Best Practice - OpenClinica

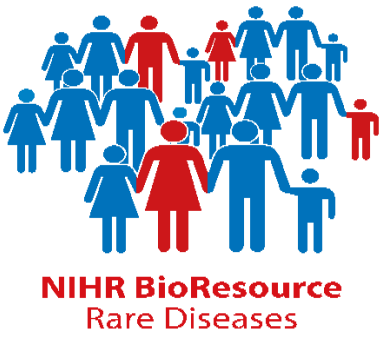
Purpose of this document: To provide best practice guidelines for each of the Rare Diseases studies to conform to the RDCIT standards which will allow the data to be stored in a secure, accessible, extendable and sustainable way. The focus of this document is the RDCIT Secure Data Service and the OpenClinica study setup.

Audience: Principal Investigator and person responsible for setting up and configuring the study in OpenClinica (hereafter referred to as designate or data manager).

Filename: Best Practice – OpenClinica v1.9
Date: 14/02/2017

Authors: Roger James, Catherine Titterton

NIHR Rare Diseases Translational Research Collaboration
Department of Haematology, University of Cambridge
National Health Service Blood and Transplant Cambridge
Long Road, Cambridge
CB2 0PT



Document control

Version	Date	Who	Changes
1.0	21/08/2014	Roger James, Catherine Titterton	First draft
1.1	27/11/2014	Julie Von Ziegenweidt	Inclusion of GEL minimum data specification – section 5; Appendix – standard lookup codes.
1.2	22/12/2014	Catherine Titterton, Roger James	RDCIT Service levels (section 1.2); updated standard CRFs in line with GEL minimum data specification; Ethics (section 3.3),
1.3	09/01/2015	Vera Matser	Updated 5.1 Site and Study Level Security, 6.1 Shared CRFs, 7. Appendix
1.4	11/01/2016	Vera Matser	1.2.5 Data Coordinators; 1.2.6 RDCIT webpages; 3.3 Access requirements; 3.4 Good Neighbour policy; 3.5 OpenClinica Data Import; 4 Clinical coding; 4.1 Data Plan Design; 5.2 Study Parameters
1.5	13/01/2016	Vera Matser	Incorporation of RDCIT study (5.1 – 5.2) and user authorisation forms (6.2.1). Updated 2. Best practice in one quick bite
1.6	05/02/2016	Vera Matser	Access methods (section 3.3.1), alteration to collection of GMC number (section 6.2.1)
1.7	11/03/2016	Vera Matser	Naming convention change OCsecure to OC live / OC import. Distinction between OC live and OC import (section 3.1). Updated section 2. Update section 1.2.4; Replace superuser with designate; Storing Patient Identifiable Data (section 3.3.5); Data import into RDCIT Secure Data Service (section 3.5.2)
1.8	01/08/2016	Ping Yu	Removed the RDCIT helpdesk email. Updated Event naming conversion (section 4.4.4), CRF versioning section (section 5.5.1), and Passwords reset (section 6.2.3). Modified common CRFs and Process CRFs naming conversion (section 4.4.2). Added a new section of testing procedure (section 7), and the previous section 7 (V1.7) became section 8.
1.9	14/02/2017	Ping Yu	Added a new section on CRF Batch migration (section 5.5.3). Added a guild line on how to remove a Subject and PHI labelling (section 6.1). Replaced Code Browser with Data Plan Composer (section 4.1). Introduced PhenoTips to store patient pedigrees information (section 6.2.5)

Check if there is a later version of this document here

Contents

1 Overview

1.1 Introduction

Patient records on NIHR BioResource Rare Disease studies and NIHR Rare Diseases Translational Research Collaboration studies are intended as a resource for research. The records need to be held in a manner which is sustainable, extendable and accessible to allow advances in science and investigation. It should be possible to longitudinally add information from multiple external sources to the patient records. In parallel it is expedient to align the patient records to the Genomics England infrastructure developments. The infrastructure being built now must, in the future, allow the possibility of direct patient access through a patient portal.

OpenClinica has been chosen as the NIHR Rare Disease patient record system and will be the master record of patient data on the Rare Diseases Genome project. OpenClinica is the entry point for clinicians recording patient data and enriching the patient record with in-depth phenotype information.

Within OpenClinica the NHS number will provide the unique identifier (index spine) to which other systems will reference each patient's record. For non-NHS England and Wales patients a similar unique index should be used.

External systems linking to OpenClinica include:

- **CiviCRM** – a patient record management system providing patient recruitment, ethics and pertinent findings feedback linked to OpenClinica. CiviCRM also provides the portal for account creation of clinicians on OpenClinica.
- **NHS systems** "lookup" from OpenClinica to extract patient medical history data when accessed by authorised clinicians (under development).
- **External ontologies** (HPO, SNOMED-CT, etc) for phenotype data input in OpenClinica.

The "full secure data system" that OpenClinica and the external systems such as CiviCRM are part of will be referred to as the **RDCIT Secure Data Service** (RDCIT SDS; hosted by AIMES Grid Services C.I.C - HSCIC registration H8121).

1.2 Rare Disease Clinical Infrastructure Team (RDCIT)

The Rare Diseases Clinical Infrastructure Team (RDCIT) are responsible for implementing the RDCIT Secure Data Service, consisting of OpenClinica and related IT infrastructure, which supports the Rare Disease studies (NIHR BioResource BRIDGE studies), East of England Genomics Medicine Centre (GMC) (recruiting participants to the 100,000 Genomes Project led by Genomics England) and the Rare Diseases Translational Research Collaboration (RD-TRC) studies. The RDCIT provide training and support for each study setting up on OpenClinica.

Access to OpenClinica services hosted by the RDCIT is determined by the project funding made available to resource the OpenClinica environment and services. In the table below the service support levels are defined as Gold, Silver and Bronze. Bronze is free of charge in contrast to Gold and Silver which are directly contributing to the infrastructure and resource costs.

The primary funding is from the NIHR BioResource – Rare Diseases (RD BioResource) who qualify for the 'Gold' service and have the prime responsibility for establishing and maintaining the OpenClinica technical, data and service environment. They are the 'first amongst equals' client of the service. The RD-TRC provides resource to the service primarily in the form of Clinical Data Coordinators to help with the data mapping, importation and local support of the system nationally. The support levels for the RD-TRC studies are designated as Silver.

Beyond the RD BioResource and RD-TRC which directly fund the RDCIT, other projects and activities can make use of the service. Without any additional or explicit funding of our activities we cannot afford for this support level to be at the same level as that provided to the RD BioResource and the RD-TRC. However we are keen to exploit and share our knowledge and capability with OpenClinica and to help other research projects. Where appropriate this will be free of charge with the option for projects to access higher levels of support if they make available to the RDCIT additional resources.

