
3. The software will attempt to download all required R components needed to run the software.
4. A web browser should now open showing the application.

Note 1: Please use the close button at the end of the masking process to close the console window.  
Note 2: On first run of a new pc when selecting to create a masking report Miktex may prompt for installation of additional component. Please allow this. In this event it may be necessary to press the download button for a second time to ensure generation of the output zip file.

# Prepare a new PC to use the Ezy De-Identifier Tool (ZIP version) with from R Studio

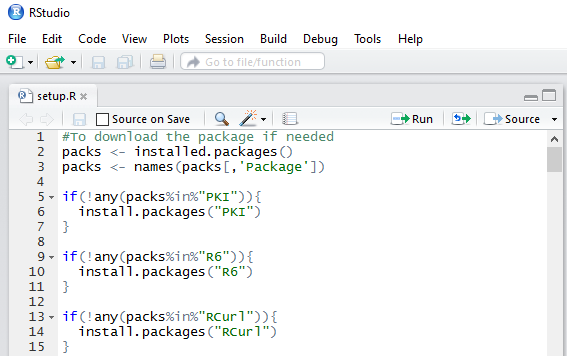
For first time use on a new PC

1. Download the latest version of the software from <https://github.com/dasasmk/EzyDeident/archive/master.zip>
2. Install R and R Studio on your PC.

<https://cran.r-project.org/bin/windows/base/>

<https://www.rstudio.com/products/rstudio/download/>

1. Start R Studio from the Start Menu.
2. Select File > Open File
3. Select ‘setup.R’
4. In the UI click the button ‘Source’



1. The software will attempt to download all required components needed to run the software. At the end two MSI files for Pandoc and MikTex will execute. In the event of a failed download these components can be manually installed using the following

Required Applications:

<https://github.com/jgm/pandoc/releases/download/1.15.2/pandoc-1.15.2-windows.msi>

<http://mirrors.ctan.org/systems/win32/miktex/setup/basic-miktex-2.9.5823.exe>

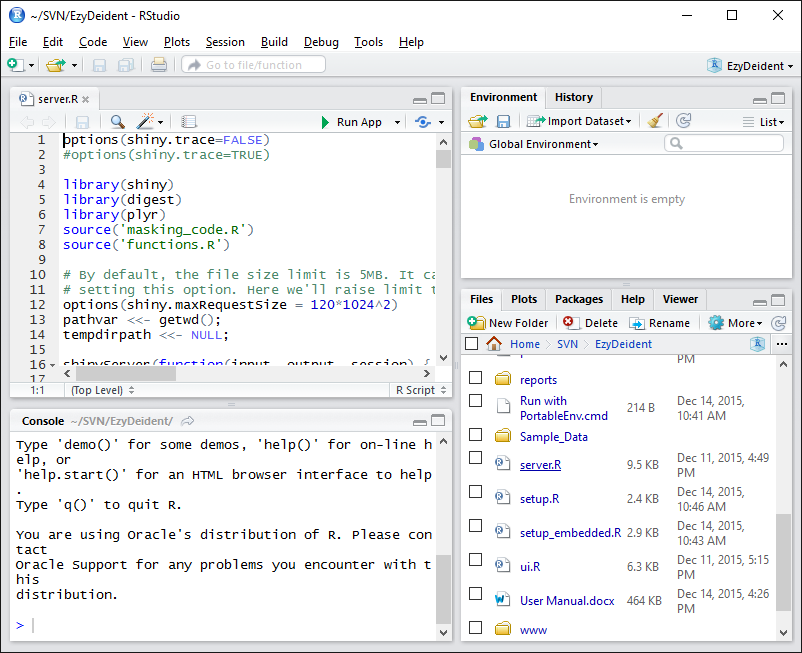
Required R Packages:

PKI,R6,RCurl,Rcpp,base64enc,bitops,caTools,digest,evaluate,formatR,highr,htmltools,httpuv,  
jsolite,knitr,magrittr,markdown,mime,plyr,rmarkdown,shiny,stringi,stringr,xtable,yaml,pander

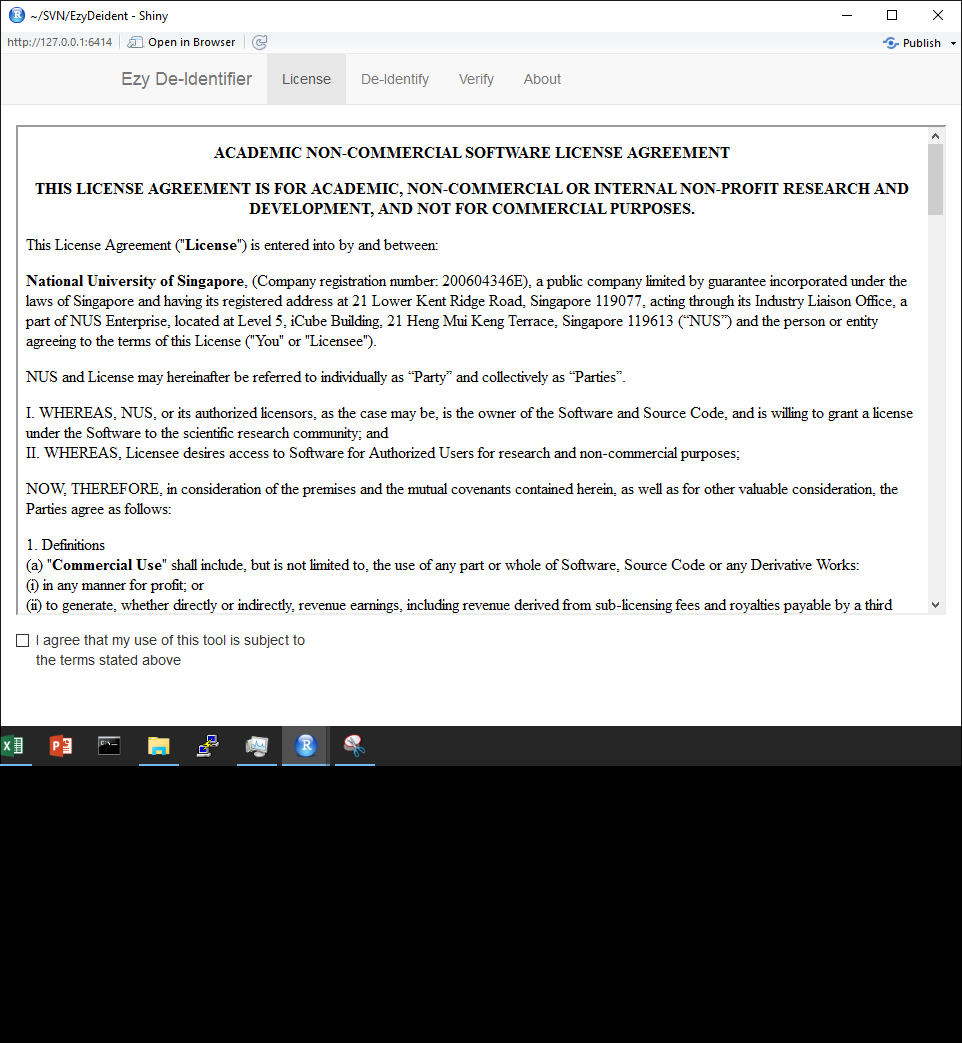
# Starting the Ezy De-Identifier Tool with from R Studio (ZIP version)

