

Appendix 8: Data linkage using pseudonymisation

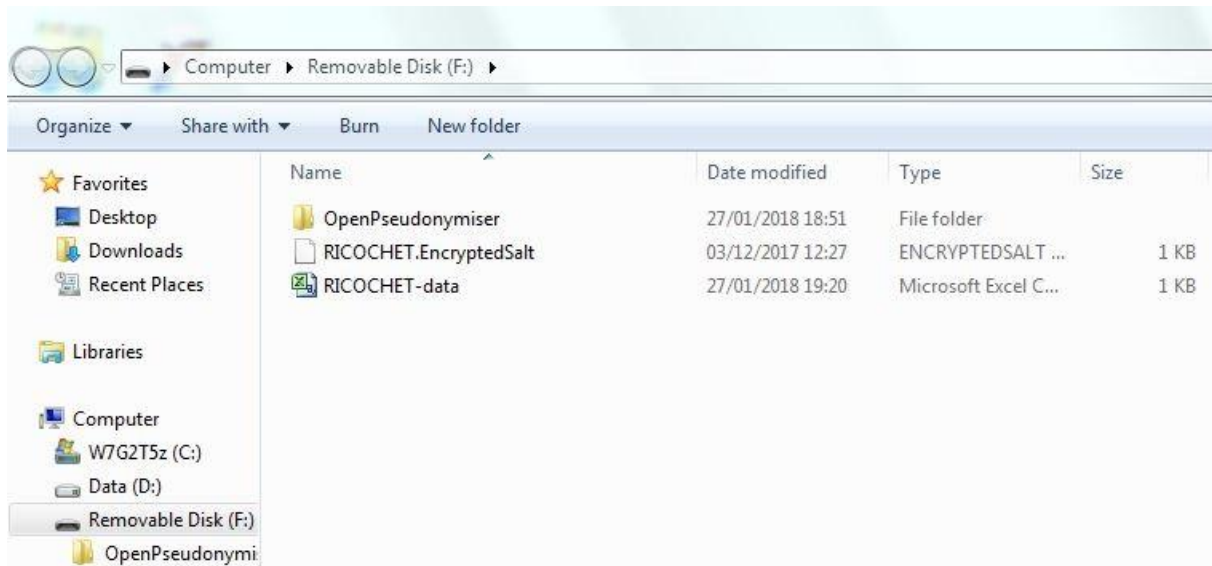
For this project to describe the entire investigative management pathway, patient data must be linked between local and specialist centres. In order to comply with the seven Caldicott Principles, The Operating Framework for the NHS in England 2011/2012, Confidentiality – NHS Code of Practice, the Data Protection Act 1998 and the forthcoming EU General Data Protection Regulation, replacing the Data Protection Directive 95/46/EC and due to be implemented on 25th May 2018, patients will be de-identified when added to the secure REDCap database. Whilst complete anonymisation is not possible by definition as it will not enable data linkage, pseudonymisation is a technique to de-identify data that is recognised by NHS Digital. Although the NHS Digital implementation of pseudonymisation is no longer active, the OpenPseudonymiser project, lead by Professor Julia Hippisley-Cox at the University of Nottingham has been developed that will aid data linkage.

OpenPseudonymiser is a free, open source, standalone Windows application, written in C# that utilises the .NET assembly as well as Java, the source code of which is freely available on GitHub for inspection, and uses the principles outlined by the Information Commissioners Office. It can be redistributed and/or modified under the terms of the GNU General Public Licence in accordance to the Free Software Foundation. OpenPseudonymiser checks the validity of the NHS Number within a Comma Separated Variable (CSV) file, adds extra data to the NHS Number using an encrypted set of characters, known as a salt, followed by cryptographic hash functions using the internationally recognised Secure Hash Algorithm 2 (SHA-256) to produce a string of output characters, known as the digest. It is the digest that will be used as the case identifier within REDCap. It can be used for linking data, as using the same NHS Number with the same encrypted salt and SHA-256 will produce the same digest. Further details can be found in the OpenPseudonymiser Desktop User Guide and instructional video at https://www.openpseudonymiser.org/OpenPseudonymiser_Docs.aspx