1.2.1 Systems and Services

- [**Gold**] access to dedicated OpenClinica System, direct support in developing study specific CRFS, (some) customisation work for new data sources (e.g. coding look-ups)
- [**Gold** / **Silver**] access to N3 hosted 'virtual PC' [data embassy] from both the N3 network (expected summer 2016) and Internet, resources to assist with the extraction, mapping and import of data sets from HER systems, patient registries etc.
- [**Bronze**] access to shared OpenClinica system, access to OC through web interface, OC account management on OCplay and the production system.
- [**Bronze**] access to all of the code and data services developed under Gold or Silver (such as a specific extract of data from EPIC).
- [**Bronze**] OpenClinica is available with 95% core time [9-5 UK] availability, in a secure data centre verified for the safe storage and use of patient identifiable data. Access to N3 hosted systems (expected summer 2016) and data services will be enabled although only specific interfaces, procured through the Gold service, will be enabled

1.2.2 Training and Support

- [**Gold** / **Silver**] local support from data coordinators and some personalised help and training
- [**Gold** / **Silver**] local support to manage the data mapping, data import and export of data collections from existing EHR systems
- [**Bronze**] Access to data import tools and training material, attendance at the 'standard' OC training on how to develop project specific OC Case Report Forms and use of the Rare Disease Minimum data set forms

1.2.3 Costing and Pricing

The Bronze service is available free of charge (to 2017) subject to a limit of 10 users, 1000 subjects and/or 20000 data points. Silver service is priced as 1 FTE per 3000 subjects or equivalent (cash or kind), Gold service is 1 FTE or equivalent per 2000 subjects.

One off requests from the Bronze service users will be considered separately and be subject to resource availability after the Gold and Silver work.

1.2.4 RDCIT team members and contact details

Name	Job Title	Email	Telephone
Dr Roger James	RD-TRC IT Director	rj352@medschl.cam.ac.uk	01223 254902
Catherine Titterton	Project Manager	ct474@medschl.cam.ac.uk	01223 254616
Julie Von Ziegenweidt	Clinical Research Data Lead	jv359@medschl.cam.ac.uk	01223 254617
Ping Yu	Clinical User Training officer	py234@medschl.cam.ac.uk	01223 254611
Csaba Halmagyi	IT Integration Specialist	ch686@medschl.cam.ac.uk	01223 254613
Paul Treadway	IT Infrastructure Specialist	pt381@medschl.cam.ac.uk	01223 254615
Coleen McJannet	Administrator	cm686@medschl.cam.ac.uk	01223 254900

1.2.5 RD-TRC Data Co-ordinators and contact details

Name	Job Title	Email	Telephone
Dr Gayatri Chavali	Senior Clinical Data Co-ordinator	gc292@medschl.cam.ac.uk	01223 254906
John Li	Clinical Data Co-ordinator	jl954@medschl.cam.ac.uk	01223 254618
Kimberley St John Green	Clinical Data Co-ordinator	kas65@medschl.cam.ac.uk	01223 254627
Kitti Czegledi	Clinical Data Co-ordinator	kc514@medschl.cam.ac.uk	01223 254627

1.2.6 RDCIT webpages

All tools, user documentation and training resources produced by the RDCIT Team will be placed on the RDCIT webpages.

User documentation - http://rdcit.org/user-documentation/

RDCIT Secure Data Service Forms - http://rdcit.org/rdcit-secure-data-service/

OpenClinica 'How To' pages - http://rdcit.org/online-training-material/

RDCIT Tools - http://rdcit.org/openclinica/rdcit-tools/

2 Best Practice in one quick bite

The follow-on sections in this document provide detailed guidance for setting up your study in OpenClinica. This section provides a more easily digestible read of the key practices. Note that in using OpenClinica as it is designed, much of these points will be taken care of.

Patients (Subjects)

- The NHS number needs to be provided for all subjects.
- Use the Person ID field to store the NHS number.
- Use an identifier which is pseudonymised as the Study Subject ID. Note if your study is part of the Genomics England 100,000 Genomes Project, a unique participant ID will be provided by GEL.
- The project decides on the study parameter options (Person ID visibility, Secondary Label (Subject Secondary ID) visibility and use; Date of Birth options) unless pre-defined by the RDCIT.
- If applicable, Patient Identifiable Data (PID – e.g. contact data) should be preferentially stored in CiviCRM or in the standardised (RDCIT-owned) PID CRF in OpenClinica (Name: COZ_PID). This CRF should not be customised by any study.

Clinicians (Users)

- Everyone who is entering or accessing OpenClinica must have their own user account
- User accounts should be requested using the RDCIT User Authorisation form (6.2.1)

Pre-study setup

- A RDCIT approved data plan should be developed (see 4. Clinical Coding)
- Prior to the study going live the RDCIT Study Authorisation form will need to be completed by the PI

CRF Design

- In CRF design where possible use the standard conventions for naming items, values displayed and values stored. (see 4.4)
- Prefix CRF names with the Unique Protocol ID (e.g. 3-character project code or BRIDGE project code)
- Only load verified CRFs to the OpenClinica live environment.

Study setup

- Provide full contact details for the Principal Investigator (Lead Applicant) on the study.
- Unique Protocol ID: Use the unique three-character project code (e.g. BRIDGE project code) as the studies Unique Protocol ID in the Study Description
- Secondary IDs: indicate whether your study is part of BRIDGE, RD-TRC, GMC, or Other (if not belonging to the previous three)
- If applicable, Patient Identifiable Data (PID – e.g. contact data) should be preferentially stored in CiviCRM or in the standardised (RDCIT-owned) PID CRF in OpenClinica (Name: COZ_PID). This CRF should not be customised by any study.

Sites

- The site Unique Protocol ID should be prefixed with the three-character project code (which is the Study Unique Protocol ID)
- For each site provide full contact details of the clinician on that site responsible for patient care (in the Facility Information section)
- If a hospital has multiple clinical leads within a study, who do not share patients, a study can be set up with separate sites e.g. Brompton: Smith and Brompton: Jones.
- Ensure all clinicians are assigned to a site (to restrict their access to patients they are responsible for)
- Ensure all patients are assigned to a site (corresponding to clinicians administering their records)
- The only users with access at study level should be the Principal Investigator and people responsible for configuring or monitoring the study (designate/ OpenClinica Data Manager).

Final Sign off Procedure

- RDCIT will provide a checklist for each study to review with the RDCIT team prior to launching their OpenClinica study (based on Data Plan).
- Study and User Authorisation forms must be completed prior to the study going live
- Users who have access to data across multiple sites will need to complete a Research Passport/Letter of Access with Cambridge University Hospitals NHS Foundation Trust (CUH).

