

Public involvement & engagement in the work of a data safe haven: a case study of the SAIL Databank

Kerina H Jones^{1*}, Sharon Heys¹, Rachel Thompson¹, Lynsey Cross¹, and David Ford¹

Submission History

| | |
|------------|------------|
| Submitted: | 02/06/2020 |
| Accepted: | 06/07/2020 |
| Published: | 24/08/2020 |

¹Swansea University, Singleton Park, Swansea, SA2 8PP

Abstract

Background

The SAIL Databank is a data safe haven established in 2007 at Swansea University (Wales). It was set up to create new opportunities for research using routinely-collected health and other public service datasets in linkable anonymised form. SAIL forms the bedrock of other Population Data Science initiatives made possible by the data and safe haven environment.

Aim

The aim of this paper is to provide an overview of public involvement & engagement in connection with the SAIL Databank and related Population Data Science initiatives.

Approach

We have a public involvement & engagement policy for SAIL in the context of Population Data Science. We established a Consumer Panel to provide advice on the work of SAIL and associated initiatives, including on proposed uses of SAIL data. We reviewed the topics discussed and provide examples of advice to researchers. We carried out a survey with members on their experiences of being on the Panel and their perceptions of the work of SAIL. We have a programme of wider public engagement and provide illustrations of this work.

Discussion

We summarise what this paper adds and some lessons learned. In the rapidly developing area of Population Data Science it is important that people feel welcome, that they are encouraged to ask questions and are provided with digestible information and adequate consideration time. Citizens have provided us with valuable anticipated and unanticipated opinions and novel viewpoints. We seek to take a pragmatic approach, prioritising the communication modes that allow maximum public input commensurate with the purpose of the activity.

Conclusion

This paper has set out our policy, rationale, scope and practical approaches to public involvement & engagement for SAIL and our related Population Data Science initiatives. Although there will be jurisdictional, cultural and organizational differences, we believe that the material covered in this paper will be of interest to other data focused enterprises across the world.

Keywords

public engagement; data safe haven

Background

The SAIL Databank was established in 2007 by the Population Data Science group at Swansea University (Wales) [1]. It was known at that time as the Health Information Research Unit; the department was renamed Population Data Science in 2019 [2]. SAIL was created in recognition of the immense, untapped research potential of individual-level data collected in the course of health and other public service delivery. Its

predicate was that making these vast arrays of data accessible safely would open up ways to answer important questions that could not otherwise be addressed without prohibitive effort and cost. SAIL is the bedrock of further Population Data Science investments hosted at Swansea University [3], and the scope of SAIL data has expanded to include administrative data that were not previously accessible in this way (such as education, housing and employment) and emerging health data types (such as genomic, free-text and imaging). In this way, SAIL

*Corresponding Author:

Email Address: K.H.Jones@Swansea.ac.uk (Kerina Jones)

is becoming an increasingly rich resource for population data science – ‘*the science of data about people*’, including wider factors that influence wellbeing [2].

Aim and scope

The aim of this paper is to provide an overview of public involvement & engagement (PI/E) in connection with the SAIL Databank and related Population Data Science initiatives. Since our work is based mainly on data from the population of Wales, we refer to the Wales and UK context. However, we believe the principles will also be of value to a wider readership.

Approach

Defining public involvement & engagement

In the context of our work, we refer to PI/E as a purposeful set of activities designed to enable an individual or an organisation to work with the public in a two-way arrangement. We primarily focus on the general public, but we include patients, practitioners and other relevant stakeholders within scope. We refer to public involvement as activities in which individuals have a specific role such as being included as co-applicants on research proposals or as members of steering groups for strategic developments. We refer to public engagement as a broader two-way process of listening to, and interacting with, the public about research, particular developments or associated issues in order to take their views on board [4, 5]. Both of these are distinct from communications to disseminate information in a one-way direction only. In this document, we will refer to PI/E collectively as a single entity covering a spectrum of activities, commonly presented as a ladder (Figure 1). The original model was developed by Arnstein [6] and it has been adapted many times, as we have done for our model. Other (closely-related) terms could have been chosen for the rungs of the ladder; the key message being to show a spectrum of levels of PI/E.

Rationale for public involvement & engagement

The perceived importance of PI/E in public initiatives, such as research and service delivery, has grown increasingly in recent years. This is good practice in a democratic society to demonstrate respect, responsibility and openness towards individuals and society. It has become an imperative for publicly funded initiatives with research councils and charities embodying this in their policies and strategies and signing up to the UK Research and Innovation (UKRI) concordat [7]. Health and Care Research Wales (HCRW) state their commitment to creating an environment in which all the health and social care research that takes place in Wales happens with the public, for the public, so that the people of Wales have a key role in improving the quality and relevance of research [8].