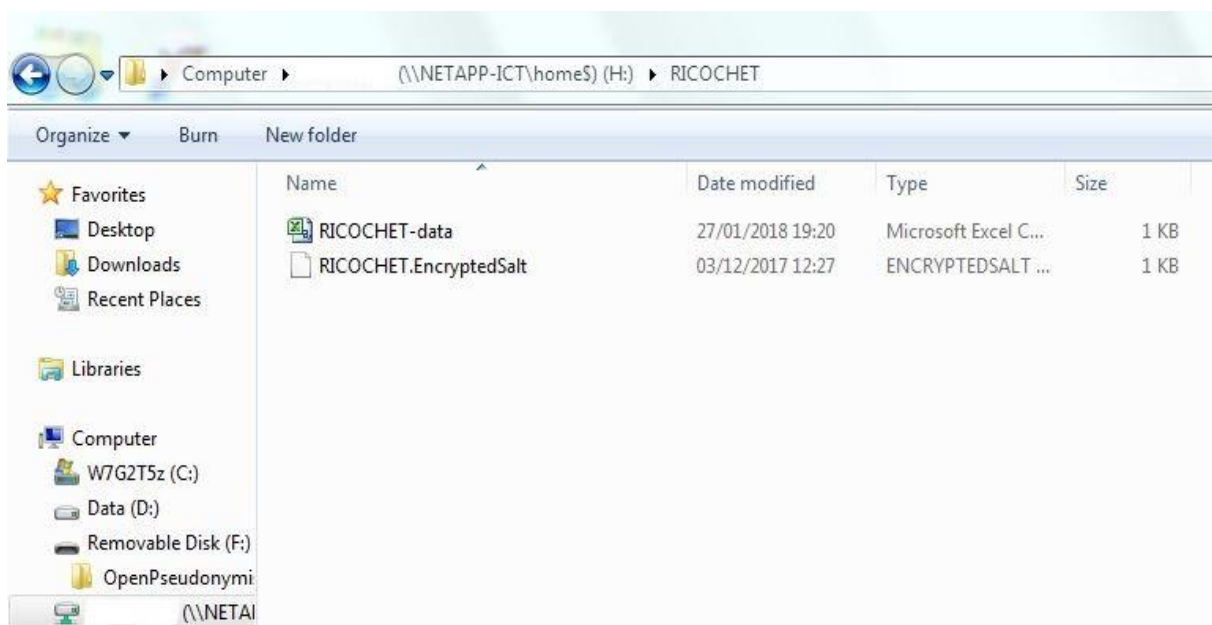


The process of pseudonymisation specific for the RICOCHET Study using OpenPseudonymiser application is as follows:

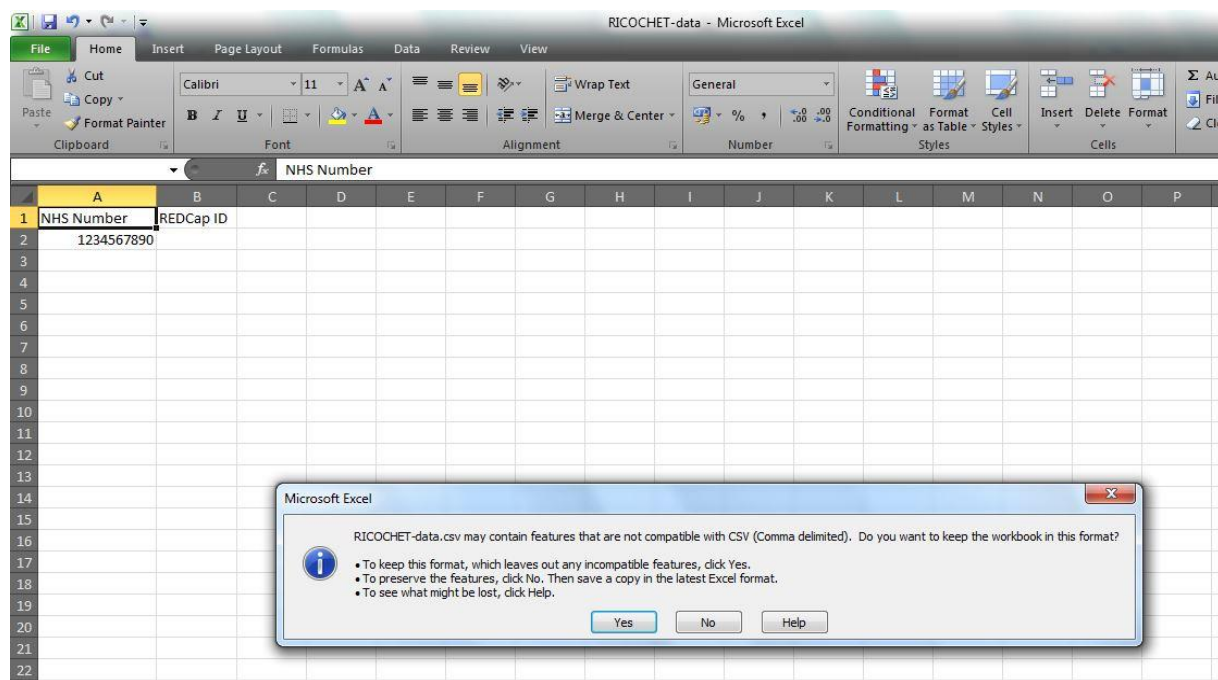
1. A USB stick containing the OpenPseudonymiser application, the encrypted salt key named 'RICOCHET.EncryptedSalt' and the CSV file named 'RICOCHET-data.csv' will be delivered to the local lead in each participating centre.