3 RDCIT hosted OpenClinica

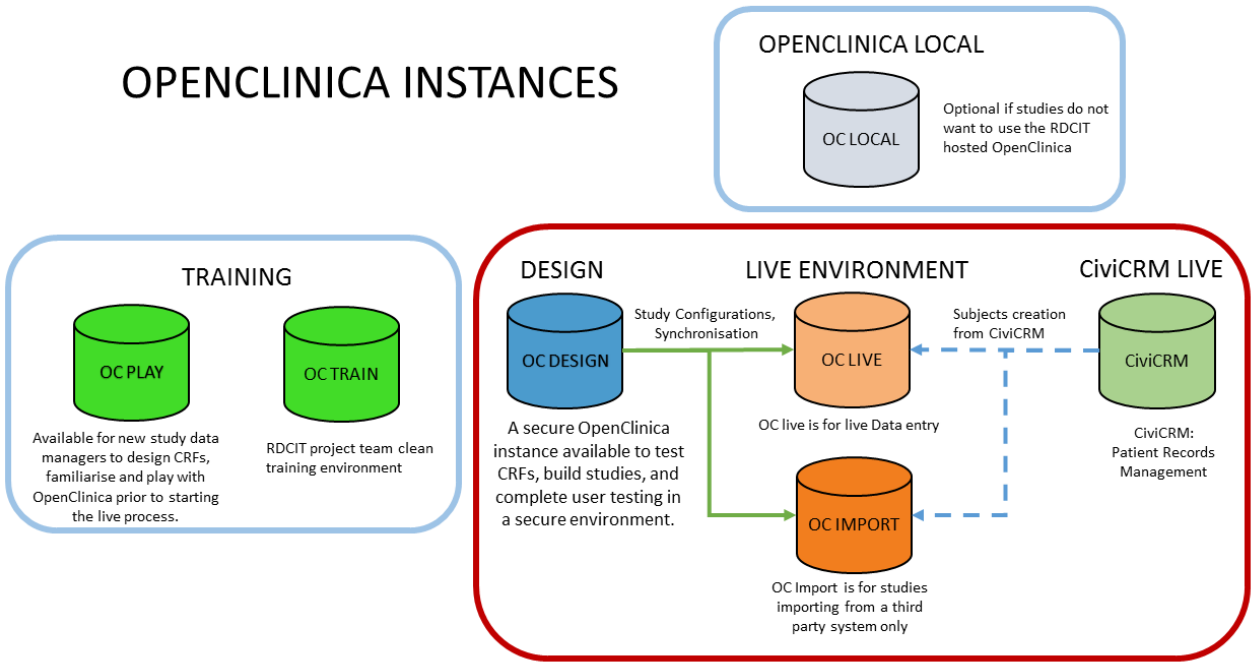
3.1 Environments

For studies using the RDCIT Secure Data Service, a number of OpenClinica environments are being provided:

1) **OC play** – available for study leads to access OpenClinica immediately to gain familiarity with the system prior to study setup on OpenClinica, and CRF design. OCplay will be regularly reset to a 'clean state'. An account can be requested from the RDCIT support team.
2) **OC train** – a 'clean' environment through which RDCIT training will be delivered
3) **OC design** – provided for CRF testing, testing user access, study process and rules logic.
4) **OC live** – provided for studies using live data entry, limited data import will be supported on this instance. Will be integrated into the NHS network, where possible provide lookups to NHS datasets and will hold patient identifiable information (under development).
5) **OC import** – provided for studies importing data from a third-party system, live data entry will not be supported on this instance.

If studies choose to install their own local instance of OpenClinica or use an alternate system, they will be required to submit their study data files in a format compatible for import into OC import.

6) **OC local** – an OC environment locally hosted by the study organisation, unsupported by RDCIT



3.2 OpenClinica features

Below is a list of features & requirements for the RDCIT hosted OC live/OC import instance. Note that features indicated with a * are expected in phase 2 of the RDCIT Secure Data Service development.

OpenClinica Features
Federated structure to share study forms and data across different hospitals, clinics, BRCs and BRUs.
Version control of CRFs
Queries attached to patient records through OpenClinica Notes and Discrepancies
Study and site level security model to maintain patient record confidentiality
Secure backup and restore of study data
Managed user accounts (access for clinicians collecting data)
Audit trail of all subject record changes
* Sample status tracking
* Linked system records automatically flagged in the case of consent being withdrawn.
* Pertinent findings flagged with email notification to PI's and referring consultants.
Links to ontologies (SNOMED CT, HPO etc)
Ability to bulk import data related to study patients (e.g. medication history, birth record information) using ODIN
* IG Compliant environment to hold personal patient data
* N3 Connection (NHS National Network)
Linked data spine which references the patients genome sequence and big data analysis files
OpenClinica Requirements
NHS numbers (or unique alias which studies must provide a linked table to the NHS number to RDCIT) must be included with patient record
All users must adhere to the RDCIT Best Practice
Studies retain responsibility for all study ethics & consent procedures even if the relevant documents are stored within the secure environment.
Data Transfer
Must meet Good Neighbour Policy and Clinical Coding Guidelines - study data and associated metadata to be submitted in a format ready for import into OpenClinica

3.3 Access Requirements

3.3.1 Access methods

Access to the RDCIT Secure Data Service will be given to individual users upon request of the study's Principal Investigator. What type of access is granted to the user depends on the connection capabilities between the user's host institution and the RDCIT Secure Data Service as well as what action the user needs to be able to perform in the secure environment.

3.3.1.1 HTML5

HTML5 access allows user to log on a virtual computer (opens in a separate browser window) in the RDCIT Secure Data Environment rather than their own host PC, and it allows user to access the work station on a virtual computer. However, actions such as copy & paste, print (which are dependent on the host PC) are not available. This type of access is possible if the browser is HTML5-comptible and the connection is allowed by the host institution. The RDCIT have developed a script which can check HTML5 Browser compatibility (http://rdcit.org/rdcit-secure-data-service/).

3.3.1.2 GATEWAY

The Gateway (direct connections) access method works through the user's own PC. Downloading any data through this connection has been technically prohibited. No software has to be downloaded to be able to use this method. Some host institutional security settings may prohibit Gateway access. The RDCIT browser detection script will be able to verify if the Gateway connection is supported (http://rdcit.org/rdcit-secure-data-service/).

3.3.1.3 SSL VPN

SSL VPN (also called traffic-light system) allows the user to work from their own PC and establish a secure connection to the RDCIT Secure Data Service. This type of access is dependent on the ability to download and install software on the host PC as well as host institutional security settings.

3.3.2 Research Passport/Letter of Access

For non-NHS staff we request that users carrying a letter of Access (LOA) or Research Passport (RP) submit a copy of their LOA/RP. Any user with access to data across more than one site (including NHS staff) will be required to have a LOA/RP with Cambridge University Hospitals NHS Foundation Trust (CUH). Note that this process takes a couple of weeks so please contact the RDCIT to start this process as early as possible.

3.3.3 Hosting

All data stored in OCplay and OCtrain is hosted on a server in the School of Clinical Medicine at the University of Cambridge. The OCplay and OCtrain instances are not certified to hold any patient data. The instances are available purely to learn to use OpenClinica and to test Case Report Form (CRFs). Please email the RDCIT support team to obtain an OCplay account. The OCplay instance is deleted every 6-months, users will be notified.

The RDCIT Secure Data Service is hosted by AIMES Grid Services C.I.C (HSCIC registration H8121), based in Liverpool.