In relation to law in connection with the use of personal data, the EU General Data Protection Regulation 2016 (GDPR) [9] and the UK Data Protection Act 2018 (DPA) [10] make it clear that uses of personal data should be transparent for the exercise of the rights of

the data subject. Furthermore, the principles of ethics for science communication recommend proper involvement and engagement with audiences, respecting their views and moving away from the idea of individuals as passive recipients to be educated [11]. Work by organisations such as INVOLVE and Understanding Patient Data (UPD), as well as numerous research studies, have shown people’s interest in, and the benefits of, PI/E [5, 12]. This is important to enable the social responsibilities of those engaging, and those being engaged, to be mutually fulfilled.

Unlike many types of research that naturally come into contact with study participants as part of the activities, research relying only on the use of data can be more remote from the individuals that provided the data. Because of this, there is a need to bridge the communication gap from the most basic to the more complex modes of PI/E. In recognition of this, an international consensus position has been set out for data-intensive research [13].

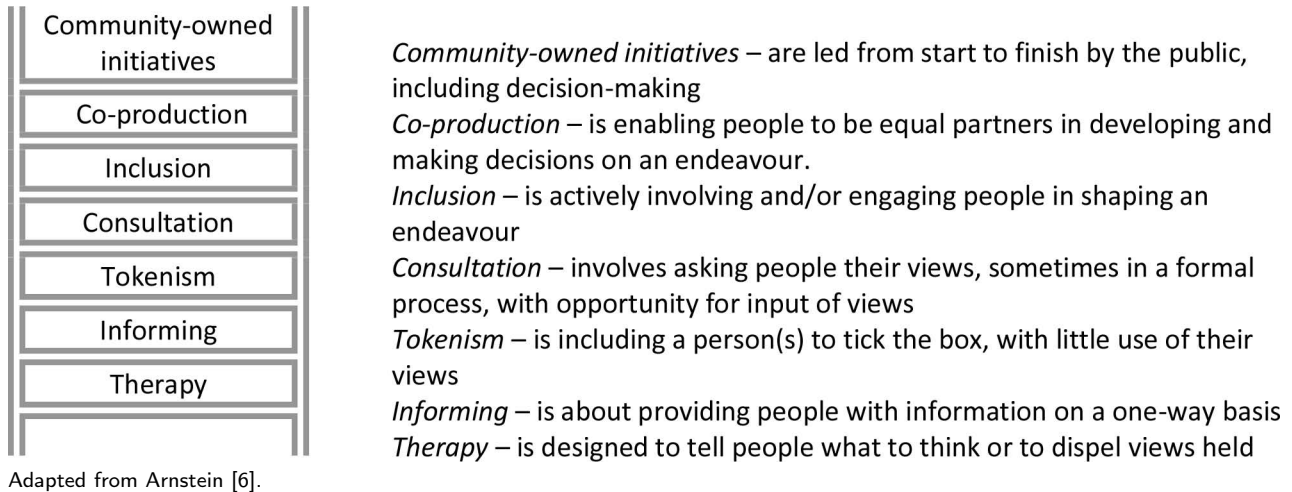
Scope of our public involvement & engagement activities

The scope of our PI/E activities is based on the ladder diagram shown in figure 1. We endeavour to take a pragmatic approach to PI/E, to maximise the use of our resources and to work most effectively with the public. We exclude *Therapy* on the basis that it is not good practice to manipulate the public towards a particular viewpoint. We exclude *Community-owned initiatives* since ultimately, we are accountable for the work of SAIL. We recognise the key importance of *Informing*, but see this as primarily a one-way mode of communication when its function is static dissemination. However, it is both an essential prerequisite of, and integral to, two-way interactions. We acknowledge *Tokenism* as sometimes being better than nothing, but we discourage it as an approach to PI/E. As such, the main two-way categories in scope for our PI/E activities are: *Consultation*, *Inclusion* and *Co-production*. Within these categories, the method of engagement varies, ranging from quick surveys eliciting top-of-mind responses to more in-depth engagement and deliberative activities. We recognise that the boundaries between the categories cannot always be clearly defined, that a given activity may include several modes of communication and may need to be flexible and change over time. We endeavour to select the mode to fit the purpose of the activity, be it, for example, advising, shaping, decision-making or reassurance on decisions already made.

Public involvement & engagement roles

We take the work forward via leadership and operational roles in PI/E. Across the Population Data Science Department, we currently have a designated associate director, a full time operational PI/E officer and a full-time research associate in information governance and public engagement. In addition, various associated projects employ staff with portions of their time dedicated to PI/E. We take this approach because PI/E is a specialist area requiring particular expertise and dedicated time. We also recognise that there are different aspects to PI/E and that different skill sets are needed depending on the nature of the activity. For example: the associate director leads on PI/E policy and strategy and a programme of PI/E

Figure 1: Ladder of public involvement & engagement activities



research; the PI/E Officer leads on running the Consumer Panel, assisting researchers with public involvement and other operational activities; and the research associate leads on particular research studies.

