



# SAIL User Forum

January 20<sup>th</sup> 2022



# SAIL User Forum - January 2022

## Agenda

---

### 1. Welcome to SAIL user forum

- 1.1 Introduction to what the SAIL user forum moving forward (frequency of forums, new group and chair, type of content and structure of forum, contact and organisation).

### 2. Talks

- 2.1 Talk 1: One Wales collaboration / Con-Cov / SAIL's response to COVID-19.
- 2.2 Talk 2: Welsh Government / ADR Wales response to COVID-19. Kathryn Helliwell
- 2.3 Talk 3: Public Health Wales / Network Data Lab Wales response to COVID-19. Jiao Song

### 3. Data acquisition and meta-data team

- 3.1 Updates on new data sources / planned acquisition to SAIL.
- 3.2 Updates around meta-data and documentation on data available in SAIL.

### 4. Governance and legal team

- 4.1 The new guidance document from SAIL "The Do's, Do Nots and What to Check"
- 4.2 Updates on SAIL policies and user agreements

---

### 5. Analytical team

- 5.1 HDR Phenotype library.
- 5.2 "New PEDW" data.

### 6. Technical services

#### 7. Marketing and comms and SAIL business updates

- 7.1 Business team – Update to SAIL policy (amendments and extensions)
- 7.2 Marketing – Updates on good news stories and recent news.  
Marketing – Contact SAIL communications team if interested in completing a blog, news stories or other dissemination around your new funding / work / outputs.

### 8. Feedback

- 8.1 Survey link and ongoing feedback for user forums moving forward.
- 8.2 Contact us if you are interested in presenting at the next / future user forum.

### 9. AOB

---

# SAIL User Forum - January 2022

1. Welcome to SAIL user forum

Some general house keeping points:



SAIL User forum's will not be recorded.



All slides will be shared after the forum via confluence.



Q&A is welcome via the chat or at designated points over video/audio.



Please mute microphones when not asking questions or presenting.

# SAIL User Forum - January 2022

## 1. Welcome to SAIL user forum

### What to expect from the “new” SAIL user forums moving forward...



Forum frequency & timings – bi-monthly and on maintenance windows.



Forum chairs – Ashley Akbari & Chris Orton.



Forum structure & content – standing agenda, input from community.



Forum contact – questions/suggestions [sailbusinessadmin@chi.swan.ac.uk](mailto:sailbusinessadmin@chi.swan.ac.uk)

# SAIL User Forum - January 2022

## 2. Talks

Three talks (*15mins each*):

Talk 1: Ashley Akbari (*on behalf of the One Wales collaboration / Con-Cov / SAIL*)

Using linked data to respond to the COVID-19 pandemic in Wales

Talk 2: Kathryn Helliwell, Tony Whiffen & Lucy Robinson

Welsh Government / ADR Wales response to COVID-19.

Talk 3: Jiao Song

Public Health Wales / Network Data Lab Wales response to COVID-19.



# SAIL User Forum - January 2022

## 3. Data acquisition and meta-data team



Cynthia Mcnerney (*Data acquisition team*)

- Updates on new data sources / planned acquisition to SAIL.

Hannah Davies (*Meta-data team*)

- Updates around meta-data and documentation on data available in SAIL.

# SAIL User Forum - January 2022

## 3. Data acquisition and meta-data team

The screenshot shows the Health Data Research Innovation Gateway website. The header includes the logo and navigation links for 'About', 'Community', 'COVID-19', and 'Sign in | Sign up'. The main content area features a search bar with 'sail' entered and filter buttons for 'All Datasets', 'All tools', 'COVID-19', and 'Cancer'. Below the search bar, statistics are displayed: 56 registered users, 454 datasets, 37 access requests, 11 tools, and 3,303 searches in the last month. A 'BETA' notice indicates this is a new service. The search results page shows a list of datasets with filters on the left for License, Physical sample availability, Keywords, Publisher, and Age band. The results include 'Active Adult Survey', 'Annual District Birth Extract', 'Annual District Death Daily', 'Annual District Death Extract', 'Bowel Screening Wales', and 'Breast Test Wales'.

