SAIL DATABANK

Ymchwil Iechyd a Gofal Cymru
Health and Care Research Wales

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

2022 - 2023
Annual report

CONTENTS

04 INTRODUCTION

06 ORGANISATIONAL CHART

08 KEY PARTNERSHIPS & COLLABORATIONS

20 ADDITIONAL PARTNERSHIPS & COLLABORATIONS

26 OUR YEAR IN NUMBERS

28 WORKSTREAMS

- 28 Workstream #1
- 34 Workstream #2
- 36 Workstream #3

38 PUBLIC INVOLVEMENT & ENGAGEMENT

INTRODUCTION

It gives me great pleasure to present our Stakeholder Report, which highlights the work and achievements of the Secure Anonymised Information Linkage (SAIL) Databank between April 2022 and March 2023.

Now in its sixteenth year of operation, the SAIL Databank is a world leading and internationally acclaimed research resource, recognised as an integral part of the national informatics infrastructure within Wales and as an innovation and thought leader across the UK. 2022 – 2023 has been an extremely successful year for SAIL Databank; we have supported the development of 64 new research projects which have in turn been successful in securing a combined total of £36 million in research funding, of which £27 million has come into Wales. SAIL is currently supporting a total portfolio of 324 live research projects covering a wide variety of topics. Projects supported by SAIL have created 25 full time posts within Wales and results from the research have generated 140 publications this year alone, in addition to a variety of other impacts and outcomes, including media articles and reports to policy making bodies such as the NHS and government departments.

SAIL is a key partner in several research programmes that provide significant long-term benefit at a whole population level. Our long-standing partnership with Administrative Data Research (ADR) Wales continues to provide Welsh Government with a vital source of research which directly translates into better informed policy decisions and more effective public services, from improving education and healthcare to tackling crime. Our cutting-edge collaboration with the Family Justice Observatory is bringing together a range of data on children and families who are subject to family law proceedings which had previously been unavailable for linked analysis, and is yielding valuable insight into gaps in the provision of services and highlighting areas where change will have the biggest impact. We are also proud to continue supporting a wide range of research into various aspects of child and adolescent mental health, following increased recognition of the effect of the COVID-19 pandemic's effect young people's emotional wellbeing. In addition, we continue to support research led by NHS clinicians from a wide range of specialties and are exploring ways to form more formal partnerships with some NHS organisations to increase the levels of support we can provide.

SAIL has always been a pioneer in the field of Data Science, and we are continue to lead in the development of innovative new data sharing and access models. In our 2021-22 Stakeholder Report, we highlighted our work to develop a number of novel technologies to meet growing demand from the research community to be able to analyse data from a number of sources, both across the UK and internationally, from a single point of access. This requirement, known as data federation, comes with formidable technical and governance challenges. In February 2023, SAIL became part of two new Data and Analytics Research Environment – UK (DARE-UK) projects to develop prototype platforms using a new secure data access process to enable researchers to work-with and analyse data held in physically separated locations. We look forward to continuing playing a leading role in these exciting projects during the next twelve months.

We would like to sincerely thank our funders, data providers and stakeholders for their ongoing support, underpinning our evolution from a tiny experimental pilot to an internationally renowned data repository. We also owe a huge debt of gratitude to our International Advisory Board, whose advice and insight have been invaluable in shaping our strategic direction. As the data science landscape continues to change within the UK, with the role of data being seen as increasingly important for informing national strategy setting, and Trusted Research Environments and Data Repositories increasingly being established as national centrally funded resources, it will be important for SAIL that our funding and operating frameworks evolve in line with these new models. We look forward to continuing to engage with our funders and stakeholders throughout the coming year to ensure SAIL continues to flourish into the future.

Prof. David Ford
Co-director
SAIL Databank

Prof. Ronan Lyons OBE
Co-director
SAIL Databank

Professor Kerina Jones
Co-investigator
SAIL Databank

FUNDERS

MANAGEMENT

OPERATIONAL

Health and Care Research Wales (HCRW)

UK Research and Innovation (UKRI) & others

SAIL Management Board

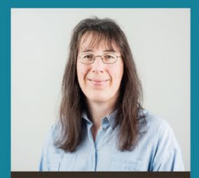
Professor David Ford
Co-Director

Ronan Lyons
Co-Director

Dr Kerina Jones
Professor of Population Data Science

SAIL Operations Group

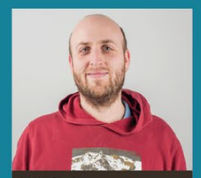
Ashley Akbari
Associate Professor Population Data Science Research & Senior User Representative



Caroline Brooks
Senior Data Scientist - Team Lead

Rob Garlick
Audit & Compliance Manager

Sharon Heys
Head of Legislation & Due Diligence



Chris Jones
Infrastructure Manager

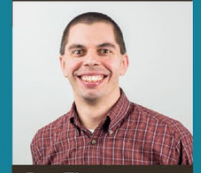
Julie Kennedy
Business Manager

Cynthia McNerney
Head of Data Acquisition & Provisioning

Chris Orton
Population Data Science Programme Manager & User Representative

Chris Roberts
Marketing & Communications Representative

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

* Staff supported by multiple funders

EXTERNAL GROUPS

Scientific External Advisory Board

Information Governance Review Panel

Consumer Panel

Data Management Group

SAIL User Forum

KEY PARTNERSHIPS & COLLABORATIONS

Regular health checks make life-saving difference for people with an intellectual disability

Researchers have found that health checks help improve survival for people with an intellectual disability, especially those with autism or Down's syndrome.

The study analysed the medical records of 26,954 people with an intellectual disability in Wales between 2005 and 2017. People with an intellectual disability experience more health conditions such as epilepsy, autism, and dental problems and are at a higher risk of leading sedentary lives and becoming overweight, subsequently developing diabetes, cardiovascular disease, and respiratory disease. They are also more likely to be exposed to poverty, poor housing conditions, unemployment, and other social causes of poor health.

Annual health checks for people with an intellectual disability were introduced in Wales in 2006 to improve early detection and treatment. However, the lack of evidence about the long-term health benefits of providing these checks has resulted in many GPs not offering them. The researchers looked to address that lack of evidence and explore whether health checks are associated with better survival and lower mortality rates.

The data revealed:

1. Having a health check was associated with reduced deaths for people with autism and those with Down's syndrome
2. There was little evidence of reduced death rates for those diagnosed with diabetes or epilepsy
3. There was no evidence to indicate that health checks improved outcomes for people diagnosed with cancer
4. Health checks are more effective the earlier people receive them
5. Many people may have received one health check but not undergone annual checks
6. There was a lower-than-expected uptake of health checks among the Welsh adult population – more than 70 per cent had no record of having a health check.

This research has resulted in further studies to make health checks more accessible for the eligible Welsh population currently not receiving one. Furthermore, NHS England are introducing health checks from 16 years old, and Wales is looking to change policy to offer health checks from a younger age.

More information on the study is available here:
https://bmjopen.bmj.com/content/12/4/e049441.full

Research into long COVID reveals an increased risk of fatigue, post-viral illness, and embolism

COVID-19 is associated with increased risks of post-viral illness, fatigue, embolism, depression, anxiety, and respiratory conditions, according to new research led by the National Centre for Population Health & Wellbeing (NCPHW) Research team based in Population Data Science at Swansea University.

There are growing concerns around the long-term symptoms experienced by people who have had COVID-19, however, there is little research into long-term COVID and its impact on healthcare. The study looked to establish whether testing positive for COVID-19 resulted in increased use of primary and secondary care in the first six months following the test compared to those who had not tested positive using SAIL Databank (SAIL) data, which includes nationwide GP data, hospital in- and out-patient records, and death records collected by the Office for National Statistics (ONS). The team studied data from over 41,000 individuals living in Wales who tested positive for COVID between 28/02/20 and 26/08/21. Researchers looked at healthcare contact in the 1-4 weeks and 5-24 weeks following a positive COVID-19 test.

The major findings were that, when compared to people who tested negative, positive individuals were 212% more likely to die in the first four weeks, 28% less likely to die between five and twenty-four weeks if they were tested in the community but, 22% more likely to die if they tested positive in hospital and more likely to attend healthcare for post viral illness, fatigue, or embolism after 5 weeks if tested in the community.

Despite the strong associations found between testing positive for COVID and healthcare use in the short and long term, the overall number of individuals attending a healthcare setting was very low.

