GD057 OpenClinica Database Hosting and Storage

This document provides guidance on the hosting and storage of study databases on OpenClinica for research studies managed by the Papworth Trials Unit Collaboration (PTUC). It also applies to studies sponsored by the Royal Papworth NHS Foundation Trust, or hosted studies where data storing responsibility has been delegated to Royal Papworth Hospital.

Data held within OpenClinica3 (OC3):

1. When a study is live, the study data is held on OC3, plus 2 external locations: Insight and DataMart
   a. OpenClinica 3 – This is "Front End" of the database, and where the majority of staff enter and interact with the data. Studies on OC3 are hosted within the OpenClinica Optimised Cloud, which operates through Amazon Web Services (AWS) with hosting in London. This hosting environment includes dedicated application servers, advanced security measures. The system meets the UK-recognised 'Safe-Haven' standard certified under ISO 27001 and SSAE 16 Type II SOC 1 security standards.
   b. Insight – This is an OpenClinica hosted module that is requested for all studies. It contains all the CRF data, plus all the Query/Discrepancy note data. For some very small studies it also includes the audit history data, but this is unstable on the OC3 version of Insight, so can disappear once the study reaches an unspecified size. This is a known bug and cannot be resolved. Insight is a data warehouse that is generated from main OpenClinica transactional database, and snapshots updates are done approximately every hour. Within Insight there are multiple tools for reviewing, reporting and extracting the data.
   c. DataMart – This is a local hosted copy of the CRF data, based upon a manually built extract from the live study built by the Clinical Trial Data Programmer. This extract is run every evening.

2. Once a database is locked (Interim or final database lock) the data is extracted for the statisticians from the DataMart. This is filed in N:\Shared\R&D DATA MANAGEMENT & STAT TEAM\Studies\<study name> unless there is a study specific reason for this to be stored somewhere else, this would be file noted. This data is in a tab delimited text formatted file and usually contains a file per CRF. Though this might differ depending on requests from the statisticians or the structure of the data in the CRF. For example, it might be sensible to break a CRF into multiple files.

3. For multi-site studies, a copy of each individual patient's casebook are printed to pdf and filed in the eTMF for all non-Papworth sites. These contain the CRF data, audit trails, notes and discrepancies. This is sent to the individual sites for their records. This is done manually for each patient, making it time consuming.

4. When it is time to archive a study, after allowing sufficient time to identify and correct any database or extract issues, the following steps should be managed by the Data Management Lead and can be delegated to the data team:
   a. The data extracts should be copied into the electronic TMF folders.
   b. The query and if available the audit trail tables from Insight should be extracted and filed in the electronic TMF folders. If this option is not available, then a list of the Notes & Discrepancies printed to pdf on OpenClinica and saved in the electronic TMF folder.
   c. A full copy of the database should be extracted from OC3 in CDISC ODM XML 1.3 with OpenClinica extensions format and filed in the electronic TMF folders. This file contains all of the metadata for the study, in an .xml format, which is technically complex, but serves as a complete record of the study.

Data held within OpenClinica4 (OC4):

As the transition to OC4 is ongoing, data archiving processes are still being finalised. The archiving processes for OC4 will however, follow a structure similar to that of OC3. Thus,

1. When a study is live, the study data is held on OC4, plus 1 external location: Insight
   a. OpenClinica 4 – Again, this is the "Front End" and is used by most staff. Studies on OC4 are hosted on the OpenClinica Virtual Cloud, which operates through Amazon Web Services (AWS) with hosting in London. The hosting environment meets the UK-recognised 'Safe-Haven' standard for secure data hosting, certified under ISO 27001, ensuring compliance with information governance and security requirements.
   b. Insight – As with OC3 Insight, this must be requested. However, the key difference is that the audit tables remain stable and available throughout the study, regardless of its file size.
   c. DataMart – This is not currently available, but it is planned to have the data locally hosted.

2. When the study is locked the data is extracted from Insight in csv formatted files and filed the same way.

3. For all studies the casebook can be provided for all sites, as they are printed to pdf for all patients in one step.

4. When it becomes time to archive a study, after sufficient time for any issues with the database or extracts to be discovered and corrected, the following steps should be managed by the Data Management Lead and can be delegated to the data team:
   a. The data extracts should be copied into the electronic TMF folders.
   b. All the tables from Insight should be extracted and filed in the electronic TMF folders.
   c. A full copy of the database should be extracted from OC4 in CDISC ODM XML 1.3 with OpenClinica extensions format and filed in the electronic TMF folders. This file contains all of the metadata for the study, in a xml format, which is a technically difficult but a complete document of the study.