3.3.4 Ethics & Consent

The principal investigator of a study retains full responsibility for all study ethics & consent procedures even if the relevant documents are stored within the RDCIT Secure Data Service. Contact the RDCIT team directly if you need any advice regarding Ethics or Consent.

3.3.5 Patient Identifiable Data

Patient Identifiable Data (PID), including but not limited to patient contact details, should be preferentially stored in CiviCRM rather than in OpenClinica. How to store PID will need to be discussed on a study-per-study basis, please contact the RDCIT Team (or RD-TRC Data Co-ordinators). If a decision is made to store PID in OpenClinica studies the standardised (RDCIT-owned) PID CRF has to be added to the study. This CRF should not be customised by the study. If changes are needed, please contact the RDCIT.

3.3.6 Data sharing

Any request for data sharing will be handled separately through the appropriate programmes data access agreement.

3.4 Good neighbour policy

We encourage studies to use OpenClinica as their primary data collection tool but acknowledge that there will be cases where this is not feasible or practical; e.g. study where a third-party system is well established or if the study is nearing the end of recruitment.

Studies that have chosen not to use OpenClinica for their data collection fall under the *Good Neighbour Policy*. The Good Neighbour Policy also applies if a study uses a locally (to their institution) installed version of OpenClinica. In short the Good Neighbour Policy states that all studies are required to submit their study data files in a format compatible for import into OC import. Studies will still need to comply with the same data quality and clinical coding standards expected of RDCIT hosted studies (section 4 – Clinical Coding). For guidance on how to prepare a dataset for import into OpenClinica see the OpenClinica Import & Export Guide.

Studies exclusively importing data from a third-party system will be using OC import in the RDCIT Secure Data Service. Live data entry is not supported on this OpenClinica instance.

3.5 OpenClinica Data Import

OpenClinica uses the CDISC ODM XML as their operational data model. The model includes the clinical data along with its associated metadata, administrative data, reference data and audit information. OpenClinica allows you to create a dataset and export/download the dataset in a number of different formats (e.g. CDISC ODM XML, HTML, Excel, Tab-delimited).

Importing data into OpenClinica is possible but only if the importer data is in CDISC ODM XML format.

3.5.1 OpenClinica Data Importer Web Application (ODIN)

The RDCIT developed OpenClinica Data Importer Web Application (ODIN) is a browser-based tool that makes it possible to import subjects to OpenClinica from a comma separated values (csv) file, schedule events for the subjects and import data from the .csv file. The application will transform the provided .csv data into a CDISC ODM XML structure ready for import into OpenClinica.

ODIN (and the ODIN user documentation) can be downloaded from the RDCIT website (http://rdcit.org/openclinica/rdcit-tools/). See the OpenClinica Import&Export Guide (http://rdcit.org/user-documentation/)

Hands-on training courses data import are available on a regular basis. Contact the RDCIT to request upcoming dates or to register on a course.

3.5.2 Data Import into RDCIT Secure Data Service

Studies exclusively importing data from a third-party system will be using OC import in the RDCIT Secure Data Service. Live data entry is not supported on this OpenClinica instance. OC live allows a small amounts of data import, mainly live data entry is supported in OC live.

4 Clinical Coding & Naming Conventions

Using a standard clinical terminology allows clinicians to record the participant's information in a consistent manner making communication between health professionals (across different systems) more effective and unambiguous.

Within the RDCIT we ask that HPO is the primary ontology used and SNOMED CT is used for anything that cannot be found in HPO. This does not preclude any study from *additionally* coding to other standards and ontologies (e.g. OMIM, ICD-0-3).

4.1 Data Plan Design

As part of the process of setting up a study in the RDCIT Secure Data Service, studies are asked to prepare a data plan. The data plan is the preliminary work for setting up your OpenClinica CRFs and Study.

Please refer to the RDCIT_DataPlanDesign for further detail on developing a Data Plan and how to integrate Clinical Coding into the OpenClinica CRF. All clinical ontology (re)coding for data items must be confirmed and signed off by a clinician related to the study who is a specialist in the specific disease area, as well as by the RDCIT Team.

4.2 RDCIT Data Dictionary

To improve standardisation across the different projects the ontology codes already identified by studies in the RDCIT are held in the RDCIT Data Dictionary which can be searched by using the online tool Data Plan Composer.

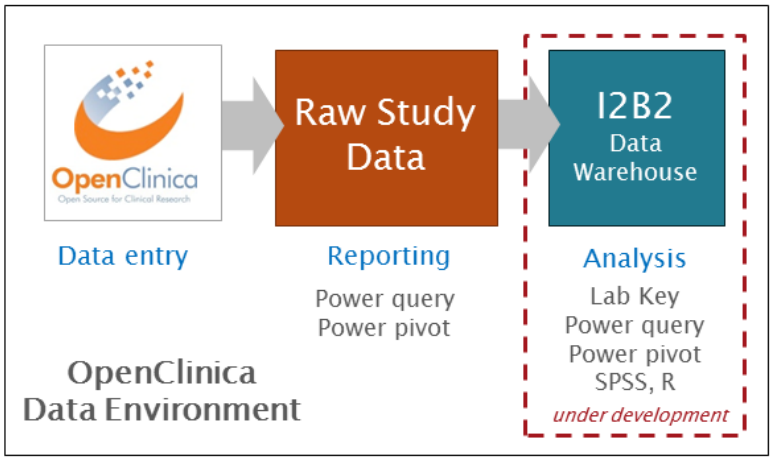
Please refer to the RDCIT_DataPlanDesign for further detail on the functionality of the RDCIT Data Plan Composer.

4.3 RDCIT Data Warehouse

The RDCIT OpenClinica Data Environment has three main components (all housed within the RDCIT Secure Data Service);

- OpenClinica: handles the data entry,
- Raw Study Data: the second component is a Reporting Data Warehouse, primarily used for mid-study reporting (e.g. recruitment figures).
- I2B2: third component is the true Data Warehouse, which is used for data analysis.

Only clinically coded data will be housed in the I2B2 Data Warehouse. Clinical coding dependant on the RDCIT Data Dictionary, as discussed below, is applied before data is moved to the I2B2 Data Warehouse.



SCHEMATIC OVERVIEW OF THE RDCIT OPENCLINICA DATA ENVIRONMENT

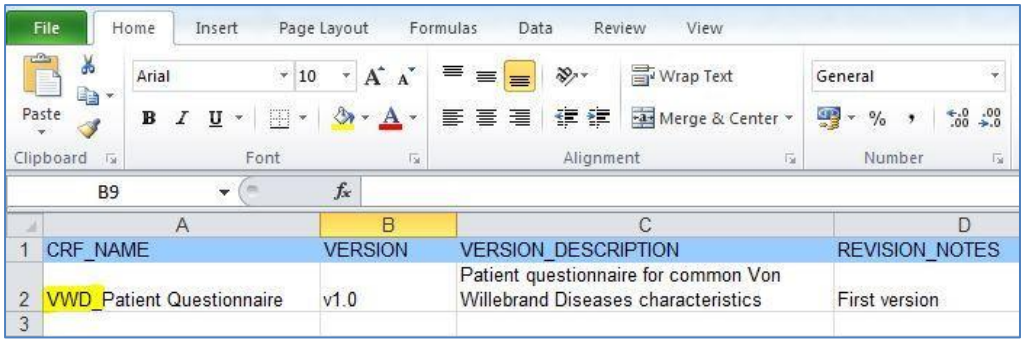
4.4 Field naming convention (Studies, Events, CRFs and Items)

4.4.1 Studies

When creating your study, the **Unique Protocol ID** must be your three character project code agreed with the Sequencing and Informatics Committee (SIC) or the RD-TRC/RDCIT.