<https://healthdatagateway.org/>

The screenshot shows the SAIL Databank website. The header features the SAIL Databank logo and a 'Menu' icon. The main heading is 'SAIL DATASETS' with the tagline 'Amongst the richest population data in the world'. The text describes the richness of the data and its use in research. A section titled 'A list of all the datasets currently available for researchers to use can be found here via our Asset Manager\*' includes a thumbnail of the Asset Manager interface. Below this, the 'ASSET MANAGER' header is visible with 'About Help Contact' links. The 'SAIL Databank' header includes a 'Login' button. The 'CATEGORIES' section shows 'Show 48 entries' and a search bar. Three category cards are displayed: 'FAMILY JUSTICE' (2 items), 'ADMINISTRATIVE' (15 items), and 'HEALTH' (43 items).

<https://saildatabank.com/saildata/sail-datasets/>

# SAIL User Forum - January 2022

## 4. Governance and Legal team



Cynthia Mcnerney, Sharon Heys & Rob Garlick (*Governance & Legal team*)

- The new guidance document from SAIL “The Do’s, Do Nots and What to Check”  
<https://youtu.be/Rlk337W5JwY>  
<https://popdatasci.swan.ac.uk/wp-content/uploads/2022/01/SAIL-USER-DOs-and-DONTs-October-20211-1.pdf>
- Updates on SAIL policies and user agreements



# SAIL User Forum - January 2022

## 5. Analytical team



Hannah Davies and Dan Thayer (*Analytical team*)

- HDR Phenotype library.

<https://phenotypes.healthdatagateway.org/>

- To load phenotypes please contact the SAIL Analytical Team via the helpdesk.
  - Functionality coming soon to allow users to add themselves.
- 
- “New PEDW” data.

# Cleaned PEDW Dataset



Daniel Thayer

Senior Data Scientist – Team Lead, SAIL Analytical Services

# Challenges of using PEDW

- Normalized structure (5 tables)
  - Complexity: Harder to write queries.
  - Many joins = increased load on database.
- Measuring admissions not straightforward (spells, provider spells, person spells).

# Solution: Cleaned version

- Partially denormalized: more duplication, but greater simplicity, performance.
- Three standalone tables that cover most common uses.
  - Each contains ALF, other key information

Admissions

Diagnoses

Procedures

# Admissions

- One row per admission (person spell)
- Start and end date of person spell
- Admission info from first spell/episode in person spell
- Discharge info from last spell/episode in person spell
- Specialty, diagnosis information from both first and last episode available (commonly used to infer reason for admission etc.).

# Diagnosis

- One row per diagnosis code
- Includes ALF
- Start and end date of person spell
- Information about spell, episode, and diagnosis number, both:
  - Original values from PEDW (epi\_num etc.)
  - Computed sequences that are guaranteed to start at 1 and be sequential (epi\_seq etc.)

*Resolves known errors where certain spells might not have an episode #1, etc.*

# Procedure (similar to diagnosis)

- One row per procedure code
- Includes ALF
- Start and end date of person spell
- Information about spell, episode, and diagnosis number, both:
  - Original values from PEDW (epi\_num etc.)
  - Computed sequences that are guaranteed to start at 1 and be sequential (epi\_seq etc.)

*Resolves known errors where certain spells might not have an episode #1, etc.*

# Example query comparison

Subset of cohort with type 1 diabetes diagnosis in first episode

## Standard PEDW

```
Select *
  from my_cohort c
  join sailpedwv.spell s on
    s.alf_pe = c.alf_pe
  join sailpedwv.episode e on
    s.prov_unit_cd = d.prov_unit_cd and
    s.spell_num_pe = d.spell_num_pe and
    epi_num = 1
  join sailpedwv.diag d on
    s.prov_unit_cd = d.prov_unit_cd and
    s.spell_num_pe = d.spell_num_pe and
    diag_cd_123 = 'E10'
```

## Cleaned PEDW

```
Select *
  from my_cohort c
  join sailpedwv.pedw_single_diag d on
    d.alf_pe = c.alf_pe and
    epi_seq = 1 and
    diag_cd_1234 like 'E10%'
```