The full pre-print is available here:
https://www.medrxiv.org/content/10.1101/2022.04.21.22274152v1

The probability of mothers involved in care proceedings returning to court shows no change in nearly a decade

New research has revealed that mothers in England and Wales who appear in a first set of care proceedings still have the same probability of returning to court for subsequent (recurrent) proceedings as they did when earlier benchmark studies were carried out in 2015, 2017 and 2020. Almost a decade on from the first study on recurrence, approximately one in four mothers are still at risk of returning to court. In addition, the actual number of mothers in recurrent proceedings has markedly increased since earlier studies, because more families are appearing in care proceedings.

Since the publication of the first research into recurrence, a range of intensive, therapeutic preventative services have been developed for women who appear in recurrent care proceedings (and increasingly their male partners), to help them stabilise their lives, address histories of trauma, and engage with physical, sexual, and mental health services. Evaluations show these services have had a positive impact – however, there are relatively few of them available, access to them is unevenly spread, some have closed due to insufficient funding or budget cuts, and many have small teams providing support to only a limited number of women and men. Ultimately, the research infers that the level of support available is insufficient to meet the need.

Due to previous limited data availability, much of the published research literature to date on care proceedings has not distinguished parents in terms of age. This study found that the risk of returning to court within a 10-year period is higher for mothers who first gave birth when young – increasing from one in four to one in three. A high proportion (41.8 per cent) of mothers appearing in recurrent care proceedings with a baby in England and Wales were estimated to be 14-19 years old when they first gave birth. In the general population of England and Wales, the mean maternal age at first birth ranged from 29.7 years in 2011 to 30.7 years in 2020.

Regional differences in recurrent proceedings were also examined for the first time, for mothers returning to court with a new child. In England, there is a divide between London and the South, and the Midlands and the North, which corresponds with related research on newborn babies in care proceedings. The starkest differences were found between London and the North East.

Reports and summaries in English and Welsh can be downloaded at:
https://www.nuffieldfjo.org.uk/resource/mothers-in-recurrent-care-proceedings-new-evidence-for-england-and-wales

Research findings reveal that only 1 in 3 women received the COVID-19 vaccination during pregnancy

This Swansea University-led study revealed that 1 in 3 pregnant women in Wales have the COVID-19 vaccine during pregnancy, despite 2 in 3 reporting they would have the vaccination given the opportunity. Vaccine hesitancy is an important consideration among vulnerable populations, especially during the COVID-19 pandemic.

The research team used SAIL to bring together and examine GP, hospital admissions, the national community child health, maternal indicators, and COVID-19 vaccination data in Wales. They found that 32.7% of women were vaccinated (at least one dose of the vaccine) during pregnancy; 34.1% remained unvaccinated throughout follow-up period, and 33.2% received the vaccine postpartum. Younger women (<30 years) and those living in areas of high deprivation were less likely to have the vaccine, as well as Asian and Other ethnic groups were slightly more likely to have the vaccine in pregnancy compared to the White group respectively.

This is counter to their findings that 69% of the participants stated they would be happy to have the vaccine during pregnancy with reasons for having the vaccine related to protecting self and baby, perceived risk level, and receipt of sufficient evidence and advice. Reasons given for vaccine refusal included lack of research about long-term outcomes for the baby, anxiety about vaccines, inconsistent advice/information, and preference to wait until after the pregnancy.

Read the full pre-print here:
https://www.medrxiv.org/content/10.1101/2022.05.09.22274769v1.full

20-year study uncovers the cost burden of cellulitis for the NHS in Wales

A 20-year population-scale study of patients with cellulitis has revealed the full economic cost burden for the NHS in Wales. The condition is caused by bacterial infection and can make skin, anywhere on the body, painful, hot, or swollen, sometimes accompanied by blistering. Cellulitis accounts for a significant financial and resource cost to the NHS as well as a large proportion of hospital bed occupancy. This research highlights the potential benefits, to the NHS, to improve diagnosis and treatment pathways for cellulitis.

The study used SAIL data, covering nearly 80% of primary care data of the Welsh population, and 100% of secondary care data. Extrapolations made from this are likely to represent a realistic estimate of the problem. The data analysis showed that estimated annual direct costs for NHS Wales are substantial (>£28 million). Extrapolated for the UK, this amounts to >£571 million. In-Patient events and length of stay costs are the main cost drivers, with annual Welsh NHS estimates of £19,664,126, with primary care events costing £8,890,212. Initiatives to identify early signs/risks of cellulitis, improving the accuracy of initial diagnosis, and improved evidence-based treatment pathways to reduce incidence and severity by even small percentages would result in major financial savings and reduce the burden on patients.

Commenting on the instrumental role SAIL had on the study, lead researcher, Ioan Humphreys, said:

“Using the SAIL Databank for this study was crucial to accurately gauging the economic burden of cellulitis. Utilising the coding to link the patients through from primary care to secondary care was invaluable to us as health services researchers. This piece of work has also led to further SAIL Databank Lymphoedema related analyses which is currently being written up by the same group of researchers.”

The full publication can be found here: https://pubmed.ncbi.nlm.nih.gov/36648008/

New study identifies link between parental health literacy with COVID-19 test results in children

The study has highlighted the importance of parental health literacy and COVID-19 testing. The findings suggested those with higher understanding and access to medical information were more likely to take their child to be tested for COVID-19 and subsequently be identified as positive.

The study collected 7,000 responses submitted prior to March 2020 and the start of the pandemic from the health and wellbeing questionnaire delivered by HAPPEN (the HAPPEN Survey) that focuses on the physical and mental health of primary school children in Wales. The responses were then examined alongside routine PCR COVID-19 testing data, collected between March 2020 and August 2021, stored in SAIL.

Findings from the study indicate associations of healthier behaviours may be related to parental health literacy; the ability to access, understand and apply medical information to make informed decisions regarding medical advice or guidelines. The study also found that children from deprived areas were also more likely to have a positive COVID-19 test result.

They also emphasise the importance of positive parental monitoring behaviours and having the resources required to support these behaviours, such as participating in physical group activities, which may increase exposure to COVID-19, but is essential for child development.

Lead researcher Dr Emily Marchant, based at Administrative Data Research Wales (ADR Wales) in Population Data Science at Swansea University and funded by the Economic and Social Research Council (ESRC), said:

"COVID-19 has brought to the forefront what individuals can do, and the measures society puts in place, for promoting positive health behaviours. Our study suggests the importance of health literacy, associated parental monitoring behaviours, and having the time, knowledge and resources to access relevant health services. As we head into the start of the new school year, we have a real opportunity to establish and build collective 'health literacy capital' for current and future generations."

Read more about the study here:
https://bmjopen.bmj.com/content/12/9/e061344

New study identifies risk factors associated with low birthweights

Multiple births, a short interval between pregnancies and mothers with a maternal physical or mental health condition are more at risk of having a low birthweight (LBW) baby according to new research. Every year 20 million children are born with a birth weight below 2,500 grams and considered LBW babies. The study looked to understand the risk factors for LBW so that resources and interventions could be scheduled effectively. The cohort study comprised 693,377 children born in Wales between 1st January 1998 and 31st December 2018. Participants were selected from the National Community Child Health database. The research team anonymously linked multiple routinely collected administrative datasets to gain a deeper understanding of the risk factors associated with LBW.

The research revealed mothers at the highest risk of having a low birth weight baby included:

Those expecting more than one baby (twins, triplets etc.);

Those who with a pregnancy interval of less than one year; and,

Those with maternal physical and mental health conditions, including diabetes, anaemia, depression, severe mental illness, anxiety, and use of anti-depressant medication during pregnancy.

This study suggests that the most important factors in reducing the risk of LBW include the following:

Address multiple births (e.g., in assisted reproduction practices)

Addressing factors associated with pre-term births (previous history of pre-term birth)

Addressing maternal health, such as reducing smoking, investing in maternal mental health, addressing substance use (alcohol/drugs),

Treating underlying health conditions (diabetes/anaemia),

Promoting pregnancy planning to give an adequate pregnancy interval and healthy weight of the mother, especially for those in deprived urban areas.

Read the study in full:
https://bmjopen.bmj.com/content/13/2/e063836

New research has revealed the impact the COVID-19 pandemic had on senior staff in Welsh schools

According to a Swansea University report, Welsh school senior staff wellbeing is not only lower than the UK average, but senior leaders also experienced moderate to high stress, with more than half displaying depressive symptoms.

More than 170 senior leaders from schools across Wales took part in the Covid-19 School Leadership Survey, which explored the burden and stress that school heads and senior leadership staff experienced during Covid-19. Their responses have now formed the basis of a new report.

