Ebola Data Platform

Privacy Impact Assessment

Version: 1.0

Release: Draft

Date: 05/06/2017

Author(s): D. Newton

Table of Contents

Introduction

The aim of a Privacy Impact Assessment (PIA) is to provide a structured process to analysing and assessing privacy risk to individuals and organisations, in the collection, use and disclosure of personal data. It forms part of a due diligence process when designing and carrying out a new project involving personal data. This is both from an ethical point of view, to minimise and protect against any potential negative effects to individuals such as an unjustified intrusion and/or stigmatisation, and also from a legal compliance, information security point of view, where failure to embed appropriate privacy protection may result in a potential or actual breach of privacy laws, damage to public trust/organisational reputation, or lead to prohibitive costs in retro-fitting systems to ensure compliance and/or address concerns about privacy.

The PIA will capture and identify data protection and privacy concerns (in the project), and provide a risk analysis of the likeliness of these events or circumstances materialising. From this assessment, it will provide recommendations for managing and mitigating these risks, and what is needed to fulfil any regulatory data protection requirements and can inform (what would be recognised as) best practice. This is whilst recognising the need for flexibility, and governance arrangements that can accommodate changes in the future. A PIA represents a snapshot in time, relevant to current domestic and international law, industry standards and accepted best practice. Such components are dynamic and evolving and therefore the PIA is an ongoing and iterative process to be carried out periodically, to ensure it can be updated and developed in line with relevant legislative, regulatory and best practice guidance in this area.

Project Background

Despite more than two-dozen outbreaks of the Ebola Virus Disease (Ebola) over the past 40 years, there has been a historically low research and development response, resulting in inadequate options for diagnosis and treatment. Although the size and consequences of the 2013-16 West African Ebola outbreak were unprecedented, empiric and scientific evidence to inform advances in diagnosis, triage, management and follow-up of suspected and confirmed Ebola patients remains inadequate.

Operational actors and researchers executed limited evaluation of the efficacy and safety of new treatments during the 2013-16 outbreak, and protocol development, gaining availability of experimental compounds and clinical trial implementation take time. The epidemic waned shortly after initiation of many of the trials, leading to a paucity of eligible participants, underpowered studies and little data. Moreover, the data that was collected used a range of formats and trial designs making them difficult to compare. In parallel, biological sample collections were conducted by different laboratories and are now preserved in several different labs with no obvious connection to clinical data from the original patient.

Subsequently, there is a wealth of clinical, laboratory, and epidemiological data, collected in the three most affected countries (i.e. Guinea, Liberia and Sierra Leone) as well as some historical data from previous outbreaks occurring in Central and East Africa in the last 4 decades. In reference to the West Africa outbreak, the data is scattered globally across patient charts and databases from Ebola Treatment Units, Community Care Centers, mobile laboratories, academic institutions, government and non-government organisations and Ministries of Health. Collectively, these data are the largest volume of information ever recorded on Ebola, but there is no means to collate them.

To date, the power and utility of this data has been limited to analysis of small batches, usually from single Ebola Treatment Units or from single organisations that responded to the outbreak. There is no common repository of clinical, laboratory or epidemiological data on Ebola, de facto undermining the ability of the research community to effectively use these resources, prioritise research and leverage

existing knowledge. These scattered collections of data are crucial to advance our understanding of Ebola and fill the existing knowledge gaps. Bringing these data together in a useful format that can be accessed by the research and humanitarian communities will maximise the potential to generate robust evidence to improve our response to future outbreaks and optimise care of Ebola patients.

The Infectious Disease Data Observatory (IDDO) will establish an Ebola Data Sharing Platform to deliver policy-changing evidence to support patient care and outbreak response. Retrospective and prospective data on individuals known or suspected to have Ebola infection will be amalgamated from health care institutions, laboratories, public health systems and academia. Data will be standardized to a uniform structure to enable pooled analysis and will be shared under an agreed governance and ethical framework.

By enabling ethical and equitable sharing of data and information on emerging infections, the Platform aims to increase the dissemination of knowledge and improve patient outcomes. The Platform's overall goals are:

- Incentivize and maximize contributions to the Platform of the widest variety of data and information, whether clinical, epidemiological, laboratory or other data of any kind related to Ebola or other emergent pathogens;
- Give the scientific, health and research community and public health authorities the ability to use such Data through a timely, transparent sharing process with as few restrictions as possible to facilitate and accelerate the research for curative, preventative and diagnostic tools and assist in the identification of new outbreaks;
- Encourage the rapid dissemination of information in an equitable manner, which respects the interests of those who collect the data;

  while

- Protecting the rights, safety and privacy of the individuals and communities from whom the data originated and ensuring compliance with ethical requirements and applicable laws and regulations.

Do we need a privacy impact assessment for the EVD Data Sharing Platform?

Privacy Impact Assessments (or Data Protection Impact Assessments) are carried out whenever there is a change that is likely to involve a new use or significant change in the way in which personal data is handled. For example, this could be for a redesign of an existing process or service, or potentially a new service, process or information asset being introduced. To identify the need for a PIA a set of screening questions are often used. Answering yes to anyone of these questions indicates a need for an assessment. The screening questions and answers for this project can be found in Appendix 1.

The Ebola Virus Disease Data Platform project aims are clearly defined to create a new service that will process personal data on behalf of contributors for the benefits of research and to provide standardisation for the future collection and dissemination of data. As much of the data was collected for treatment purposes only, the project entails a new use of the data that was not originally intended and new parties accessing the data who previously did not have access. Additionally, the project wishes to scrutinise the proposed platform through a privacy impact assessment to ensure further design and development of the solution minimises any potential privacy concerns and can provide a secure and controlled way for research to be conducted safely, ethically and legally.

In summary, the PIA will:

1. Document the data flows of the EVD Platform in terms of what data is being processed, where it is coming from and where it is going to.
2. Identify the risks to individual privacy in terms of security and as potential threats to confidentiality, integrity or availability.
3. Clarify the legal basis (and requirements) for data processing and data sharing
4. Identify and evaluate the privacy risks
5. Provide recommendations to mitigate and manage these risks

Purposes for Data Processing

The EVD Data Sharing Platform Initiative processes data for three purposes:

1. De-identification* of contributed Ebola data;
2. Curation of data into a standardised form;
3. Managing access requests and sharing de-identified Ebola datasets;

Organisations who agree to the Terms of Submission and who have data relating to Ebola Virus Disease will be able to submit their datasets to the EVD platform. The platform will facilitate de-identification and curation of these data into a standardised Ebola database. The initiative also provides a governance process to manage and control research requests to access the Ebola database.

*Please see the Glossary at the very end of this document for definitions of this and other information governance terminology.

Data Collected

The data collected by the EVD Platform comes from organisations who have provided interventions to patients suspected of Ebola in West Africa. This includes hospitals, treatment centres, community centres, clinical trials units or other structures mobilised for this emergency. The data is a mix of information from health records, clinical trial and epidemiological data. See Appendix 2 for a data dictionary illustrating all the types of data being collected.

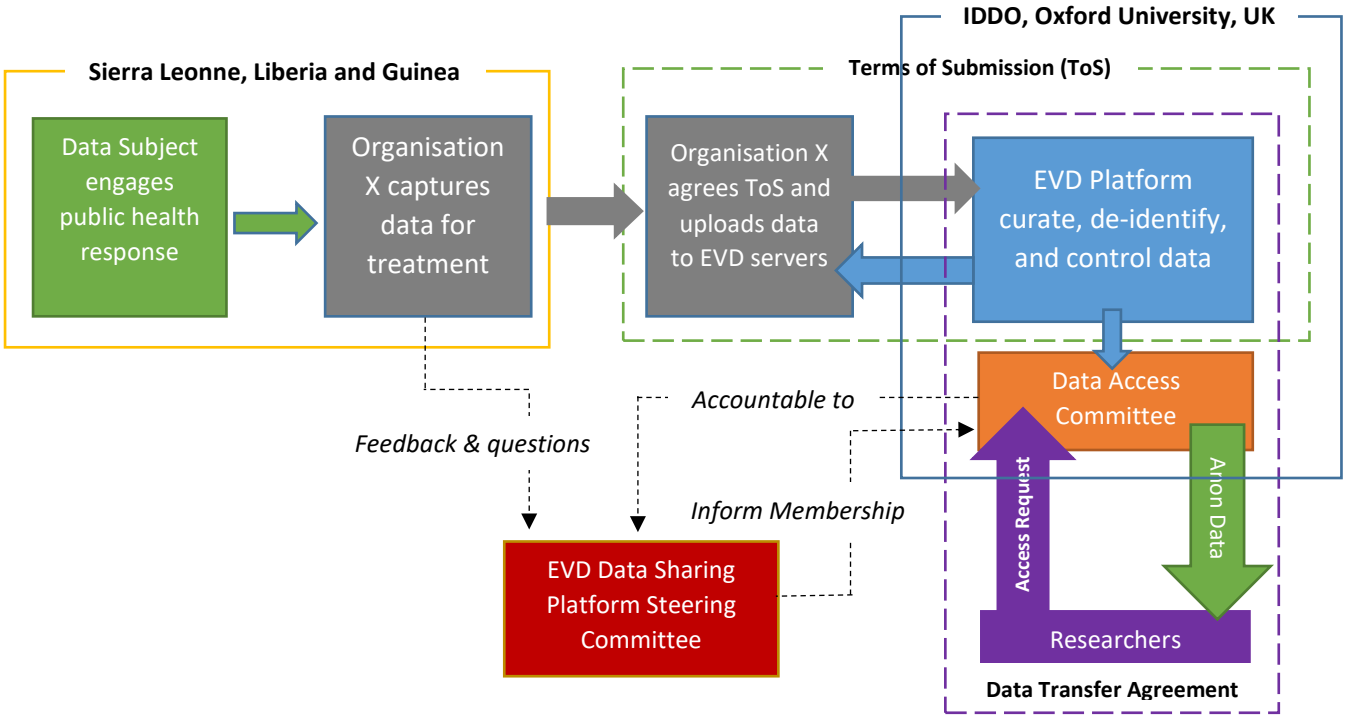
Privacy Risk 1

Potential risk that data being released could lead to subject identification on its own or through combining data with auxiliary information (known as data intrusion, the mosaic effect, or Jigsaw identification). A risk analysis can be carried out on the data types to assess the likeliness of this occurring and an acceptable risk threshold. Appropriate statistical disclosure controls can then be applied to datasets that do not fall within these limits to reduce any potential risk.

