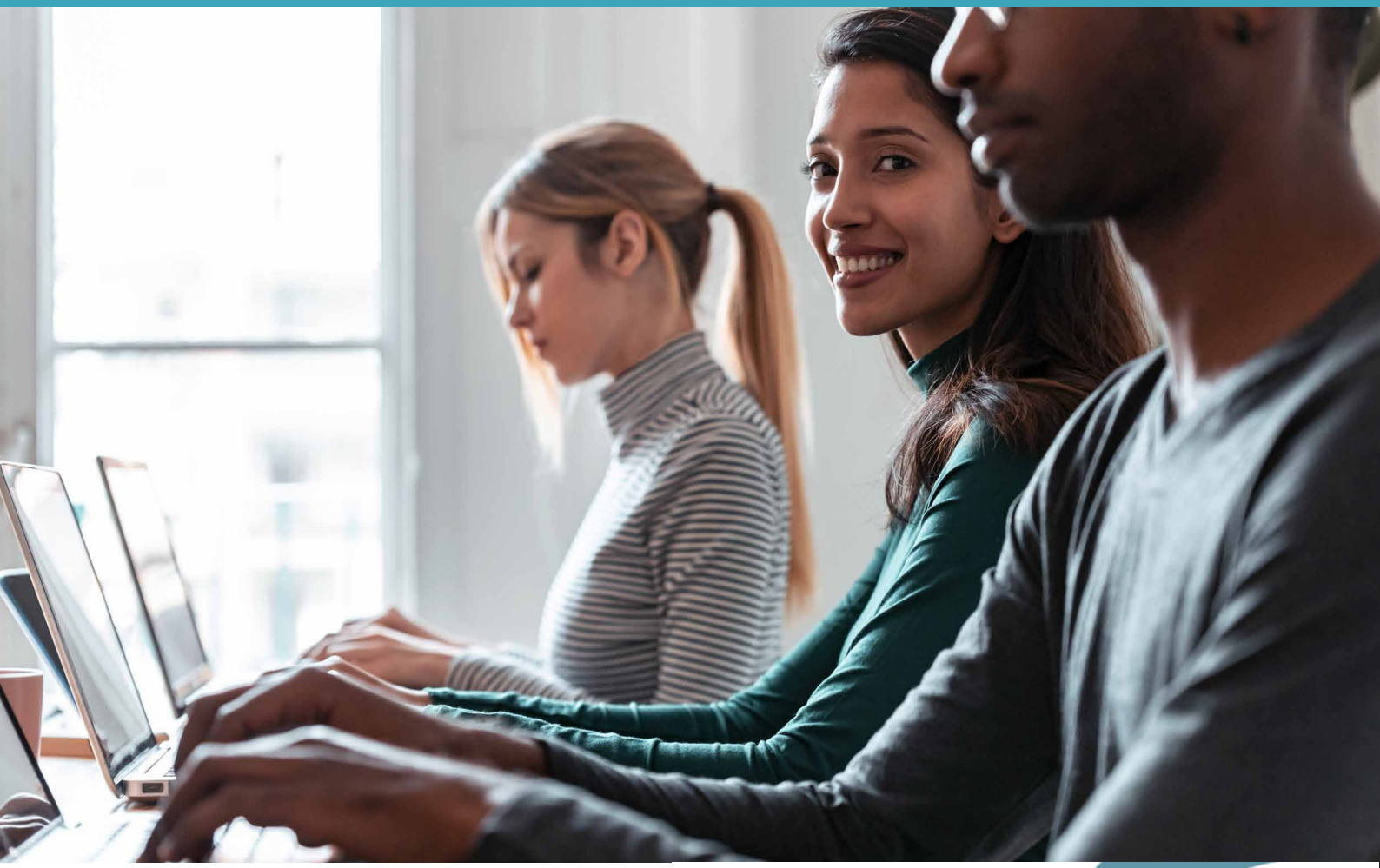
SAIL DATABANK



Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Ariennir gan **Lywodraeth Cymru**
Funded by **Welsh Government**

2020 - 2021
Annual report

Foreword

Professor David Ford
Director, SAIL Databank

This year has been without doubt the most challenging period in the existence of SAIL Databank. The COVID-19 pandemic has led to unprecedented demands for data to support government and NHS emergency planning, and also to support longer term research into all aspects of the virus. For most of the year we were receiving double the usual levels of requests for our services. As we observed in our previous Annual Report, the SAIL Databank, with its rich data resources and established world-leading technology and governance structures, was uniquely prepared to address this need, rapidly ingesting a range of new data sources and making these available (along with existing SAIL data resources) to meet the needs of government, the NHS and the research community. By June 2020, SAIL Databank was supporting nearly two thirds of all active UK projects focussing on COVID-19, having secured access to sixteen new data sources, and converted nearly half of all enquiries received into live projects, the most urgent of which received data access within five working days of application. SAIL was also a key partner in establishing the One Wales COVID-19 response team comprising colleagues from Health Data Research (HDR) UK, Administrative Data Research (ADR) Wales, SAIL Databank, Adolescent Data Platform (ADP), BREATHE (HDR UK hub for respiratory health), Welsh Government, Public Health Wales and Digital Health and Care Wales.. Using SAIL data, the One Wales team was able to track the development of health conditions in individuals and nested in households and school populations, monitor the development and spread of the virus and evaluate the impact of exposures and the effects of treatments on outcomes.