The Covid-19 Senior Leadership Study Report 2022 was conducted through the HAPPEN Wales primary school network, part of a wider international study through the Covid-19 Health Literacy Network which spans more than 30 countries.

Previous research had looked at the impact of school closures on pupils' health and wellbeing and the learning and development challenges for teaching and support staff, but this study looked to explore the gap in evidence surrounding the effects of the pandemic on headteachers and senior leadership figures.

The survey revealed that during the pandemic, 75 per cent of the staff said they worked at a level they knew wasn't good for them, 93 per cent put in extra hours and the majority admitted to feeling mentally exhausted by the job. It also asked school leaders to rate the health needs of their staff and when it came to mental health, 91 per cent said the most important issue was stress and coping. Internalised problems such as anxiety and depression were also rated highly.

The report recommends:

- A more strategic approach to supporting the wellbeing of educational leaders in Wales
- Greater clarity on the extent and quality of leadership development provision to specifically support leaders' wellbeing
- Further research charting changes over time in leaders' experience of their wellbeing which could contribute to strengthening the evidence base in this area.

Read the full report here:
https://happen-wales.co.uk/wp-content/uploads/2022/11/FINAL-COVID-19-School-Leadership-Study-Full-Report39.pdf

Estimating mental health service use in children and young people from Welsh Gypsy and Traveller communities

A new feasibility study has shown that it is technically possible to assess the needs of young people who have lived in Gypsy and Traveller sites using routinely collected, linked data. However, the study exposes an under-representation of these ethnic groups within health and administrative datasets which presents a major barrier to being able to accurately assess the needs of this population.

Previous studies have shown severe socio-economic inequalities and poorer physical and mental health in Gypsy and Traveller ethnic groups than in other populations. Anonymised NHS data from Wales was used to examine the mental health of children and young people, as well as code lists for defining mental health conditions, and algorithms to estimate measures of incidence and prevalence. This data was linked to demographic and health (primary and secondary care) data within SAIL.

Using geographic spatial analysis techniques, the team were then able to link this data securely and anonymously to Welsh Government's biannual Welsh Gypsy and Traveller Caravan Count; a census of caravans located in both authorised sites and unauthorised sites.

The study suggests the lack of visibility for Gypsies and Travellers in routine health and administrative datasets represents a significant barrier to recognising the need and provision of services, particularly considering suggested changes to key sources of data used for commissioning, such as the UK census. Furthermore, it suggests that health service providers in Wales should ensure that plans to record Gypsy and Traveller ethnicity are fully implemented across all datasets to ensure underserved groups have a voice in service provision and reconfiguration.

The full publication is available at:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0281504#:~:text=and%20Traveller%20site.-,Conclusion,in%20Gypsy%20and%20Traveller%20sites

SAIL Databank helps spearhead DARE-UK projects

In February 2023, SAIL became part of two new Data and Analytics Research Environment – UK (DARE-UK) projects. Over £1 million of funding has been made available to support these projects, one of which Prof. Simon Thompson, SAIL's Chief Technical Officer, has been named principal investigator for.

This project, DARE-Teleport, will explore how to make data federation much simpler, but just as safe and secure, by automating the federation process. The UK has a wealth of electronic health record data that is used by researchers in a multitude of ways, including studying the COVID pandemic, understanding disease, and improving the many services the NHS provides. Because data is generated and held in many places, TREs have been developed to store data coming from the NHS and other public organisations. However, researchers often want to study data from the whole of the UK. One way of bringing all this data together is the aforementioned data federation – a process that has very specific approvals associated with it. A new novel model of pop-up TREs will be demonstrated by federating anonymised individual-level health data from Wales and Scotland. A key part of this work will be to explore how similar but different approval processes can be brought together so that the data controllers – the people who are ultimately responsible for the data – can have the confidence it is being used appropriately and securely by researchers and in line with the use cases they have approved, but by using a new secure data access process to enable researchers to work-with and analyse data held in physically separated locations.

The other project, DARE-TREFX, aims to show through a real reference implementation how we can use secure Research Objects to move between TREs while still supporting the Five Safes principles that govern and protect patient data; all overseen by patient representatives. Researchers often need to perform analysis over multiple TREs where certain geographical or governance boundaries exist, such as the devolved nature of healthcare in the United Kingdom. Analysis across a federation of TREs would enable timely analysis of UK wide scattered data to answer urgent questions, as we needed in the COVID-19 pandemic. The technologies and standards needed to do this are available now and DARE-TREFX is assembling leading technology providers from ELIXIR-UK and HDR-UK, with three TRE providers and two leading analysis platforms to enable this. The impact will be a step change for how researchers can safely combine data from many sources, and for how data providers from any sector can safely implement this using technology and standards we already have today.

SAIL Databank involved in creation of a near real-time monitoring dashboard for medication dispensing during COVID-19

Measures for keeping the spread of the COVID-19 virus under control resulted in many face-to-face activities moving to online and telephone consultations unless necessary. Many individuals experienced this change in their routine interaction with clinical services such as blood tests, repeated medication, general practice consultations etc. These changes may have resulted in missed opportunities in diagnosing and treatment of some groups and supporting healthcare delivery planning and public health strategies.

In the first study using the Welsh Dispensing Dataset available in SAIL Databank to generate insights on dispensing patterns over the course of the COVID-19 pandemic, a team of researchers led by Prof. Ronan Lyons analysed drug dispensing records for the population of Wales to provide direct insights into how health services, such community pharmacies, are being used across the Welsh population before and after COVID-19. The research team developed an interactive dashboard that illustrates the total number of medications that was dispensed to individual patients from 2016 to 2021. A simple comparison of the total number of items, revealed changes in routine dispensing patterns at the start of the first COVID-19 lock down in March 2020. In Wales, most people who were on repeated prescriptions have been given a supply for 2 or 3 months to keep required interactions under control.

> **Lead authors, Ashley Akbari and Fatemeh Torabi, said, "Many diseases get treated or controlled by drugs; therefore, we think the three main avenues of potential impact for this resource are: 1) drug dispensing patterns can provide novel insights into disease management. 2) It can also be used as a proxy measure for monitoring and evaluation of the clinical and health impact of changes in treatment patterns. 3) It is also important in identifying areas of unmet clinical need."**

This work forms part of the Con-COV initiative; Controlling COVID-19 through enhanced population surveillance and intervention.

The full publication can be found here
https://ijpds.org/article/view/1715

A new study looks at the uptake of COVID-19 booster vaccines by health care workers in Wales

From September 2021, Health Care Workers (HCWs) in Wales began receiving a COVID-19 booster vaccination. This was the first dose beyond the primary vaccination schedule. Due to the emergence of new variants and increasing vaccination hesitancy, there is a need to understand booster vaccine uptake and subsequent breakthrough in this high-risk population. Health care workers (HCWs) have typically been prioritised to receive a vaccine due to their increased risk of exposure to COVID and their potential to transmit infection to patients. It has been suggested that ethnicity, sentiment towards the organisation they work for, disability status, flu vaccine uptake, social pressure, and information from trusted sources were associated with first and second dose uptake for HCWs.

This study, led by researchers from Swansea University, took a cohort of over 73,000 HCWs and looked at the uptake of COVID-19 booster vaccinations from September 2021 to February 2022, comparing them against the uptake of the initial primary vaccination schedule. They also analysed booster breakthrough in the cohort.

They found that sociodemographic characteristics are associated with lower vaccination uptake in HCWs, along with higher risk of breakthrough infections. HCWs who were younger, residing in a more deprived area, living with children, or of a Black or Mixed ethnicity were less likely to receive a booster vaccination. Likewise, similar characteristics were at a higher risk of a breakthrough infection, particularly after the second dose.

The full paper is available at:
https://www.sciencedirect.com/science/article/pii/S0264410X23000348?via%3Dihub

ADDITIONAL PARTNERSHIPS & COLLABORATIONS

SAIL supports new DATA-CAN Centre of Excellence

This UK-wide partnership aims to improve care and outcomes for people with cancer by making high quality health data more accessible through secure and responsible pathways for cancer researchers and health professionals. SAIL Databank will support DATA-CAN researchers to use routinely collected NHS health and clinical trial data related to cancer to answer specific questions and challenges. By working alongside other TRE's from across the four UK home nations, SAIL will also help support the Cancer Data Network (CDN) to improve cancer outcomes for patients by monitoring treatments and interventions over time. More information here: https://www.data-can.org.uk/

SAIL Databank academic selected as Senior Research Leader for Wales

Health and Care Research Wales selected SAIL co-director, Prof. Ronan Lyons, to join their cohort of Senior Research Leaders. Prof. Lyons will act as an ambassador for research into diseases, treatments and services that can change people's lives and drive improvements in patient care. Health and Care Research Wales Senior Research Leaders are among the most prominent and prestigious health and care researchers in the country, and they play a vital role in leading and developing the health and care research community in Wales.