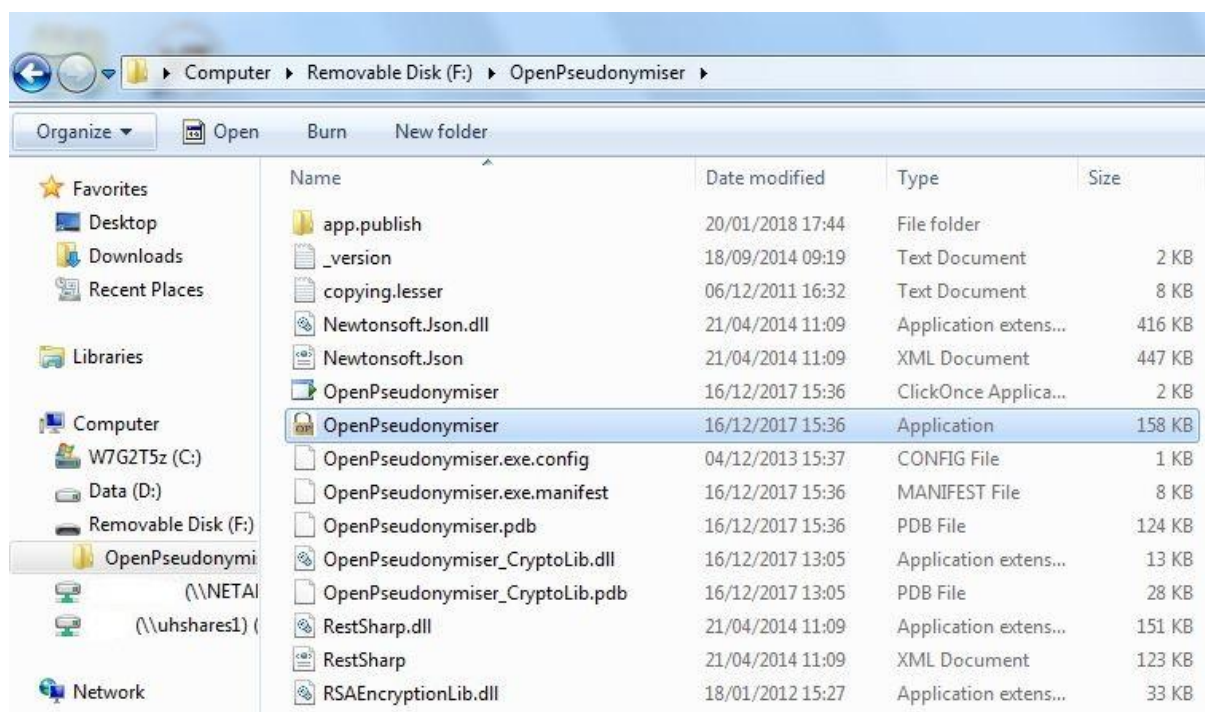
2. On receipt and insertion of the USB stick, 'RICOCHET.EncryptedSalt' and 'RICOCHET-data.csv' should be copied to the same location on a local computer. If possible, copy the files to a network drive for access on different computers within the same organisation using a personal login to the computer.