> ***SAIL Databank was supporting nearly two thirds of all active UK projects focussing on COVID-19...***

The insight delivered to Welsh Technical Advisory Group (TAG) and UK Scientific Advisory Group for Emergencies (SAGE) by One Wales has added to understanding and helped to shape decisions on lockdown guidance and policies in Wales. SAIL Databank's world leading technology platform was also in unprecedented demand to facilitate safe and secure sharing of data across geographical and organisational boundaries to enable research to fight COVID-19. Involvement in key international programmes such as International COVID-19 Data Alliance (ICODA), BREATHE Data Hub and the National Core Studies Connectivity Programme has led to SAIL receiving formal recognition from HDRUK, who commended "the responsiveness and commitment of your team over the last 6 months in managing a complex and highly dynamic programme of work."

Whilst providing a platform for critical COVID-19 response work has been a major part of our workload this year, SAIL has continued to support a diverse range of other research, with over 301 projects currently using our data. SAIL-supported projects issued 38 key publications this year on a wide range of topics including the impact of air pollution, dementia, epilepsy, cystic fibrosis, asthma and idiopathic intracranial hypertension. New SAIL supported research has secured more than £48 million funding this year, with over £11 million being attracted into Wales, and 49 new jobs being created.

Being a Director of SAIL Databank during these momentous times has been an immensely gratifying experience, seeing first hand the contribution we have been able to make to the global fight against COVID. It has also been immensely humbling to witness the heroic efforts of our team, who have juggled a double workload with all the challenges of home working to deliver such outstanding results.

Professor Ronan Lyons
Co-Director, SAIL Databank

Professor Kerina Jones
Co-investigator, SAIL Databank

FUNDERS

MANAGEMENT

OPERATIONAL

EXTERNAL GROUPS

Health & Care Research Wales

UKRI & others

SAIL Management Board

Professor David Ford
Co-Director

Ronan Lyons
Co-Director

Dr Kerina Jones
Professor of Population Data Science

SAIL Operations Group

Ashley Akbari
Senior Research Manager & Data Scientist

Lee Au-Yeung
Data Manager

Caroline Brooks
Senior Data Scientist - Team Lead

Sharon Heys
Head of Legislation and Due Diligence

Chris Jones
Infrastructure Manager

Julie Kennedy
Business Manager

Cynthia McNerney
Head of Data Acquisition and Provisioning

Jon Smart
Chief Operating Officer

Dan Thayer
Senior Data Scientist - Team Lead

Professor Simon Thompson
Chief Technical Officer

Analytical Services

Business, Governance & Administration

Marketing, Communication & Engagement

Scientific External Advisory Board

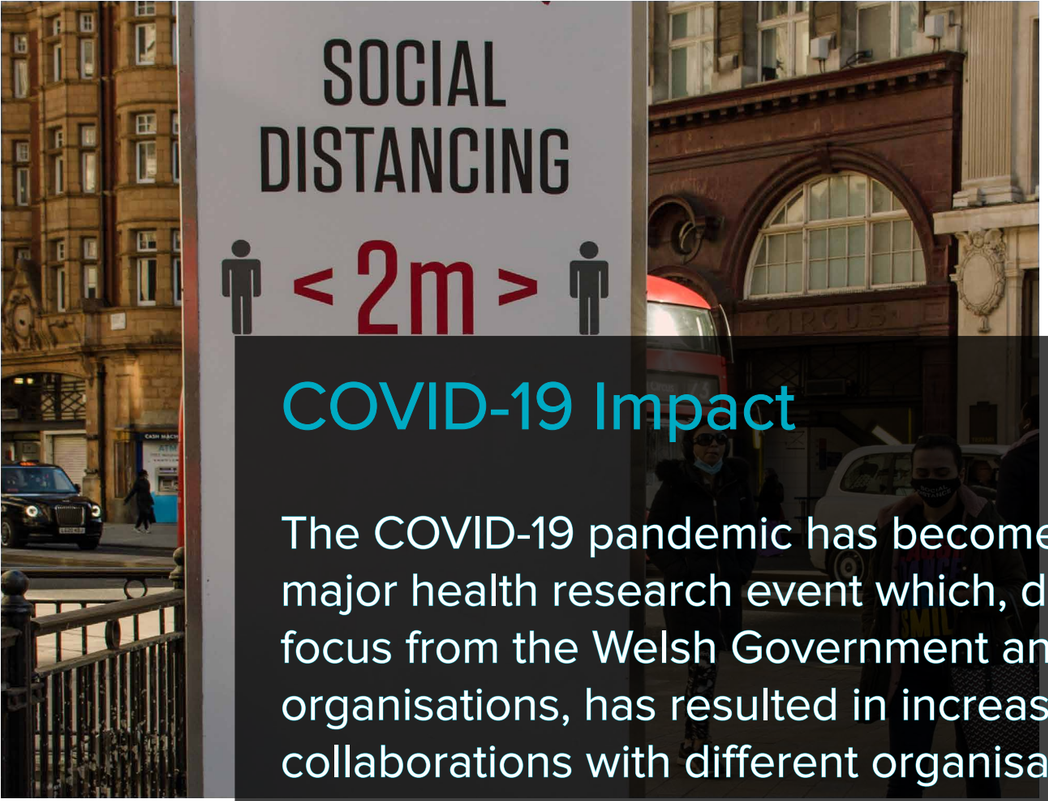
Information Governance Review Panel

Consumer Panel

Data Management Group

Analysts Forum

* Staff supported by multiple funders



COVID-19 Impact

The COVID-19 pandemic has become a major health research event which, due to focus from the Welsh Government and other organisations, has resulted in increased collaborations with different organisations as well as increased opportunities.