Data Flows in the EVD Platform

The data relates to subjects in Ebola affected states and has been gathered by various organisations (e.g. NGOs, academic groups, charities, military structures and government). These organisations can then submit this data to the Ebola platform after agreeing the Terms of Submission (ToS). The EVD Platform, is hosted by IDDO at the University of Oxford and all data processing takes place over the University network of servers and computer terminals. These machines are based in Oxford, United Kingdom. Data is curated and de-identified before being mapped into a new Ebola Database. Legitimate researchers can then request access to data held in the database to carry out related studies. The flow diagram below outlines at a high level the flow of data.

Figure 1 High Level Summary of Data Flow, Participants and Agreements

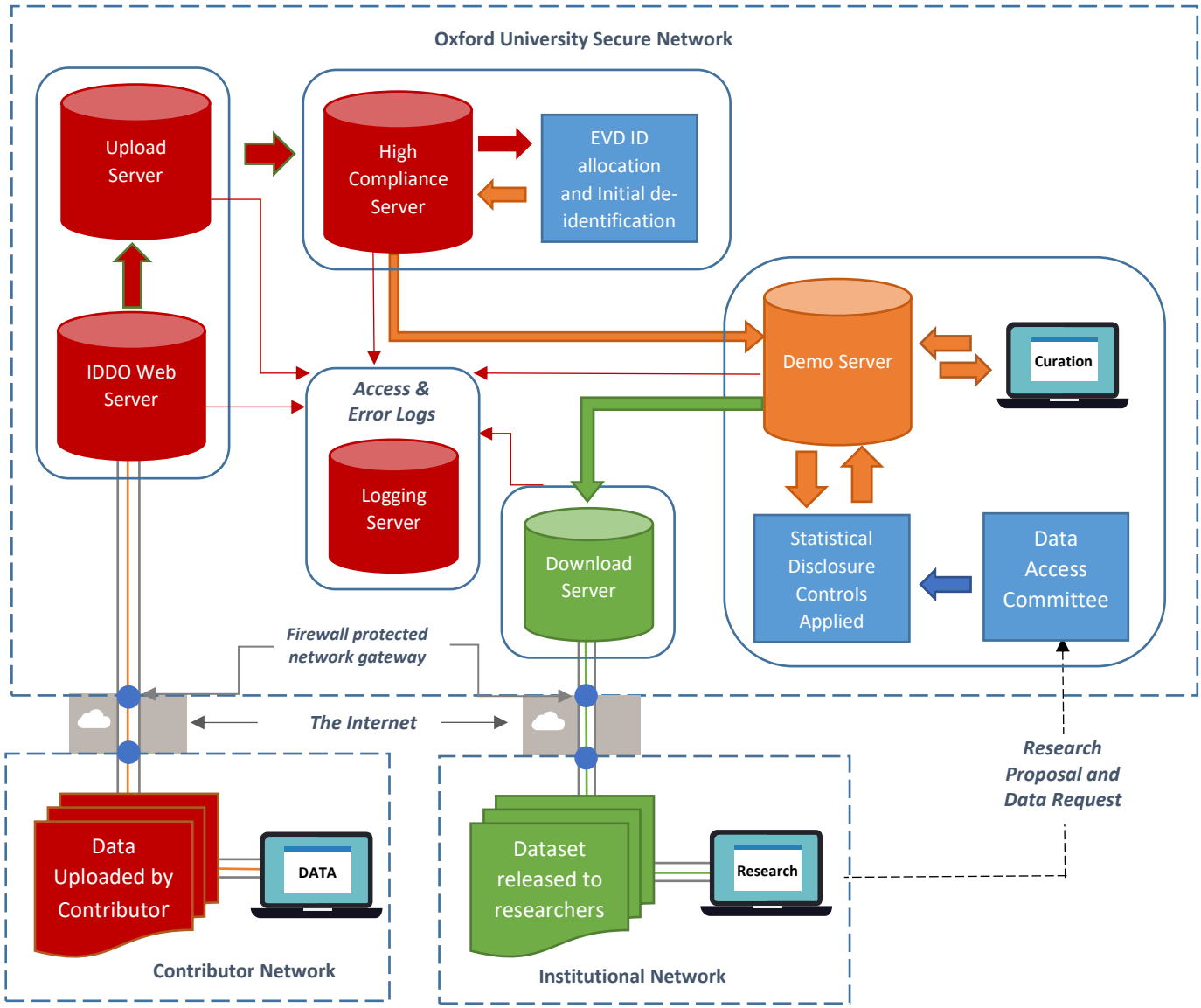


Data States and flow

The table below and Figure 2 provide further detailed information on data flows and the state the data is in at each point of processing.

	**RED**: This data **DOES** identify a person. This data would be utilised only for the purpose in which consent of the data subject was given or under certain data protection exceptions for data processing.
	**AMBER**: This data **MIGHT** allow the identification of a person if the person is well known to the viewer or is linked to further data, but processing is lawful and identification is prevented by contract and technical means to avoid a privacy breach.
	**GREEN**: This data **CANNOT** identify a person and has been de-identified in accordance with a recognised de-identification process and/or statistical disclosure control(s) applied.

Figure 2 Data Flow in the EVD Platform



Data Processing in the EVD Platform

Data is uploaded by a contributor to the EVD secure upload server. This file may contain personal data including protected health identifiers (PHI). Data is encrypted using a 256bit private and public key pair for the upload and once delivered to the upload server it is transferred to the '*High Compliance Server*'. When data is encrypted it is no longer considered identifiable as if intercepted the data would be meaningless requiring decryption to be identifiable again. When data is at rest i.e. being stored, it is encrypted, i.e. on the server. When it is being processed/accessed it is decrypted and identifiable. The data is then decrypted on the High Compliance Server by a data team member who will then apply the 'anonymisation standard operating procedure'. This involves removal of 16 key identifiers based on the Safe Harbor method. Details of the identifiers removed can be found in appendix 3. The data managers carry out initial data cleansing to look for any errors or internal discrepancies. The data manager uses Microsoft Excel or OpenRefine for this initial activity and any issues or clarifications are sought directly from the data contributor. The de-identified data is then transferred securely to the *Demo Server* for storage and management, and securely deleted from the High Compliance Server.

The curated dataset is downloaded to a secure network drive (restricted access to authorised users only) from the *Demo server* (trained, authorised staff only – username and password access) and worked on via an Oxford computer terminal. Any new data types and tables are added to a master data dictionary alongside any required definitions, validations, and controlled vocabularies. Following this, using STATA, a data analysis and statistical software application, a .do file for each table in the master data dictionary is created. For each table, a mapping process takes place to tell STATA which columns in the contributed dataset match which columns in the table being worked on. With the mappings in place, STATA can populate the table columns with data from the corresponding columns in the contributed dataset. Data validations and cleaning is done during this process, with all actions logged to provide an audit trail of changes to the original source data. Once all tables for a given data submission are populated it is uploaded back to the Demo Server, and the merged tables form the IDDO Ebola Database.

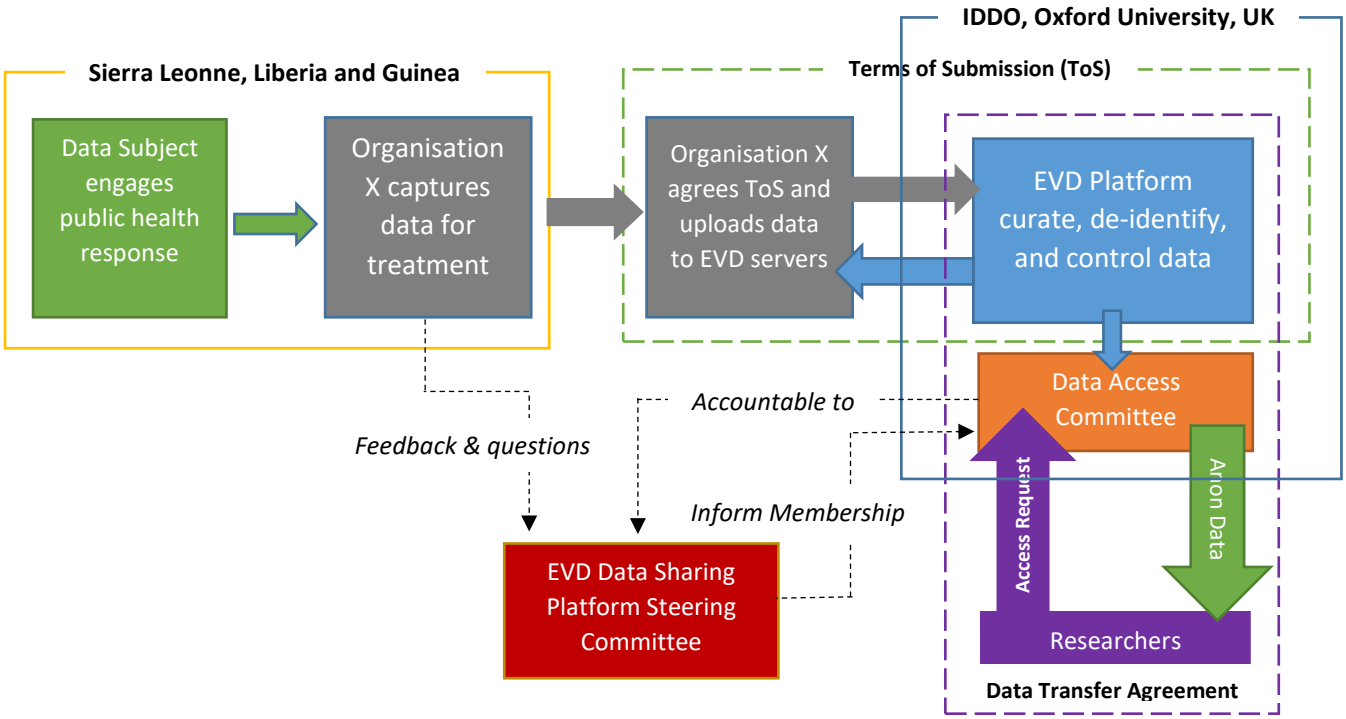
Researchers can request access to de-identified data via the Data Access Committee. The committee will review the submitted research proposal and decide on releasing data along with any further de-identification/statistical disclosure controls that might be needed before release. A Data Transfer Agreement will also need to be signed by both the researcher/responsible person on behalf of their institution and the nominated representative of the EVD Platform/University of Oxford.