# Example queries

Find hospital admissions for hangnail

```
Select *  
  from my_cohort c  
  join sailpedwv.pedw_admissions a on  
    a.alf_pe = c.alf_pe and  
    first_primary_diagnosis = 'L030'
```

Count hospital admissions within a period

```
Select alf_pe, count(*) as admissions  
  from sailpedwv.pedw_admissions  
  where start_date between  
    '2000-01-01' and '2010-01-01'  
  group by alf_pe
```

Find hospital admissions of at least a week duration with a procedure

```
Select *  
  from my_cohort c  
  join sailpedwv.pedw_single_oper o on  
    o.alf_pe = c.alf_pe and  
    days(end_date) - days(start_date) > 7
```

# The Way Forward

- Provided to all new SAIL projects
- Available to existing projects on request
- A supplement, not a replacement for existing PEDW.
- Metadata in HDR UK Gateway:  
[www.healthdatagateway.org](http://www.healthdatagateway.org) (search for PEDW)
- Feedback welcome! (Use SAIL helpdesk)

# Next up: Cleaned GP

- Already implemented: add ALF to event table
- Next: transform event data based on dimensional modeling.
  - Replace event code with numeric ID in main event table.
  - Provide codes with metadata in lookup table.
    - Is code a valid V2 or V3 Read code?
    - Description
- Future possibilities:
  - Categorisation of types of codes.
  - Cleaned versions of frequently used measurements/test results.
  - Additional information for medications.

# SAIL User Forum - January 2022

## 6. Technical services



Richard Hier & Simon Thompson (*Technical services team*)

# SAIL User Forum - January 2022

## 6. Technical services

### SAIL Technology – Stability

- VDI
  - Windows 10 & phasing out of dedicated desktops.
  - Performance improvements to date:
    - Replacement storage.
    - Refinement of the desktops applications.
    - Increased RAM for standard desktops.
- System Availability
  - New hardware for backup solution – ensuring that backups are not running over into working day.
- Helpdesk
  - Dedicated individual to allocate helpdesk tickets and to follow up on older helpdesk tickets.

# SAIL User Forum - January 2022

## 6. Technical services

### SAIL Technology – Additional Components

- High Performance Computing cluster.
- Jupyter Hub to be made available for SAIL users.
- VDI - Replacement hardware.
- Migration away from DB2:
  - Support faster ingestion of data.
  - Consideration being given to end user during the migration.

# SAIL User Forum - January 2022

## 7. SAIL business updates and Marketing & comms



Julie Kennedy (*Business team*)

- Update to SAIL policy (amendments and extensions).
  - Main update already announced at the end of last year is the new amendment policy. This is now one form and one process for all changes (data, users, dates of access, desktop upgrades etc.) and can be access via the helpdesk:

<https://jira.hiru.swan.ac.uk/servicedesk/customer/portal/6/create/206>

Chris Roberts (*Marketing & comms team*)

- Updates on good news stories and recent news.
- Contact SAIL communications team if interested in completing a blog, news stories or other dissemination around your new funding / work / outputs.



# SAIL Databank Communications Update

Chris Roberts

Senior Impact & Engagement Officer





Types of communications

- Your research

- SAIL research

Via Web, Socials, Media,

- Partnerships and collaborations

Print, Events

- Funding awards and investments

All our communications output can be found here –

<https://popdatasci.swan.ac.uk/news/>

# RESEARCH BLOG: THE PEOPLE & PROJECTS USING SAIL DATABANK IN HEALTH & CARE RESEARCH WALES' £6.5M FUNDING ANNOUNCEMENT



**Dr Leigh Sanyaolu, Cardiff University**

**IM**proving Prophylactic Antibiotic use for Recurrent urinary Tract Infection (**IMPART**): mixed-methods study to address evidence gaps and develop a decision aid

## Please tell us about yourself

My name is Dr Leigh Sanyaolu. I am a General Practitioner and a Health and Care Research Wales NIHR Doctoral Fellow at Cardiff University. My current research is funded by the Welsh Government through Health and Care Research Wales.

