



SAIL - Address-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 a Residential Anonymous Linking Field (RALF). 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 1R” dataset containing sensitive address-level demographics data which is sent to NHS Wales Informatics **and not sent to the SAIL Team**. “File 1R” data is processed by NHS Wales Informatics, who match and anonymise the data, and then send to us.
2. “File 2” containing environmental data or other non-identifiable data and is sent directly to us **and not sent to NHS Wales Informatics**.

File 1R: (Address Identifiable Data) contains unique address-level information.

The table below describes the required file structure for the “file 1R” 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 identifier
UPRN	Integer	UPRN (from Address Layer of OS Master Map)
OS_X	Decimal(31,8)	X Co-ordinate, British National Grid co-ordinate system.
OS_Y	Decimal(31,8)	Y Co-ordinate, British National Grid co-ordinate system
ADDRESS_1	varchar(255)	Address



Field Name	Data Type	Description
ADDRESS_2	varchar(255)	Address
ADDRESS_3	varchar(255)	Address
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"
ENV_METRIC_1	Varchar(50)	Environmental Metric 1 , Additional non-identifiable data related to a location
ENV_METRIC_2	Varchar(50)	Environmental Metric 2 , Additional non-identifiable data related to a location
ENV_METRIC_3	Varchar(50)	Environmental Metric 3, Additional non-identifiable data related to a location

Presence of RM_ADD_KEY or UPRN or / and OS_X, OS_Y will ensure good address to RALF matching. Please leave any of the unavailable fields blank.

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. You can generate this field by simply creating a unique number for each row.

File 2: (Environmental Metrics or non-identifiable data related to a location) It comprises of a delimited extract for all the tables containing environment metrics or non-identifiable data related to a location.

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)
.....	Other Environmental metrics, non-identifiable data related to location



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. ENVCD_0230_20140112.csv
2. Data present in csv (comma delimited file) file format. For massive data quantities, this format is most suitable.
3. Character fields enclosed in double quotes

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 1R, 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 a secure file upload you can upload files containing Environmental Metrics or non-identifiable data related to a location directly to HIRU. An account will be created for you, and you can login and upload files.

Website: 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.



Secure File Upload

Contact:
SAIL DATABANK
College of Medicine
Swansea University
Singleton Park
Swansea
SA2 8PP

Tel:
+44 (0) 1792 606131

Email:
SAILDatabank@swansea.ac.uk

Login

Account Name:

Password:

Application: WAST

Login

Welcome to the Swansea University Switching Service.
137.44.141.196

Powered by SAIL

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 1R: (Location Source File)

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

FOR FILE 2: (Environmental Metrics or non-identifiable data related to a location)

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