Privacy Risk 2

Use of language particularly 'Anonymisation'. The word 'anonymise' has variable meanings across different sectors and jurisdictions. It can easily create confusion and be misinterpreted. A consistent use of language is very important when dealing with sensitive personal data. In the example of the EVD Platform, the data being processed contains personal information and at the earliest stage numerous identifiers are removed. Even after this removal or 'de-identification' the subjects can be re-identified due to the EVD Platform holding access to a link between a unique identifier and the original personal data. Therefore, in the hands of the EVD Platform it is always **personal data**. For those who cannot access the key or link, and have no other information to facilitate re-identification, it is **anonymous data**. See the glossary for further information on terminology. Efforts should be made in project documentation to ensure consistency in the use of language and make clear what state the data is in and the appropriate controls are in place to handle this data.

Legal Basis for Data Processing and Data Sharing

To clarify the legal basis of the EVD platform it is important to understand the role of each participant in the movement of data and the context of what the project is trying to achieve. The diagram below outlines the parties involved and the current agreements in place to manage the transfer of data.



The data being processed relates to citizens from Sierra Leone, Guinea and Liberia (the *data subjects*). The organisations who gathered the data are institutions from various parts of the world operating as non-governmental organisations, government ministries, public health agencies, charities and Universities. The EVD Platform (where data is processed) is hosted by The University of Oxford in the United Kingdom. Figure 1 outlines the relationships between these parties. The platform has taken the approach that due to much of the data being taken without consent for research (only for treatment), that it will apply techniques to de-identify the data and render it in a state whereby an individual is not identifiable. The terms of submitting data to EVD and subsequent processing of data into a de-identified state for release to approved researchers need to satisfy any relevant data protection law in the country of operation.

Basis for Submission and Responsibility of the Data Controller

At the time of writing Sierra Leone and Liberia do not have any specific domestic laws relating to personal data protection[1]. Each country (including Guinea) does make several provisions in their domestic *Telecommunication Acts* to protecting the right to privacy and personal data[2,3,4] and are all signatories to the Universal Declaration of Human Rights, an international agreement and cornerstone of international human rights law, protecting the individual right to privacy[5]. There is evidence however that the data protection landscape is changing, and Liberia are potentially in the process of transposing elements of the sub-regional Supplementary Act A/SA.1/01/10 on Personal Data Protection within ECOWAS (2010) [6] into domestic legislation[7]. Sierra Leone have also signed up to the

[1] Data protection regulations and international data flows: Implications for trade and development (2016)
[2] The Telecommunications Act of Liberia (2007), Article 3 (g) and Article 51 (2)
[3] The Telecommunications Act of Sierra Leone (2006), Article 37 (c) and Article 47
[4] The Telecommunications Act of Guinea (1995), Article 42
[5] United Nations Universal Declaration of Human Rights
[6] Supplementary Act A/SA.1/01/10 on Personal Data Protection within ECOWAS (2010)
[7] International Telecommunications Union (2014): Cyber wellness Profile Liberia

ECOWAS agreement alongside the African Union Convention on Cyber Security and Personal Data Protection[8] (AU CCSPDP). Both these regional agreements aim to support the creation of legal frameworks for cyber security and personal data protection, and whilst do not constitute domestic legislation, represent the direction of travel and a collective agreement (and commitment) of the countries involved to harmonise cyber security and personal data protection legislation at national and regional levels to a set of minimum standards, and to align these with the international community (there are many similarities with the UK Data Protection Act[9] and the European Union's General Data Protection regulation[10]). The Republic of Guinea, have moved this a step further and have very recently passed a law on Cyber Security and Personal Data Protection (RG CSPDP)[11]. Therefore, this assessment will use the Cyber security law in Guinea, ECOWAS PDP and AU CCSPDP agreements to clarify lawful basis of submission and any compliance requirements for the processing of citizen data from Sierra Leone, Liberia and Guinea in the EVD platform.

To help clarify the basis for submission it is important to define what is involved and the role of each participant in the data flow. The data being submitted to the EVD Platform relates to health record information, clinical trial and epidemiological data collected from citizens of Sierra Leone, Guinea and Liberia, during the Ebola outbreak response (in the context of a public health emergency as declared by the World Health Organisation). This type of information is defined as '*Personal Data*' (and 'Sensitive Data' by the ECOWAS and African Union) by the different agreements and legislation*:*

Supplementary Act A/SA.1/01/10 on Personal Data Protection within ECOWAS (2010):

Personal Data

Any information relating to an identified individual or who maybe directly or indirectly identifiable by reference to an identification number or one of several elements related to their physical, physiological, genetic, psychological, cultural, societal, or economic identity;

Sensitive Data

*Means **personal data** relating to religious, philosophical, political, trade union opinions or activities, to his sexual life, racial origin or health, relating to social measures, proceedings and criminal or administrative sanctions.*

African Union Convention on Cyber Security and Personal Data Protection (2014):

Personal Data

Any information relating to an identified or identifiable natural person by which this person can be identified, directly or indirectly in particular by reference to an identification number or to one or more factors specific to his/her physical, physiological, mental, economic, cultural or social identity;

Sensitive Data

*Means all **personal data** relating to religious, philosophical, political and trade union opinions and activities, as well as to sex life or race, health, social measures, legal proceedings and penal or administrative functions.*

[8] African Union Convention on Cyber Security and Personal Data Protection (2014)
[9] UK Legislation: Data Protection Act (1998)
[10] European Union General Data Protection Regulation (2016)
[11] Law on Cyber Security and Data Protection (2016)

Republic of Guinea Law on Cyber Security and Data Protection (2016)

Personal Data

Any information of any kind and regardless of medium, including sound and image, relating to an identified or identifiable natural person directly or indirectly, by reference to an identification number or to one or more factors specific to his physical, physiological, genetic, mental, cultural, social or economic;

The data has been collected and submitted under the *Terms of Submission* by Organisation X to the University of Oxford (the EVD Platform) for processing. Based on the definitions provided in the different agreements, Organisation X constitutes a '*Data Controller*':

Supplementary Act A/SA.1/01/10 on Personal Data Protection within ECOWAS (2010):

Data Controller

Any public or private individual or legal entity, body or association, who alone, or jointly with others decides to collect and process personal data and determines the purposes for which such data are processed;

African Union Convention on Cyber Security and Personal Data Protection (2014):

Data Controller

Any natural or legal person, public or private, any other organisation or association which alone or jointly with others, decided to collect and process personal data and determines the purposes;

Republic of Guinea Law on Cyber Security and Data Protection (2016)

The (Data) Controller

the person or entity, public or private person or any other organization or association which alone or jointly with others, decides to collect and process personal data, and determines the purposes;

The University of Oxford processes data into a standardised format (via a set of curation processes) and removes personal identifiers to render the data anonymous. Processing is carried out under the terms of submission agreement between Organisation X and the University. Based on the definitions from the regional agreements and legislation the University of Oxford is a *'data processor'*:

Supplementary Act A/SA.1/01/10 on Personal Data Protection within ECOWAS (2010):

Data Processor

Any public or private individual or legal entity, body or association who processes data on behalf of the data controller;

African Union Convention on Cyber Security and Personal Data Protection (2014):

Sub-Contractor (Data Processor)

Any natural or legal person, public or private, any other organization or association, that processes data on behalf of the data controller;

Republic of Guinea Law on Cyber Security and Data Protection (2016)

Subcontractor (Data Processor)

Any natural or legal person, public or private, or any other organization or association, which deals with personal data on behalf of the controller of the data.

This assessment will consider both the responsibility of the EVD Platform (the University of Oxford) and Organisation X (a contributor). This is because, crucial to the EVD Platform achieving its aims will be to ensure contributors can readily provide information. Therefore, providing all necessary support and information for those contributors to discharge their responsibilities as data controllers, will enable the EVD Platform to better achieve its aims and operate in an ethically responsible manner.

Responsibilities as a Data Controller

Article 5 of the RG CSPDP, Article 10 of the AU CCSPDP and Article 6 of the ECOWAS agreements outline the required formalities for processing personal data of citizens (the *data subjects)*. For purposes, such as processing personal data for health research, which is the aim of submitting data to the EVD Platform, requires authorisation from the National Protection Authority (see RG CSPDP, Article 3, AU CCSPDP, Article 10, paragraph 4(a) and ECOWAS PDP Article 5). In addition, if a data controller wishes to transfer personal data to another country for processing, authority must be requested from the National Protection Authority (RG CSPDP Article 28, AU CCSPDP Article 14, paragraph 6(b) and ECOWAS PDP Article 36, paragraph 2). At the time of writing, this assessment could not identify a National Protection Authority within Sierra Leone nor Liberia. The Republic of Guinea has legislation in place to create a data protection authority but an authority was not found in this review. The legislation provides information on a *National Centre of Information Systems Security*, and it would be important to potentially signpost contributors towards this body to seek information and advice in this area prior to contribution.

Privacy Risk 3

The agreement between the contributor (data controller) and the EVD Platform (data processor) should make clear that any necessary authority or approval has been acquired to have data processed by the platform. Signposting to any known information would support controllers in fulfilling this obligation e.g. Guinea's *National Centre of Information Security*. This is a key responsibility however of the data controller, and particularly relates to the terms of the contracts/agreements they make with data processors/subcontractors. In terms of the data protection landscape in all countries contributing, it will be important to keep up to date with any changes regarding domestic law. Terms of Submission should be reviewed as legislation is approved in this area and national protection authorities established. This is to ensure contributors can readily use the EVD Platform for data processing, and all processing that takes place, complies with domestic law. It is therefore recommended this Privacy Impact Assessment is repeated on an annual basis.

Following authority for processing there are further rules around processing data in non-ECOWAS countries. Under Article 36, paragraph 1 of ECOWAS PDP, Article 14, paragraph 6 (a) of the AU CCSPDP and Article 28 of the RG CSPDP legislation, the data controller transferring data to another country must ensure a higher or equivalent level of protection of the privacy, freedoms and fundamental rights of persons whose data are being processed. Therefore, the organisations submitting data to the EVD Platform must be satisfied with the data security model the EVD Platform has in place to process personal data.

Privacy Risk 4

Whilst this is the responsibility of the institution submitting data, to support contribution to the EVD Platform, a full data security model should be made available to contributors detailing the end to end process and all technical and procedural controls in place. This will ensure the IDDO EVD

Platform/University are operating transparently and providing contributors all the information they require to make a decision. Additionally, this will ensure the EVD platform is meeting its obligations as a data processor (detailed later in this assessment).

In all agreements and legislation, the AU CCSPDP, ECOWAS PDP and RG CSPDP outline a series of principles and articles for guiding the processing personal data. The language used is very similar across each document and will only be defined separately where there is a significant difference in wording.