4.4.2 CRFs

When saving your CRF at the design stage in MS Excel, it is not necessary, but you will find it easier to save your Excel file with the same name as your CRF, which is defined on the CRF sheet in cell A2.



The first 3 characters of a CRF name must be the three-character project code (Unique Protocol ID) (above example is VWD_ for the project Von Willebrands Disease). This avoids any overlap between studies when naming CRFs and makes it clear which CRFs are being used on your study. The prefix COZ is reserved for common Clinical Data CRFs (e.g. COZ_PID for Patient Identifiable Data). The Z is reserved for process data CRFs (e.g. Z_Sample_process for common Sample Process CRF). Both naming convention CRFs will be applied to all studies.

4.4.3 Data Items

Where possible we recommend that studies use standardised item names. The standardisation will be based on the RDCIT Data Dictionary. Please refer to the Data Plan Design Guide for more information.

4.4.4 Events

Where possible keep your event names short. Long event names risks expanding the subject matrix showing events beyond the standard screen width, with the need to scroll sideways.

Note: Events with associated study level (hidden) CRFs are still shown on the subject matrix, but clinician staff at site level will still not be able to access this data.

4.4.5 Rules

The rule OIDs are recommended to be the prefixed with the three-character project code (unique protocol ID) as rule OIDs need to be unique within a server (e.g. AAA_LXORLTICK includes the prefix 'AAA').

5 Study set-up

5.1 Study authorisation

Prior to being given access to the RDCIT Secure Data Service the study Principal Investigator (PI) will need to complete the RDCIT Study Authorisation form to certify that they wish their study to be housed in the RDCIT Secure Data Service (hosted by Aimes Grid Services C.I.C; registration H8121). The PI will need to confirm that all users associated with the study will conform to RDCIT Best Practice.

Study authorisation form is available on the RDCIT webpages

5.2 Authorising a designate

The study PI can choose to authorise a designate (OpenClinica Data Manager) for the Study to handle study setup and all further communication and user authorisation. The study authorisation form allows the PI to select which activities the designate should be authorised to do (Study Setup / User authorisation / Study Reporting and Monitoring Requests / Patient withdraw).

Users who have access to data across more than one site will need to complete a Research Passport/Letter of Access with Cambridge University Hospitals NHS Foundation Trust (CUH).

5.3 Study and site contact details

For each study, the Principal Investigator details and institutional address, needs to be included in the *Study Description and Status* and *Facility Information* section. For each site set up within the study, separate contact details of the Lead consultant and site address details, must be added. See the 'RDCIT Study Setup Guide' for detailed steps.

5.4 Study Parameter Configuration

The study setup process in OpenClinica has 2 main phases: the initial study setup (completed by RDCIT Team) and full study configuration (completed by the Study). The Study Parameter Configuration options required for

RDCIT compatible studies are shown below. Please see the 'RDCIT Study Setup Guide' for full details on how to set up your study.

Parameter	Value	Note
Collect subject Date of Birth	Yes	Full date or year as determined by study protocol. [note: year of birth collection sets birth date to 01.01.xxxx]
Allow Discrepancy Management	Yes	Required for queries on patient records and phenotypic data entered into OpenClinica
Sex required?	-	Decision of study, can be a CRF item
Person ID required?	Optional	NHS number in OC live / OC import
Show Person ID on CRF header	No	
How to generate the Study Subject ID	Manual entry	Required to enable data import using ODIN
When performing data entry Interviewer name required?	-	Recommended to only use if interviewer is different from OpenClinica User. If the same this is unnecessary as the username is logged
Interviewer name default as blank	-	Decision of study; dependent on previous setting
Interviewer name editable	-	Decision of study; dependent on previous setting
Interview date required	-	Decision of study
Interview date default	-	Decision of study; dependent on previous setting
Interview date editable	-	Decision of study; dependent on previous setting
Secondary Label viewable	-	Decision of study (this changes visibility on the study subject record)
Forced reason for change	Yes	Required for full audit trail
Event location required	Not used	'Not used' option required to enable data import using ODIN

5.5 Case Report Form (CRF) Design

It is not possible to delete CRFs from OpenClinica entirely. During the design phase OpenClinica can quickly become cluttered with test copies of CRFs. You will need to create, review and modify your CRFs in OCplay (or a local version of OpenClinica). Once you have verified your CRFs are accurate and ready to collect data, upload them to the secure environment (OCDesign) for user testing with real patient data, and then you can finally upload the CRFs to the live environment (OClive or OCimport) for live data entry or data import. If a CRF has to be amended after user testing, OCDesign (or OCplay) is available to all studies for this purpose. Under no circumstances should real patient data be entered into OCpaly.

The quality of the study data relies on the quality of data collection instruments i.e. the CRFs. When designing your CRF(s) in OpenClinica, ensure that you collect the data as specified by the study protocol and Data Plan. Avoid collecting any extraneous "just in case" data. Ensure the data collected is in the precise format requested (e.g. measurement scales, correct decimalisation point, date format etc.) using data input validation.

Refer to the CRF Design Guide and RDCIT Data plan Design Guide for detailed instructions

5.5.1 CRF version management

When uploading a new version of a CRF to OpenClinica (OCplay) ensure:

- Major version changes where data is impacted i.e. v2.0 to v3.0
- Minor version change for cosmetic changes i.e. v2.0 to v2.1
- Keep versions to two levels v2.1 or v2.2. So no versions 2.11 or 2.20
- Version format in the CRF includes the letter v - so v1.0 not 1.0

Change the version number of the new CRF in the CRF tab before uploading or delete the currently uploaded version before uploading.

Why this is important: OpenClinica systematically tracks version, however it does not restrict people designing CRFs to a specific version format. A common format across all studies will facilitate version tracking in the analysis phase.

In addition events are associated with CRF versions. Uploading a new CRF version does not automatically update the version being used by the event. This is a second separate step. As all CRFs are held in a common list, this avoids the risk of another study loading a new version of a CRF you are using and it being automatically applied to your event records.

5.5.2 CRF version in the live environment

It is possible to upload a new version of a CRF to OCdesign when there are minor changes in the CRF, and it should follow the same rules when upload a new version of CRF in OCplay. After testing the CRF in OCdesign, the CRF starts from Version 1.0 when uploading the CRF to OClive or OCimport in order to keep a clean OpenClininca instance in the live environment. If there is a requirement to upload a new version of CRF in the live environment, the next version will be V2.0 no matter major changes or minor changes in the CRF. The version pattern follows like this: V1.0 to V2.0, or V2.0 to V3.0, etc.