1. There has been an increased frequency of data updates of certain key datasets, e.g. Welsh Demographic Service Dataset (WDSD) has moved from monthly to weekly updates, Welsh Longitudinal GP (WLGP) has recently moved from quarterly to monthly.

2. The overall breadth of data now offered has increased with new datasets being added on an ongoing basis, many of these being refreshed on a daily basis, e.g. COVID Test Results & COVID Vaccination data.

3. Collaborations with organisations studying and collecting data related to COVID, e.g. ZOE App data.

4. The overall volume of work has significantly increased across all aspects of SAIL's services: scoping, IGRP reviews, project provisioning, file out reviews, data quality checking & metadata curation.

5. Since March 2020 we have had 27 new datasets introduced into the SAIL Databank

Key Partnerships & Collaborations

ICODA Project

The International COVID-19 Data Alliance (ICODA) is sponsored by the COVID-19 Therapeutic Accelerator Program led by the Gates Foundation, Mastercard, Wellcome, Minderoo and a growing list of supportive funders, to address urgent questions relating to COVID-19. It is convened by Health Data Research UK. SAIL has been working as strategic partner within the ICODA programme since early December 2020, and the first scientific outcomes from the initial pilot projects are expected around Q2 2021. Outputs from the project to 31st March 2021 include:

ICODA
International COVID-19 Data Alliance

1. New data contributor agreements developed to enable ICODA projects to bring data into SAIL. These were built in conjunction with the Bill & Melinda Gates Foundation and facilitate data analysis by ICODA Researchers working in the ICODA Workbench trusted research environment.

2. Technical work to enable federated data analysis from the ICODA Workbench to the SAIL databank has resulted in a successful first execution of a federated analysis pathway

3. Data hosting - 86 data contributor agreements have been issued across 40 countries to bring on data for ICODA's International perinatal outcomes driver project, of which to-date 23 are signed. The SAIL databank is also the strategic target for data contributed as part of ICODA's Gates Grand Challenges pilot projects

4. Sharing of SAIL expertise around information governance and joint working of governance processes

National Core Studies Connectivity Programme

The National Core Studies Data & Connectivity Programme is a central part of the UK's response to the COVID-19 pandemic, and aims to link SAIL and other Trusted Research Environments (TREs) across the UK to accelerate key research questions by enabling streamlined data access and analysis. Through this programme, which brings £ 1,134,000 funding for a variety of teams with Data Population Health at Swansea University, high value data assets have been made available for research in a secure and sustainable way, including the National Immunisation Management System (vaccine datasets) for the whole of England, Wales and Scotland; the NHS Digital ONS Joint Public Data Asset; and the HDR UK BHF Data Science Centre trusted research environment that has linked data on 54 million people in England. The programme has received recognition from leading national scientific advisors.

On 31st March 2021, SAIL received a letter of commendation from HDRUK, recognising the outstanding contribution SAIL had made to the Programme. 'SAIL is an important and valued delivery partner in this work. We are grateful for the responsiveness and commitment of your team over the last 6 months in managing a complex and highly dynamic programme of work. Key achievements over the last 6 months include the significant increase in datasets being onboarded to SAIL at pace and scale, including new assets such as vaccine datasets, and the increased impetus to build infrastructure required to support aspects of the programme such as federated capabilities. The subsequent increase in numbers of researchers accessing these datasets has been of huge value and will continue to provide real support to urgent COVID research. Despite the considerable ambiguity and uncertainty over this period, you have worked in a collegiate way, with a strong focus on making a positive difference to people's lives during the pandemic.'

> *"Thank you, without HDR UK's work a lot of what we have achieved this year during COVID would not have been possible"* - Chris Whitty (NIHR meeting March 2021)

> *"What you are doing on data linkage, governance and coverage across the 4 nations is crucial and impressive. Your work is absolutely making a difference and it is being noticed. It's not going to get quieter though, it's only going to get busier."* - Sir Patrick Vallance (Data & Connectivity Delivery Group Feb 2021)

BREATHE Data Hub

In addition to being one of the Directors of the SAIL Databank, Professor David Ford is also Chief Data Officer for BREATHE, the Health Data Research UK hub for respiratory health, a key partner of SAIL Databank in delivering the project's work programme. During 2020-21, BREATHE and SAIL delivered a number of key collaborative projects supporting the national emergency response to COVID-19.