Once the system has been prepared

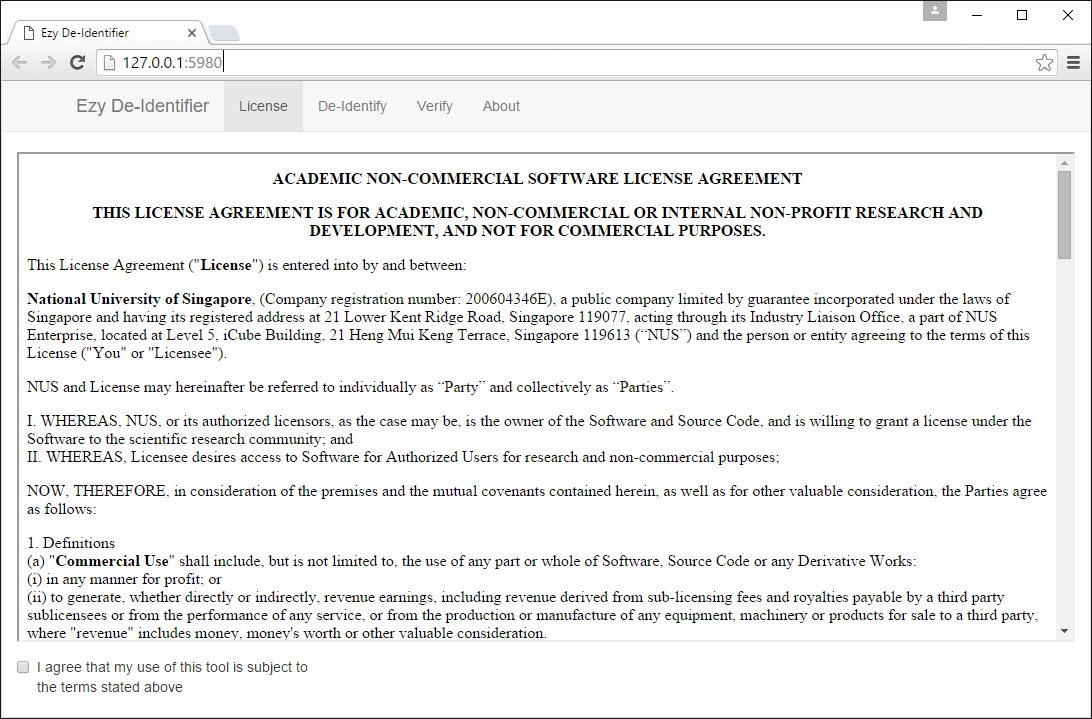
1. Start R Studio from the Start Menu.
2. Select File > Open Project
3. Navigate to the EzyDeident folder
4. Select to open MUI.Rproj
5. Click on Server.R in the bottom left of the R Studio Window



1. Click ‘Run App’
2. A new window will open showing the first page of the application, however do not use this. Click the button which reads ‘Open in Browser’ to maximize compatibility.



1. The application will launch in the web browser and data can be processed.



# Using the Ezy De-Identifier Tool for use with PortableEnv

To create a portable USB stick of the software

1. Download the latest version of the software from <https://github.com/dasasmk/EzyDeident/archive/master.zip>
2. Download the PortableENV from <http://bit.ly/1ZodFNl>
3. Unzip the downloaded files and move the whole PortableEnv folder to the root of a USB stick.
4. Copy the De-identification Tools EzyDeident folder to the root of a USB stick.
5. In the EzyDeident folder run the command ‘Download Components for PortableEnv.cmd’. The PortableEnv will be customized for use with the Ezy De-Identifier tool (internet required).
6. Execute the file ‘Run with PortableEnv.cmd’ to start the tool. In future you can start the application using Run with PortableEnv.cmd without needing an internet connection