Principle 1/Article 13/Article 18: Principle of consent and legitimacy of personal data processing

The first principle relates to data being processed only when a subject has given consent for that purpose. This requirement however may be waived where the processing is necessary for certain activities. In the case of a data controller submitting data to the EVD Platform and the platform itself, the requirement for consent can be waived based on the exceptions in AU CCSPDP Article 14, Paragraph 1(b), ECOWAS PDP, Article 23, Paragraph 2(b) and the RG CSPDP, Chapter VIII, Article 18(b):

> *14, 1 (b): Performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or in a third party to whom the data are disclosed;*
>
> *23, 2 (b): For the implementation of a public interest mission or relevant to the exercise of public authority that is vested in the data controller or the third party whom the data is disclosed;*
>
> *18, (b): or for the performance of a mission of public interest or in the exercise of the public authority vested in the data controller or the third parties to whom the data are disclosed;*

For clarity, this exception applies to data controllers processing data or whom they subcontract/agree to process data on their behalf – a *data processor*. The data cannot be released to a recipient on this basis. In the case of the EVD Platform, the agreement between the controller and processor is to de-identify the data to render it anonymous to any potential recipient, and therefore when released it is no longer personal data and these protections no longer apply. Further detail is outlined later in the assessment.

Principle 2/Article 24/Article 19: Principle of lawfulness and fairness of personal data processing

The second principle relates to the collection, recording, processing, storage and transmission of personal data being undertaken in a lawful, fair and non-fraudulent manner. In the case of the EVD Platform, data controllers submitting data need to satisfy themselves that they have the authority to submit the data, the principles for processing are satisfied and the platform has the necessary security controls in place to protect the privacy and rights of the data subjects.

Principle 3/Article 25/Article 20: Principle of purpose, relevance and storage of processed personal data

Under all agreements and legislation four specific areas are outlined regarding this processing principle:

> *1) Personal data collected shall be obtained for specific, explicit and lawful purposes and not be used in a manner incompatible with such purpose;*
>
> *2) Data collection shall be adequate, relevant and not excessive in relation to the purposes for which they are collected and further processed;*

3) Data shall be kept for no longer than is necessary for the purposes for which the data were collected or further processed;

4) Beyond the required period, data may be stored only for the specific needs of data processing undertaken for historical, statistical or research purposes under the law.

The data controller submitting data must satisfy themselves these four areas of principle 3 are being met, whilst as referenced in Privacy Risk 4, the EVD Platform must provide specific and transparent details on the agreed purpose for processing, data to be collected, retention periods, processing involved and any further intended purpose for the data. This should be documented in a full data security model and made available to data contributors.

Principle 4/Article 26/Article 21: Principle of accuracy of personal data

This principle refers to data that is collected should be accurate and, where necessary, kept up to date. Every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified.

The data controller will need to agree the processes the EVD Platform has in place are suitable and can support it to comply with this principle. The platform curates' data submitted to correct for any mistakes or inaccuracies where possible and practical, following a standard operating procedure for curating data. Details of this process should be made available to contributors in a published security document as reference in Privacy Risk 4.

Principle 5/Article 27/Article 22: Principle of transparency of personal data processing

The principle of transparency requires the data controller to disclose information on personal data processing. The agreements do not define if information disclosure is required only on request or this must be communicated from when processing begins. Pragmatically, a notice on the contributor's public facing website that they are contributing data to the EVD Platform and a mechanism for individuals to find out more would fulfil either interpretation of the principle. The EVD Platform has a public facing website.

Privacy Risk 5

A notice should be provided informing the contributor of this principle and the responsibility they have when submitting data. The IDDO EVD Platform could also provide this and more detailed information on its public website that contributors could then reference.

Principle 6/Article 28 & 29/Article 23 & 24: Principle of confidentiality and security of personal data processing

a) Personal data shall be processed confidentially and protected, in particular where the processing involves transmission of the data over a network;

b) Where processing is undertaken on behalf of a controller, the latter shall choose a processor providing sufficient guarantees. It is incumbent on the controller and processor to ensure compliance with the security measures defined in this Convention.

Under Principle 6 the data controller must satisfy themselves adequate protections are in place to protect confidentiality and that processing of data occurs in line with the principles of processing and domestic law. The Privacy Impact Assessment is a process that will assess the security and controls in place for the EVD Platform and provide any recommendations to ensure compliance and good

practice. See *Privacy Risk 4, 5*, 6 and 7 for action to support contributors in fulfilling due diligence regarding this principle.

Article 14/Article 31: Specific Principles for Processing Sensitive Data

The RG CSPDP does not set out specific principles for processing sensitive data. The ECOWAS and African Union agreements however, set out rules relating to processing of sensitive data (of which health data falls within) and specifically highlights the following:

> *State Parties shall undertake to prohibit any data collection and processing revealing racial, ethnic and regional origin, parental filiation, political opinions, religious or philosophical beliefs, trade union membership, sex life and genetic information or, more generally, data on the state of health of the data subject.*

However, both agreements provide an exception to this for the purposes of submitting data to the EVD Platform to carry out health research:

> *AU CCSPDP, Article 14, paragraph 2(f): Processing is necessary in the public interest, especially for historical, statistical or scientific purposes;*
>
> *ECOWAS PDP, Article 31, paragraph 6: Processing is necessary in the public interest, especially for historical, statistical or scientific purposes;*

In addition, all agreements and legislation make an exception under the following articles:

> ***AU CCSPDP Article 14, paragraph 4/ECOWAS PDP Article 32/RG CSPDP Article 25:***
>
> *Personal data processing for journalistic purposes or for **the purpose of research** or artistic or literary expression **shall be acceptable** where the processing is solely for literary and artistic expression or **for professional exercise of** journalistic or **research activity**, **in accordance with the code of conduct of these professions**.*

Privacy Risk 6

To support compliance with these exceptions, the EVD Platform should have the project and its security model reviewed by an appropriate ethics board(s) to provide feedback on the project and the proposed security framework. The outcome should be published on the public facing website that the project has been reviewed, the outcome, the body and when it took place.

Data Subjects Rights

All three agreements (African Union, ECOWAS and Guinea) provide consistent definition on the data rights of data subjects. Each are provided and reviewed below:

Article 16/Article 38/Article 30: Right to Information: *The data controller shall provide the natural person whose data are to be processed with the following information, no later than the time when the data are collected, and regardless of the means and facilities used, with the following information:*

> *a) His/her identity and of his/her representative, if any;*
>
> *b) The purposes of the processing for which the data are intended;*
>
> *c) Categories of data involved;*
>
> *d) Recipient(s) to which the data might be disclosed;*
>
> *e) The capacity to request to be removed from the file;*
>
> *f) Existence of the right of access to and the right to rectify the data concerning him/her;*

g) Period for which data are stored;

h) Proposed transfers of data to third countries.

Whilst the contributor will have likely provided this information when data was collected for treatment purposes, personal data is now being processed for a different purpose (health research). There is limited information in the agreements on what fair or good practice looks like but a data controller in this context has a responsibility to make sure this information is publicly available. Currently the EVD Platform has a public facing webpage with some information relating to their identity and the purpose of processing. In addition, an inventory of the types and categories of data available on the EVD Platform will be available via the data inventory on the public facing website. It is unknown if data contributors provide any information independently, but they will be advised to link their own information regarding the platform to the EVD Platform website.

To support compliance of Article 16, 38 and 30 a privacy notice placed on the EVD Platform website to explain each of these outlined areas that contributors could reference would be supportive (see Privacy Risk 5). Making contributors aware as well that it would be necessary to place a notice on their public facing website to provide the required information about sharing data with EVD (and a link to EVD to find out more) would support compliance.

Article 17/Article 39/Article 31: Right to Access: *Any natural person whose personal data are to be processed may request from the controller, in the form of questions, the following:*

a) Such information as would enable him/her to evaluate and object to the processing;

b) Confirmation as to whether or not data relating to him/her are being processed;

c) Communication to him/her of the personal data undergoing processing and any available information as to their source;

d) Information as to the purpose of the processing, the categories of personal data concerned, and the recipients or categories of recipients to whom the data are disclosed.

Article 18/Article 40/Article 32: Right to Object: *Any natural person has the right to object, on legitimate grounds, to the processing of the data relating to him/her. He/she shall have the right to be informed before personal data relating to him/her are disclosed for the first time to third parties or used on their behalf for the purposes of marketing, and to be expressly offered the right to object, free of charge, to such disclosures or uses.*

Article 19/Article 41/Article 33: Right of rectification or Erasure: *Any natural person may demand that the data controller rectify, complete, update, block or erase, as the case may be, the personal data concerning him/her where such data are inaccurate, incomplete, equivocal or out of date, or whose collection, use, disclosure or storage are prohibited.*

Article 34 of the RG CSPDP: adds further detail that those legitimately related to data subjects who are deceased can ask the data controller to update their information with the relevant information.

Article 35 of the RG CSPDP adds further reasons for a subject to request information, or removal of personal data:

- *Data is no longer needed for the purpose under which the data was collected or processed;*
- *The person concerned by the treatment of their personal data withdrew consent on which collection and processing is based, the retention period has expired, and there is no other legal grounds to processing of the data;*

- *The person concerned by the treatment of the personal data where there is no legal ground for data processing.*
- *The treatment of the personal data is not in line with the provisions of this Act; or*
- *For any other legitimate reason*

Article 36 of the RG CSPDP: *If personal data is made public, the data controller responsible for the data must take all reasonable, including technical measures to inform third parties dealing with such data of the person affected to delete all links to these given person(s) or any copy or reproduction of these.*

(The de-identification process being used by the EVD Platform would fulfil this obligation for the data controller).

Article 37 of the RG CSPDP: *The controller must proceed without delay to remove data (as requested) except where conservation of personal data is necessary:*

- *To the exercise of the right to freedom of expression*
- *Is for reasons of general interest in the field of public health, in accordance with law*
- *In compliance with a legal obligation to the personal data provided under this Act or any other legislation governing the personal data in the Republic of Guinea, and which the controller of said data is submitted.*

Article 38 of the RG CSPDP: *The responsible of personal data processing must put in place appropriate mechanisms to ensure the implementation of respect for the right to be forgotten and erasing such data.*

- *The controller must consider periodically the need or not to keep the personal data, according to the provisions of this Act.*
- *When erasure of personal data is carried out, must not proceed with any other processing of this data.*

Article 39 of the RG CSPDP: *The Data Protection Authority adopt measures and guidelines, for the purposes of state:*

- *The conditions of the removal of links to these personal data, copies or reproductions of existing publicly accessible electronic communications*
- *The conditions and criteria for the limitation of the treatment of personal data*

Article 40 of the RG CSPDP: *Where the personal data that has been processed into a structured format commonly used, the person concerned by the data processing has the right to obtain from the person responsible for processing such data, a copy of the data in the commonly used structured electronic format to allow the re-use of these data by the person concerned.*

When the person responsible for data processing has provided data, and that the processing is based on sound prior consent or contact, (the subject) has the right to pass these personal data and any other information provided and stored by a system to another system in the electronic format commonly used without the controller.