After uploading the CRF to the OClive or OCimport, and the study is up running in OClive or OCimport, you will need to remove the CRF or the entire study from OCdesign to avoid the CRF version confusions.

5.5.3 CRF Batch migration

Batch CRF version migration is introduced in OpenClinica version 3.12. this allows you to migrate from one version of a CRF to another for a group subjects and/or events. Please read CRF Migration for more details

6 Clinician and patient records

6.1 Site and study level security

In OpenClinica patients are called *subjects* and clinicians are called *users*.

Clinicians:

- Assign users (clinician collecting data) to the correct study site level in order that the appropriate site clinicians can view and edit patient records under their care.

Patients:

- Must all be added at site level. Patients can be re-assigned from study level down to a site, they cannot be moved back up to study level
- A subject can be reassigned between sites on the same study if necessary.
- A subject assigned to a site still remains visible at study level to the Study Manager and Study Director.
- To remove a subject, please add a discrepancy note on the Person ID field (for audit trail purposes) before removing the subject.

CRFs:

- Mark the data items as "**PHI**" for those Protected Health Information. Please read Protected Health Information for more details
- Put any protected data in a study level CRF. A study level CRF is created by 'hiding' the CRF from sites when adding it to an event, which makes the data visible at study level only and invisible to site level users.

Why this is important: OpenClinica Sites provides a security model built on *user roles*, *sites* and *CRF visibility*. OpenClinica roles are divided between study level and site level users. User roles define what a user is able to do, but do not limit what users are able to view i.e. all patient data is visible to all users, unless both patients and clinicians are assigned to sites. Users must be set up at site level and subjects assigned to a site in order to maintain patient confidentiality between study sites.

6.2 Clinician User Accounts

Accounts on RDCIT hosted secure OC instances will be created for each study by the RDCIT support team. The sections below provides an overview of the procedures and information required to create OC user accounts.

6.2.1 RDCIT User authorisation form

For security reasons all requests for user accounts on the live OpenClinica environments (OC live / OC import) will need to go through the Principal Investigator of the study (or by the PI authorised designate).

During the initial study setup the RDCIT User authorisation form should be used to request the initial user account needed for the study (includes all users at study and site level). Any further user account request can be made via a separate RDCIT User Authorisation form but needs to be signed by the PI/Designate. The form will be updated by the RDCIT.

The RDCIT User authorisation form is available on the RDCIT webpages

If access is no longer needed for any user on the study the authorisation form (or email) should be used to revoke the user's access; this should be done without delay.

User account sharing is strictly prohibited. If a user has forgotten his or her username/password a request will need to be sent to the RDCIT support team to retrieve it. For security reasons, users that frequently ask for a password reset, may have his or her account disabled permanently. Therefore, it is vital to have a memorable and secure password.

OpenClinica field	Example	Note
Username	jsmith	Where possible the username will be the local-part of the users institutional email address (e.g. jsmith for jsmith@institution.com). OC requires at least 5 alphanumeric or underscore characters and does not accept a full stop in the username. Full stop in the local-part will be omitted.
First / Last name		Please provide full names, not initials
Email		Provide clinicians primary NHS email account if they have more than one. For non-clinical users provide main institutional email account
Institutional affiliation	University of Cambridge	Main institution (matching the primary email address)
Role		See below section on user roles
User type	User	Always set accounts to 'User'. Setting accounts to 'Business Administrator' overrides role permissions to give system level access to the study
Authorise SOAP services	No	Set to 'No' as default. A "local" OpenClinica account will be necessary to bulk importing data using ODIN

Why this is important: All changes and updates made to a patient's record in OpenClinica are logged in the audit log against the clinician entering the data. If there is a query on the data or additional data is required, the clinician who entered the data will be the person queried. All queries are raised through Discrepancy Notes generated by OpenClinica which will automatically be sent to the email account provided. Providing the GMC number enables us to trace the clinician within the NHS should they move institution and change email address.

6.2.2 User Roles

The only users at study level should be the Study Director (Principal Investigator) and the Data Manager (person responsible for configuring the study in OpenClinica. Referred to as designate if additionally authorised by the PI).

If a site has more than one lead consultant, and they do not share patients, then create a site for each consultant e.g. BROMPTON: Jones; BROMPTON: Smith. The Clinical Research Coordinators can be assigned to both sites. There are duplicate roles in OpenClinica (same privileges, different name), we anticipate studies using the following User Roles in OpenClinica.

User Role	Study / Site Level	Note
Study Director	Study	Principal Investigator (PI)
Data Manager	Study	Person configuring the study and designing CRFs. Designate if authorised by PI
Clinical Research Coordinator or Data Entry Person	Site	Clinician/data coordinator entering data

A full overview of all the OpenClinica user roles is available on the OpenClinica webpages (here).
Please discuss with the RDCIT Team directly if you would like to use additional roles on your study.

6.2.3 Passwords

We have implemented all necessary requirements on password security within OpenClinica. In addition, there are a number of recommendations that we strongly suggest users heed when creating their passwords, although these are optional.

Mandatory?	Policy
Yes	Password length must be at least 8 characters
Yes	Passwords should contain a mixture of uppercase and lowercase letters, numbers **and symbols.** Note that whitespace is considered a symbol, as well as punctuation marks
Yes	Password requires changing after first log-in
Yes	A user is allowed 4 attempts at a password before being locked out. After being locked out, an administrator must re-set the password
	Passwords should never contain the username of the individual or be a recognisable dictionary word.

The above policies apply to OCplay and OCdesign only. For the live environment (OClive and OCimport), the RDCIT support team create an user's account with a default password, and the user needs to reset the password via RDCIT Secure Data Service – Self-service password. The password should contain at least one upper case, one lower case, one number and one symbol to meet requirements of complexity.

6.2.4 Patient records (subjects)

Patient confidentiality is maintained by assigning patients to a site and restricting access to their record to clinicians within that site. NOTE: A patient who is on multiple studies will have a record on each study site. The NHS number will provide the linkage to the patient's record on different studies. The user doing the original data entry is logged in the audit trail and will be the person queried (using inbuilt OpenClinica discrepancy notes) for any clarification on the patient record.

Field	Use for	Note
Person ID	NHS number	In OC live / OC import this must be the NHS number
Study Subject ID	Study specific ID or the GEL Participant ID	Study Subject ID can be allocated as the study determines. Name, initials, code or a combination. If the study is recruiting as part of the 100,000 Genomes Project then the Study Subject ID must be the Participant ID from Genomics England.
Secondary ID	Family identifier	If the study is recruiting as part of the 100,000 Genomes Project then the Family ID from Genomics England must be used.
Date of birth	Required	Full date or year as determined by study protocol. [note: year of birth collection sets birth date to 01.01.xxxx]
Subject Group Class	Optional	Subject Group Class can be used to capture patient subcategories

Bulk data import using ODIN is possible for studies already holding large numbers of patient records (See the 'Import and Export Guide').