In April 2020 a collaboration between SAIL, BREATHE, Kings College London and technology company ZOE developed a secure, anonymised data pipeline to deliver information from ZOE's COVID-19 symptom tracking app into the NHS, supporting the response to the pandemic. Before the roll out of effective COVID-19 testing and Track and Trace programmes, the data from this app offered one of the only ways of predicting the prevalence of the virus in a community. Demand for the data was huge and eventually resulted in 55 individual research projects using the data to support national emergency planning.

SAIL is supporting the BREATHE-led COVIDENCE study which will help identify who is most at risk of contracting COVID-19 and why some people become more ill than others with the disease. The study aims to recruit a diverse group of at least 12,000 people, including those who have already had proven or suspected COVID-19, and a mixture of people both with and without underlying conditions (such as diabetes, lung disease, heart disease and high blood pressure). The information gathered will help scientists to understand why certain people appear to be at greater risk. The study, which will run for up to five years, will also create a platform to fast-track future trials of preventative treatments, to see if they help to protect against COVID-19.

A collaboration between SAIL and BREATHE secured £4 million to undertake the Co-Connect project which will address fundamental questions about the development of immunity to COVID-19. SAIL Databank will support CO-CONNECT researchers by providing access to a range of data including COVID-19 antibody data from NHS Wales' 'pillar 3' serology testing.

An asthma study reported that people from deprived areas in Wales have worse outcomes and increased risk of death: The new research published in the journal, PLOS Medicine, was conducted by Swansea University's Wales Asthma Observatory in collaboration with Asthma UK Centre for Applied Research, Applied Research Collaboration North West Coast (ARC NWC) at the University of Liverpool and BREATHE - The Health Data Research Hub for Respiratory Health. Jointly funded by Health and Care Research Wales and Swansea Bay University Health Board, the study discovered that people with asthma from socially and economically deprived areas of Wales have less control of their asthma, suffer from more asthma attacks and are at a higher risk of death. The authors used SAIL Databank data for the period of 2013 to 2017 to investigate the link between GP care data, emergency hospital admissions, prescriptions and asthma deaths together with geographical and socioeconomic measures for ranked areas of deprivation. The results of the study will be used to inform public health messaging and intervention within deprived communities to better empower patients to manage their asthma.

BREATHE Health Data Research Hub

Key Partnerships & Collaborations

Non COVID related research supported by SAIL has also resulted in tangible impacts during 2020-21.

Uncovering private family law: Who's coming to court in England?

Published by the Nuffield Family Justice Data Platform (Nuffield FJO), the Uncovering private family law: Who's coming to court in England? study shows that private law children cases in England disproportionally involve people living in deprived areas. In 2019/20, 30 per cent of applicants lived in the most deprived quintile (by definition representing 20 per cent of the wider population). Just 13 per cent lived in areas in the least deprived quintile.

The report also reveals that parents' need for support and assistance from the family courts varies by region. In 2019/20, private law application rates in the northern regions ranged between 79 and 81 per 10,000 families with dependent children, but were just 44 per 10,000 in London and 59 per 10,000 in the South East.

The project is part of SAIL Databank's ongoing flagship programme – the Family Justice Observatory - which has enabled a variety of population level data to be linked to court proceedings data, enabling the effects of family court decisions on the lives of the families which use their services to be effectively evaluated for the first time ever.

Dr Linda Cusworth from the Centre for Child and Family Justice Research at Lancaster University, lead author of the report, said: ***"Understanding who uses the family courts, and what they might need and want, will provide a more balanced and effective approach to policy and practice development and can avoid mistaken assumptions. It may also result in more carefully targeted interventions, both within and outside the court."***

Nuffield Family Justice Observatory

Idiopathic Intracranial Hypertension

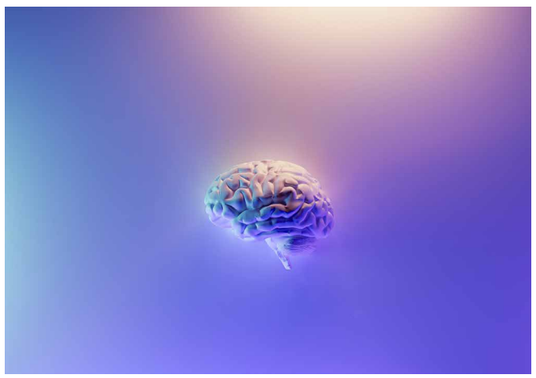
Dr Owen Pickrell, (consultant neurologist and clinical associate professor at Swansea University) and his research team have been widely cited in the media for their SAIL databank based research into idiopathic intracranial hypertension.