## Please tell us about the research

It focuses on the management of recurrent UTIs (Urinary Tract Infections) in women and improving long-term antibiotic use through the development of an evidence-based decision aid. The evidence informing the decision aid will come from existing research trials and studies, anonymised routine data available from SAIL databank and qualitative interviews with patients and healthcare professionals working in primary care. I will be collaborating with the Primary Care and Interventions unit at the UK Health and Security Agency and the HDR UK team at Population Data Science, Swansea University, on this research, with support from Bladder Health UK.

## Please tell us how this research will benefit people

The decision aid will give patients evidence-based options to support the shared decision-making process regarding their recurrent UTIs in primary care. This could lead to significant impact through reduced acute infections, long-term antibiotic use, antibiotic resistance patterns and NHS costs.

Funding awarded - £576,735



**Dr Helen Hodges, CASCADE / Cardiff University**

Linking survey and administrative data to enhance understandings of risky behaviours and potential protective factors in children receiving social care: A feasibility study.

## Please tell us about yourself

I did my PhD in Criminology at Swansea, graduating in December 2018 and I'm currently working towards an MSc in Health Data Science. As a student, I was a member of the Consumer Panel for Data Linkage Research and I am also involved with the Observatory on Human Rights of Children.

## Please tell us about the research

The project will use data from the School Health and Wellbeing Survey conducted by the School Health Research Network (SHRN) to investigate young people's participation in a range of 'risky' behaviours including smoking, drinking, drug use, truanting, exclusion, bullying/being bullied, 'sexting' and sexual health. By linking responses to SAIL's routinely collected social services, education and health data, I hope to reveal patterns of participation in these behaviours amongst those receiving care and support from social services and identify potential protective factors. I will draw on the expertise of colleagues from CASCADE and DECIPHer as well as advice from a diverse range of young people, including those with lived experiences using surveys.

## Please tell us how this research will benefit people

By linking to social services and family court data there is an opportunity to take a more tailored approach rather than treating looked-after children as a homogenous group which can be used to inform policy in Wales. We can also observe individual changes in participation to inform the development of the tailored approach.

Funding awarded - £367,377

Find out more... [CASCADE](#) / [SHRN](#) / [DECIPHer](#)

# RESEARCH BLOG:

## COVID-19 PROJECTS FUNDED BY UKRI'S NATIONAL CORE STUDIES TO USE SAIL DATABANK



**Dr Sara Khalid**  
University of Oxford

**Improving methods in health technology to reduce inequalities, particularly ethnicity bias, using COVID-19 as an example.**

"We know there are inequalities in healthcare and patient outcomes between ethnic groups and the COVID-19 pandemic has highlighted the problem. Health technology can predict a person's future health risk but any biases in the data and models could lead to patients ultimately getting the wrong care or no care. Working with routinely collected data from SAIL Databank, and elsewhere around the UK, and by building ethnicity-specific models, we hope to tackle some of these underlying biases. By doing so we hope to move towards realistic, reproducible and representative models that are suitable and fair for all ethnicities to guide better care for COVID-19 patients."



Inequality in health has been made worse by the COVID-19 pandemic. People from minority ethnic backgrounds are more likely to become very sick or die from COVID-19. If there is bias in the data or bias in the model, patients from ethnic minority backgrounds could potentially received the wrong care or no care. This study aims to improve existing technology for predicting personalised future risk of health conditions, particularly those affecting overlooked groups of patients. Find out more > <https://www.hdruc.ac.uk/projects/improving-methods-in-health-technology-to-reduce-inequalities-particularly-ethnicity-bias-using-covid-19-as-an-example/>



**Dr Olivia Swann**  
The University of Edinburgh

**Do children and young people need extra follow up care after having SARS-CoV-2 infection?**

"Our project aims to find out whether children and young people need extra follow up care after having SARS-CoV-2 infection. Using population-level data is key to reducing bias in answering this question. The ability to use high quality linked primary and secondary care data from SAIL will allow us to carry out this study across three nations (Wales, Scotland and England) to ensure it is as representative as possible."