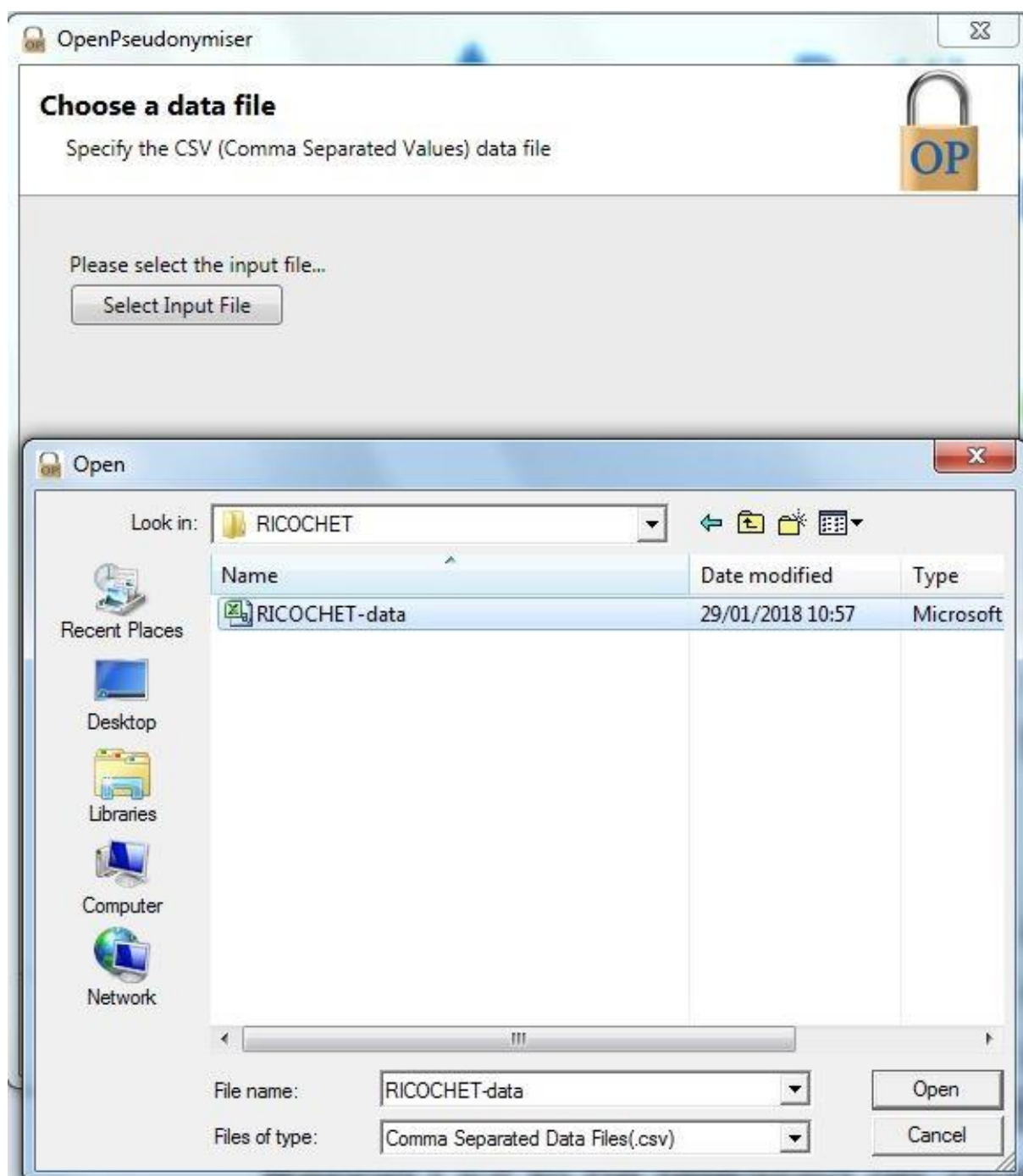
- When a patient has been identified for the RICOCHET Study, open 'RICOCHET-data.csv' in Microsoft Excel, enter the patient's NHS Number underneath the 'NHS Number' header and REDCap identifier underneath the 'REDCap ID' header. Save the CSV file in the same format. Click 'Yes' to 'RICOCHET-data.csv may contain features that are not compatible with CSV (Comma delimited). Do you want to keep the workbook in this format?' Close 'RICOCHET-data.csv'.