Children younger within the school year are more likely to be treated for ADHD, suggesting immaturity may influence diagnosis

Children younger within the school year are more likely to be treated for Attention-Deficit Hyperactivity Disorder (ADHD), suggesting immaturity may influence diagnosis. The research team looked to explore the impact of holding back children and whether flexibility in school starting dates masks or reduces the relative age effect. The full article can be found at: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-022-13453-w#Ack1

BBC Wales Live with Bethan Roberts investigate whether some young people in Wales are using illegal drugs to manage their mental health problems

In June 2022, Sarah Rees, a Senior Data Scientist from SAIL Databank working on Adolescent Mental Health Data Platform (ADP), was invited on to BBC Wales Live to discuss the findings from an ADP research project led by Prof Ann John which looks at the link between substance use and mental health problems

Health Data Research UK investment boost for Population Data Science Centre of Excellence to unite UK's health data

Health Data Research (HDR) UK has recently announced a £2.5m investment as part of its strategic mission to unite the UK's health data. The fund will support twelve collaborative projects across the UK; SAIL Databank will be an integral part of three of these. Further details of these projects can be found at:

https://popdatasci.swan.ac.uk/health-data-research-uk-investment-boost-for-population-data-science-centre-of-excellence-to-unite-uks-health-data/

SAIL supports new research study to explore potential of ultrasound scans for older people with suspected heart failure and any associated improvements in care quality

The project will assess whether an advanced nurse practitioner can accurately perform a focused ultrasound scan of the heart and lungs in a residential setting and whether this improves the quality of the community pathway for older people. Data in SAIL will be used to understand the scale and context of the problem in Wales, including the real-world pathway, diagnoses, outcomes, resource use and costs for a cohort of older people with breathlessness. In the latter parts of this study, the data gathered from the effectiveness trial will be linked within SAIL to wider Welsh population data sources, paving the way for future research that could impact the broader system of care.

Professor Sir Ian Diamond visits Population Data Science at Swansea University

Professor Sir Ian Diamond, UK's National Statistician, visited SAIL Databank and was particularly complimentary of SAIL Databank, praising it by saying:

> You are working in an incredibly important area. I remember thinking years ago, if only we could link data from different disciplines together, we could begin to answer some very important questions that could impact people's lives. I would urge you to be bold and radical...Your work really can change people's lives.

SAIL Databank Wins Health and Care Research Wales Impact Award

In October 2022, the Health and Care Research Wales' 8th Annual Conference was held in Cardiff and it was announced that SAIL Databank were joint winners of this year's Impact Award. The award recognises the role of SAIL Databank in the One Wales collaborative response to COVID-19

SAIL Databank hosts interns from Nuffield research placements scheme

SAIL Databank hosted two interns from the Nuffield Research Placements scheme during Summer 2022. The scheme culminated in a showcase event to celebrate the achievements of the students from across Wales who were awarded Nuffield Research Placements in 2022. For further information and read more about the proposals the interns developed: https://popdatasci.swan.ac.uk/sail-databank-hosts-interns-from-nuffield-research-placements-scheme/

SAIL Databank at the forefront of case study looking at harmonising electronic health records for reproducible research

The UK-wide Cardiovascular COVID-19 research collaboration known as the CVD-COVID-UK Consortium has published new research describing the implementation of an efficient, transparent, scalable, and reproducible health data harmonisation method. This new method enables multi-nation collaborative research across the UK's Trusted Research Environments (TREs), highly secure computing environments that provide remote access to health and administrative data for approved researchers that can be used in research to improve lives.

SAIL continues to collaborate and lead the Welsh delivery component of BHF Data Science Centre

The BHF Data Science Centre, part of a partnership between HDR and BHF, was created in January 2020 and works with a range of partners to assist in the commencement of research using health data into the causes, prevention and treatment of all heart and circulation-related diseases. SAIL Databank has been a long-term collaborator of the data science centre, using its status as the foremost data repository in Wales to provide data and a trusted research environment for numerous studies to be carried out. More information on the BHF Data Science Centre can be found at: https://www.hdruk.ac.uk/helping-with-health-data/bhf-data-science-centre/

SAIL supports 4 projects as part of AIM £12m funding

The National Institute for Health and Care Research (NIHR) awarded almost £12m worth of funding to support new research as part of the first Artificial Intelligence for Multiple Long-Term Conditions (AIM) funding call. SAIL Databank will be accessed by 4 of the projects funded by AIM – DECODE, MELD-B, AIM-CISC, and AI-MIXED. More information on the AIM funding call and the projects funded is available at: https://www.nihr.ac.uk/news/nihr-awards-12-million-to-artificial-intelligence-research-to-help-understand-multiple-long-term-conditions/28581

SAIL supports establishment of the Alan Turing Institute's Research Support Facility

The new Research Support Facility (RSF), led by the Turing Institute and partnered with Swansea University and the University of Edinburgh, will offer AI and advanced data science support to the research teams funded by AIM. The facility aims to impress upon the collaborations funded by AIM the best practices in data security, data standards, reproducibility, and public and patient engagement. SAIL's role is as theme lead for secure and reproducible infrastructure, supporting those using SAIL, making the most of SeRP-led technology developments such as GPU rollout and federation, and more widely working on creating guidance, outputs, and training in supporting current and future research groups in the MLTC research space using data across the UK

Our Year In Numbers

£36m

Research income secured

CV

25 Jobs

Created in Wales

Core Metrics

Reporting period: **2022/2023**

Health and Care Research Wales infrastructure award to the group

Direct funding awarded

£908k

Jobs created through direct funding

CV

11.5

Grants won during reporting period

Grants won	Led by group	Group collaborating
Number	4	60
Value	£13.4m	£22.7m
Funding to Wales	£13.4m	£13.9m
Funding to group	£11.2m	£810k
Additional jobs created for Wales	25	0
Additional jobs created for group	10	0

25

Number of publications

2

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

Workstream 1

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

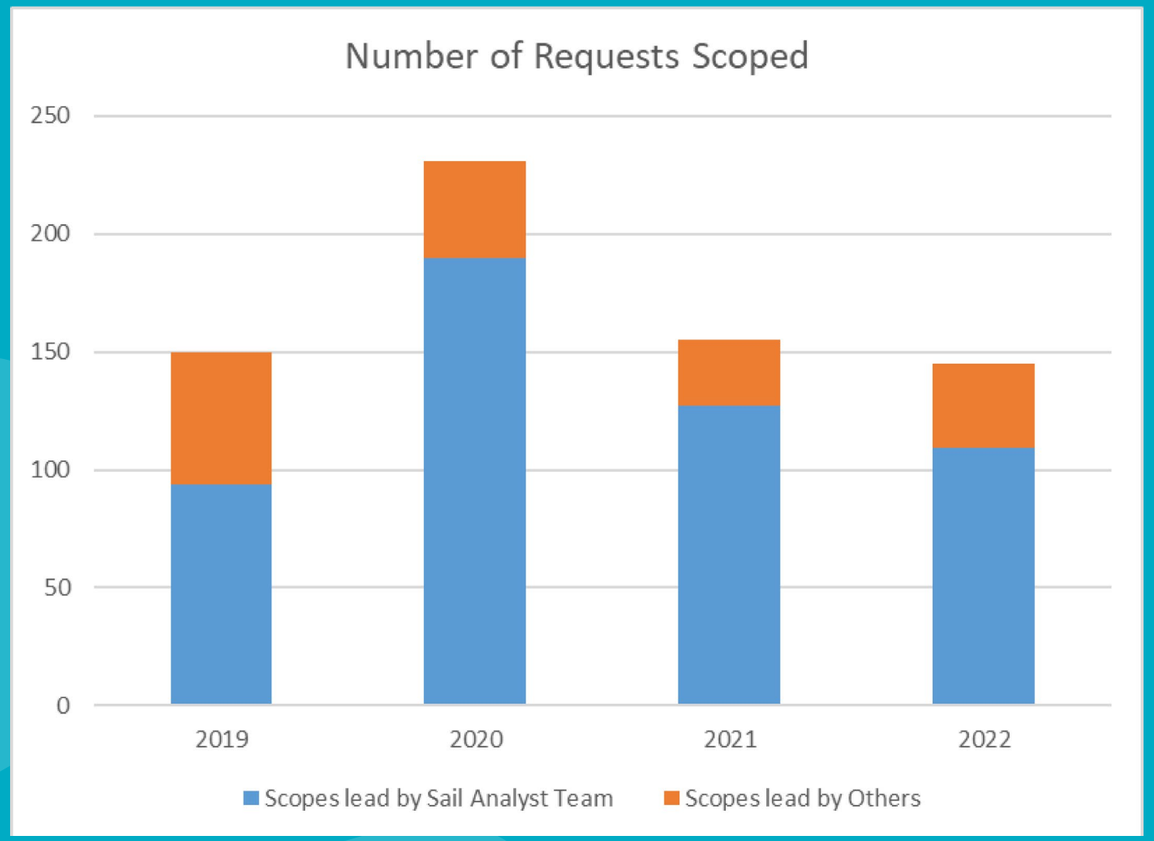
Throughout the year, the SAIL team has continued to offer a comprehensive range of support services to a wide portfolio of research projects. Key activities include:

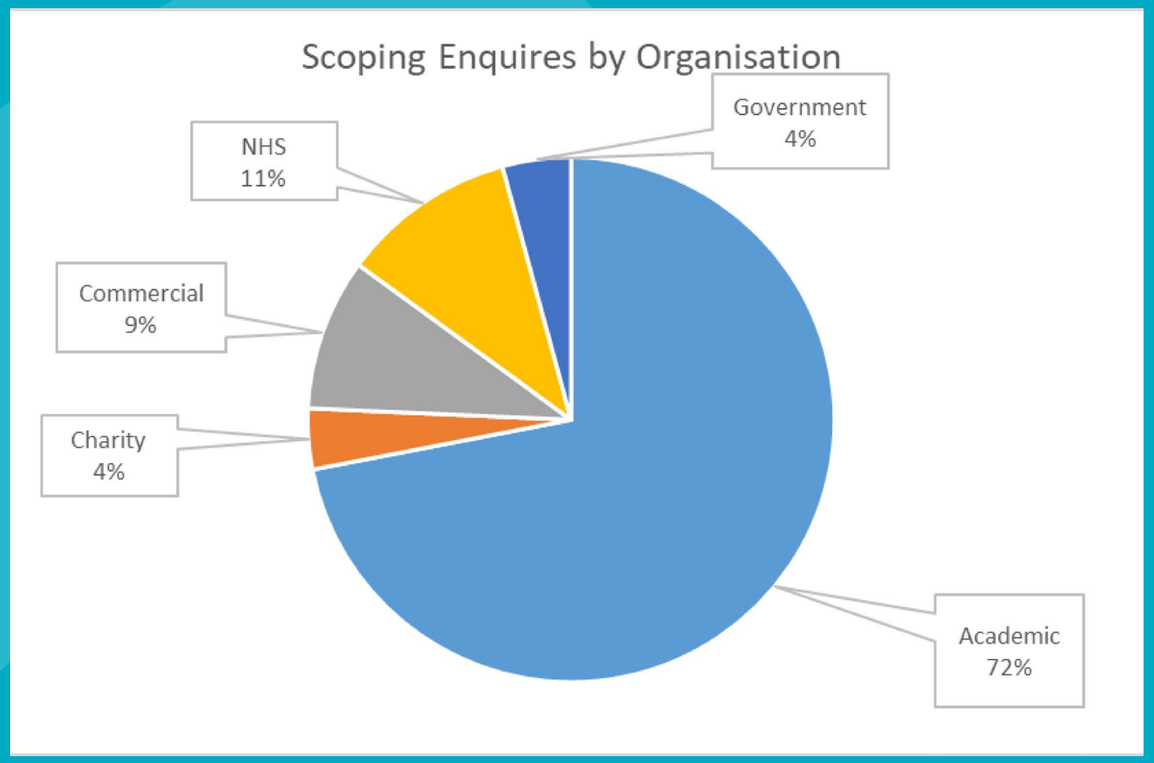
Project Scoping

The first contact that researchers have with SAIL is generally when they approach us with a project which requires scoping. Exceptional scoping volumes were seen during 2020 as a consequence of the pandemic, showing the critical role SAIL had during the pandemic. 2022 saw a minor drop from 2021 in overall scoping requests but remained around the same level as pre-pandemic.

The number of scopes led by the SAIL Analyst team as opposed to other research teams based in Swansea University's Data Science Building still remains more than pre-pandemic suggesting that more researchers are coming directly to SAIL for support for their project, possibly reflecting more complex or wide-ranging requirements.

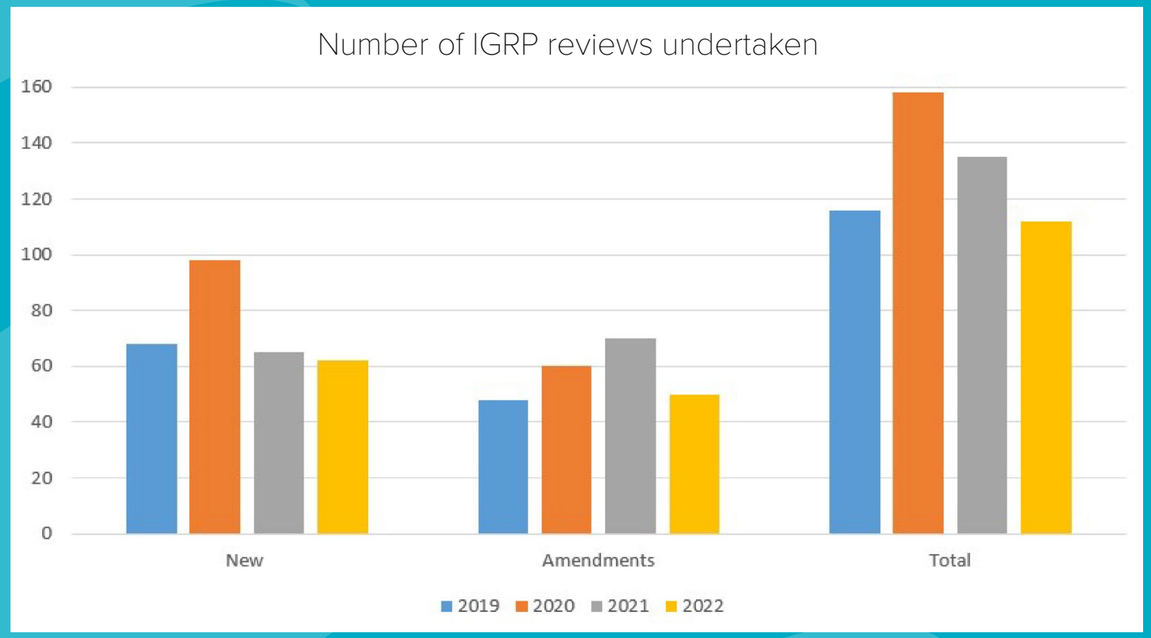
Exceptional scoping volumes were seen during 2020 as a consequence of the pandemic





Governance Review

When researchers want to use SAIL data for their project, they must make an application to gain approval from our independent Information Governance Review Panel (IGRP) and they must apply again if any aspect of their project changes. Whilst in 2023 the number of new applications and requests for amendments is lower than in previous years, the value of the projects approved was greater than in the previous year. This indicates that post-COVID, SAIL is increasingly providing support to bigger and more complex projects.