Researchers want to uncover the true burden of healthcare use for children and young people due to COVID-19. To find answers, the team will bring together data from GPs and hospitals on all children and young people in Scotland, Wales and England. Using state-of-the-art computing approaches to map out patterns of childrens and young peoples healthcare use before and after admission/infection with COVID-19. These will then be compared to those children and young people uninfected. Find out more > <https://www.hdruc.ac.uk/projects/do-children-and-young-people-need-extra-follow-up-care-after-having-sars-cov-2-infection/>



**RESEARCH BLOG:**  
**SAIL DATABANK TO SHED LIGHT ON SUN SAFETY POLICY IN WELSH PRIMARY SCHOOLS**



**Sunproofed:** A Mixed Methods Evaluation of Sun Safety Policies in Primary Schools in Wales



**Dr Julie Peconi**, PhD, is the Chief investigator of the Sunproofed study and is based at Swansea Trials Unit in Swansea University Medical School.

**WHY IS THIS RESEARCH NEEDED?**

**WHAT DO YOU HOPE TO LEARN?**

**HOW WILL SAIL DATABANK HELP?**

**WHO IS INVOLVED IN RESEARCH?**

**POTENTIAL BENEFITS AND IMPACT?**



# SAIL DATABANK REVEALS LINK BETWEEN HOUSEHOLD MENTAL ILL-HEALTH AND DEVELOPMENTAL DISORDERS IN CHILDREN

© 1 September 2021

New research, published in [BMC Medicine](#), has discovered the risk of children developing mental ill-health increases significantly when they have lived with someone who also has a common mental health disorder.

The study, led by Cardiff University, showed children who had grown up living with someone with mental health difficulties were 63% more likely to experience any mental health issue, which includes but is not limited to anxiety, depression, anti-social behaviour and personality disorders.



## STUDY INTO LINK BETWEEN MENTAL HEALTH AND SUBSTANCE ABUSE ISSUES ON YOUNG PEOPLE'S MORTALITY

14 January 2022

Young people with a history of mental health and drink and drug issues are more likely to die than those with a mental illness or who use substances alone. New research also revealed that males and those from the most deprived areas are at the greatest risk. A study carried out by the Adolescent Mental Health...

[Read More](#)



## SAIL DATABANK'S ANALYTICAL SERVICES TEAM SUPPORTS INNOVATIVE UK CATALOGUE OF DISEASE DEFINITIONS

13 December 2021

Health Data Research (HDR) UK recently launched the HDR UK Phenotype Library; a new resource supporting the sharing of electronic phenotypes (definitions of how health data can be used to measure real-world concepts relevant to research and clinical care). HDR UK is the national institute for health data science that unites the UK's health data to...

[Read More](#)



## WELSH RESEARCH FINDS COVID-19 INFECTION RATES LOWER AMONGST PEOPLE WHO HAD EXPERIENCED HOMELESSNESS THAN THE GENERAL POPULATION

7 December 2021

The research, published in the International Journal of Population Data Science, was carried out by ADR Wales, one of the Centres of Excellence based in Population Data Science at Swansea University Medical School, housing and homelessness lead academics Dr Ian Thomas and Dr Peter Mackie, using linkage between anonymised data from healthcare-related services and Covid-19 testing results. The research...

[Read More](#)

Top Latest People Photos Videos



Population Data Science at Swansea University @PopDat... · Jan 11 ...  
Professor in Public Health at Swansea University recognised in New Year's Honours 🏆 [popdatasci.swan.ac.uk/professor-in-p...](https://popdatasci.swan.ac.uk/professor-in-p...) #ProfRonanLyons @HDR\_UK @ADR\_Wales @DementiasUK @NCPHWR\_Wales #SAILDatabank @ResearchWales



popdatasci.swan.ac.uk  
Professor in Public Health at Swansea University recognised in New Y...  
A public health expert recognised by the Queen says his honour is a reflection of just how highly the work of Swansea University's SAIL ...