Where studies include non NHS England and Wales patients (i.e. international patients), Principal Investigators need to ensure a truly unique Person ID is used. Please discuss with the RDCIT support team.

Why collecting the NHS number is important: The purpose of the NHS number is to provide a unique medical identifier. The NHS number is designed specifically for the task of being a central index of patients. Outside of the NHS core systems, the NHS number is restricted to registered medical professionals; there is no online 'look up' facility of NHS numbers. Using the NHS number as the OpenClinica Person ID enables us to check for uniqueness across studies and link records for patients who are on more than one study. When OC live links into NHS systems, the NHS number will be fundamental for deep phenotyping patient records in OpenClinica.

6.2.5 Patient Pedigrees

Pedigree relationship – If your study collects pedigree information we recommend that you have a look at PhenoTips, please contact RDCIT for more details.

7 Testing procedure

After developing a data plan and CRFs for a specific study, you will move to the next step: Testing CRFs. Here are the broad guidelines on how to test CRFs:

Step 1: Operational Testing (OCplay only)

- Add dummy data to cover all the CRF options
- Create and export a dataset
- Check that all codes have been correctly implemented, else, please go back to modify the CRF until all codes are brought into OpenClinica

Note: The Data coordinator managing the study can request feedback from RDCIT about the CRFs and the Data plan

Step 2: Usability Testing (OCdesign only)

- Upload the CRFs to OCdesign and create the study. Request feedback from RDCIT regarding the study setup
- Give access to selected users to test the CRFs with real clinical data
- Collect all the feedback from users and modify the CRFs based on the feedback
- Create and export a dataset
- Test data mapping for import studies
- Formal sign-off of the final CRFs by RDCIT and the study team

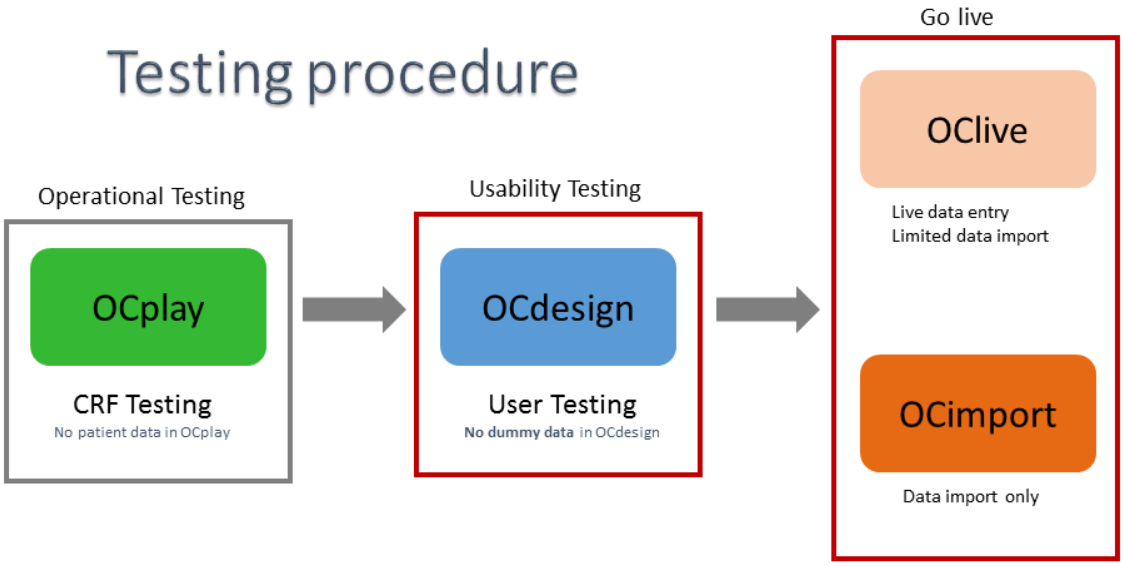
Note: No dummy data in OCdesign, Best practice sign-off should be done by RDCIT in OCdesign

Step 3: Go live

Once the CRFs have had final sign – off by both RDCIT and the study team, you can move the study to OClive

- Create the study in OClive (live data entry and limited data import) or OCimport (Data import only) based on the Best practice sign- off
- Request user account from RDCIT
- Go live

Note: Prior to the study going live the RDCIT Study Authorisation form and User Authorisation form will need to be completed by the PI.



8 Appendix

The Best Practice Guide has been set up to provide you with the necessary data structure that if implemented within your study, will meet the requirements as set out by Genomics England. In the below appendix the Genomics England minimum data specification is given as well as the Standards Lookup Codes, NHS speciality codes and standard item conventions that should be used if possible.
As the NIHR Rare Diseases studies need to be aligned to the Genomics England infrastructure, its OpenClinica database is expected, at the very least, to hold the minimum amount of data that has been stipulated by Genomics England in order to provide a level of consistency across the entire program and in-depth phenotyping for patients with rare diseases.

8.1 Genomics England minimum data specification

The Genomics England minimum data specification can be viewed on the link below

http://rdcit.org/genomics-england-minimum-data-specification/

8.2 Standard Lookups Codes

8.2.1 Information Source

CODE	DESCRIPTION
1	EHR (electronic health records)
2	PATIENT QUESTIONNAIRE
3	CIVICRM (participant management system)
4	PAPER
5	MS EXCEL
6	MS ACCESS

8.2.2 NHS Number Status

CODE	DESCRIPTION
01	Number present and verified
02	Number present but not traced
03	Trace required
04	Trace attempted - No match or multiple match found
05	Trace needs to be resolved - (NHS Number or Patient detail conflict)
06	Trace in progress
07	Number not present and trace not required
08	Trace postponed (baby under six weeks old)

8.2.3 Family Relationship

CODE	RELATIONSHIP TO PROBAND	SNOMED-CT CODE
1	Parent	444294004
2	Mother	65656005
3	Father	9947008
4	Sibling	60614009
5	Brother	60614009
6	Sister	73678001
7	Twin (identical)	313415001
8	Twin (non-identical)	313416000
9	Maternal Aunt	442051000124109
10	Maternal Uncle	442031000124102
11	Maternal Grandmother	394859001
12	Maternal Grandfather	394857004
13	Paternal Aunt	442061000124106
14	Paternal Uncle	442041000124107
15	Paternal Grandmother	394858009
16	Paternal Grandfather	394856008
17	First Cousin	4577005
18	Other - blood relative	125679009