The folder structure created during this process should be as follows

USB ROOT

|

──EzyDeident-master

────|Sample\_Data

────|www

──PortableEnv

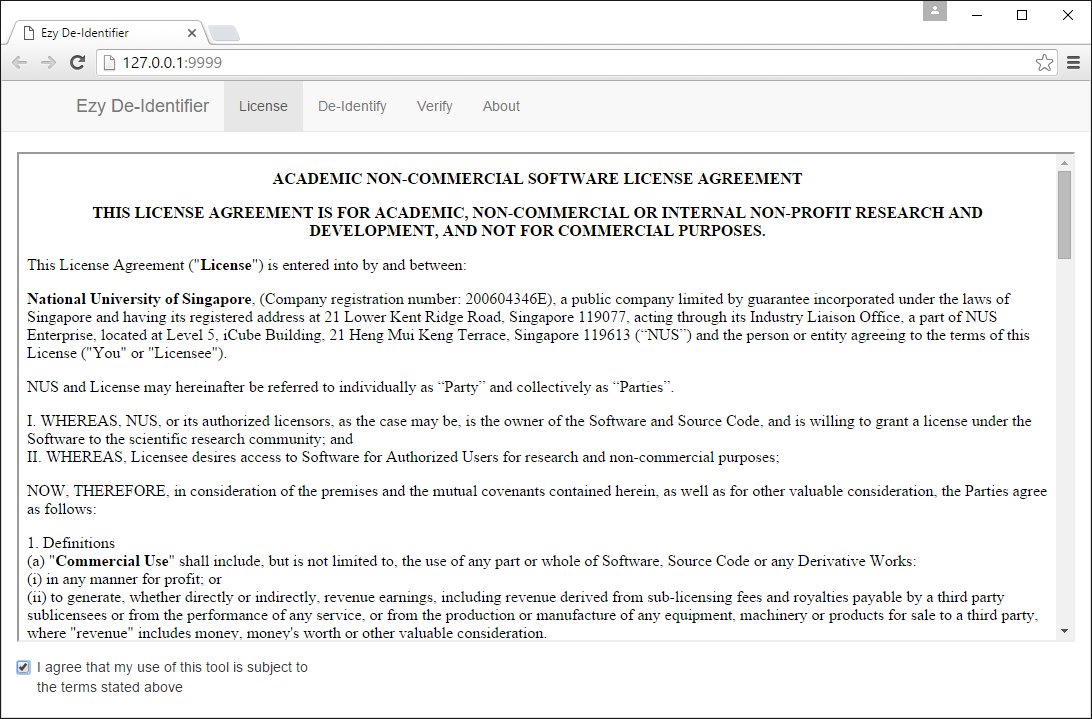
────|miktex-portable-2.9.5719

────|Pandoc

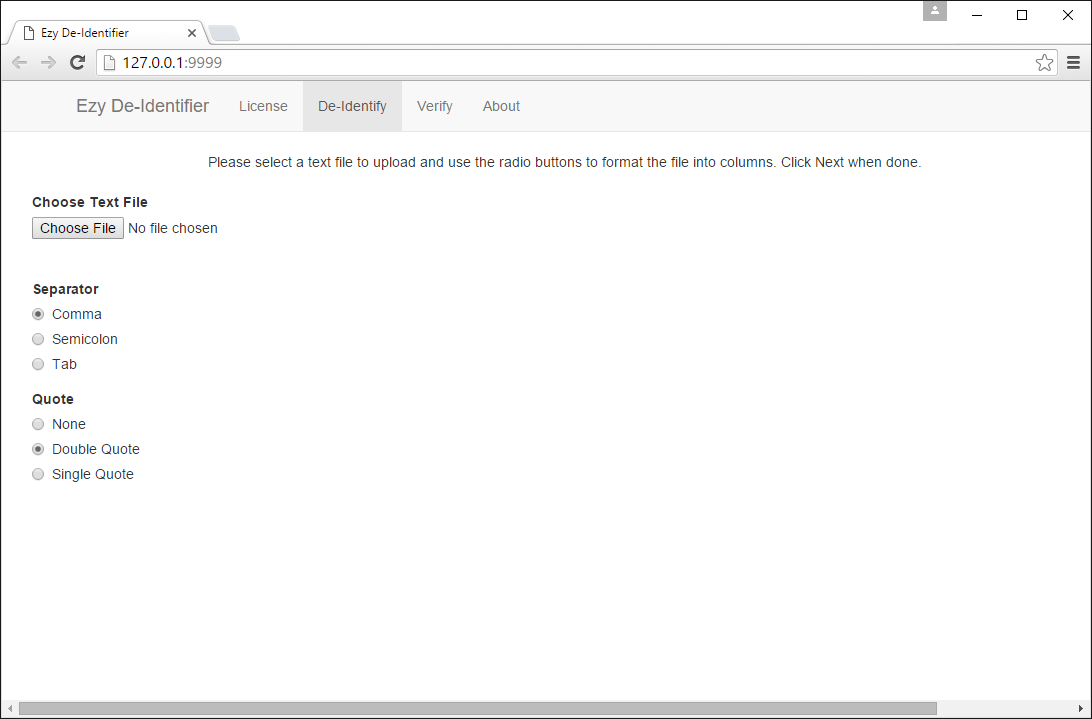
────|R-Portable

# Protecting Data using the Ezy De-Identifier Tool

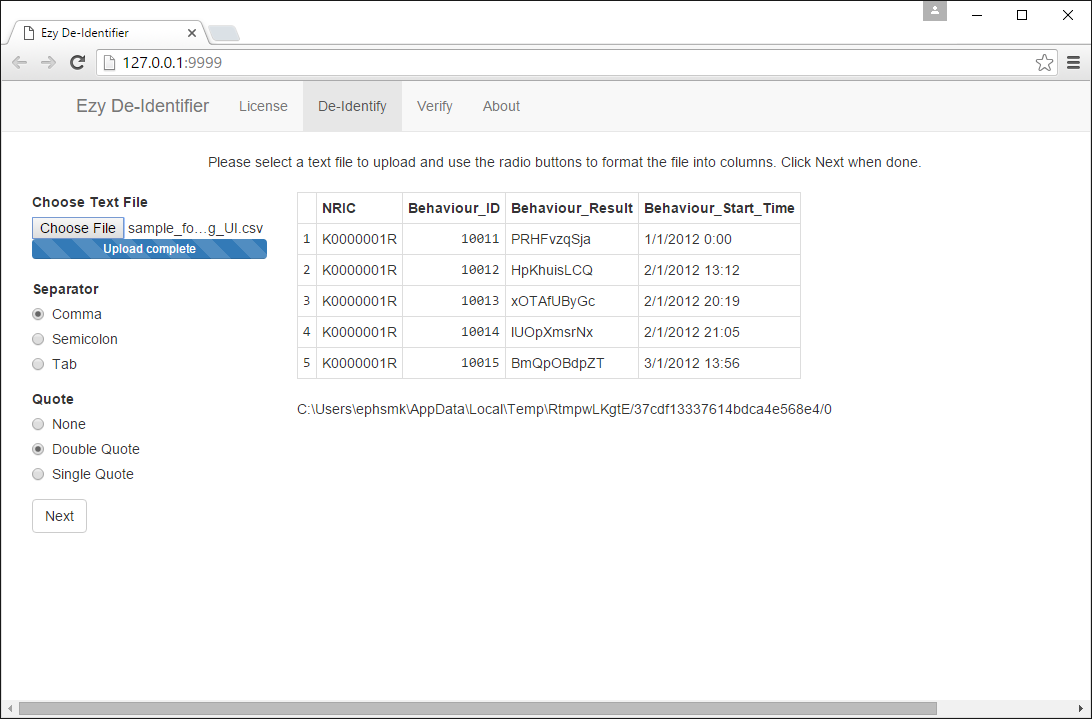
1. Read and accept the license agreement on the first page of the application