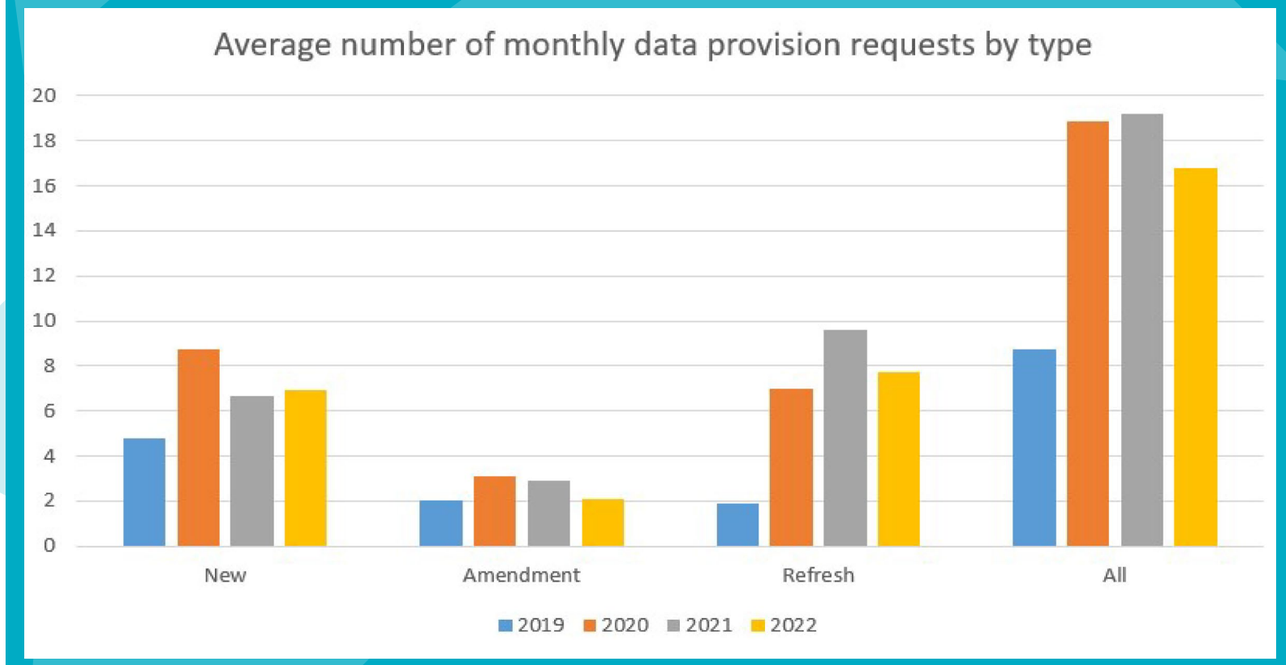


Data Provisioning

Once a project is approved, we provide an internal service to construct the project data views before they are made available to the researchers. This process ensures that the researchers only receive the data needed and approved within their IGRP application and adhering to the data provider's requirements. We have standard project view quality checks which are performed against all project views.

We aim to provision project views within 30 days of application approval, where data is available and approved. In 2022, 97% (175/181) of the project data provisions completed were done within 30 days of approval &/or data available.

2022 has seen a slight reduction in the number of project provision requests compared to 2021, however, the demand is still much higher than pre-2020.



Project Audits

The project audits assure compliance with SAIL governance and with external accreditation such as the Digital Economy Act. In total, 20 projects were audited as part of the annual SAIL projects audit programme. These were selected at random from a report of active projects at the end of 2021. The projects are assessed based on rigorous criteria covering the project lifecycle, and completion of each audit requires input from across the SAIL team.

Some key actions were identified and have been implemented:

1 Audit of non-project folders accessible within SAIL secure environment.

2 Implementation of induction pack and welcome pack to help new users get set up with SAIL and ensure their compliance with governance.

3 Improving processes for monitoring and tracking project audit actions

Findings are reported to the SAIL Operations Group and implementation of actions is monitored by the internal audit committee.

The SAIL analyst team presented an overview of Project Audits at the SAIL User Forum. A copy of the presentation sides can be found here in the main SAIL Forum slide deck https://docs.hiru.swan.ac.uk/display/SUF/SAIL+User+Forum+2022-09-15

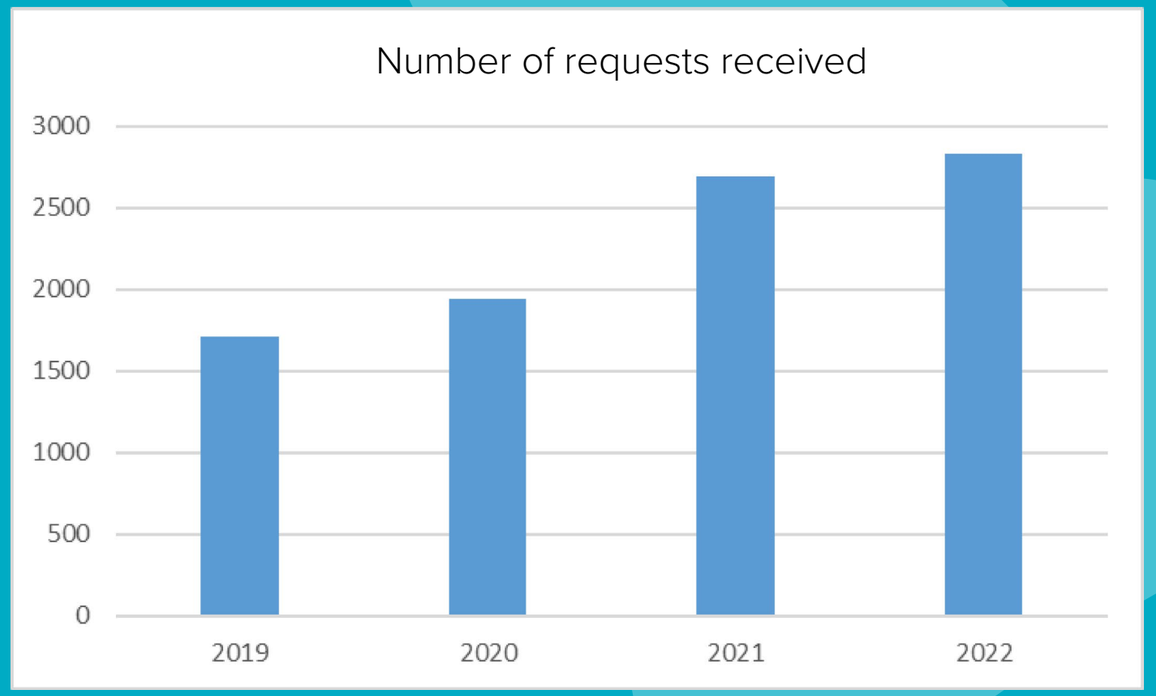
Project audits for 2023 are underway.

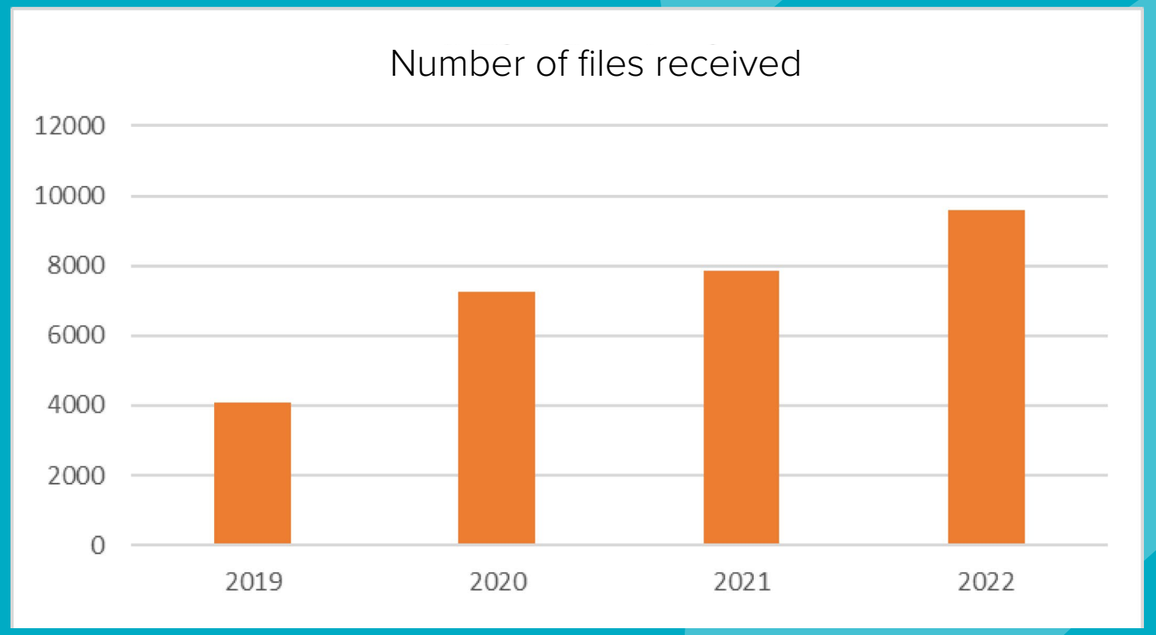
File out Reviews

Researchers access the SAIL data within the secure SAIL Gateway and once they have prepared outputs that they would like released from the SAIL Gateway these outputs are reviewed by our File Out Reviewers to ensure that they comply with the SAIL Output Policy.