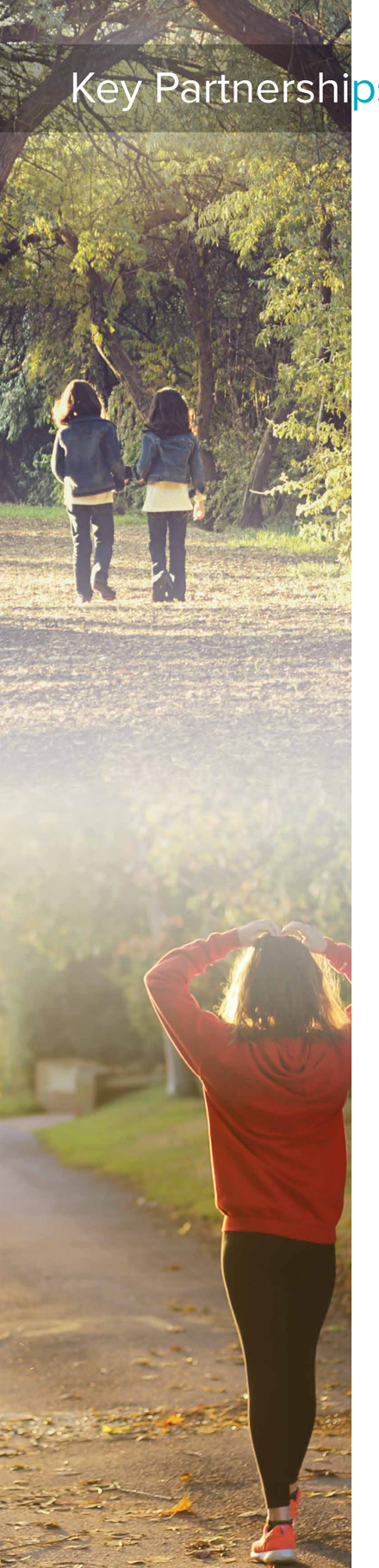
This poorly understood condition causes increased pressure in the fluid surrounding in the brain and can lead to severely disabling headaches as well as vision loss, which can be permanent.

The team used data within SAIL to analyse 35 million patient years of data spanning the period 2003 to 2017 – the first time such a large scale analysis of this patient group has taken place. They identified a total of 1,765 people within Wales living with IIH during that period, 85% of whom were women.

Key findings from the research included identified prevalence of the condition (rising from 12 in 100,000 people in 2003, to 76 in 100,000 in 2017), gender mix of the affected population (85 per cent female) and correlations between the condition and other factors (there were strong links to high body mass index and social deprivation).

More details on the study can be found at https://n.neurology.org/content/early/2021/01/20/WNL.0000000000011463. An example of the tabloid coverage can be found at https://www.thesun.co.uk/living/3321901/people-who-are-too-fat-or-too-thin-are-more-likely-to-suffer-from-migraines/





Key Partnerships & Collaborations

Identifying children with Cystic Fibrosis

SAIL supported research into cystic fibrosis resulted in publication of a research paper Identifying children with Cystic Fibrosis in Population-Scale Routinely Collected Data in Wales. This is the culmination of a five year project involving Liverpool and Swansea Universities, Public Health Wales and The Cystic Fibrosis Trust, which involved the linkage of data from the UK Cystic Fibrosis Registry with SAIL data to test whether there was a disparity between numbers of children identified within CF Registry data and those with a CF diagnosis in other health records.

The study found that 257 out of 352 diagnosis in SAIL data were true cases and only 11 children with CF had no recorded diagnosis within SAIL. The study also found that if children had had CF codes in multiple eHR data sources (primary care, secondary care or the congenital anomaly register), they were more likely to be true cases, highlighting the benefits of using a resource such as SAIL which now brings together a number of eHR datasets with CF Registry data. The CF Registry Data will remain in SAIL and is now available to other researchers. The paper can be found at https://ijpds.org/article/view/1346

Impact of schizophrenia genetic liability on the association between schizophrenia and physical illness: data-linkage study

A collaboration between Swansea and Cardiff Universities published its findings into the association between genetic predisposition to schizophrenia and poorer physical health by linking genetic data to other SAIL datasets. The research team found that individuals with schizophrenia had a life expectancy of 20 years less than the average, with increased rates of neurodevelopmental disorders (epilepsy, intellectual disability and congenital disorders), smoking, type 2 diabetes mellitus and ischaemic heart disease compared with the general population. However, the research suggested that whilst there was a clear association between schizophrenia and increased risks of a variety of life limiting conditions, the link may not be directly due to schizophrenia genetic liability.

An exciting aspect of the research was that it the first study to successfully link genetic data to other datasets within SAIL, opening up the potential for further studies using a combination of existing SAIL resources and genetics data. The publication can be found at https://www.cambridge.org/core/journals/bjpsych-open/article/impact-of-schizophrenia-genetic-liability-on-the-association-between-schizophrenia-and-physical-illness-datalinkage-study/

Key Partnerships & Collaborations

ADR Wales recognised as exemplar for data collaboration

Following a seven year ongoing partnership with SAIL Databank, Administrative Data Research (ADR) Wales has been recognised as an exemplar for data collaboration in the new Digital Strategy for Wales. The Strategy, which sets out the roadmap for digital services and inclusion in Wales highlighted the role played by the ADR Wales partnership which brings together Government and academia to provide data driven analysis to areas of Welsh Government policy.

We want to improve the services provided by working together and make sure that all data is used and shared effectively, has consistent standards, is protected and gets to where it needs to go. This can include the innovative use of data and data analytics to radically transform the way we deliver public services by delivering new insight. The ADR Wales Partnership has already shown what can be achieved by working together in using data for research. We will champion greater use of data linking and enhance data science capability in the public sector to transform how our information is used for public benefit. Focussing particularly on early years, housing, social care, mental health, wellbeing, skills and employment, and most recently rapid analysis to aid Covid-19 response in Wales, the data linked and analysed by ADR Wales aims to address the priority areas for action in Wales, as identified in the Welsh Government's national strategy Prosperity for All. ADR Wales' work enables the Welsh Government to understand far more about the relationship between different areas of service provision and to better understand the experiences of people as they move through different services. Previous analysis has already contributed to Welsh Government policy interventions.