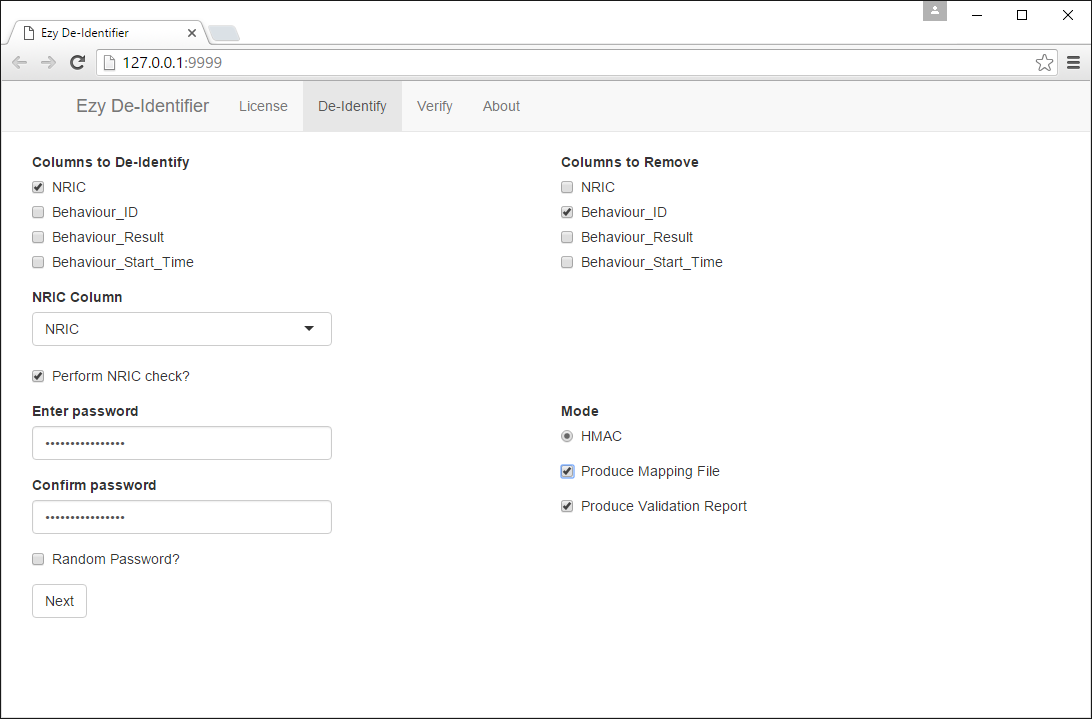
1. Click the De-Identify tab



1. In the top left click the Choose File Button, a dialog will appear allowing you to select a text file to import. Once uploaded a preview of the first 5 lines of data will be shown. Use the Separator and Quote radio buttons to change the input settings to match the file uploaded. Once the data is shown as a table click Next