Over the last year, the amount of files being requested to be released from SAIL has continued to grow. For the period of January 2022 to December 2022 the team received over 2800 requests (an increase of ~5% on the previous year) and reviewed over 9500 files (an increase of ~22% on the previous year) with ~ 93% of reviews being completed or the researcher being messaged within 2 working days of the request submission. This equates to an average of around 190 files reviewed per week.

2022 also saw the SAIL analyst team introduce audits of the reviews undertaken, where a third reviewer inspects a selection of reviews carried out in the previous week to ensure that they are meeting the required disclosure policy.



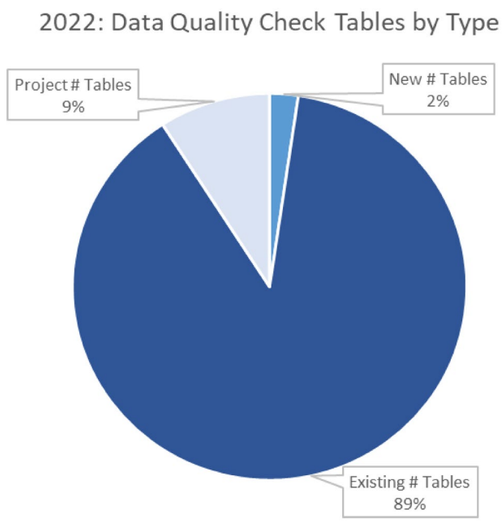


Workstream 2

Data acquisition to meet expressed requirements of our user base

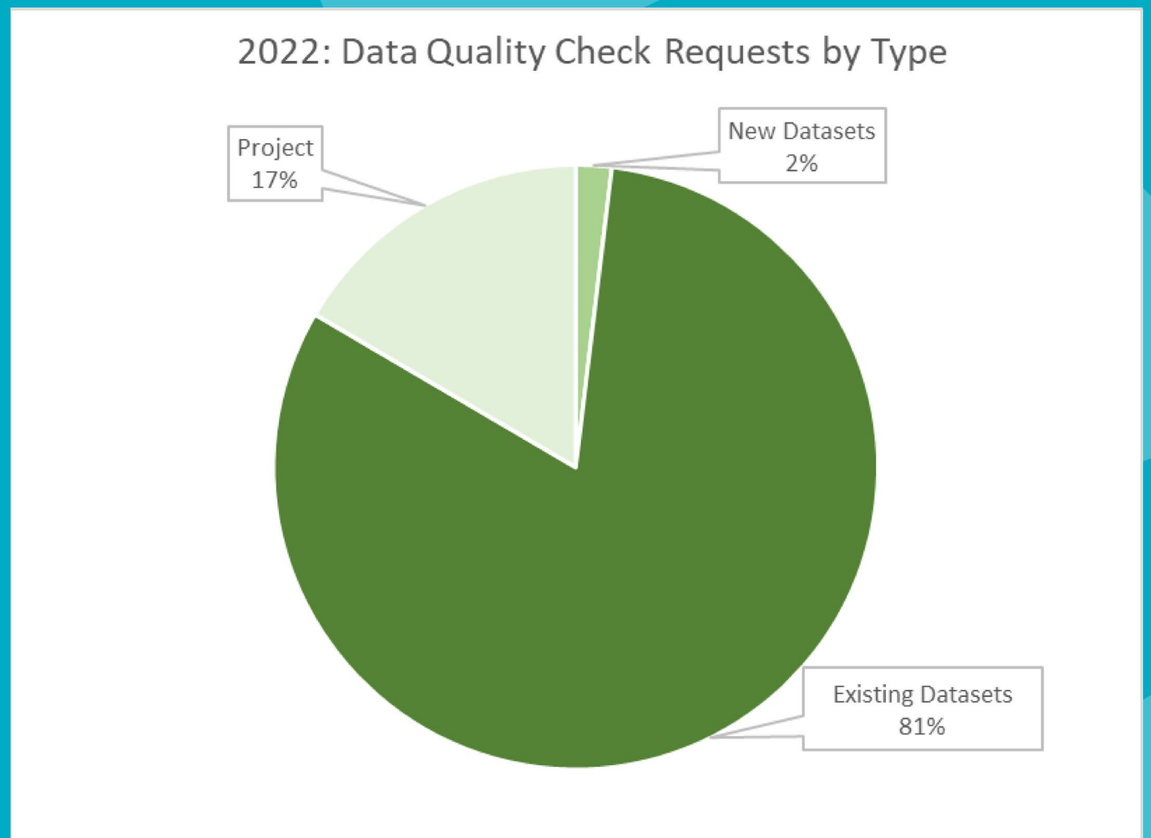
In the period April 2022-March 2023, SAIL continued to improve its data acquisition and views to meet the demands to the researchers using our service for their projects. Over this period SAIL secured access to the following new datasets:

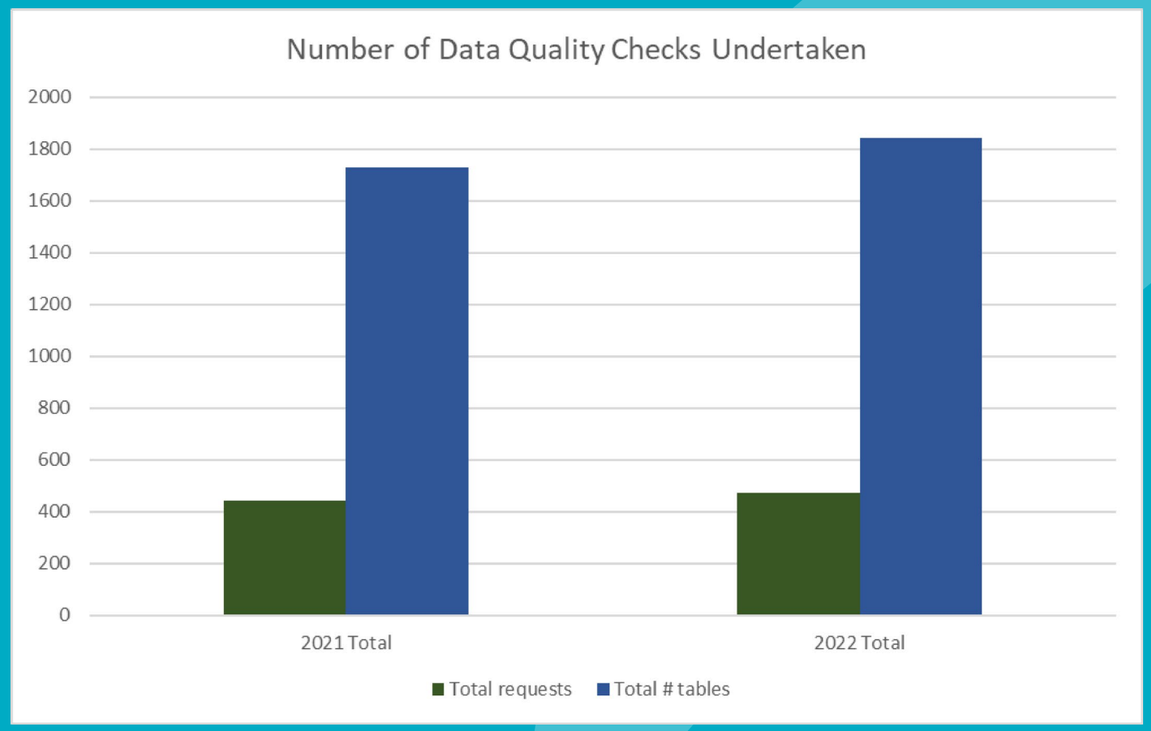
- Ministry of Justice datasets made available under the Data First Programme – Magistrates and Crown Court data along with Prisons and Probation data
- A new Anti-viral dataset has been acquired from Digital Health and Care Wales.
- Census 2021 England and Wales
- School Health Research Network –Student Health and Wellbeing Surveys 2017, 2019 and 2021



As well as these acquisitions, SAIL has updated several key datasets and resolved some known quality and completeness issues:

- Welsh Cancer Intelligence and Surveillance Unit (WCISU) has been updated to include data for diagnosis years 1990 to 2019.
- Public Health Wales have also provided an annual update of Bowel, Cervical and Breast Screening data.
- Cafcass England and Cafcass Cymru have supplied updated data which will be able to be linked to the Family Man dataset recently acquired from Ministry of Justice.
- Welsh Government have re-supplied Education data including PLASC and HESA to address issues identified with previous uploads
- Congenital Anomaly Register and Information Service (CARIS) data has been updated to include data up to the end of 2020.
- A data sharing agreement with Welsh Government for the School Workforce Annual Census has been renewed.