The Data Protection Authority may specify the electronic format, as well as technical standards, modalities and procedures for the transmission of personal data.

These rights relate to obligations the data controller would need to fulfil. To do so, and to allow the data controller to use the EVD Platform, information relating to how the platform would fulfil any request relating to the *Right to Access*, the *Right to Object* and the *Right to Rectification or Erasure* should be made available to the data controller. The platform has a documented standard operating

procedure in place to handle these requests and should share this information with the data controller by adding it to a published information security document (see Privacy Risk 4). Contributors should also include the required information through a privacy notice as recommended in Privacy Risk 5.

Obligations of the Data Controller

The agreements set out consistent obligations for a data controller and cover four separate areas as follows (in the order of African Union, ECOWAS and Guinea):

> ***Article 20/Article 42/Article 41: Confidentiality Obligations***
>
> *Processing of personal data shall be confidential. Such processing shall be undertaken solely by persons operating under the authority of a data controller and only on instructions from the controller.*
>
> ***Article 21/Article 43/Article 43: Security Obligations***
>
> *The data controller must take all appropriate precautions, according to the nature of the data, and in particular, to prevent such data from being altered or destroyed, or accessed by unauthorized third parties.*
>
> ***Additional detail in the Republic of Guinea Article 43:*** *The controller must implement all technical measures and organizing appropriate, in order to ensure the protection of the personal data it processes against the destruction, accidental or illegal, accidental loss, alteration, dissemination or access unauthorized data, especially when the data management includes data transmissions in a network, as well as against other forms of unlawful treatment.*
>
> ***Republic of Guinea, Article 44:*** *The controller of personal data is required to prepare an annual report to the attention of the Data Protection Authority with information respecting their compliance with the provisions in this Act.*
>
> ***Article 22/Article 44/Article 45: Storage/Preservation Obligations***
>
> *Personal data shall be kept for no longer than is necessary for the purposes for which the data were collected or processed*
>
> ***Article 23/Article 45/Article 46: Sustainability/Durability Obligations***
>
> *1) The data controller shall take all appropriate measures to ensure that processed personal data can be utilized regardless of the technical device employed in the process.*
>
> *2) The processing official shall, in particular, ensure that technological changes do not constitute an obstacle to the said utilization.*

These reiterate the obligations in the six principles for processing personal data outlined earlier. The data controllers contributing data must satisfy themselves the EVD Platform (the data processor) has in place all technical and procedural measures to secure the data being contributed, to maintain confidentiality and protect the privacy of data subjects. They must be confident the security model in place will prevent any unauthorised third parties access to personal data, there will be no unauthorised use of the data (other than which it has instructed for processing) and appropriate controls are employed throughout the various stages of processing, including reasonable mechanisms to ensure compliance (e.g. audit and reporting processes). As the data processor, the EVD Platform should ensure this detail is in the security model (see privacy risk 4).

Responsibilities as a Data Processor (the EVD Platform and Oxford University)

Under the definitions provided in the West African agreements and Guinea legislation, the University acts as a 'data processor'. Therefore, it is the responsibility of the data controller to satisfy themselves the data processor has all the necessary security arrangements in place to process sensitive personal data in accordance with data protection law and these conditions are reflected in the agreement between parties (the *Terms of Submission*). These terms and conditions operate effectively as a data processing agreement between the data controller and data processor. As a data processor, the University must abide by these terms and conditions for processing sensitive personal data. Any diversion from these, unauthorised use or unauthorised release of data would be a breach of the agreement, and the data controller entitled to take action against the data processor (as defined in the agreement between parties).

The University of Oxford is based in the United Kingdom and has a legal obligation to follow domestic data protection law for processing personal data. As the EVD Platform does not collect any further participant data, and the Terms of Submission define the purpose and manner of processing, the platform operates as a data processor under the UK Data Protection Act (1998)[12]. Data Processors under the data protection act have no direct liability[13] and the data controller must ensure the proposed security and processing arrangements meet their requirements (and reflect these in data processing contract/agreements). However, it must support the data controller in discharging its responsibilities and act only under instruction from the data controller and in compliance with the written agreement between the parties. The incoming General Data Protection Regulation in May 2018 will also add significantly more responsibility to data processors, and it will be possible for a regulator and data subject to bring actions directly against data processors[14]. Therefore, the EVD Platform will need to ensure it continues to review its security arrangements so they are in line with current data protection legislation, policy and practice and can continue to meet its obligations as a data processor. The EVD Platform Secretariat have a communications team who are well placed to support Data Controllers to remain aware of the current requirements and best practice in data protection legislation and policy.

Security Standards and UK Policy on Processing International Data

Specific standards for data security are not outlined in the agreements and legislation. Only that it is the obligation of a data controller to ensure these are appropriate for the personal data being processed. Elements of this detail would often come at a domestic level and be provided by a regulatory body (a data protection authority e.g. Information Commissioners Office) or government agency (such as the National Cyber Security Centre in the UK) in the form of guidelines. There are standards however, provided by the International Standards Organisation (ISO) that are globally recognised (membership across 193 countries) alongside national organisations leading in this area such as the National Institute of Standards and Technology (NIST). Additionally, the UK Government has produced a Security Policy Framework for processing international personal data[15][16]. This is primarily to guide how government departments and agencies process international data and whilst the University is an educational charity and not bound by this policy the guidance sets out a principle of reciprocity that data will be secured to 'at least' the same standard as UK data:

> *Where specific reciprocal security agreements / arrangements are in place with foreign governments or international organisations, equivalent protections and markings must be*

[12] http://www.legislation.gov.uk/ukpga/1998/29/contents
[13] https://ico.org.uk/media/1546/data-controllers-and-data-processors-dp-guidance.pdf
[14] http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:2016:119:FULL&from=EN
[15] http://www.eurim.org.uk/activities/ig/idg/SecurityPolicyFramework.pdf
[16] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/251480/Government-Security-Classifications-April-2014.pdf

recognised and any information received must be handled with AT LEAST the same degree of protection as if it were UK information of equivalent classification.

The EVD Platform is processing data subject health information, which is marked as OFFICIAL under the UK Government classification. The National Cyber Security Centre and the Health and Social Care Information Centre (HSCIC – now NHS Digital) provides guidance in the UK for how to protect data of this marking (personal and health information). These along with internationally recognised standards such as ISO, NIST and Federal Information Processing standards (FIPS) provide a good benchmark to assess the current technical security controls in place for data transfer, storage, access, availability, and removal.

Platform Hosted Environment

The Platform is hosted on a virtual datacentre, provisioned through the Oxford University Private Cloud, a highly available, enterprise grade environment with hardware physically located in Oxford, United Kingdom[17]. This utilises vCloud Air technology by VMware, a secure and robust infrastructure platform technology, which is set up and mirrored across two sites (South Parks Road and Banbury Road). In the case of a disaster, or service disruption it can failover to the alternative environment[18]. Oxford University and VMware have an established Information Security Management System (ISMS) based on ISO 27001 standards to manage risks relating to confidentiality, integrity, and availability of information[19][20]. The technology in use is a robust choice and has achieved a number of cloud hosting accreditations including the Cloud Security Alliance (CSA) Cloud Controls Matrix (CCM) v3.0 and SSAE-16 / ISAE 3402[21].

Data in Transit

IDDO Client to Web/Upload Server: Data Transfer using HyperText Transfer Protocol Secure (HTTPS). The connection on the website (IDDO.ORG) is encrypted and authenticated using a Transport Layer Security (TLS), a strong key exchange (ECDHE_RSA with P-256), and a strong cipher (AES_128_GCM).
Web/Upload Server to High Compliance Server: Data is transferred between servers using a Linux command known as 'rsync'[22]. Before transfer the source data is encrypted using a hybrid cryptography approach, combining symmetric and asymmetric key algorithms. An AES 256 compliant symmetric key is generated using SecureRandom (to generate a 256-bit character string) and used with AES to encrypt the data (AES is a symmetric algorithm/cipher and therefore the key that encrypts data is also used to decrypt data). The private AES key is then encrypted using an asymmetric cipher/algorithm (RSA) and the generated public key. The private component of the RSA key is stored on the High Compliance Server and is only accessible via a controlled register of approved users. Both the data and its method of decryption (the AES 256 private key) are therefore securely encrypted. The dataset is transferred using *rsync* to the High Compliance Server and the source data deleted from the upload server and cleared from memory (deletion method described below).
High Compliance Server to Demo Server: The data is decrypted and de-identified using the adapted Safe Harbor process. Following initial curation activity, the de-identified data is then securely transferred to the Demo Server using a web browser (secured by HTTPS).
Demo Server to Oxford Desktop Terminal: To download the de-identified data for curation (and applying any statistical disclosure controls to the data) an Oxford desktop terminal is used to connect to the demo server via a web browser. Access to the server is restricted (username and password required) and only trained, authorised users can access the application. The connection is secured by

[17] http://www.it.ox.ac.uk/data-centres/university-private-cloud
[18] http://help.it.ox.ac.uk/internal/sld/private-cloud
[19] https://www.brightline.com/certificate-directory/Y8eJUUXMjLA3/
[20] https://www.infosec.ox.ac.uk/sites/default/files/Cloud%20Security%20Guidance%200.04.docx
[21] http://vcloud.vmware.com/uk/service-offering/cloud-compliance
[22] http://linuxcommand.org/man_pages/rsync1.html

a HTTPS connection with TLS authentication, strong key exchange (RSA) and using a strong cipher (AES). Initial cleansing is carried out locally using a desktop application called OpenRefine and Microsoft Excel. Further curation activities are carried out using STATA. All connections for moving data are encrypted and secured using HTTPS.

Demo Server to Download Server: Requested (anonymous) datasets are transferred to the download server using a secure transfer protocol.

Data at Rest

Server storage: Physically isolated and encrypted using hybrid cryptography approach

Oxford Network Drive and Desktop Terminal: Isolated but not encrypted (policy controls on storage and movement).

The encryption methodology being employed to secure data in transit and at rest is compliant, and exceeds (in some parts the UK standards for processing health related information. HSCIC/NHS Digital guidance recommends the use of the AES 256 cipher/algorithm, with a minimum key length of 256 bits[23]. In the EVD Platform to maintain comparable algorithm strengths between the different cryptographic classes (AES and RSA), the key generated for RSA is 15,360 bits in length exceeding the NHS guidance but falling in line with international standards laid out by National Institute of Standards and Technology (NIST)[24]. Locally, the University of Oxford also provides guidance on securing and encrypting data stored on servers and removable media. This recommends using the Federal Information Processing Published Standards 140 (FIPS PUB 140-2). This standard provides requirements for cryptographic approaches and modules, covering 11 areas, which the hybrid cryptography approach implemented demonstrates FIPS PUB 140-2 level 1 capability[25] .