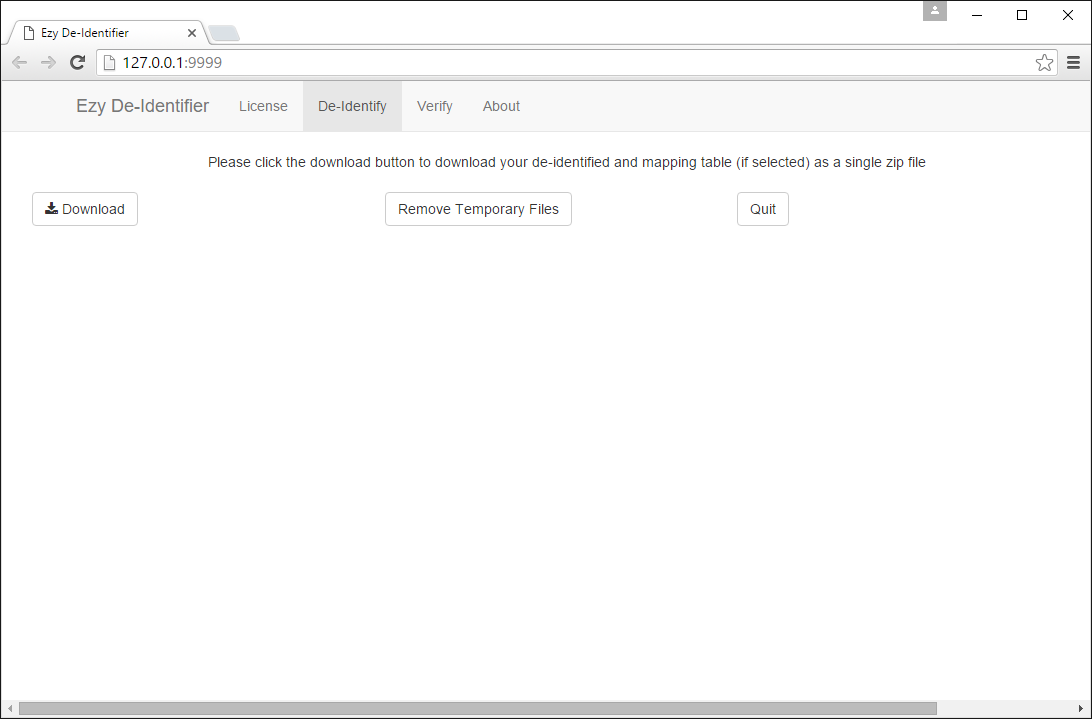


1. On the next screen you must specify how the application will de-identify the data, the requirements for this is stated in the table below. Once the application is configured click next.



|  |  |
| --- | --- |
| **Item** | **Requires** |
| Columns to de-identify | At least one |
| Columns to remove | Zero or more |
| NRIC Column / Perform NRIC Check | Optional |
| Enter Password | Required (recommend long password) |
| Confirm Password | Required (recommend long password) |
| Random Password? | Optional |
| Produce Mapping File | Optional |
| Produce Validation Report | Optional |

1. On the final page click the download button, after some time a zip file will be downloaded by the browser. The time to generate this file is dependent on the size of the data and speed of the PC being used.



## NRIC Column Check

If the NRIC column check is enabled an additional column of data will be added to the output utilizing the following coding system

|  |  |
| --- | --- |
| **Condition** | **Result** |
| NRIC Meets all Checks | 0 |
| NRIC is not starting with S,T,F,G | 1 |
| NRIC starts with X | 2 |
| NRIC is not 9 digits | -1 |

## Random Password

If the random password tick box is checked the software will automatically generate a 16 character password which includes special characters. For maximum security in a scenario where the data de-identification does not need to be recreated selected this option.

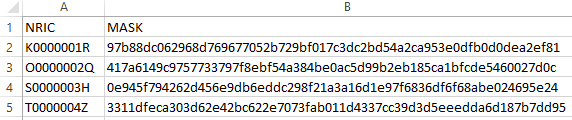
## Product Mapping File

When selected enabling this option will create mapping tables to allow the reverse identification of protected data. In the example below the user has selected to de-identify the column ‘NRIC’. This column will be protected in the main output product in the field NRIC\_masked. An additional file will be created called Mapping-NRIC.csv which will contain the unique combinations of the original value of NRIC and the de-identified value now present in the output.csv. For security do not generate a mapping table if there is no requirement to recover the original data.

Output.csv

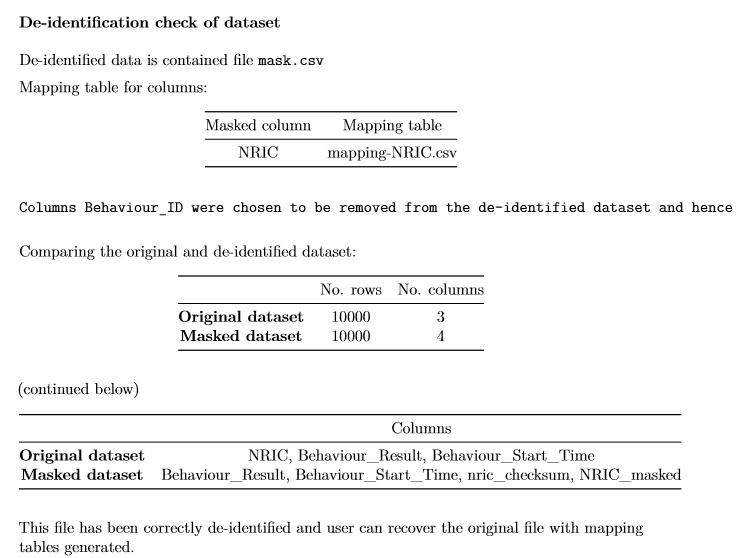


Mapping-NRIC.csv



## Produce Validation Report

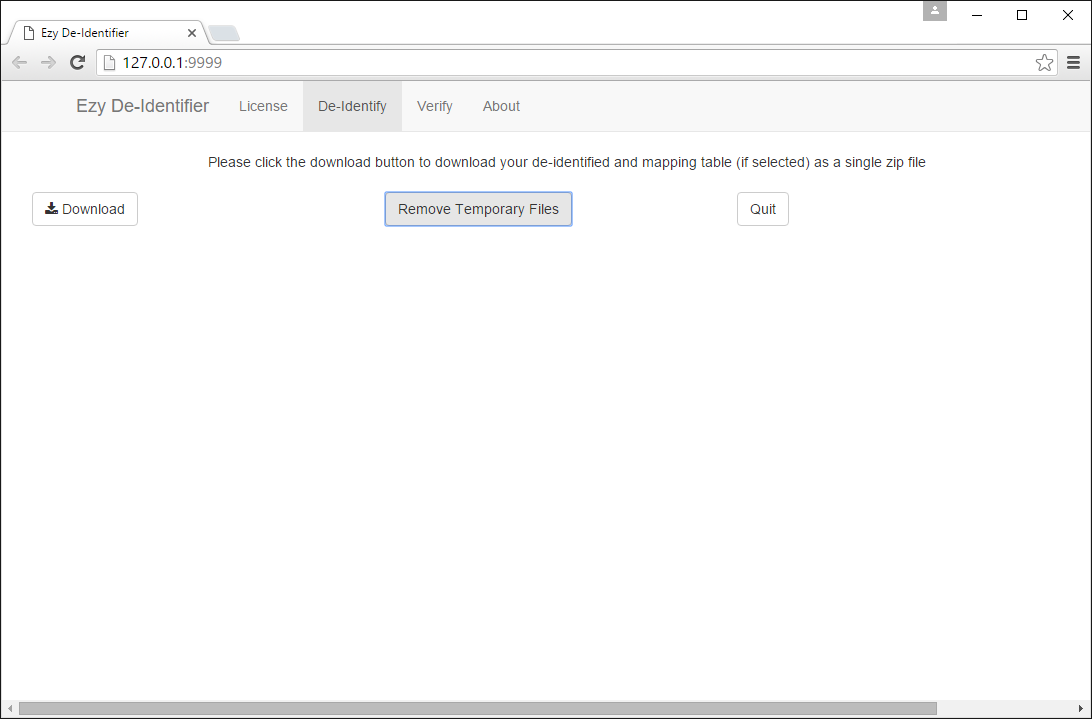
When selected enabling this option will generate a PDF report to validate that the software has been able to reconstruct the input data using the de-identified data and mapping tables. Selecting this option provides a check that the protection has been applied correctly, but will increase processing time.

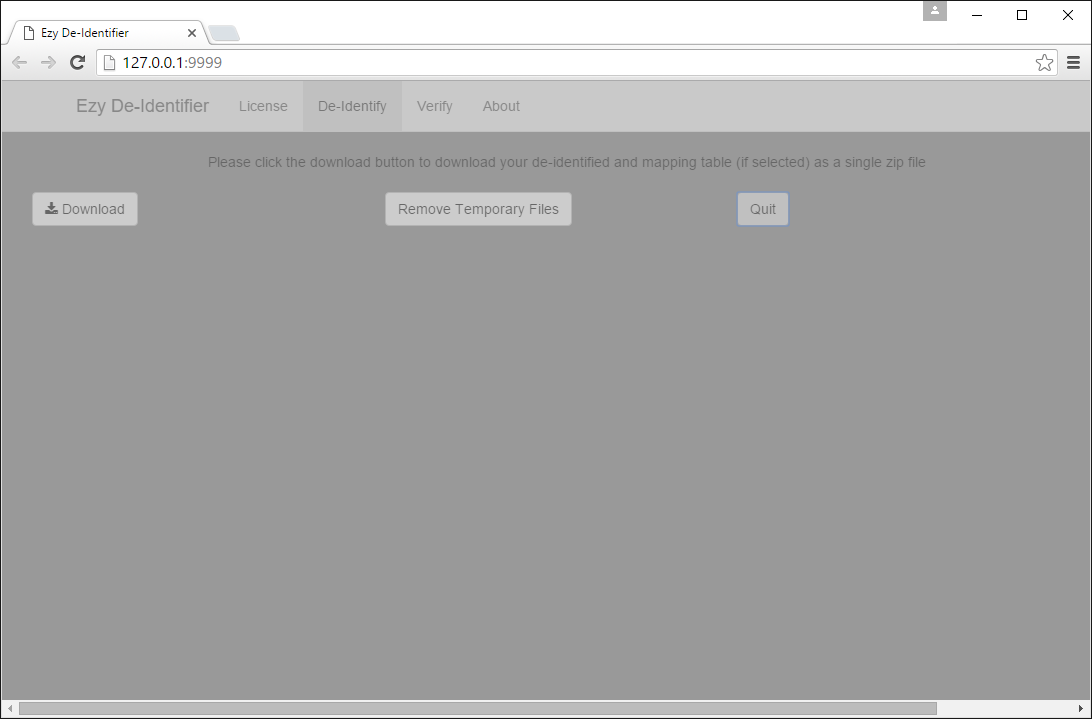


# Shutting Down the Ezy De-Identifier Tool

To close the application and remove temporary files from the system

1. Once the zip file of de-identified data has been downloaded click on the button to Remove Temporary Files.
2. Click the Quit button
3. Once the browser window dims, close your web browser.





# Ezy De-Identifier Tool End User License Agreement

**ACADEMIC NON-COMMERCIAL SOFTWARE LICENSE AGREEMENT**

**THIS LICENSE AGREEMENT IS FOR ACADEMIC, NON-COMMERCIAL OR INTERNAL NON-PROFIT RESEARCH AND DEVELOPMENT, AND NOT FOR COMMERCIAL PURPOSES.**

**This License Agreement ("License") is entered into by and between:**

**National University of Singapore, (Company registration number: 200604346E), a public company limited by guarantee incorporated under the laws of Singapore and having its registered address at 21 Lower Kent Ridge Road, Singapore 119077, acting through its Industry Liaison Office, a part of NUS Enterprise, located at Level 5, iCube Building, 21 Heng Mui Keng Terrace, Singapore 119613 (“NUS”) and the person or entity agreeing to the terms of this License ("You" or "Licensee").**

**NUS and License may hereinafter be referred to individually as “Party” and collectively as “Parties”.**

1. **WHEREAS, NUS, or its authorized licensors, as the case may be, is the owner of the Software and Source Code, and is willing to grant a license under the Software to the scientific research community; and**
2. **WHEREAS, Licensee desires access to Software for Authorized Users for research and non-commercial purposes;**

**NOW, THEREFORE, in consideration of the premises and the mutual covenants contained herein, as well as for other valuable consideration, the Parties agree as follows:**