ADR UK
Data-driven change

Networked Data Lab Wales

The Networked Data Lab (NDL) Wales released a report in January 2021 that details work undertaken by NDL Wales analysts within Public Health Wales, using SeRP & SAIL Databank to further study the group of people within Wales who are clinically extremely vulnerable to COVID-19 and to identify common socio-economic factors such as deprivation, employment, location and ethnicity within the group.

> *"This first short report provides valuable insights into the diversity of the shielded populations in Wales and across the country, reflecting their complex health and care needs. Further in-depth analyses are underway to help ensure future health and care support meets the needs of this clinically extremely vulnerable group."*

Additional Partnerships & Collaborations

EUROLinkCAT
This is a Europe-wide project aimed at establishing a linked cohort of children with congenital anomalies. SAIL is the lead organisation for Wales. Our team transforms the SAIL data using reusable research methodologies so that it is ready for research. https://www.eurolinkcat.eu/

ConcePTION
This study involves 20 European nations, and is examining the safety of medicines use pre-conception, during pregnancy and breastfeeding. Building on the knowledge and methodologies developed in EUROLinkCAT. SAIL is the lead organisation for Wales. https://www.imi-conception.eu/

UK-REACH
The SAIL team have a significant supporting role in the delivery of this project, which is examining the impact of COVID-19 on healthcare workers, depending on their ethnicity. https://uk-reach.org/main/

ADP
The SAIL team are supporting the Adolescent Mental Health Data Platform which is a programme to develop a specialist research environment focusing on the mental health of children and young people.
https://popdatasci.swan.ac.uk/centres-of-excellence/adolescent-mental-health-data-platform/

RECOVERY
The SAIL team are providing fortnightly data updates of the individuals recruited to the RECOVERY (Randomised Evaluation of COVD-19 Therapy) trial. Having recruited over 28,000 patients with COVID-19 from 176 hospitals around the UK, RECOVERY is the largest study in the world to test treatments for people admitted to hospital with COVID 19. It is identifying which treatments are effective and which are not. https://www.recoverytrial.net/

UK Biobank
The SAIL team are providing monthly data updates of the individuals recruited to the UK Biobank. https://www.ukbiobank.ac.uk/

Millennium Study Cohort
The SAIL team have provided an update of the individuals recruited to the Millennium Study Cohort. https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/

GenOMICC – COVID-19
The SAIL team will be providing monthly data updates of the individuals recruited to the GenOMICC COVID-19 study. https://covid.genomicc.org/

HealthWise Wales
The SAIL team have a significant supporting role in the delivery of this project, including provision of technical infrastructure, linking the study data to SAIL datasets, provisioning data and reviewing file out requests.
https://www.healthwisewales.gov.wales/

SIMPLIFIED registry trial
The SAIL team are providing biannual data updates of the individuals recruited to the SIMPLIFIED registry trial.
https://cctu.org.uk/portfolio/core/trials-open-to-recruitment/the-simplified-registry-trial

TIME (Treatment in Morning vs Evening)
The SAIL team are providing biannual data updates of the individuals recruited to the TIME study.
https://www.isrctn.com/ISRCTN18157641

ConCOV
Professor Ronan Lyons has secured 828,353 from the UK Research and Innovation (UKRI) COVID-19 Rapid Response Call to help understand and address the challenges of the pandemic. The project, Controlling COVID-19 through enhanced population surveillance and intervention (ConCOV), will run for 12 months and will provide a platform for research to inform evidence-based strategies to control the virus, safeguard the general population and help bring the UK out of lockdown.

PHIRI
Professor Lyon's team has also secured over €220,000 of Horizon 2020 funding to join 'Population Health Research Infrastructure' (PHIRI) - a network of over 41 partners from over 31 countries. The Swansea team, supported by SAIL databank, will focus on measuring the impact of the COVID-19 on vulnerable populations and also to evaluate the effectiveness of counter-measures.

Our Year In Numbers

£50.4m Research income secured

45 Jobs — Created in Wales

Core Metrics

Reporting period: **2020/2021**

Health and Care Research Wales infrastructure award to the group

Direct funding awarded **£856k**

Jobs created through direct funding **35**

Grants won during reporting period

Grants won	Led by group	Group collaborating
Number	14	59
Value	£8.5m	£42m
Funding to Wales	£8.3m	£2.9m
Funding to group	£4.5m	£1m
Additional jobs created for Wales	42	3
Additional jobs created for group	33	3

38 Number of publications

5 Number of public engagement events

12 Number of public involvement opportunities

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Llywodraeth Cymru
Welsh Government

SAIL Databank Work Packages

The SAIL Databank team offer a wide range of services and support, both internally within SAIL and externally to research projects. Whilst the majority of the year has been focused on responding to the COVID situation, we took the opportunity to improve our processes to ensure a more pro-active response to the impact on our services

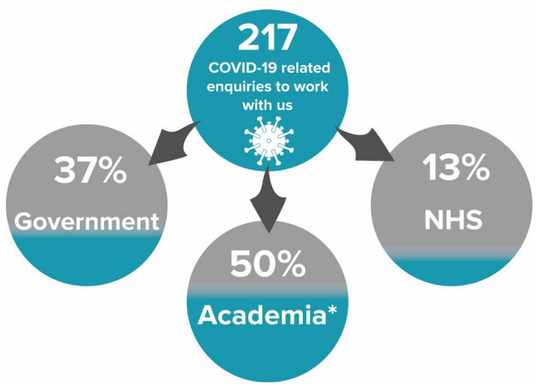
SAIL is focussing on the delivery of three key workstreams to deliver during the 2020-25 period.