The National Cyber Security Centre (NCSC) provides guidance on securing communications between applications and users (e.g. web browsers and server applications) and recommends the use of the TLS protocol (version 1.2) to secure connections[26], alongside detailed information on potential cryptographic profiles to use. The methods in use for the EVD Platform meet and in places exceed all these criteria. Policy controls are in place to restrict movement and access to the network drive used to store data for curation. Whole disk encryption could be applied to further secure the data but in the context that this information has already been de-identified (via the adapted Safe Harbor method), the controls in place are sufficient.

Privacy Risk 7

Encryption key strengths have a lifetime and therefore will require review periodically as part of ongoing privacy impact assessments.

Access Controls

Physical and Environmental Controls: Access to the building is controlled via a secure entry system and only authorised staff with a swipe card can enter the building and office space.

Network controls: Firewall controls in place to prevent unauthorised access to Oxford University network and resources (including curation network drives).

User Access Management: Servers are only accessible by approved system administrators who are registered and have been issued logon credentials. These are accessed via remote desktop and/or VPN. Desktop terminals and University network resources are accessed by approved users who have

---

23 http://webarchive.nationalarchives.gov.uk/20160729133355/http://systems.hscic.gov.uk/infogov/security/encryptionguide.pdf

24 http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r4.pdf

25 http://csrc.nist.gov/groups/STM/cmvp/documents/fips140-2/FIPS1402IG.pdf

26 https://www.ncsc.gov.uk/guidance/tls-external-facing-services

been issued with a University network username and password. Quality and ageing controls are applied to passwords.

Annex A of ISO 27001 sets out requirements for controlling and securing data[27]. These include a range of physical, legal, user and organisational methods for controlling and safeguarding information. In particular, it sets out requirements for physical, network and user based controls all of which are in use in the EVD Data Sharing Platform Initiative.

Data Retention

The IDDO EVD Platform has in place a defined retention policy. This states that personal identifiers relating to data subjects will be retained only for the purposes of linking records with any further associated data received. The retention period will expire at the point all known linked data has been received and participants records matched, or after 5 years, whichever occurs first. The personal identifier will then be securely destroyed as per the *Data Sanitisation* process. This represents a transparent and clear process and falls in line with guidance on data retention published by the Medical Research Council and the NHS which recommend a 10-20 year retention period[28][29].

Data Backup Policy

Data is backed up on a routine basis using the Oxford University backup service[30] and the NSMS services agreement[31]. The Platform has a standard operating procedure in place for carrying out backups. This includes daily, weekly and monthly backups of all critical data. Full tape backups are taken of the server file-stores and kept indefinitely.

Annex A10.5 of ISO 27001 provides a brief statement on backup controls: *"Back-up copies of information and software shall be taken and tested regularly in accordance with the agreed backup policy".* Whilst the National Institute for Standards and Technology guidance adds a focus on balancing business need, potential impact, and accounting for feasibility and cost[32]. The EVD Platform backup policy represents a robust and appropriate balancing of resource requirements against, the business need and the potential impact a loss of availability might have.

Data Sanitisation and Removal

Data on the Upload and High Compliance Servers is securely removed using a secure 'srm' command on the Linux system. This is a command line programme that overwrites, truncates and renames file contents before unlinking from a directory. By default, 'srm' uses 35 passes to overwrite data[33].

The standard recommended by the Department of Health is to use either a clearing or purging process to remove/delete data. This involves using a three or seven pass overwrite process. Typically using sequential writes of patterned data. It advises to ensure historical data is removed to make as many passes as practicable as the likelihood of total data eradication is proportional to the number of passes[34]. These processes are also reflected in the *Guidelines for Media Sanitisation* from the National Institute of Standards and technology, which based on the nature of the data, the need to reuse

[27] http://gender.govmu.org/English/Documents/activities/gender%20infsys/AnnexIX1302.pdf
[28] https://www.mrc.ac.uk/publications/browse/good-research-practice-principles-and-guidelines/
[29] https://digital.nhs.uk/media/1158/Records-Management-Code-of-Practice-for-Health-and-Social-Care-2016/pdf/Records-management-COP-HSC-2016
[30] https://help.it.ox.ac.uk/internal/sld/hfs
[31] http://help.it.ox.ac.uk/nsms/slas/smsla#Backups
[32] http://csrc.nist.gov/publications/nistpubs/800-12/handbook.pdf
[33] http://srm.sourceforge.net/srm.html
[34] Disposal & Destruction of Sensitive Data, Department of Health (2007)

storage media, and environmental factors, recommend a purge process be used, and the project is adopting[35].

Data Curation

The data submitted, following de-identification, is curated using several tools to ensure maximum scientific utility. This includes standardisation, imputing missing values and modification/identification of outliers.

The data controller contributing data to the EVD Platform will need to assure themselves from the information provided by the EVD Platform that these obligations will be met. As recommended in Privacy Risk 4, the EVD Platform will need to create a Security Model document to evidence the processes in place.

Common Law Duty of Confidence

At the time of writing this assessment could not find any available information relating to common law for data protection. In the UK, there is common law (that which is based on previous court cases, sometimes known as case law) that if information is given in circumstances where it is expected that a duty of confidence applies, the information cannot normally be disclosed without the information provider's consent[36]. This PIA has not found any similar case law in the subject countries and therefore no precedent for this project to consider. In any case, the EVD Platform removes identifiers at the earliest opportunity and all processing is carried out by staff who owe a contractual duty of confidence when dealing with sensitive information. Data is subsequently processed into a state which minimises all practical chances of an individual being identified before it is released to a recipient (see below) and there are several procedural controls to manage and govern the release of data (see Submission and Data Access Control).

De-identification and Anonymisation

One of the purposes of data processing in the EVD Platform is to de-identify the information to protect individual privacy and minimise any risks of a subject being identified when a dataset is disclosed to an approved researcher. The RG CSPDP (2106), AU CCSPDP (2014) and the ECOWAS PDP (2010) do not contain information or recommendations on processes or techniques for de-identification. Currently the platform has adopted the processes for de-identification as outlined in the Section 164.514(b), of the Privacy Rule of the Health Insurance Portability and Accountability Act (1996). It is a two-step approach, the first utilising an adapted Safe Harbor method, removing 16 (instead of 18) identifiers and the second using a process known as the *Expert Determination method* to remove any further identifiers from a dataset or apply statistical disclosure controls before it is released to a recipient. It is important to note the data is effectively '**de-identified'** after the de-identification process as the subjects can be reidentified due to the EVD Platform holding access to a link between a unique identifier and the original personal data. In the hands of the EVD Platform it is still personal data and all such protections to privacy and principles for processing must be enforced. For those however who cannot access the key or link, it is **anonymised** data. The EVD Platform do not release the link between de-identified data and the original personal data and therefore the datasets released can be considered 'anonymous' (whereby data protection law no longer applies).

Data de-identification currently utilises the following standard operating procedure:

Stage 1: Removing direct and indirect identifiers using an adapted Safe Harbor Process

1. At the point of submission, the data is immediately encrypted and automatically copied from the *Upload Server* to the *High Compliance Server.*

[35] Guidelines for Media Sanitisation, National Institute of Standards and Technology (2014)
[36] Information Provided in Confidence (2015)

2. The data file is decrypted on the server and transformed into a CSV file. A unique patient ID is randomly generated and assigned to each record, and a link table maintained so that updates to the record may be possible in the future (this table associates a unique Ebola Data Sharing Platform ID with a set of identifiers).
3. The file can then be reviewed by the curation team to identify protected health identifiers (PHI) in accordance with HIPAA safe harbour. 16 of the 18 Safe Harbour variables are removed while retaining dates and geographic locations for patient matching/data linkage purposes (see appendix 1 for details). An audit trail is kept of the identifiers removed.
4. Following initial error checks in the data is then transferred from the *High Compliance Server* to the *Demo Server* for further curation activity. The decrypted data on the High Compliance Server is deleted using the secure eraser function (see Data Sanitisation for details).

Stage 2: Expert Determination

Before a dataset is written to the Download Server for an approved researcher to download, the DAC may request and specify certain statistical disclosure controls to be applied to the dataset. This would be in a scenario where the requested dataset contains a series of indirect/quasi identifiers, which when combined, increase the risk of subject re-identification above the DAC's acceptable risk threshold (a *'mosaic/jigsaw effect')*. A statistical control is applied to the data to mitigate this disclosure risk so the dataset falls within the acceptable threshold. This has been benchmarked at a probability threshold of 0.2, or what is often referred to as an equivalence class of 5, based on the literature relating to risk thresholds for re-identification of sensitive health data[37][38][39]. This means for any given query of the released dataset, there are a minimum of 5 possible individual records that match the query. Each data request will be considered individually based on their scientific justification and the potential risk of disclosure presented.

Disclosure controls are applied to items such as dates, geographic locators and any other variables identified as being of potential risk such as pregnancy status and very high/low ages. These are applied to the data to bring the re-identification risk below the Platform stated threshold for the dataset (or a given threshold prescribed by the DAC) and include processes such as: **omission/suppression**(e.g. removing data), **masking** (e.g. replacing data with a mask such as XXXXX), **truncation** (e.g. removing the day from a date leaving only mm/yyyy), or **generalisation/aggregation** e.g. transforming geographic locators to a regional, country level or ages into ranges such as 0-16, 16-24, 25-39, 40-65, 66-85,85+.

The approach is adopting an internationally recognised standard to safeguard privacy, that has significant evidence demonstrating efficacy[40][41]. It represents a practical approach to de-identification where risk of disclosure and data utility are balanced to find an acceptable level of risk, as noted in the Information Commissioners Office Code of Practice on Anonymisation [42][43]. It supports the removal of identifiable data at the earliest opportunity in processing, minimising any potential exposure, and applies a secondary checkpoint through the Expert Determination process, for review and quality control, ensuring all known risks are managed as far as possible.