**1. Definitions**

**(a) "Commercial Use" shall include, but is not limited to, the use of any part or whole of Software, Source Code or any Derivative Works:**

**(i) in any manner for profit; or**

**(ii) to generate, whether directly or indirectly, revenue earnings, including revenue derived from sub-licensing fees and royalties payable by a third party sublicensees or from the performance of any service, or from the production or manufacture of any equipment, machinery or products for sale to a third party, where "revenue" includes money, money's worth or other valuable consideration.**

**(b) "Software" means the object code/executable version of the software entitled “Ezy De-identifier”, together with related on-line or electronic documentation (if any) which may accompany the Software. Software includes any updates, error corrections, or new versions of the Software that are made available to Licensee free of charge by NUS (in its sole discretion), provided that NUS will have no obligation to make any such Software updates, error corrections, or new versions of the Software available.**

**(c) "Source Code" means the human-readable source code for the Software that is made available by NUS under this License.**

**2. Grant of License**

**2.1 Subject to Licensee’s compliance with the terms and conditions of this License, NUS hereby grants Licensee, and Licensee accepts:**

**(a) a non-exclusive, personal, non-transferable, non-sublicensable, royalty-free, fully paid-up, revocable, right to install, use and copy the Software in object code form only for the following purposes:**

**(i) testing and evaluation purposes; or**

**(ii) non-Commercial Use, including for academic or internal non-profit research and development purposes,**

**(collectively, "Purposes").**

**(b) a non-exclusive, personal, non-transferable, non-sublicensable, royalty-free, fully paid-up, revocable, right to install, use, copy, modify, prepare derivative works of, and compile into object code form the Source Code ("Derivative Works") for the Purposes only.**

**2.2 Licensee shall:**

**(a) not authorise, or permit the use by, or distribute, sublicense or otherwise transfer, or disclose, the Source Code to any third party, other than to Licensee’s contractors solely for the Purposes licensed herein;**

**(b) take all reasonable steps to ensure that the Source Code remains confidential and is protected against unauthorized disclosure, use or release, and will treat it with at least the same level of care as Licensee would use to protect and secure Licensee’s own confidential information, but in no event less than reasonable care.**

**(c) notify NUS of any Derivative Works created by the Licensee by sending an email notification to (i)** [**ephtcs@nus.edu.sg**](mailto:ephtcs@nus.edu.sg)**, (ii)** [**ephsmk@nus.edu.sg**](mailto:ephsmk@nus.edu.sg)**, and (iii) ephdx@nus.edu.sg.**

**3. Licensee acknowledges that any and all of the copyrights, patents, trademarks, design rights, know-how and other intellectual property and proprietary rights subsisting in or used in connection with the Software or the Source Code shall be and remain the sole property of NUS and its authorized licensors. Licensee must reproduce all titles, trademarks, disclaimers and copyright and restricted rights notices on to any copies of the Software or the Source Code as the case may be. Licensee's rights in the Software and the Source Code will be limited to those license rights expressly granted in this License, and NUS reserves all rights and licenses in and to the Software and the Source Code not expressly granted to Licensee herein. In the event that Licensee or any of its affiliates initiates any action: (a) challenging the ownership or validity of any NUS patents, copyrights, or other intellectual property rights in the Software or the Source Code; or (b) alleging that any use or exploitation of the Software or the Source Code infringes any patents, the rights and licenses granted herein shall terminate as of the date any such action is initiated.**

**4. The Software, Source Code and the Derivative Works shall remain in Licensee’s control and shall not be published, distributed, licensed or otherwise transferred or made available to any third party by Licensee, other than as provided in this Clause and to Licensee’s employees, academic staff and registered students involved in research or testing and evaluation of the Software, Source Code and Derivative Works under Licensee’s supervision, who are only permitted to a limited use of the Software, Source Code and Derivative Works for the Purposes in accordance with the terms of this License. Licensee may distribute the Derivative Works to any third party for non-Commercial Use only, provided that a) Licensee notifies NUS of such distribution in advance by sending an email notification to (i)** [**ephtcs@nus.edu.sg**](mailto:ephtcs@nus.edu.sg)**, (ii)** [**ephsmk@nus.edu.sg**](mailto:ephsmk@nus.edu.sg)**, and (iii)** [**ephdx@nus.edu.sg**](mailto:ephdx@nus.edu.sg)**; and b) the following legend is included in the About section of the Derivative Works: “This is a derivative work of Ezy De-identifer - Copyright 2015, originally co-developed by the Saw Swee Hock School of Public Health and Yong Loo Lin School of Medicine at the National University of Singapore and National University Health System in Singapore.”**

**5. The Software, Source Code and Derivative Works shall not be used in any manner or for any purpose other than the Purposes as expressly permitted by this License. If Licensee wishes to obtain the Software or Source Code for any other use or purposes, including any Commercial Use, Licensee will need to execute a separate licensing agreement with NUS on terms and conditions to be agreed.**

**6. Subject to the underlying copyrights in the Software and Source Code owned by NUS and its authorized licensors, Licensee retains all title, copyright and other proprietary interests in the Derivative Works developed by Licensee and copies thereof. Licensee hereby grants NUS a royalty-free, fully paid-up, non-exclusive, perpetual, worldwide, irrevocable license to use, copy, modify, sublicense, display publicly, distribute and prepare derivative works of the source code, object code and executable form of any Derivative Works in whole or in part, in stand-alone or as incorporated into or bundled with the Software and Source Code and to permit others to do so.**

**7. For the purposes of clarity, the license grant-back of Derivative Works pursuant to section 6 and the restrictions for use of the Derivative Works in this License shall only apply to works which translate, adapt, transform, modify or arrange the Software and Source Code licensed in this License. Further, the license grant-back of Derivative Works pursuant to section 6 will not apply if the Derivative Works are made by Licensee under any Commercial Use License under the Software and Source Code obtained from NUS.**