8.2.4 Feedback Options for Participant

CODE	DESCRIPTION
1	Pertinent Findings
2	Incidental Findings
3	Both
9	Unknown

8.2.5 Certainty of Phenotype Observation

CODE	DESCRIPTION
	To be confirmed

8.2.6 Prominence of Phenotype

CODE	DESCRIPTION
0	Not
1	Secondary
2	Mild
3	Moderate
4	Severe

8.2.7 Mode of Inheritance

CODE	DESCRIPTION	HPO CODE
1	Autosomal Dominant Inheritance	HP:0000006
2	X-Linked Dominant Inheritance	HP:0001423
3	Autosomal Recessive Inheritance	HP:0000007
4	X-Linked Recessive Inheritance	HP:0001419
5	Complex Inheritance	HP:0001426
6	Unknown	

8.2.8 Status of Family Member

CODE	DESCRIPTION
1	Affected
2	Unaffected
3	Known Carrier

8.2.9 Related to Probands Condition?

CODE	DESCRIPTION
1	Related
2	Unrelated
9	Not known
0	Not applicable

8.2.10 Death Location

CODE	DESCRIPTION
1	Hospital
2	NHS hospice / Specialist Palliative Care Unit
3	Voluntary hospice / Specialist Palliative Care Unit
4	Patients own home

5	Care Home
6	Other

8.2.11 Participant Sampling Type

CODE	DESCRIPTION
1	Singleton
2	Duo
3	Trio

8.3 NHS Speciality Codes

Surgical Specialities	
100	GENERAL SURGERY
101	UROLOGY
110	TRAUMA & ORTHOPAEDICS
120	ENT
130	OPHTHALMOLOGY
140	ORAL SURGERY
141	RESTORATIVE DENTISTRY
142	PAEDIATRIC DENTISTRY
143	ORTHODONTICS
145	ORAL & MAXILLO FACIAL SURGERY
146	ENDODONTICS
147	PERIODONTICS
148	PROSTHODONTICS
149	SURGICAL DENTISTRY
150	NEUROSURGERY
160	PLASTIC SURGERY
170	CARDIOTHORACIC SURGERY
171	PAEDIATRIC SURGERY
180	ACCIDENT & EMERGENCY
191	PAIN MANAGEMENT (Retired 1 April 2004)
Medical Specialties	
190	ANAESTHETICS
192	CRITICAL CARE MEDICINE
300	GENERAL MEDICINE
301	GASTROENTEROLOGY
302	ENDOCRINOLOGY
303	CLINICAL HAEMATOLOGY
304	CLINICAL PHYSIOLOGY
305	CLINICAL PHARMACOLOGY
310	AUDIOLOGICAL MEDICINE
311	CLINICAL GENETICS
312	CLINICAL CYTOGENETICS and MOLECULAR GENETICS (Retired 1 April 2010) *
313	CLINICAL IMMUNOLOGY and ALLERGY
314	REHABILITATION

315	PALLIATIVE MEDICINE
320	CARDIOLOGY
321	PAEDIATRIC CARDIOLOGY
325	SPORT AND EXERCISE MEDICINE
326	ACUTE INTERNAL MEDICINE
330	DERMATOLOGY
340	RESPIRATORY MEDICINE (also known as thoracic medicine)
350	INFECTIOUS DISEASES
352	TROPICAL MEDICINE
360	GENITOURINARY MEDICINE
361	NEPHROLOGY
370	MEDICAL ONCOLOGY
371	NUCLEAR MEDICINE
400	NEUROLOGY
401	CLINICAL NEURO-PHYSIOLOGY
410	RHEUMATOLOGY
420	PAEDIATRICS
421	PAEDIATRIC NEUROLOGY
430	GERIATRIC MEDICINE
450	DENTAL MEDICINE SPECIALTIES
451	SPECIAL CARE DENTISTRY
460	MEDICAL OPHTHALMOLOGY
500	OBSTETRICS and GYNAECOLOGY †
501	OBSTETRICS
502	GYNAECOLOGY
504	COMMUNITY SEXUAL AND REPRODUCTIVE HEALTH
510	ANTENATAL CLINIC (Retired 1 April 2004)
520	POSTNATAL CLINIC (Retired 1 April 2004)
560	MIDWIFE EPISODE
600	GENERAL MEDICAL PRACTICE
601	GENERAL DENTAL PRACTICE
610	MATERNITY FUNCTION (Retired 1 April 2004)
620	OTHER THAN MATERNITY (Retired 1 April 2004)
Psychiatry	
700	LEARNING DISABILITY
710	ADULT MENTAL ILLNESS
711	CHILD and ADOLESCENT PSYCHIATRY
712	FORENSIC PSYCHIATRY
713	PSYCHOTHERAPY
715	OLD AGE PSYCHIATRY
Radiology	
800	CLINICAL ONCOLOGY (previously RADIOTHERAPY)
810	RADIOLOGY
Pathology	
820	GENERAL PATHOLOGY
821	BLOOD TRANSFUSION
822	CHEMICAL PATHOLOGY

823	HAEMATOLOGY
824	HISTOPATHOLOGY
830	IMMUNOPATHOLOGY
831	MEDICAL MICROBIOLOGY AND VIROLOGY
832	NEUROPATHOLOGY (Retired 1 April 2004)
833	MEDICAL MICROBIOLOGY (also known as MICROBIOLOGY AND BACTERIOLOGY)
834	MEDICAL VIROLOGY
Other	
900	COMMUNITY MEDICINE
901	OCCUPATIONAL MEDICINE
902	COMMUNITY HEALTH SERVICES DENTAL
903	PUBLIC HEALTH MEDICINE
904	PUBLIC HEALTH DENTAL
950	NURSING EPISODE
960	ALLIED HEALTH PROFESSIONAL EPISODE
990	JOINT CONSULTANT CLINICS (Retired 1 April 2004)
Notes:	
†	Code 500 is not acceptable for Central Returns including Hospital Episode Statistics
*	Code 312 is retained for CONSULTANTS qualified in this Main Specialty prior to 1 April 2010.

8.4 Example standard item conventions

STANDARD OPTIONS (Source reference: RD-TRC standards)

ITEM NAME	OPTION	VALUE
GENERIC RESPONSE	Yes	1
	No	0
	Unknown	9
	Positive	1
	Negative	0
	Never Used	3
	Unsure	9

SUBJECT / PERSON DATA OPTIONS (Source reference: http://www.datadictionary.nhs.uk)

ITEM NAME	OPTION	VALUE
SEX AT BIRTH	Unknown	0
	Male	1
	Female	2
	Not specified	9
ETHNICITY	White - British	A
	White - Irish	B
	White - Other	C
	Mixed - White & Black Caribbean	D

	Mixed - White & Black African	E
	Mixed - White & Asian	F
	Mixed - Other	G
	Asian - Indian	H
	Asian - Pakistani	J
	Asian - Bangladeshi	K
	Asian - Other	L
	Black - Caribbean	M
	Black - African	N
	Black - Other	P
	Chinese - Other	S
	Not Stated	Z
MARITAL STATUS	Single	S
	Married/Civil	M
	Divorced / Person whose Civil Partnership has been dissolved	D
	Widowed/Surviving Civil Partner	W
SMOKING STATUS	Current smoker	1
	Ex-smoker	2
	Non-smoker - history unknown	3
	Never smoked	4
	Not Stated	Z