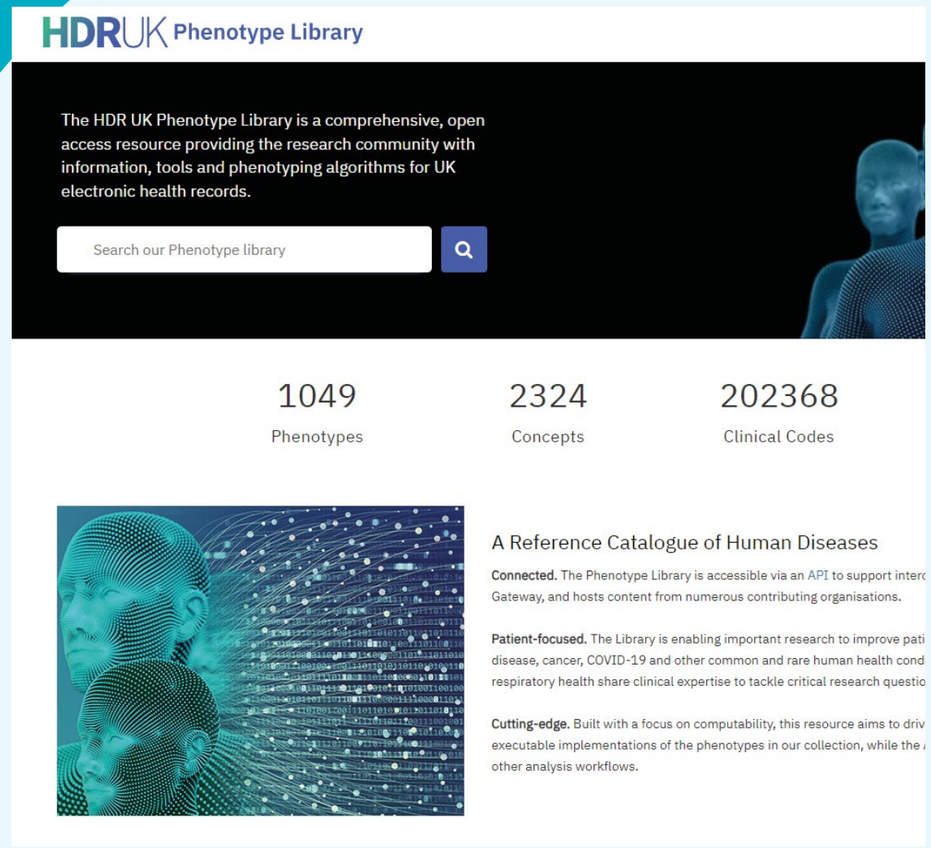
Workstream 3

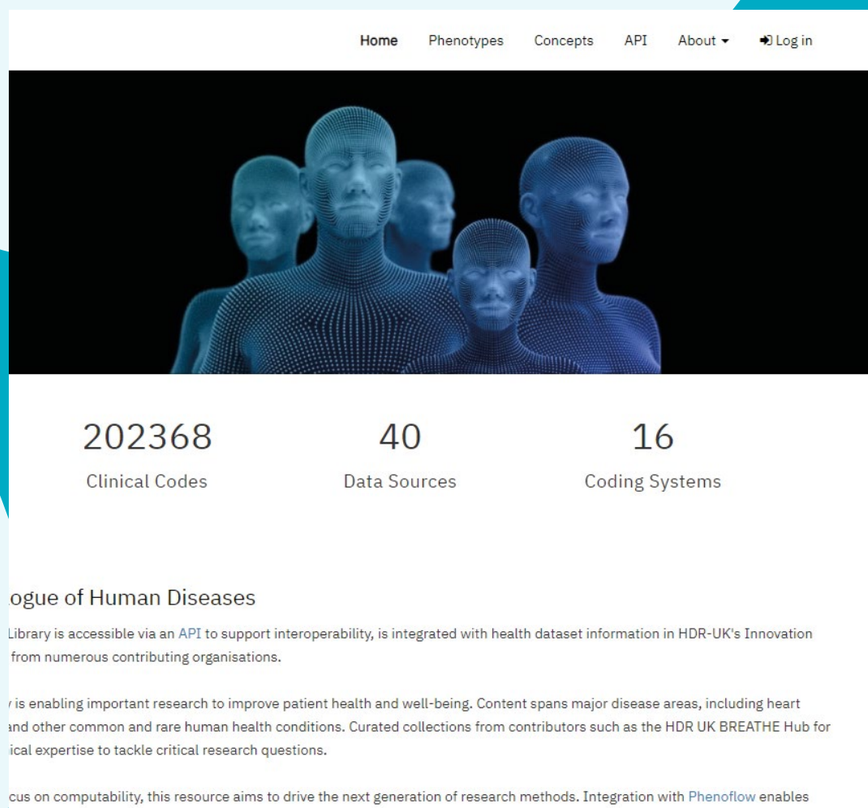
Increase research speed and efficiency by developing a set of re-usable tools (Concept Library) that will enable significantly reduce the time and effort required to conduct a complete analysis of a research question within the SAIL Databank.

We continue to develop the Phenotype Library in conjunction with HDR UK. Built on SAIL's Concept Library codebase and originally released in 2021, this resource has grown into a world-leading tool allowing researchers to share methods and algorithms created for use in health data. Currently, the Library holds over 1040 phenotypes defined against 40 datasets. They represent the work of numerous contributors, including several SAIL projects. Some key developments in 2022 included:

- Adding the ability for researchers to submit their own content
- Further development of an R package that researchers can use to import codes directly from concept library into research workflows
- WDSD demographic dataset was launched along with cleaned research ready views
- The research ready PEDW dataset that was developed started to have an impact on the SAIL Databank user base enabling research. During 2022, 105 projects were provisioned the cleaned PEDW, totalling 435 views.
- Partnerships with projects such as ADP and BREATHE were developed to add their phenotypes to the Library and plan features that meet their needs.

Finally, our key role in delivering the Library as part of HDR UK's Phenomics workstream has been recognised with increased responsibility and funding in the next phase of the project, starting April 2023.





PUBLIC ENGAGEMENT & INVOLVEMENT

SAIL Databank's long-running Consumer Panel is now into its 12th year and continues to be at the heart of our commitment to public engagement and involvement in our work. The panel is comprised of 16 people, but recruitment is ongoing for additional members to maintain the high demand for meetings, which included an extra meeting in November 2022 due to researcher demand. All panel members are members of the public from varying professions. Panel members are invited to assist in many areas of the SAIL Databank process. They provide help and advice to researchers in developing their project ideas, shape funding applications by ensuring that all research includes a clear plan for maximising public benefit and advise researchers on good practice in the dissemination of their research findings. Some members of the Consumer Panel are also members of the Information Governance Review Panel, the independent body which makes decisions on applications to use SAIL data.

During the Covid-19 pandemic the decision was made to move all consumer panel meetings online via Zoom and this remains the case now as it makes the meetings easier to attend for both researchers and panel member.

In addition to the ad hoc support to researchers described above, during the April 2022-March 2023 period the Consumer Panel provided support to several key initiatives, highlights of the research supported includes:

- The Better Outcome through Linked Data (BOLD Substance Misuse Pilot presented to the panel in April 2022

  The BOLD Programme was awarded £19.7m of funding from HM Treasury's Shared Outcomes Fund in May 2021 to deliver a 3-year data linking programme aimed at joining up anonymised datasets across social policy departments to support vulnerable people.

  The programme is divided into four pilot focusing on different policy areas – Victims of Crime, Reducing Reoffending, Substance Misuse and Homelessness. Welsh Government is working with Public Health Wales, the Office for Health Improvement and Disparities, Department for Health and Social Care and ADR Wales on the SM Pilot.

  https://www.gov.uk/government/publications/ministry-of-justice-better-outcomes-through-linked-data-bold/ministry-of-justice-better-outcomes-through-linked-data-bold

- Networked Data Lab (NDL) Wales continue to use the Panel to present their research and the following was published last year after presenting to the panel before the commencement of their study and then presented their findings last November:

  https://saildatabank.com/new-study-examines-differences-in-mental-health-hospital-attendances-for-children-young-people-in-wales/

  https://phw.nhs.wales/news/new-data-highlights-important-differences-in-mental-health-crisis-presentation-amongst-children-and-young-people-across-wales/

- Dr Emily Lowthian visited the Panel to present her study focussed on academic attainment and educational pathways for looked after children in Wales.

  https://saildatabank.com/educational-pathways-for-looked-after-children/

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