



## SAIL - Person-level Export File Structure & Data Transfer

### Introduction

The SAIL Databank is a central repository containing anonymised person-level and address-level data drawn from operational and national systems. Using a novel anonymisation process, the SAIL Databank links datasets together to form a rich information base which is a national resource for e-health research and evaluation. All datasets are securely transferred into SAIL using the “Split-file” process with the support of the NHS Wales Informatics Service our trusted third party. During this process person-level demographics are translated to an Anonymous Linking Field (ALF). This document describes the required file formats for person-level data and methods of data transfer.

### Split-file Process

The original dataset is split into two types of files:

1. “File 1” dataset containing sensitive person-level demographics data which is sent to NHS Wales Informatics **and not sent to the SAIL Team**. “File 1” data is processed by NHS Wales Informatics who match and anonymise the data and then send it to us.
2. “File 2” containing clinical data or other non-identifiable data and is sent directly to us **and not sent to NHS Wales Informatics**.

**File 1: (Person Identifiable Data)** contains unique person-level information.

The table below describes the required file structure for the “file 1” that will be sent to NHS Wales Informatics for matching and anonymisation. *(To be delivered to NHS Wales Informatics only)*

Field Name	Data Type	Description
SYSTEM_ID	varchar(50) (Unique)	Unique identifier
NHS_NUMBER	varchar(10)	NHS Number ( with no spaces)
SURNAME	varchar(50)	Surname
FORENAME	varchar(50)	Forename
ADDRESS_1	varchar(255)	Address
ADDRESS_2	varchar(255)	Address
ADDRESS_3	varchar(255)	Address



Field Name	Data Type	Description
ADDRESS_4	varchar(255)	Address
ADDRESS_5	varchar(255)	Address
POSTCODE	varchar(8)	Post Code, where possible in formal space separated format i.e. 4 & 3 = "YYYY ZZZ" 3 & 3 = "YYY ZZZ" 2 & 3 = "YY ZZZ"
DATE_OF_BIRTH	varchar(10) dd.mm.yyyy (EUR format)	Date of Birth: dd.mm.yyyy (EUR format)
GENDER_CD	char(1) Where 1=male, 2=female, 8= not specified, 9=unknown	Gender in codified format as 1=male, 2=female, 9=unknown
CREATE_DATE	Varchar(20) Ideally dd.mm.yyyy (EUR format) (without time) But flexible e.g. will take text version of date such as '10 <sup>th</sup> August 2007'	File creation date: dd.mm.yyyy (EUR format)
FIELD_1	char(1)	Additional non-identifiable data
FIELD_2	varchar(50)	Additional non-identifiable data
FIELD_3	varchar(50)	Additional non-identifiable data
FIELD_4	varchar(50)	Additional non-identifiable data

SYSTEM\_ID is the unique join key that will be used to link the final two files back together. This join key can be generated as part of the split or an existing unique field can be used. This field can be generated by simply creating a unique number for each row.

**File 2: (Clinical or other non-identifiable Data)** This comprises of a delimited extract for all the tables containing clinical information or other non-identifiable data.

The required file structure for the "file 2"s that will be sent to the SAIL team.  
(To be delivered to SAIL only)



Field Name	Data Type
SYSTEM_ID	varchar(50) (Unique)
.....	Data provider specific information

## Formatting preferences:

1. File Names should follow following naming convention. Please contact your project lead for SAIL Project Number.  
<tablename>\_<sail project number>\_<todays dateYYYYMMDD>.csv  
e.g. GPREADCD\_0230\_20140112.csv
2. Data presented in csv (comma delimited file) file format. For massive data quantities, this format is most suitable.
3. Character fields enclosed in double quotes
4. NHS Number should not contain any spaces. If your exported csv file has spaces please open it in excel and use this code below, to remove spaces on all rows.
  - a. Open excel, Select NHS number column -> then go to Home -->
  - b. Find & Select --> Replace
  - c. In "Find What" box enter single space --> Keep "Replace with" box empty --> Replace All
  - d. Save file.

## Data Transfer

**Never send data files via e-mail.** E-mail is not sufficiently secure for transferring potentially sensitive personal information.

### Method 1:

For File 1, secure electronic data transfer facility at NHS Wales Informatics is available.

If your organisation is on the NHS DAWN / NHS network in England, use website: <https://nwdss.wales.nhs.uk/NwdssSFU/>

If your organisation is not within NHS network use website: <https://www.nwdss.wales.nhs.uk/NwdssSFU/sfuLogin.aspx>

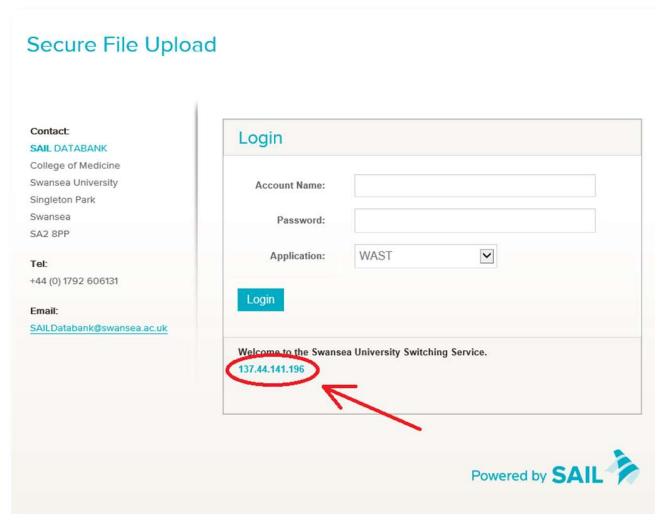
The Data Acquisition Team at NHS Wales Informatics can set up an account for new users to upload location data.



For File 2, using secure file upload you can send “file 2” data files directly to SAIL. An account will be created for you, and you can login and upload files. Website: [https://ccs-hiru-fe1.swan.ac.uk/hiru\\_su/](https://ccs-hiru-fe1.swan.ac.uk/hiru_su/)

If you intend to use this method, please let us know the following details so that we can set up an account for uploading files to SAIL.

1. IP address(es) of the PC(s) that will be used when uploading the files as shown on the upload site.



2. Please provide a name, email address and phone number for the official contact within your organisation, regarding delivery of SAIL data.

### Method 2:

Alternatively if your organisation has a secure file download service, we could login to your website and download relevant data from there.

### Method 3:

If neither of the above methods are possible please contact the SAIL team to discuss alternative secure methods of file transfer.

### Data Dictionary

If a data dictionary is available, please e-mail a copy to the SAIL team.



## Key Contact Details

FOR FILE 1: (Identifiable Data)

Data Acquisition Team  
NHS Wales Informatics Service  
Tŷ Glan-yr-Afon  
21 Cowbridge Road East  
Cardiff  
CF11 9AD  
Email : [PDIT@wales.nhs.uk](mailto:PDIT@wales.nhs.uk)

FOR FILE 2: (Clinical Data or other Non-Identifiable Data)

Data Acquisition Team  
ADRC-Wales & SAIL Databank  
Swansea University Medical School  
Data Science Building  
Singleton Park  
Swansea  
SA2 8PP  
Email: [daq@chi.swan.ac.uk](mailto:daq@chi.swan.ac.uk)