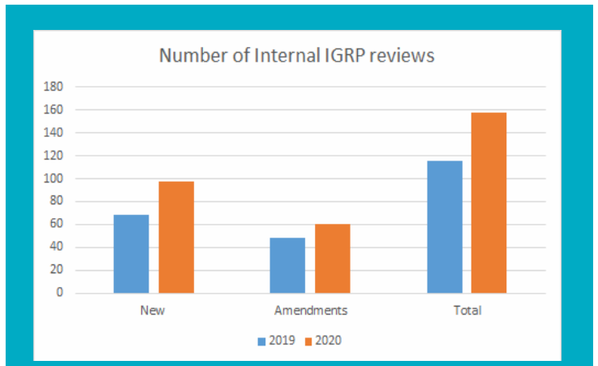
Workstream 1

To run the core SAIL service, providing a comprehensive support service to research projects to enable access to SAIL data.

During the period April 2020 – March 2021, SAIL has supported or led the development of 161 applications for research funding, of which 70 have been successful. This has resulted in over £48 million research income being secured, £11.5 million of which has been for Welsh based research teams. A total of 44.75 WTE posts within Wales are funded from this grant income, showing that SAIL continues to be an effective magnet for drawing resources into Wales.

SAIL has continued to play a pivotal role in supporting the response to the COVID-19 pandemic, both in terms of making data available for government / NHS planning and also facilitating a wide variety of research projects. At one stage SAIL Databank was supporting two thirds of all COVID-related research using annoymised data. By the end of March 2021 the position was as follows:





2020-21 has seen an exponential rise in demand for SAIL services. Project scoping requests have doubled and the numbers of projects seeking IGRP approval (either for their initial access to data, or for additional data as part of a project amendment) has increased by 35%.

All projects have been subject to SAIL's rigorous information governance and data security safeguarding processing, but a large number have also been expedited due to their importance in supporting the emergency response to the pandemic. This has resulted in a huge increase in pressure on both the internal SAIL team and our external IGRP reviewers, and we owe them a profound debt of thanks for their efforts during this extraordinary year.

**** 93 approved individual projects. The majority of these projects are working with or directly informing NHS or government, and have enabled decision makers to have data available to inform the development of strategies to manage the pandemic.***

Workstream 2

Data acquisition to meet expressed requirements of our user base.

During 2020-21 a total of 27 new data sources have been deposited within SAIL Databank to aid Covid-19 analysis. Core data flows from NWIS (except GP data) are now established in their monthly pattern.

SAIL has also continued to expand its data catalogue in non-COVID specific areas, and is now in receipt of Domiciliary Social Care Workers data from Social Care Wales. We have signed an agreement for Live Attendance Education Data - from Welsh Government (this is in the process of being loaded and will be available very shortly). We have also signed our first agreement with Ministry of Justice to acquire some data on offenders who have been identified from prison (P-NOMIS) administrative data systems.

The requirement for increased access to data during the COVID-19 pandemic has resulted in significant demand from government agencies and organisations such as HDRUK for data to be available across organisational and sometimes geographic boundaries. Projects including ICODA and National Core Studies are now requiring SAIIL to develop new data security and governance processes to enable federated access to and querying of data, to enable a more joined up approach to data resources across the UK and internationally.

We are playing a leading role in these projects and programmes and are committed to ensuring that SAIL remains in the forefront of technological advances in improving data accessibility, without compromising our rigorous data security standards.

THE FOLLOWING DATASETS ARE NOW AVAILABLE FOR WIDER USE BY RESEARCHERS

...and are not only restricted to contributing to compilation of reports to Welsh Governments' Technical Advisory Group (TAG)

COVID 19 LATERAL FLOW DATA – acquired from NWIS January 2021

COVID 19 VACCINE DATA – acquired from NWIS January 2021

DAILY SITUATION REPORT DATA

COVID-19 TEST RESULTS

COVID-19 SHIELDED PEOPLE LIST

CONSOLIDATED DEATH DATA SOURCE

CONTACT TRACING
(Test Trace Protect)

Workstream 3

Increase research speed and efficiency by developing a set of re-usable tools (Concept Library) that will enable significantly reduce the time and effort required to conduct a complete analysis of a research question within the SAIL Databank.

In 2020 SAIL officially launched the Concept Library, a novel in-house developed tool for managing, publishing and sharing definitions used in research. This product has been recognised as a leading tool in the Data Science community, and adopted by both the Adolescent Mental Health Data Platform and HDR-UK. In 2021 a UK-wide solution built on the Concept Library will be released by the latter organisation, with SAIL having been a key collaborator in this work.