[37] El Emam K, Dankar F, Vaillancourt R, Roffey T, Lysyk M (2009) Evaluating the risk of re-identification of patients from hospital prescription records. The Canadian Journal of Hospital Pharmacy62: 307
[38] El Emam K. (2008). Heuristics for de-identifying health data. IEEE Security & Privacy. 6(4), pp. 72-75.
[39] Cancer Care Ontario data use and disclosure policy. Toronto (ON): Cancer Care Ontario; 2005
[40] El Emam, K. El et al., 2011. A Systematic Review of Re-Identification Attacks on Health Data. 6,(12).
[41] Benitez, K., Loukides, G. & Malin, B., 2010. Beyond Safe Harbor: Automatic Discovery of Health Information De-identification Policy Alternatives. Proceedings of the ACM Int. Conf. Health informatics, pp.163–172
[42] A De-identification Strategy Used for Sharing One Data Provider's Oncology Trials Data through the Project Data Sphere® Repository. Bradley Mahlin, Ph.D (2013)
[43] https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

The Platform should be sure to be make this information available to data controllers thinking about using the platform in a published security document (as referenced in Privacy Risk 4).

Submission and Data Access Control

The EVD Platform has put in place a robust set of procedural controls to control the submission and sharing of data. Figure 1 provides an illustration of the participants in this initiative and the agreements the University has in place to date. The processes are as follows:

Contributor Registration and Verification

For an organisation to submit data to the platform they must register with the Infectious Diseases Data Observatory first (the group that are hosting the EVD Platform). This requires them to provide a range of information about who they are and their institution. Email verification is then required to validate the account before they can submit any data.

Terms of Submission

The contributor will next need to agree to the 'Terms of Submission'. Details of the Terms of Submission can be found in Appendix 4. This sets out the rights of the contributor, their obligations, how the data will be used. The contributor must digitally sign, and click to accept these terms before they can contribute their data to the platform.

Privacy Risk 8

Terms of Submission should be reviewed to ensure they reflect the data subjects' country of origin data protection requirements. This would include the recommendations cited earlier to make contributing organisations aware of their responsibilities as data controllers and ensure they are sighted/signposted to documentation on the security arrangements the data processor has in place for processing data.

Staff Contracts and Training

The EVD Platform as the data processor carries out a variety of curation tasks as part of personal data processing. These include data standardisation, mapping and de-identification. All staff accessing and curating data are employees of the University of Oxford and have an employment contract in place governing how they operate. This includes clauses relating to data privacy and that sensitive information must be kept confidential. Additionally, they will have received training in Information Security (University of Oxford), Research Data and Confidentiality (Medical Research Council)[44] and Human Subjects Protection (National Institute for Health)[45]. They are also required to follow the University policies and procedures in these areas (including the University Information Security Policy[46]). A specific additional security policy relating to the Infectious Diseases Data Observatory (which includes the EVD Platform) has been developed that all staff have reviewed and hold a duty to follow. See Appendix 5 for a copy of the security document.

Data Access Committee

The release of data relating to the Ebola database is managed and governed by the Data Access Committee. The committee is made up of a selection of experts from various fields and is selected by the EVD Platform Steering Committee. The committee reviews all applications to ensure appropriate use of the platform and evaluate any risks associated with the study alongside potential benefits the study might bring to science and society. The committee can also make decisions on further de-identification and/or statistical disclosure controls to apply to the dataset before release. See appendix 7 for the Terms of Reference for the Data Access Committee. This committee provides a

[44] http://byglearning.co.uk/mrcrsc-lms/course/category.php?id=1
[45] https://phrp.nihtraining.com/users/login.php
[46] https://infosec.ox.ac.uk/guidance-policy

further procedural control to assess the requested dataset is adequately de-identified, minimising the chances of identification. This function provides a further layer of control and accountability to the movement of data.

Data Transfer Agreement

When a project application has been approved by the Data Access Committee, the researcher (on behalf of their institution) will need to sign a Data Transfer Agreement with the EVD Platform/Oxford University. This agreement is legally binding, and sets out the agreed purpose of sharing, terms of use, liability, termination and legal governance of the agreement. This must be signed by both parties before data can be released. See Appendix 8 for an example of the Data Transfer Agreement.

EVD Platform Steering Committee

A steering committee has been established for the platform made up of representatives of the founding institutions. The membership of the committee is planned to evolve to include more representation form Ebola endemic countries. The role of the committee will be to establish the Data Access Committee, review and approve all platform related policies and procedures, and provide strategic, financial and ethical input to all related matters in support of the development and promotion of the platform and its membership. The committee has been set up to support input and comments from any interested party and will review any questions that are raised by the larger community of stakeholders. Appendix 6 contains further details on the Steering Group Charter, governance structure, membership and institutional stakeholders. The committee provides a level of oversight for the programme and can hold groups such as the Data Access Committee to account, providing a robust mechanism to ensure the ethical, legal and equitable processing and sharing of data are followed.

Conclusion and Recommendations

It is evident personal data is being carefully controlled throughout the submission, curation and release processes of the EVD Platform. At the point of transfer ('data sharing') to a researcher, personal data is rendered anonymous and therefore data protection law no longer applies (as it only applies to data relating to a living individual who can be identified). Safeguarding of ethical standards for sharing (de-identified) data are applied by the Data Access Committee (DAC) and enforced through a contractual Data Transfer Agreement with the researcher (and their institution). The DAC and the EVD platform are overseen by the EVD Steering Committee, providing a robust and accountable governance framework for the platform.

This assessment has found no fundamental issues with the proposed project in terms of compliance with law, policy or established practice. The privacy risks surfaced (see summary table in Appendix 9) can be managed and mitigated through the recommendations described in the Privacy Solutions table in Appendix 10. There are three main areas of recommendation:

1. Carry out a risk assessment of the data types involved to develop an acceptable risk threshold and statistical disclosure approach to minimise any potential jigsaw/mosaic type effects that could lead to re-identification.
2. Create a Security Model document for the EVD Platform covering all technical and procedural controls and standard operating procedures.
3. Develop a privacy notice for the public facing website to inform the public (and potential data subjects) about how and what data is used, what this is for, who is involved, how they can find out more, and who to contact if they wish to have their data removed.

The current technical and procedural security measures in place alongside implementation of the recommended solutions will support the EVD Platform (and its contributors) to meet the current, and likely future domestic legislative data protection requirements in the subject West African countries (subject to review) whilst also demonstrating that the platform is safeguarding personal data in compliance with standards in the UK and internationally.

Appendix 1: Privacy Impact Assessment Screening Questions

If you answered **Yes** to any of the following, then you are advised to complete a detailed PIA.		**Yes**	**No**	**Comments**
i	Is the data about individuals likely to raise privacy concerns or expectations e.g. health records, criminal records or other data people would consider particularly private?	✓		Data involved is health information. Data subjects need opportunity to exercise their rights to object if they have concerns about their privacy. Assurance of security measures needs to be provided to contributors and public. Robust governance policies and procedures in place – independently assured. Legally binding data transfer agreements in place.
ii	Will the initiative involve the collection of new data about individuals?		✓	No. The project receives data from contributors.
iii	Are you using data about individuals for a purpose it is not currently used for, or in a way it is not currently used?	✓		The data will be processed to render it anonymous to those using the data.
iv	Will the initiative require you to contact individuals in ways which they may find intrusive[47]?		✓	Minimum personal data will be processed for this purpose. Access is limited to authorised personnel.
v	Will data about individuals be disclosed to organisations or people who have not previously had routine access to the data?	✓		The data processor will de-identify personal data at the first opportunity (to the extent that de-identified data remains personal data in law). Researchers will only ever access anonymous data.
vi	Does the initiative involve you using new technology which might be perceived as being privacy intrusive e.g. biometrics or facial recognition?		✓	No.
vii	Will the initiative result in you making decisions or taking action against individuals in ways which can have a significant impact on them?		✓	No. However staff awareness is required that a breach resulting in reidentification of a subject could have a significant negative impact on an individual leading to stigmatisation and discrimination.

[47] Intrusion can come in the form of collection of excessive personal information, disclosure of personal information without consent and misuse of such information. It can include the collection of information through the surveillance or monitoring of how people act in public or private spaces and through the monitoring of communications whether by post, phone or online and extends to monitoring the records of senders and recipients as well as the content of messages

Appendix 2: Example data dictionary

ebola.csv

Appendix 3: Safe Harbor Initial Protected Health Identifier Column Removal

Number	Column
1	Names
2	Telephone numbers
3	Fax Numbers
4	Email
5	Social Security Numbers
6	Medical Record Numbers
7	Health Plan Beneficiary Numbers
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

Appendix 4: Terms of Submission

IDDO Terms of Submission Ebola 18(

Appendix 5: IDDO Information Security Policy

IDDO Information Security Policy.docx

Appendix 6: Ebola Data Sharing Platform Initiative Steering Committee

Ebola DSP Steering Committee Charter 2(

Appendix 7: Terms of Reference for the Data Access Committee.

T_C of EVD Data Platform Access 13JU

Appendix 8: Data Transfer Agreement

DTA-WWARN.doc

Appendix 9: Privacy Risks

The risks captured in the privacy impact assessment relate to individuals, compliance and corporate risks. These are explained below:

Risks to individuals

- Risks that may lead to subject identification or disclosure of personal data, leading to an unlawful or unjustified intrusion on individual privacy.
- Risks that do not allow an individual to exercise their rights as a data subject.

Compliance risks

- Non-compliance with any international or domestic statutory, regulatory or common law
- Non-compliance with any regulations, sector specific legislation or standards.

Associated organisation/corporate risks

- Non-compliance with any legislation or regulation which could lead to sanctions, fines and reputational damage.
- Data losses which damage individuals could lead to claims for compensation.
- Public distrust about how information is collected, treated and used can damage an organisation's reputation and lead to loss of business, engagement with the organisation.

