IDDO Data De-identification Procedure
Clinical Trials

Introduction

The purpose of the document is to provide an overarching view of the data de-identification framework adopted by IDDO to preserve the privacy and data protection rights of the participants included in the research studies shared with IDDO. An outline of the steps undertaken to preserve the privacy and data protection rights of research participants during the following stages of data flow is presented:

a) upon submission of data by the investigators to IDDO's secure data platform;
b) during data curation process;
c) during external sharing of the curated, de-identified data for research purposes.

The platform is committed to protecting the rights, safety and privacy of the individuals and communities from whom the data originate and ensuring compliance with ethical requirements and applicable laws and regulations.

The platform complies with UK and European laws and regulatory frameworks, in particular, but without limitation, these include rules and regulations governing the protection of personal data, clinical trials, research and the protection of human rights. The platform adheres to the European Union's General Data Protection Regulation (GDPR)[1] through the UK Data Protection Act 2018 (UK GDPR),[2] and the Universal Declaration of Human Rights,[3] an international agreement and cornerstone of international human rights law, protecting the individual right to privacy. The platform also abides by University of Oxford policies reflecting these UK and European laws and regulations, which can be found on the Oxford's Central University Research Ethics Committee website.[4]

Those who request access to data on the IDDO platform are contractually required to observe the highest standards of ethics and integrity in the course of their research using platform data, regardless of the location in which the research is carried out. Requests for access to data are reviewed by the original Data Contributors or an independent Data Access Committee overseen by TDR, the Special Programme for Research and Training in Tropical Diseases at the WHO in accordance with the instructions of the Data Contributor. Following approval, a Data Use Agreement is executed between the University of Oxford on behalf of IDDO (as the legal entity and platform host) and the Data Recipient in advance of any transfer of data approved by the Data Access Committee or original Data Contributor.

De-identification upon intake of data to prevent identification

The data shared with IDDO through the IDDO secure upload portal are initially stored on a high-compliance server offering specific security features and limited access to select administrators (see Appendix 1). All servers are hosted and controlled by the University of Oxford. Contributors depositing their data are asked to only share pseudonymised data (as

defined by GDPR). In addition to this request, an initial check is performed to remove any direct identifiers from the submitted dataset that may have accidentally been submitted before the data enter the curation workflow. IDDO follows best practice established by the U.S. Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor method of de-identification, where 16 of the 18 protected health identifiers as defined by HIPAA are removed (see **Table 1** below).

Table 1 HIPPA Safe Harbor variables removed by IDDO

Safe Harbor Data Variables	
Number	**Column**
1	Names
2	Telephone numbers
3	Fax Numbers
4	Email
5	Social Security Numbers (Tax Identification Numbers)
6	Medical Record Numbers
7	Health Plan Beneficiary Numbers (Health System ID)
8	Account Numbers
9	Certificate/License Numbers
10	Vehicle identifiers and serial numbers including license plate numbers
11	Device identifiers and serial numbers
12	Web universal resource locators (URLs)
13	Internet Protocol Addresses
14	Biometric identifiers, including finger and voice prints
15	Full face photographs and any comparable images
16	Any other unique, identifying number, characteristic or code
17*	Address (geographic subdivisions smaller than state/province)
18*	All elements of dates related to an individual

* Retained for curation and minimised as outlined below in 'Further minimisation to remove identifiers.

Further minimisation to remove identifiers

Once HIPAA Safe Harbour is implemented, the submitted data then enters the curation workflow. Free text such as comments are examined for any identifying information and removed if found. Date of birth is removed and only age retained. Country of data collection is retained. Institution names are provided to the data recipient (but not linked to individual patient data) for the purpose of appropriate participation and attribution of Contributors. Upon completion of curation, the data is stored in a secure server.

Sharing of data and replacement of existing patient ID using a random key

Data domains within the available IDDO dataset are shared in a tailored manner according to the variables requested by the researcher.

New patient keys will be generated using a random mechanism that will replace the original patient ID before externally sharing the database. This will ensure that multiple requests from the same research group will not be able to link the databases together.

Researchers are contractually obliged to not use, attempt to use or permit use of the dataset to re-identify any individual, community of medical institution associate with the Dataset, nor link, attempt to link or permit a third Party to link the dataset with any other data in a manner that may enable re-identification of individuals, communities or medical institutions.

References

1. European Union General Data Protection Regulation (2016)

2. UK Data Protection Act 2018 (legislation.gov.uk)

3. United Nations Universal Declaration of Human Rights

4. www.admin.ox.ac.uk/curec

5. www.hhs.gov. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [Internet]. [cited 2020 Aug 20]. Available from: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard