<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Introduction</td>
</tr>
<tr>
<td>5</td>
<td>Foreword</td>
</tr>
<tr>
<td>6</td>
<td>Lay Summary</td>
</tr>
<tr>
<td>7</td>
<td>Work Programme</td>
</tr>
<tr>
<td>9</td>
<td>Key Achievements</td>
</tr>
<tr>
<td>12</td>
<td>Into The Future</td>
</tr>
</tbody>
</table>
The SAIL Databank is partially funded by Welsh Government through Health and Care Research Wales.

The SAIL Databank brings together anonymised person level data about the population of Wales from a variety of health and other public services, and provides this as a resource to the research community via a secure portal. Behind this deceptively simple description is an outstanding Welsh success story. SAIL today is the successful culmination of a decade of world leading innovation by a small team of data scientists, information governance experts and IT professionals at Swansea University, funded by Health and Care Research Wales. Their mission was to create a research database which provides the widest possible access to linked data without compromising data and privacy protection.

The solutions developed to meet this complex requirement are unique and world leading. The data security, information governance, and technical models developed to underpin SAIL are regarded as international exemplars of good practice. The privacy-protecting infrastructure developed to host SAIL data has become a product in its own right (UKSeRP) and is used by leading national research programmes such as MRC’s Dementia Platform UK, UK BioBank and the UK Multiple Sclerosis Register.

SAIL data is used for a variety of research purposes, all of which are independently tested to ensure that they have the potential to improve public services or create other tangible public benefit. The richness and breadth of SAIL data is unparalleled anywhere in the world at the present time, and is currently supporting research worth an average of £11,000,000 per annum, most of which is based within Wales.
We are delighted and proud to present our Annual Report, which highlights the work and achievements of the Secure Anonymised Information Linkage (SAIL) Databank between April 2018 and March 2019.

This year has been one of significant achievement and growth for SAIL, with several long-term programmes of work featured in previous Annual Reports now well established, and beginning to yield significant benefits to the research community and create positive social and economic impact. This document describes these milestone achievements, in addition to giving a brief overview of some of the other projects that we have supported, and their impact in improving health, wellbeing and public services for the population of Wales and beyond.

We would like to express sincere thanks to our funders, data providers and stakeholders for their ongoing support, which is vital to our continued existence. We also owe a huge debt of gratitude to the research teams who use SAIL, particularly those in the Administrative Data Research (ADR) Programme, the Health Data Research (HDR) UK Programme and the National Centre for Population Health and Wellbeing Research, as well as the many groups working as parts of the Welsh research infrastructure, who in addition to using SAIL data, engage with us to improve and develop the facilities available through SAIL.

There have been many successes to celebrate this year, but as always, we are aware that there is more to do. Technology advances now make it possible to link existing data in new ways, and to exploit the research potential of new information resources such as genomics or imaging data, but each advance requires development of complex legal, governance and operational frameworks. Welsh Government and other public service organisations explicitly acknowledge the importance of unlocking the full potential of data to improve public services, and are keen to use the increasingly sophisticated data resources we are now working to develop. The next few years look both demanding and rewarding, and we look forward with excitement to meeting the challenges of the future.
The SAIL Databank is based at Swansea University and is a unique and internationally renowned resource of data mainly about the population of Wales, which is used by research groups within Wales and internationally to improve public services and to increase health and wellbeing. SAIL data comes from a wide variety of sources. Each time any of us use public services such as the NHS, Social Services or education services, records are created of why we engaged with the service, and what happened to us as a result. This information has great value and can be used in a variety of ways to improve public services, for example this sort of data may be used to track people who have been on a health promotion scheme such as smoking cessation, to see how much their health improves over time. This helps Welsh Government and the NHS decide if it is worthwhile continuing to invest money in particular schemes, and to check that they are producing the benefits which were expected. All the data in SAIL is de-identified via a non-reversible technical process, so individuals can never be identified and privacy is fully safeguarded. Anyone wanting to access SAIL data for research have their requests assessed by an independent panel (which includes members of the public) who ensure that the research is in the public interest and that there is no risk to privacy and anonymity.

From the very beginning of the SAIL Databank, involving the public was an important priority for us, and we remain fully committed to this. We have an extensive public engagement programme which underpins the work we do, and we involve members of the public in helping to shape research using SAIL data, and approving every instance of data which is made available to researchers. For more details about the SAIL databank, and how you can become involved, see our website [www.saildatabank.com](http://www.saildatabank.com)
The SAIL Databank is based at Swansea University and includes a unique and world famous resource of data about the population of Wales, which is used by researchers to improve public services and to increase health and wellbeing. Our aim is to provide a safe and trusted means of harnessing population-scaled data to;

1) increase the quality and quantity of research to improve services and support better policy making, practice and citizen health
2) innovate and develop data linkage methods and technology to support the worldwide data science research community by working collaboratively with data guardians, academics, members of the public, practitioners and policymakers from Wales, across the UK and internationally.
3) Ensure that Wales remains at the forefront of health informatics research by continuing to develop new and more comprehensive datasets to answer emerging research questions.

We have three main goals, each of which an associated programme of work, which form our core areas of activity as described below:

Supporting more and better research using existing SAIL Data

SAIL has been in operation for twelve years now, and during the first seven / eight years we concentrated almost entirely upon acquiring data and supporting research related to health. This is still a major priority for us – as the diagram shows during the last year SAIL supported a wide range of research with a combined funding value in excess of £11,000,000, the majority of which was health related. However as a result of our collaboration with Welsh Government and the Administrative Data Research (ADR) Programme we have also continued to develop and increase the flows of routine health data into the SAIL databank, in consultation with NHS organisations, public sector bodies and other data-providing organisations.

The quality and depth of SAIL data means that during the last year it has been used by a wider range of research programmes and groups than ever before, including the NHS, UK and international universities and other academic institutions, charities and commercial organisations such as pharmaceutical companies. We have formal collaboration arrangements with many major research programmes, such as Health Data Research (HDR) UK, a major initiative dedicated to uniting the UK’s health data to make discoveries that improve people’s lives. SAIL is a recognised partner in the HDR UK programme and works particularly closely with the Wales and Northern Ireland site to support their research themes of Modernising Public Health and Enabling Precision Medicine. We also work closely with Healthwise Wales, the UK BioBank, the UK Dementia Platform and the UK MS Register. All of these national programmes use SAIL data on a regular basis to support their programmes of research.
Improving data linkage methods and technology

As part of the establishment of SAIL, specialist software and technology had to be developed to de-identify data at source, to load and link it to other data about the same person and to store it securely, whilst allowing approved researchers to access it for their projects. The products developed by the SAIL team for these purposes were recognised as being world leading pieces of innovation, and other research groups with their own collections of data began to express interest in using the technology. About three years ago the technology developed within SAIL was packaged together as a product called the UK Secure Research Platform (SeRP), and this has now been adopted by twelve major research programmes within the UK (including Dementia Platform UK, The MS Register and UK BioBank). Two deployments of the technology have been made internationally, with discussions on others currently in progress. Making the SeRP technology available worldwide is important and valuable as it enables more data to be made available to support medical and other research, in a safe and privacy protecting environment. In addition to deploying more instances of UK SeRP technology over the past year, we have also continued to develop and enhance novel methods and techniques to link, manage and utilise large-scale datasets such as narrative data to ensure that Wales remains at the forefront of health informatics research.

Developing new and more comprehensive datasets to support research

Over the last five years, Welsh Government have recognised an increasing need to ‘join up’ data about the population of Wales, to enable researchers to identify and study how a wide and complex range of factors interact to affect people’s health and wellbeing. In our 2017 Annual Report, we described how these datasets had been used by Welsh Government to assess the effectiveness of several of their national social care programmes, and had informed decisions as to whether further funding would be made available to continue them. Following this success, this year we have been involved in a number of projects to secure a wider range of non-health data and to make this available in conjunction with existing SAIL health data to create fuller ‘blended’ datasets to continue to meet this emerging national requirement. During 2018-19 this expanded range of SAIL data has continued to be been used by Welsh Government to assess the effectiveness of national programmes and services, such as Flying Start. We are working closely with Local Authorities across Wales to put in place the information governance and data security frameworks necessary to allow them to share more data with us on an ongoing basis. During 2018 we also secured funding from the Nuffield Trust for a 4.5 year project to bring in data from the Children and Family Court Advisory and Support Service (CAFCASS) in both England and Wales. The objective of the project is to enable a range of assessments to take place of how effectively the family justice system is working.

“SAIL has set the global benchmark for innovation, security and quality of population health research based on linked data records. The benefits of SAIL will be seen for generations to come.”
Professor Belinda Gabbe, Head of Pre-hospital, Emergency and Trauma Research, Monash University
Building the Research Community.

Training researchers of the future: One of the ongoing challenges facing the research community in Wales is attracting the top quality young researchers of the future. A career in as a Data Scientist requires a diverse range of skills and knowledge which requires ‘hands on’ training and mentoring to acquire. SAIL has launched an ambitious student internship programme to attempt to bridge this ‘skills gap’. During the 2018 academic summer recess period, SAIL provided supported employment placements to ten student interns, who were given the chance to work alongside the SAIL team. The programme was a huge success, with three of the ten now working part time in the SAIL team, whilst continuing their studies, and a fourth in a full time post within SAIL after graduating.

The programme is being re-run for the 2019 summer recess, and we are optimistic that we will achieve similar positive results. We are delighted at the success of the internship programme so far, as recruiting appropriately skilled information analysts to SAIL has been a challenge for a number of years. The student internship programme offers practical knowledge transfer for students, allowing them to gain experience of SAIL and improve their employment prospects.

Working with industry

SAIL has an excellent track record of engaging with the commercial sector, and has established working relationships in place with several key industry partners.

We have previously worked on a number of projects with pharmaceutical company, Novartis, and this year were delighted to be involved in a collaboration with them and Cardiff University on the theme of psoriatic arthritis. Our relationships with Bristol Myers Squibbs and Janssen continue to flourish, with projects in the areas of cardiology and prostate cancer ongoing this year. All three projects will generate recommendations for policy makers such as the NHS and Welsh Government and have the potential to directly improve patient care and experience.

We are in discussion with the Welsh Government Life Sciences Hub about a framework arrangement to give the Life Sciences sector team access to SAIL data for feasibility studies and business cases to support large-scale funding bids by Welsh Life Sciences businesses. Increasing the share of funding secured by Welsh Life Sciences companies is a long term strategic goal of Welsh Government and we look forward to being involved in achieving this.

Increasing research income into Wales

Between April 18 – March 2019, SAIL Databank supported projects with a combined funding value of over £11 million pounds. Without SAIL data, most of these projects would be been impossible to undertake.

In addition to this amount, we leveraged £10.3 million direct investment into Wales for funded programmes such as the ADR and the Family Justice Observatory, which created a combined total of 25 full time new jobs with an employment period of over two years.

KEY ACHIEVEMENTS
Health Related Research

Western Australia Research Platform Launched

Helix is a flagship Technology Research Platform launched on 15th Oct 2018 by Monash University, Western Australia. It will spearhead a new healthcare research era, paving the way for new and faster medical discoveries. Helix offers researchers a ground-breaking new environment underpinned by high-performance, secure and specialised IT infrastructure operated by expert staff. Helix was developed in partnership with SAIL Databank using SeRP technology.

Professor Ian Smith, Vice-Provost (Research and Research Infrastructure), Monash University, Australia said: "The Helix platform provides a tool that will allow the research community to take a huge step forward on the journey towards a healthier society."

"The Helix platform is world-leading and will pave the way towards enabling the earlier detection of some of the most crippling and chronic diseases facing Australians, and indeed the world. Fast-tracking research brings us closer to better understanding disease prevention and ultimately finding cures."

"We have partnered with Swansea University because their experience in health data security is the best in the world."

Health Data Research UK (HDR UK)

The Data Science building at Swansea University (in a collaboration with Queen University Belfast) hosts one of only six HDR UK sites in the country. HDR UK is a new, national organisation that will unite health data assets across the UK to make health data research and innovation happen at scale.

A key aspect of securing funding for establishment of HDR UK Wales and Northern Ireland was access to the high-quality, linkable health and population data within the SAIL databank. Focussing research priorities upon modernising public health and enabling precision medicine, the site will bring together people with expertise in medicine, maths, engineering, statistics, computer science and GIS to form interdisciplinary teams to solve complex health challenges.

Research findings and Publications

During the last year, research outcomes published from SAIL projects included prevention of asthma crisis, effects of air pollution on educational attainment and new treatments for severe liver disease, Graves disease and cardiovascular health.

The UK MS Register was launched in 2011 by the MS Society with the aim of capturing more real world data about living with multiple sclerosis in the United Kingdom. The Register does this in two ways: via simple questionnaires; and linkage of consented participants’ medical records with their questionnaire responses. The MS Register is a large scale user of SAIL data and runs on a SeRP technology platform.

The Register has an ongoing rolling programme of research, with results of studies being published on their website. An example is the 2018 publication “Do people with multiple sclerosis want to know their prognosis?” which presents the results of a three year study into how much information people want about their condition and prognosis at different stages of their MS. The study highlighted discrepancies between how much information people want and how much they actually receive, and includes recommendations to support health professionals in estimating and communicating an individual’s long term prognosis.

During 2018 -19 the Register received £1.2million refunding from the MS Society, and the pioneering work of the platform was recognised by the AMRC in their report “Saving Lives With Patient Data Registries”.

MS Register
Local Authorities and Welsh Government collaborations

SAIL Collaboration with Welsh Government set to continue
During 2018 SAIL Databank (in collaboration with Welsh Government) secured Economic and Social Research Council (ESRC) funding for the next phase of the ESRC’s investments in Administrative Data Research. This followed Wales previous, highly successful, Administrative Data Research Centre Wales programme launched in 2015. This latest award is part of the ESRC Big Data Initiative and secures £5,390,771 income into Wales for a programme of work to be delivered through the SAIL Databank to bring in wider and more varied social datasets. The programme is already achieving results, having obtained data from all 22 Welsh Local Authorities on the Supporting People housing scheme. Data sharing agreements for the national family support scheme Flying Start are currently being negotiated, with some data beginning to flow. Both datasets will be used by Welsh Government to conduct evaluations of the effectiveness of these two national schemes.

Family Justice Data Platform and Analytics Service Project Launched
SAIL, in partnership with Lancaster University, the Children and Family Court Advisory and Support Service (CAFCASS) and the Nuffield Family Justice Observatory, has been awarded funding by the Nuffield Trust to develop a Family Justice Data Platform. The project will bring together Family Court records from England and Wales, and hopes to secure Ministry of Justice and Department for Education data. A key message from a recent consultation across England and Wales is that there is currently insufficient intelligence about how effectively the family justice system is working. This project will enable the Nuffield Family Justice Observatory to systematically monitor performance, for example by tracking the effects of child custody decisions on the ongoing mental and physical health of the child, and related outcomes such as school attendance and attainment. This data will inform future decision-making and enable service improvements.

“Our partnership with SAIL, and its role in ADRC-Wales, has provided ground-breaking research on social and health issues using innovative, safe and secure technology that has directly informed Welsh Government programmes. It’s a crucial part of our research infrastructure here in Wales.”
Glyn Jones Chief Statistician, Welsh Government

“The new data partnership with the SAIL Databank and Lancaster University will be central to the Nuffield FJO’s mission to improve the lives of children and families through better use of data and research evidence. It will improve understanding of how the family justice system is working and enable us to share that understanding with professionals who are making decisions about vulnerable children.”
Rob Street, Director of Justice Nuffield Foundation
SAIL has proved itself highly effective at attracting investment resource for research into Wales, and we aspire to continue this success. We have set ambitious economic targets for the year ahead for research outputs created, research income attracted, jobs created and numbers of projects supported. These will become increasingly challenging to deliver. The academic and public sectors face very tough economic times, and the potential impact of Brexit on access to European funding streams is likely to exacerbate this situation for the foreseeable future.

Despite these challenges we remain focussed on continuing to develop and improve. Over the coming year we hope to continue to develop our very successful programmes of engagement with the pharmaceutical sector. We also look forward to working with the Welsh Government Life Sciences team to promote SAIL to a wider range of companies. SAIL also hopes to become involved in providing evidence to support nationally co-ordinated funding bids, and we are in discussion with Life Sciences Hub about how we can engage with this process most effectively.

We plan to continue to provide ongoing support to Welsh Government in their national services planning and evaluation programme, by using existing data in novel ways and by securing new data through our recently secured ESRC investment. This programme directly supports Welsh Government strategies and policies expressed in Prosperity for All, A Healthier Wales: our Plan for Health and Social Care and The Wellbeing of Future Generations Act and we anticipate that it will continue to have considerable impact on government investment policy. We also look forward to supporting flagship national programmes such as HDR UK in improving health data accessibility across the UK and supporting the conduct of world leading science to effect service change and improvement. We are proud that SAIL is now acknowledged as a partner of choice in such initiatives, and we look forward to continuing to support national and international Data Science developments.

We will continue our long-standing programme of working with Health and Care Research Wales Research Groups and the wider researcher communities within Wales and further afield. We will also seek to continue our positive relationship with NHS Wales, providing support either at individual project or strategic programme level. We have good existing relationships with NHS organisations and hope to continue to build upon these, despite difficult economic times. We are intent on capitalising where opportunities present themselves, and we will continue to work in support of our clinical colleagues across Wales. In particular we are committed to identifying opportunities to work more collaboratively with the NHS Wales Informatics Service (NWIS).

We anticipate the year ahead will be potentially the most challenging of SAIL’s existence to date in terms of financial pressures, but it also offers exciting challenges. This is particularly true in terms of data acquisition, where we are hoping to consolidate on earlier work on new data sources such as narrative data, genomics and imaging, and actually begin to unlock some of the value of these relatively untapped resources. 2019 – 20 promises to be a challenging but rewarding year, and we look forward to it with a great deal of excitement.

Into the Future