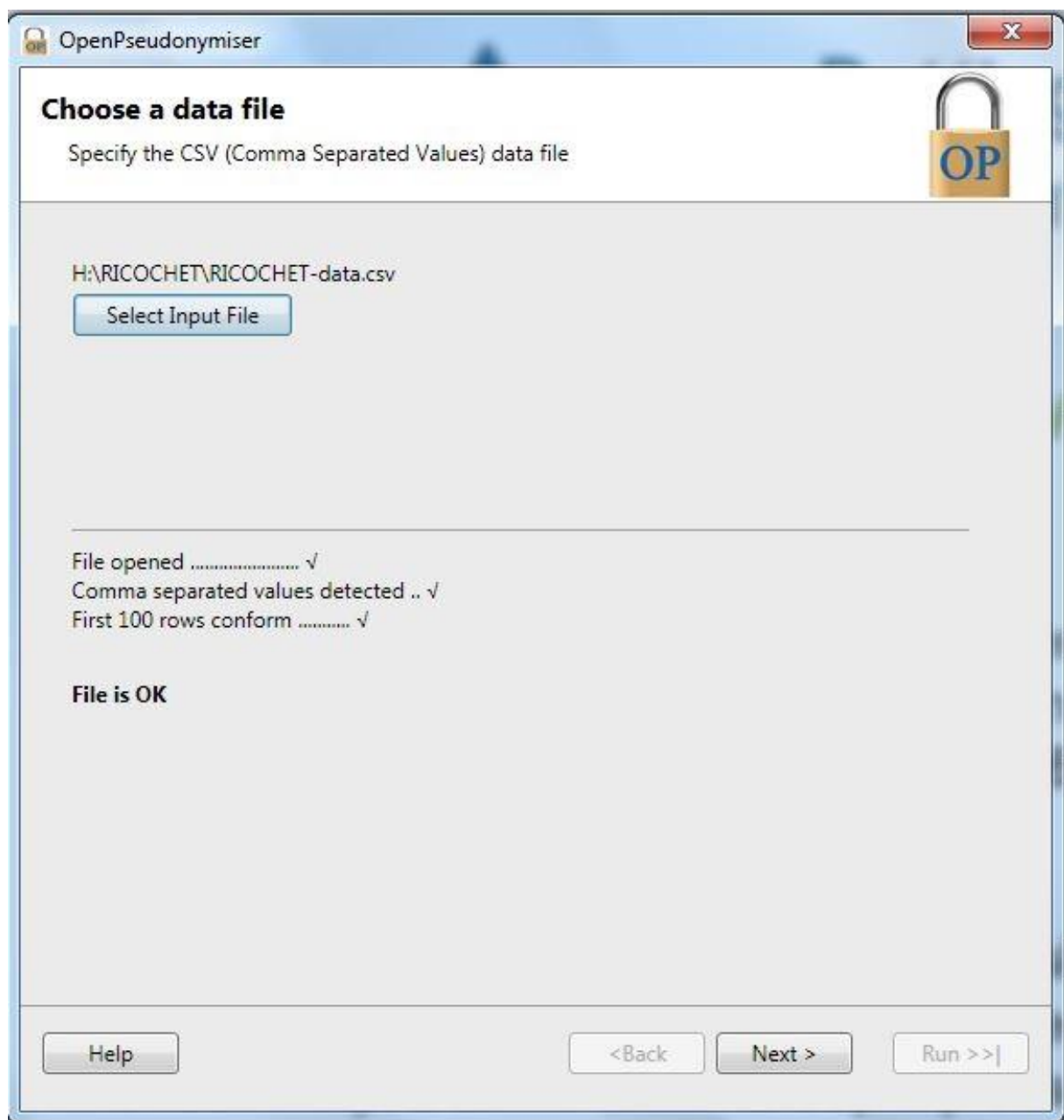


4. Double-click the OpenPseudonymiser application from the USB stick (Drive letter:/OpenPseudonymiser/OpenPseudonymiser.exe) and at first run read and accept the user licence.