**8. Licensee may provide NUS with feedback on the use of the Software and Source Code in Licensee’s work, and Licensee agrees that NUS is permitted to use any information provided by Licensee for any purposes, including for making changes to the Software, Source Code or Derivative Works. Should Licensee have any bug reports, technical questions or improvements, Licensee may send these to the developers (i)** [**ephtcs@nus.edu.sg**](mailto:ephtcs@nus.edu.sg)**, (ii)** [**ephsmk@nus.edu.sg**](mailto:ephsmk@nus.edu.sg)**, and (iii) ephdx@nus.edu.sg. Licensee acknowledges that neither NUS nor its developers are obliged to provide any maintenance or support whatsoever for the Software or Source Code.**

**9. If use of the Software or Source Codes results in outcomes which will be published, please specify the version of Software Licensee used and cite the reference set out in the Software webpage on the NUS website.**

**10. Any risk associated with using the Software, Source Code or Derivative Works (if any) shall be borne solely by Licensee. This Software and Source Code are experimental in nature and are made available as a courtesy and on an "AS IS" basis, without obligation by NUS to provide any warranties or accompanying services or support.**

**NUS AND ITS AUTHORISED LICENSORS EXPRESSLY DISCLAIM ANY AND ALL WARRANTIES REGARDING THE SOFTWARE, SOURCE CODE AND DOCUMENTATION (IF ANY), WHETHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, WARRANTIES PERTAINING TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR COMPLIANCE WITH A PARTICULAR DESCRIPTION, OR ANY IMPLIED WARRANTY ARISING FROM THE COURSE OF PERFORMANCE, COURSE OF DEALING, USAGE OF TRADE OR OTHERWISE, TO THE FULLEST EXTENT PERMITTED BY LAW. IN PARTICULAR, NUS MAKES NO WARRANTY THAT THE SOFTWARE , SOURCE CODE AND DOCUMENTATION WILL MEET LICENSEE’S REQUIREMENTS OR THAT THE OPERATION OF THE SOFTWARE, SOURCE CODE AND DOCUMENTATION WILL BE UNINTERRUPTED OR ERROR-FREE OR THAT ALL ERRORS IN THE SOFTWARE, SOURCE CODE AND DOCUMENTATION CAN BE CORRECTED.**

**IN NO EVENT SHALL NUS AND ITS AUTHORISED LICENSORS HAVE ANY LIABILITY FOR ANY LOSS OR DAMAGE OF ANY KIND OR NATURE, INCLUDING ANY DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL LOSS, PUNITIVE OR EXEMPLARY DAMAGES, LOSS OF PROFITS, LOSS OF DATA FOR ANY REASON WHATSOEVER, WHETHER SUCH LIABILITY IS ASSERTED ON THE BASIS OF CONTRACT, TORT (INCLUDING NEGLIGENCE OR STRICT LIABILITY) OR OTHERWISE, EVEN IF A PARTY HAS BEEN WARNED OF THE POSSIBLITY OF SUCH LOSS OR DAMAGES.**

**11. Licensee shall indemnify, defend and hold harmless NUS, its trustees, agents, officers and employees and the Software developers against any and all claims, suits losses, damage, costs, fees, and expenses arising out of or in connection with this License or the use of the Software and Source Code. Licensee shall pay all costs incurred by NUS in enforcing this provision, including reasonable attorney fees on a full indemnity basis.**

**12. Licensee acknowledges that any and all of the copyrights, patents, trademarks, design rights, know-how and other intellectual property and proprietary rights subsisting in or used in connection with the Software and Source Code, shall be and remain the sole property of NUS and its authorised licensors. Licensee shall not during or at any time after the term of this Licence in any way question or dispute such ownership by the NUS and its authorised licensors.**

**13. Licensee shall retain in the Software, Source Code and documentation of any Derivative Works that Licensee creates, all copyright, patent, or trademark notices from the Software, Source Code and documentation of the original Software as well as any notices of licensing or descriptive text.**

**14. This Licence shall terminate automatically if Licensee fails to abide by any of the terms of this Licence or if the License is terminated. Upon termination of this License, Licensee shall uninstall the Software, Source Codes and Derivative Works and shall erase all copies of the Software, Source Codes and Derivative Works under Licensee’s control and stored on any medium, and destroy all documentation. Notwithstanding any provision in this License, Sections 1, 3, 6, and 8 to 17 shall survive termination of this License.**

**15. The failure or delay of NUS to assert any of its rights under this License shall not be deemed to constitute a waiver of NUS' rights thereafter to enforce each and every provision of this License in accordance with its terms. No waiver by NUS of any breach or default by Licensee shall operate as a waiver of any succeeding breach or other default or breach by Licensee. A waiver must be specific, irrevocable and in writing, to be effective.**

**16. This License shall be subject to, governed by and construed in accordance with the laws of Singapore, without regard for its conflict of laws provisions and subject to section 17, the Parties submit to the jurisdiction of the courts in Singapore.**

**17. Any dispute arising from this License shall be resolved by arbitration in the English language at the Singapore International Arbitration Centre (“SIAC”) in Singapore before a single arbitrator appointed by the Chairman of the SIAC in accordance with the Arbitration Rules of the SIAC, whose decision shall be final and binding on the Parties. Judgment upon any arbitration award may be enforced in any court of competent jurisdiction. Each Party shall bear its own expenses but the Parties shall share equally the expenses of the arbitral tribunal. Nothing in this Section shall preclude either Party from seeking provisional remedies, including, *inter alia,* temporary restraining orders and preliminary injunctions from any court of competent jurisdiction in order to protect the Party's rights pending arbitration.**