🗨️ 6 ❤️ 18 📤 📺



Population Data Science at Swansea University @... · Dec 17, 2021 ...  
#SAILDatabank's Analytical Services with colleagues in @SwanseaHSC and @PublicHealthW are proud to contribute to influential pan-European network for population medicines management. Cited for their #CongenitalAnomalies adverse drug identification research [encepp.eu/standards\\_and\\_...](https://encepp.eu/standards_and_...)



🗨️ 2 📤 📺

**Contact SAIL communications team if interested in completing a blog, news stories or other dissemination around your new funding / work / outputs.**

**[Christopher.Roberts@Swansea.ac.uk](mailto:Christopher.Roberts@Swansea.ac.uk)**

**All our communications output can be found here –**  
**<https://popdatasci.swan.ac.uk/news/>**

# SAIL User Forum - January 2022

## 8. Feedback



Survey is OPEN!

<https://hiru.typeform.com/to/BgGLB8HS>



4 questions will only take a maximum of 5 minutes.

We will request feedback after each forum, we are keen to hear how we can improve the forums moving forward and tailor them to be more impactful & useful to the SAIL community.

Survey will close Friday 4<sup>th</sup> February. Feedback after this by email please.



# SAIL User Forum - January 2022

## 8. Feedback



Are you interested in presenting at a future SAIL user forum?

Do you have any requests for areas of research / developments you would like to hear about at the next forum or in the future?

Please contact SAIL by February 10<sup>th</sup>

[sailbusinessadmin@chi.swan.ac.uk](mailto:sailbusinessadmin@chi.swan.ac.uk)

# SAIL User Forum - January 2022

## 8. Feedback

SAIL User Forums for 2022 will be on *(other SAIL/SeRP maintenance window dates on the website):*

Thursdays (10:00 - 12:00)

- 17<sup>th</sup> March 2022
- 19<sup>th</sup> May 2022
- 21<sup>st</sup> July 2022
- 15<sup>th</sup> September 2022
- 17<sup>th</sup> November 2022



Any further questions...

Thank you for coming



# SAIL User Forum - January 2022

## 8. Questions from the forum

- Is there information available on the Maternal ALF?

Yes the field is available in MIDS & NCCH data sources, and meta-data is available on the innovation gateway.

- What new data is coming soon / will be updated soon?

Lots coming soon some data sources of note include CARS (CARIS) which will bring the data up to end of 2020 coverage, as well as discussion with Ministry of Justice around courts and prison data. Any further information needed please contact Cynthia via the helpdesk.

- Have you responded to your ONS declaration for births and deaths data use in SAIL?

This is a requirement for SAIL to provide to the Office for National Statistics (ONS) so please if you have not responded do so, as SAIL team will need to continue to chase for a response whether to confirm or remove access.

- Is the phenotype library available inside SAIL as well as the link publically?

Yes this is an extension of the SAIL concept library and inside SAIL is available: [conceptlibrary.sail.ukserp.ac.uk/HDRUK](https://conceptlibrary.sail.ukserp.ac.uk/HDRUK)

# SAIL User Forum - January 2022

## 8. Questions from the forum

- Can I access new PEDW even if I do not need a refresh of data / new project?

Yes please contact the helpdesk and this can be made available without charge to any projects.

- Is the size of columns really impacting performance of the WLGP and other data – or just number of rows?

Yes the number of rows is the primary performance challenges, but we have found performance is improved by both the available size/format of columns and number of rows.

- Can a message be sent out to SAIL users when there are known stability issues?

We'll use the Helpdesk confluence page to post about known issues and proposed workarounds.

- What is being considered as a replacement for DB2 in SAIL?

We're reviewing options for this. However, we're absolutely considering the impact on end users and we're currently proposing to have some middleware between users and the database so that queries do not need to be amended. We will share more news at the next forum.

## Funded by



Ymchwil Iechyd  
a Gofal Cymru  
Health and Care  
Research Wales



## Certified by



IS632731

## Supported by



## In partnership with



## Accredited by



[www.SAILDatabank.com](http://www.SAILDatabank.com)

## Powered by