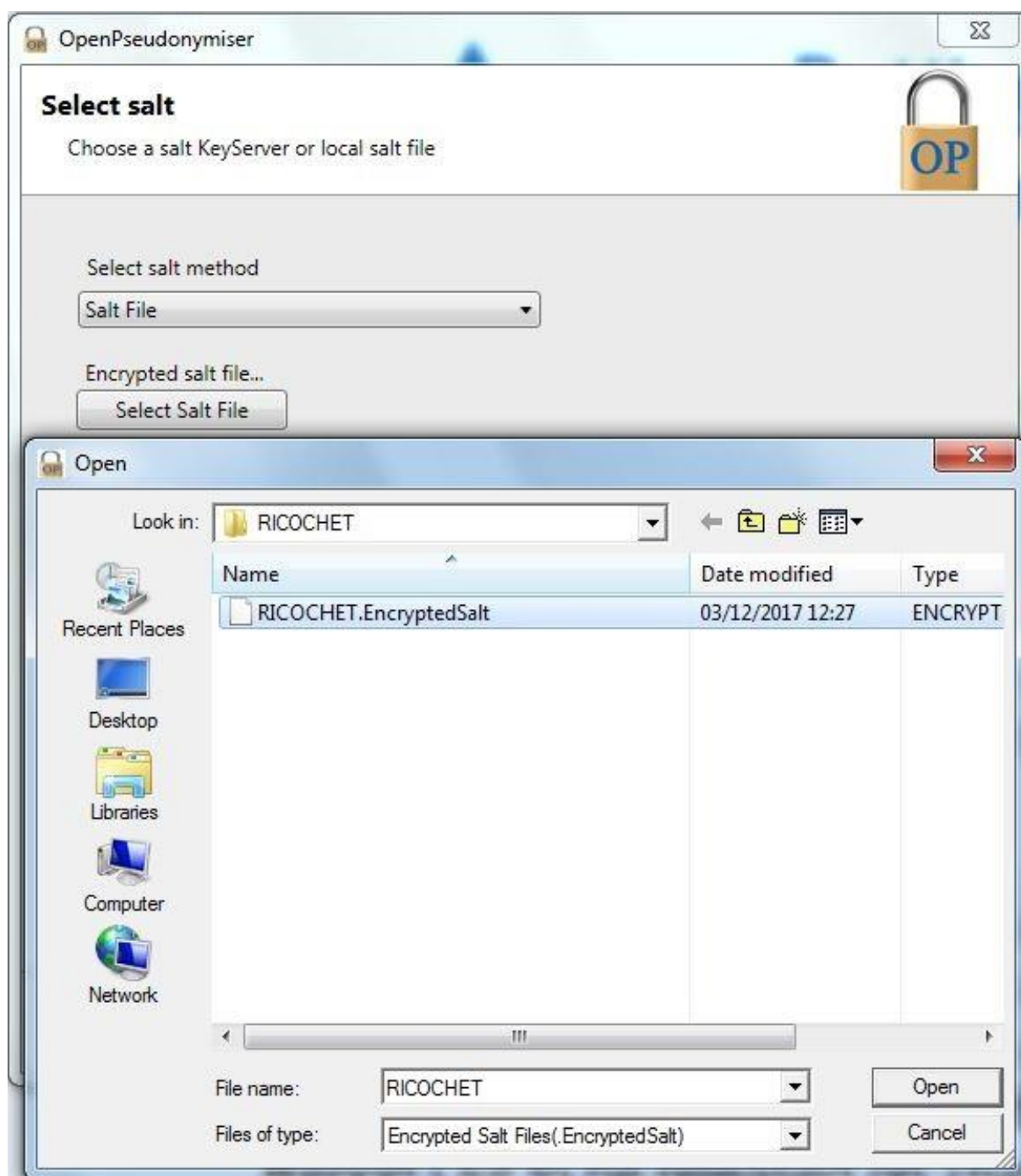


- Click 'Select Input File', navigate to and select 'RICOCHET-data.csv' on the computer and allow OpenPseudonymiser to perform some integrity checks on the file before clicking on 'Next' at the bottom right of the window. This check includes the validity of the NHS Number.





6. Change the 'Select salt method' drop-down to 'Salt File', click 'Select Salt File' and navigate to and select 'RICOCHET.EncryptedSalt', from the computer, before clicking 'Next' in the bottom right of the window.



- On the next screen, underneath 'Use in Digest' click on the checkbox for the NHS number. Ensure that the checkboxes underneath 'Use in Output' are ticked for NHS Number and REDCap ID. From the 'NHS Number Field' drop-down menu, select 'NHS Number' before clicking 'Next' in the bottom right of the window.