As the volume of data SAIL has increased, new tools have been developed to streamline the process of data management from initial receipt within SAIL to making it available to individual projects. The requirement to provide daily refreshes of data to many COVID projects has resulted in the development of standardisation and automation of our data loading processes, and this has now been rolled out to a number of other datasets unrelated to COVID research.

In response to the demand for more data to be available more quickly, we are currently developing standardised automated procedures for dataset QA, sign off and issue resolution to improve efficiency whilst continuing to safeguard data quality within SAIL.

Although the demands of COVID have reduced development capacity within the team, a number of useful tools to support researchers using SAIl have been developed and launched during 2002-21. These include a structural optimisation to the GP dataset that improves performance for researchers working with it and an automated tool which can identify key characteristics within a cohort of data subjects and produce a matched control group.

While capacity for development was somewhat reduced by meeting the needs of the COVID-19 response, our team progressed a number of important developments to support researchers.

In 2020, we officially launched the Concept Library, our flagship tool for managing, publishing, and sharing definitions used in research. This has been recognized as a leading tool in the domain, and adopted as a solution by both the Adolescent Mental Health Data Platform and HDR-UK. In 2021 a UK-wide solution built on the Concept Library will be released by the latter organization. We are key collaborators in this work, leading a multi-site development team.

We progressed research-ready datasets, through standardising and automating processing of a number of datasets, both SAIL core datasets (e.g. GP data) and derived data (standardized coverage measures). We also developed automated processing to support daily GP data feeds used for COVID-19 priority work.

We developed an R version of our code that creates a matched control group.

Ongoing development and improvment of our processes of data provision and dataset QA, to support greatly increased velocity due to COVID-19.

Over the next year the SAIL team will be focussing on...

Continuing to work closely with government, the NHS and the research community to combat the effects of COVID-19.

Maintaining and developing national and international partnerships to deliver a diverse range of research to improve health, wellbeing and public service.

Securing new data and delivering research ready datasets to help researchers access the information they need more easily.

And further developing enhanced data analysis tools to support and assist projects in their research work.

SAIL DATABANK

Public Involvement

SAIL Databank has a long standing Consumer Panel which was established in 2011. It currently has 16 members (all members of the public) with on-going recruitment.

Panel members are involved in all elements of the SAIL Databank process, from developing ideas, advising on bids through approval processes (via the independent Information Governance Review Panel), to disseminating research findings.

Every project which is considered for data access by our IGRP has at least one member of the public on the review panel. This includes all the expedited COVID emergency response projects which were reviewed within 48 hours of IGRP application.

Public Engagement

COVID restrictions have restricted the amount of public engagement the SAIL team has been able to engage in during 2020-21. We have taken advantage of any available opportunities, working virtually wherever possible in line with government guidelines.

Use My Data webinar (Nov 2020)

Use My Data is an organisation that supports and promotes the protection of individual choice, freedom and privacy in the sharing of healthcare data to improve patient treatments and outcomes. It endeavours to highlight to the general public the many benefits that appropriate usage of healthcare data can make, to save lives and improve care for all.

SAIL co-director Professor Kerina Jones gave a presentation at the above event, which was virtually attended by members of the public. She talked through how Trusted Research Environments (TREs) work, why they are potentially so important and where the patient voice sits in their design, development and running.

As a result of feedback received from the audience at this event, SAIL now publishes a full list of current projects and outputs, and statistics around dataset usage, areas of research undertaken and types of users on its website.

Data4Good Festival (March 2021)

Data4Good Festival 2021 is an online programme of events involving more than 20 partner organisations. The festival is designed to help charities and social sector organisations make better use of data. SAIL Director Professor David Ford presented SAIL Databank as a means to allow these organisations to use their data for the benefit of their service users.

HAPPEN programme

Health and Attainment of Pupils in a Primary Education Network (HAPPEN) is a programme monitoring levels of activity in schools and whether this results in tangible positive outcomes for pupils. HAPPEN uses SAIL data and infrastructure to conduct its research. During the periods of lockdown, HAPPEN conducted a number of surveys of school children and their families to assess the impact of lockdown on children and their carers' mental health and physical activity levels.

Networked Data Lab Wales

The Networked Data Lab (NDL) Wales have conducted a number of public facing surveys to gauge the feelings and needs of people with clinical vulnerabilities during the COVID pandemic.

Secure
Anonymised
Linkable

Containing billions of person-based records, SAIL Databank is a rich and trusted population databank. It improves lives by providing researchers with secure, linkable and anonymised data that can be accessed and analysed from anywhere in the world.

Funded by

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

Ariennir gan Lywodraeth Cymru
Funded by Welsh Government

UKRI Economic and Social Research Council

Supported by

Swansea University
Prifysgol Abertawe

Certified & accredited

bsi. ISO/IEC 27001 Information Security Management
IS632731

UK Statistics Authority
Awdurdod Ystadegau'r DU

In partnership with

GIG CYMRU NHS WALES
Gwasanaeth Gwybodeg
Informatics Service

Powered by

SeRP
ISO certified and UK Statistics Authority Accredited

www.SAILDatabank.com