Number	Privacy Risk	Risk to individuals	Compliance risk	Associated organisation / corporate risk
1	Potential risk that data being released could lead to subject identification on its own or through combining data with auxiliary information (known as data intrusion, the mosaic effect, or Jigsaw identification). A risk analysis can be carried out on the data types to assess the likeliness of this occurring and an acceptable risk threshold. Appropriate statistical disclosure controls can then be applied to datasets that do not fall within these limits to reduce any potential risk.	✓	✓	✓

2	Use of language particularly 'Anonymisation'. The word anonymise has variable meanings across different sectors and jurisdictions. It can easily create confusion and be misinterpreted. A consistent use of language is very important when dealing with sensitive personal data. In the example of the EVD Platform, the data is 'de-identified' as even after de-identification the subjects can be reidentified due to the organisation holding access to a link between a unique identifier and the original personal data. Therefore, in the hands of the EVD Platform it is still personal data. Whilst for those who cannot access the key or link, it is anonymised data. Review other project documentation for consistency.		✓	✓
3	It will be important to keep up to date with the developing data protection landscape in the subject countries. Terms of Submission should be reviewed as legislation is approved in this area and national protection authorities established. This is to ensure both contributors and the EVD Platform are aware of their responsibilities as data controllers and process data in compliance with the relevant domestic law. It is recommended this Privacy Impact Assessment is repeated on a biannual/annual basis.	✓	✓	✓
4	Whilst this is the responsibility of the institution submitting data, to support contribution to the EVD Platform, a full data security model should be made available to contributors detailing the end to end process and all technical and procedural controls in place. This will ensure the University are operating transparently and providing contributors all the information they require to make a decision. Additionally, this will ensure the EVD platform is meeting its obligations as a data controller (detailed later in this assessment).		✓	
5	A notice should be provided informing the contributor of this principle and the responsibility they have when submitting data. The IDDO EVD Platform could also provide this and more detailed information on its public website that contributors could then reference.		✓	
6	To support compliance with these exceptions, the EVD Platform should have the project and its security model reviewed by an appropriate ethics board(s) to provide feedback on the project and the proposed security framework. The outcome should be published on the public facing website that the project has been reviewed, who by and when.		✓	

7	Key strengths have a lifetime and therefore will require review periodically as part of ongoing privacy impact assessments.	✓	✓	✓
8	Terms of Submission should be reviewed to ensure they reflect the data subjects' country of origin data protection requirements. This would include the recommendations cited earlier to make contributing organisations aware of their responsibilities as data controllers sharing data to another data controller.		✓	✓

Appendix 10: Privacy Solutions

Privacy Risk/Issue		**Solution**	**Risk Status**	**Evaluation** Is the final impact on individuals a justified, compliant and proportionate response?
1	Potential risk that data being released could lead to subject identification on its own or through combining data with auxiliary information (known as data intrusion, the mosaic effect, or Jigsaw identification). A risk analysis can be carried out on the data types to assess the likeliness of this occurring and an acceptable risk threshold. Appropriate statistical disclosure controls can then be applied to datasets that do not fall within these limits to reduce any potential risk.	Carry our risk analysis of data types included in data dictionary to measure likely risk of mosaic/jigsaw effects that could lead to identification. Define minimum acceptable threshold and apply any necessary disclosure controls on this basis.	Reduced	Justified, compliant and proportional: Any risk of identification to an individual or community is reduced to the minimum possible.
2	Use of language particularly 'Anonymisation'. The word anonymise has variable meanings across different sectors and jurisdictions. It can easily create confusion and be misinterpreted. A consistent use of language is very important when dealing with sensitive personal data. In the example of the EVD Platform, the data is 'de-identified' as even after de-identification the subjects can be reidentified due to the organisation holding access to a link between a unique identifier and the original personal data. Therefore, in the hands of the EVD Platform it is still personal data. Whilst for those who cannot access the key or link, it is anonymised data. Review other	Consistent and clear use of language across all documentation to ensure understanding of data state and any entailed obligations.	Eliminated	Compliant: Risk is eliminated through consistency, education and training.

	project documentation for consistency.			
3	It will be important to keep up to date with the developing data protection landscape in the subject countries. Terms of Submission should be reviewed as legislation is approved in this area and national protection authorities established. This is to ensure both contributors and the EVD Platform are aware of their responsibilities as data controllers and process data in compliance with the relevant domestic law. It is recommended this Privacy Impact Assessment is repeated on a biannual/annual basis.	Repeat Privacy Impact Assessment on annual basis to ensure compliance with data protection law as it develops.	Reduced	This is an ongoing risk to be managed by the project. The action proposed is a proportionate and valid approach to managing this.
4	Whilst this is the responsibility of the institution submitting data, to support contribution to the EVD Platform, a full data security model should be made available to contributors detailing the end to end process and all technical and procedural controls in place. This will ensure the University are operating transparently and providing contributors all the information they require to make a decision. Additionally, this will ensure the EVD platform is meeting its obligations as a data controller (detailed later in this assessment).	Create EVD Security Model documenting the end to end process of data processing and sharing. Including all controls and procedures in place.	Eliminated	Compliant and Proportional: Risk is eliminated through publishing document and is in support of data controllers' decision making.
5	A notice should be provided informing the contributor of this principle and the responsibility they have when submitting data. The IDDO EVD Platform could also provide this and more detailed information on its public website that contributors could then reference.	Privacy notice provided on website and noted in Terms of Submission to encourage contributor to do the same	Eliminated	Compliant and Proportional: Risk is eliminated through publishing notice and is in support of data controllers' decision making.
6	To support compliance with these exceptions, the EVD Platform should have the project and its security model reviewed by an appropriate ethics board(s) to provide feedback on the project and the proposed security framework. The outcome should be published on the public facing website that the project has been reviewed, who by and when.	Have the EVD Data Sharing Platform Initiative reviewed by a research ethics board	Eliminated	Justified and Proportional: Ethical review will provide a second opinion on the platforms impact on individual's welfare and privacy, adding a further level of safeguarding.
7	Key strengths have a lifetime and therefore will require review periodically as part of ongoing privacy impact assessment.	Annual Privacy Impact Assessment review to ensure adequate encryption standards are being applied	Reduced	This is an ongoing risk to be managed by the project. The action proposed is a proportionate and valid approach to managing this.

8	Terms of Submission should be reviewed to ensure they reflect the data subjects' country of origin data protection requirements. This would include the recommendations cited earlier to make contributing organisations aware of their responsibilities as data controllers sharing data to another data controller.	Annual review of the Terms of Submission	Reduced	This is an ongoing risk to be managed by the project. The action proposed is a proportionate and compliant approach to managing this.

Appendix 11: EVD Platform Risk Register

Each risk will be RAG rated by taking the likelihood and impact scores, and using the matrix below:

Likelihood					
Score	**1**	**2**	**3**	**4**	**5**
Descriptor	**Rare**	**Unlikely**	**Possible**	**Likely**	**Almost Certain**
Frequency – How often might it happen?	This probably will never happen/recur	Do not expect it to happen/recur, but it is possible it may do so	Might happen or recur occasionally	Will probably happen/recur, but is not a persisting issue or circumstance	Almost certain to happen/recur, possibly frequently
Impact					
Score	**1**	**2**	**3**	**4**	**5**
Descriptor	**Very Low**	**Low**	**Medium**	**High**	**Very High**
Impact should it happen?	Unlikely to have any impact	May have an impact	Likely to have an impact	Highly probable it will have a significant impact	Will have a major impact

Using the risk RAG rating system for scoring risks means risks can be ranked so that the most severe are addressed first. Decisions can then be made as to what mitigating action can be taken to alleviate the risk.

Likelihood / Impact	1	2	3	4	5
1	1	2	3	4	5
2	2	4	6	8	10
3	3	6	9	12	15
4	4	8	12	16	20
5	5	10	15	20	25

ID	RISK DESCRIPTION	RAG SCORE	MITIGATION	DATE RAISED	TARGETTED RESOLUTION DATE	UPDATE	STATUS
1	Mosaic effects leading to subject identification	12	Risk analysis of data types	20/02/17	31/03/2017	Risk analysis to be completed	OPEN
2	Misunderstanding of data state due to language leads to (potential) breach.	4	Consistent and clear use of language across all documentation	22/03/17	31/03/2017	Review all documentation	OPEN
3	Data protection changes lead to non-compliance and regulatory action	10	Update PIA annually to ensure compliance	20/02/17	31/03/2018	Ongoing risk managed via PIA	OPEN
4	Contributors do not think platform will fulfil their legal obligations as data controllers.	4	Share IDDO EVD Data Security Model & SOP doc with contributors	22/03/17	31/03/2017	Complete Security Document	OPEN
5	Subjects unaware of data processing or how to exercise their rights.	10	Review with contributors' communications to the public to ensure adequacy	20/02/17	31/03/2017	Agree communications plan with Steering committee	OPEN
6	No ethics approval to sign off and validate platform	6	Ethics review of IDDO EVD Data Sharing Platform	22/03/17	31/03/2017	Submit research protocol to ethics boards	OPEN
7	Encryption keys become vulnerable	10	Review key strengths annually as part of PIA	20/02/17	31/03/2018	Ongoing risk managed via PIA	OPEN
8	Terms of submission out of line with domestic legislation	10	Review Terms of Submission annually as part of PIA	20/02/17	31/03/2018	Ongoing risk managed via PIA	OPEN

Appendix 12: Risk Analysis of Data Types

Insert Risk Assessment of Data Types in EVD

Appendix 13: General Project Information

EVD Data Platform Project Proposal_V3_2

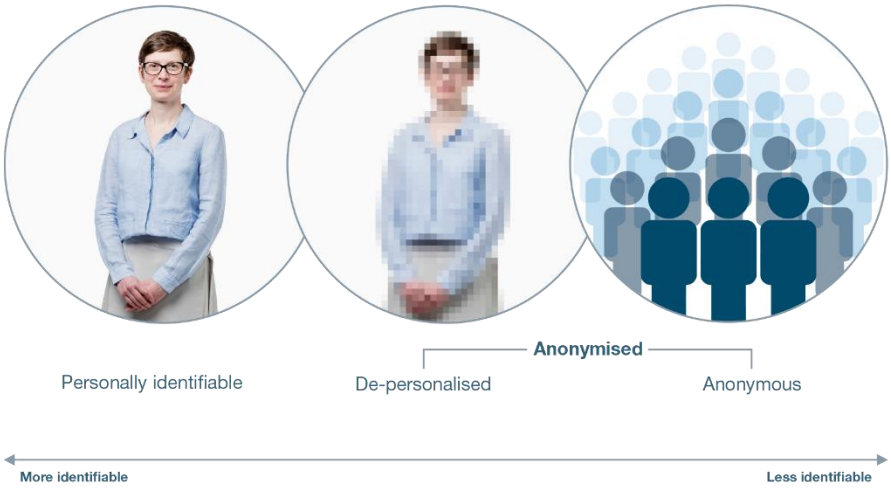
Glossary

Anonymous Data: This is information where an individual cannot be identified. This can be when data has had all identifiable elements removed, so that it would not be possible to identify an individual, or when data from many people is combined (presented often as statistics).

De-identified/ Pseudonymous /masked/anonymous in context/: This is information that does not identify an individual, because identifiers have been removed or encrypted. However, the information is still about an individual person and so needs to be handled with care. It might, in theory, be possible to re-identify the individual if the data was not adequately protected, for example if it was combined with different sources of information[48].

Personal Data/Personally Identifiable/Confidential Information: This is information that identifies a specific person. Identifiers include: name, address, geographic locators, date of birth or national health number48[48].

[48] https://understandingpatientdata.org.uk/sites/default/files/2017-04/Identifiability%20briefing%205%20April.pdf


Anonymised
Personally identifiable
De-personalised
Anonymous
More identifiable
Less identifiable