OpenPseudonymiser

Select columns

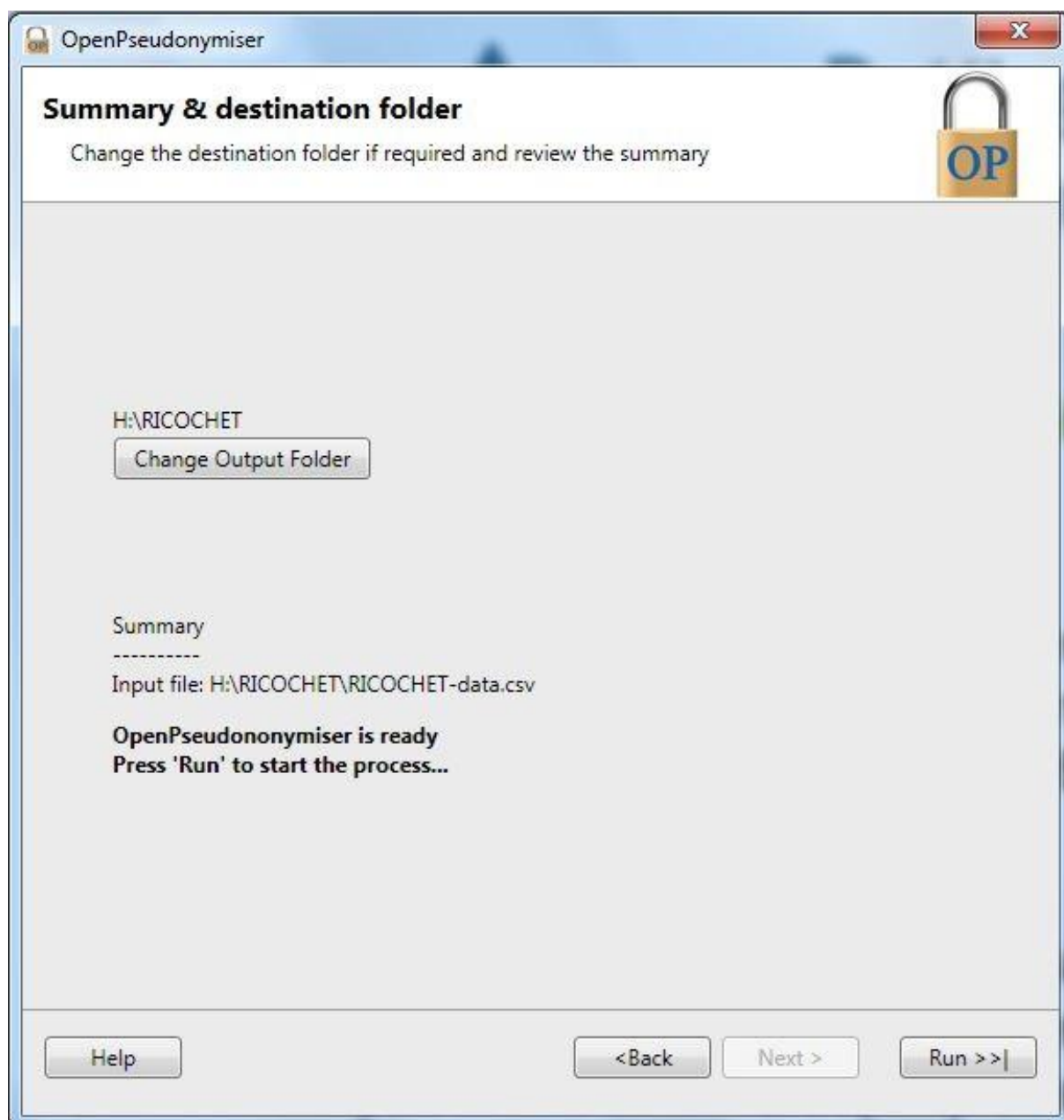
Specify which columns to use for the Digest, and which ones to output

NHSNumber field: NHS Number

Use in Digest	Use in Output	Process as Date	Column Heading
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	NHS Number
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	REDCap ID

Help <Back Next > Run >>|

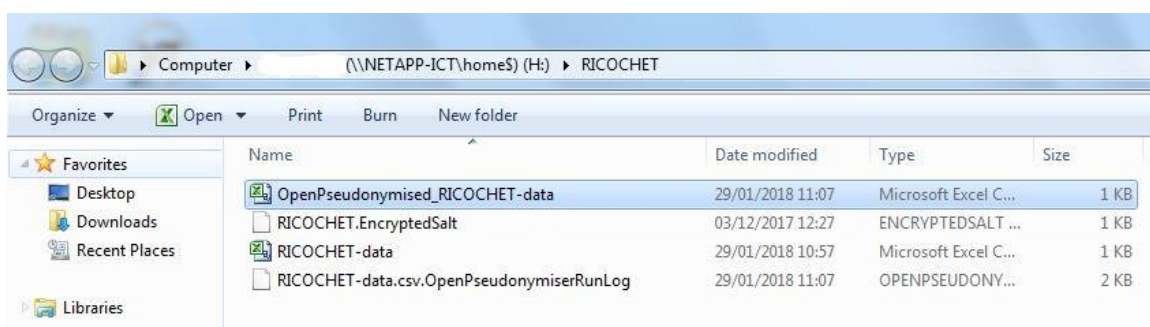
8. Ensure that the path to the output location is the same as the location of 'RICOCHET-data.csv'. If it is not, then click on 'Change Output Folder' and navigate to the same location as 'RICOCHET-data.csv'.



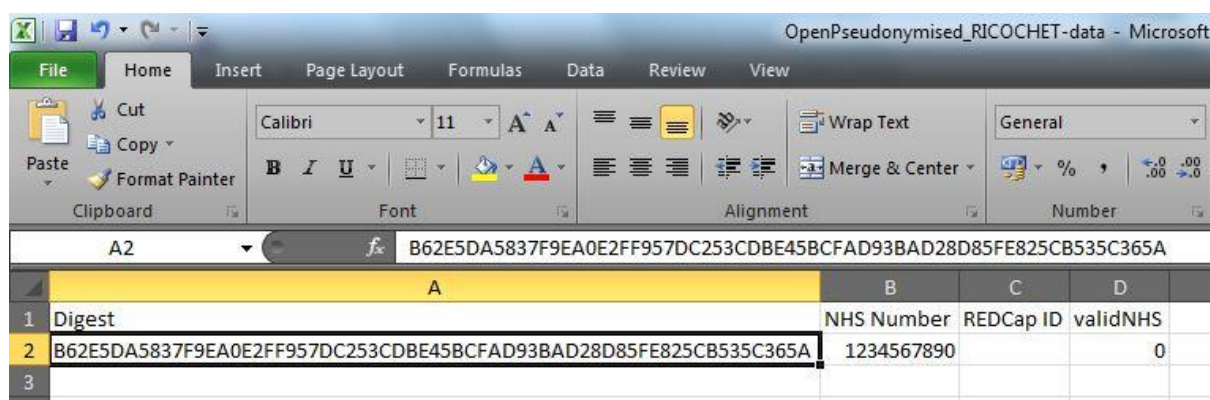
9. Click 'Run' in the bottom right of the window.



- Once pseudonymisation is complete, click on 'Open Output Folder' and in the new window open 'OpenPseudonymiser_RICOCHET-data.csv'. Microsoft Excel will open 'OpenPseudonymiser_RICOCHET-data.csv' containing the pseudonymised output digest.



- Check that the previously entered NHS number is valid by inspecting the 'validNHS' cell. A '1' is the valid NHS number whilst '0' is an invalid NHS number and should be rechecked before restarting at Step 3. If it is a valid NHS number, copy the digest and paste into REDCap when asked 'What is the Open Pseudoanonymiser digest?'.



12. When further patients are identified, repeat steps 1-11. OpenPseudonymiser will add the NHS Number and its corresponding digest as well as the REDCap ID to the existing 'OpenPseudonymiser_RICOCHET-data.csv'. It is **imperative and the responsibility of the Local Consultant to ensure that this file is backed up** as it is the only way of linking local data to REDCap data

	A	B	C	D
1	Digest	NHS Number	REDCap ID	validNHS
2	B62E5DA5837F9EA0E2FF957DC253CDBE45BCFAD93BAD28D85FE825CB535C365A	1234567890		0
3	FE1033A0F2983D9BDCD9BF3255024FFAF190069AD8839DC2C68C4CE2D6C3DE2F	9876543210		1
4				