Public involvement & engagement policy

Following the development of Population Data Science [2] and the subsequent renaming of our department, we developed a new PI/E policy for SAIL and related initiatives. Owners of particular initiatives and projects will have their own responsibilities in relation to PI/E, but may wish to make use of the SAIL policy. We developed a set of principles and policy statements for PI/E. Because of the nature of our work, our funding and our location, the principles accord with the statements of UK Research and Innovation [7] and Health & Care Research Wales [8], the UK Standards for Public Involvement [14], and the Consensus Statement on Public involvement & engagement with Data-Intensive Health Research [13]. The draft policy was reviewed by the SAIL operational staff and a group of international stakeholders (the SAIL Scientific External Advisory Board), before sign-off by the SAIL Management Board (the SAIL executive function). Plans are underway for a corresponding PI/E strategy to be developed in collaboration with our Consumer Panel before a similar review and sign-off process. We set out the six principles and corresponding policy statements below; the full policy is available on the SAIL website [15].

1) SAIL has a duty to work with the public and, consequently, has a strategic commitment to public involvement & engagement

- We embed public involvement & engagement at all levels of our work and include members of the public in strategic planning.
- We have positioned strong leadership to champion public involvement & engagement and we foster a culture of public involvement & engagement amongst our staff.

- We invest in dedicated posts to work on public involvement & engagement activities and support all staff members to work with the public on relevant areas of their work.

2) SAIL is committed to excellence in active and meaningful public involvement & engagement

- We endeavour to understand public involvement & engagement, as appropriate to context, which is shared and used consistently across the organisation.
- We promote high standards by providing advice and guidance information to staff on good practice in public involvement & engagement.
- We provide access to specialist training on public involvement & engagement for staff and members of the general public.

3) SAIL respects heterogeneity and provides members of the public with inclusive opportunities to engage and be involved

- We aim to include members of the public at all levels of our work, including strategic direction and specific data-intensive research studies.
- We seek to be inclusive of all ages, ethnic groups, cultures, socioeconomic levels, lifestyles and other definable interests in our public involvement & engagement activities.
- We recognise that there are limits to people's interest and seek to encourage, but not to coerce, inclusion.

4) SAIL is committed to transparency, clarity of purpose and fostering respectful relationships in public involvement & engagement activities

- We aim to converse with the public in a clear way and to be transparent about our activities and the use of data in research studies.

- b) We recognise the importance of clarity of purpose in public involvement & engagement activities to promote mutual understanding and so that suitable, realistic expectations can be set.
- c) We engender respectful, non-hierarchical relationships in our public involvement & engagement activities.

5) SAIL recognises and values the contribution of our staff for their public involvement & engagement activities

- a) We encourage managers to include scope for public involvement & engagement, as appropriate, in workload planning and in induction for new staff.
- b) We promote the benefits of public involvement & engagement to all staff at all stages of their careers.
- c) We communicate and celebrate our staff successes in public involvement & engagement, and through this we encourage others to work with the public.

6) SAIL assesses, reports on and acts on the impact of involving and engaging the public

- a) We promote public involvement & engagement activities that are designed to produce impact in order to value the input of all who contribute.
- b) We assess the impact of our activities and make modifications for improvement to maximise impact.
- c) We are accountable for our decisions and ensure the outcomes of activities are duly taken into account and acted upon.

Beginning with the work of the Consumer Panel, we set out the main practical aspects in our approach to PI/E below.

Consumer Panel

Roles

We established a Consumer Panel in 2011 to provide a public voice and gauge social acceptability on the work of SAIL. It remains an essential attribute of SAIL, but also provides public views on the work of associated initiatives [16]. The Consumer Panel is currently comprised of 16 members of the Welsh general public. Its role includes [15]:

- Discussing proposals for research and the implications of findings
- Providing views on data protection issues
- Reviewing information designed for a lay audience
- Guiding on how to recruit people to study steering groups
- Advising on how best to engage with the wider public
- Acting as advocates for data linkage research

In addition to being members of the Consumer Panel, three people are part of the independent Information Governance Review Panel (IGRP) which reviews all proposals to use

SAIL data for research. They represent the public voice on data use alongside professionals from the Welsh Government, the British Medical Association, Public Health Wales, the Research Ethics Service, the NHS Wales Informatics Service and Swansea Bay Health Board. This membership ensures a range of angles and responsibilities is covered in the review process [17].

Discussion topics

The Consumer Panel meets routinely for a whole day on a quarterly basis. A summary of the main topics discussed in the past three years (2017–2019) is shown in Table 1. It shows the topic and the purpose of the engagement with the Panel. We provide further details on some examples, including specific guidance provided by the Panel as illustrations.

The use of mobile phone data for health research

This study was funded by the Natural Environment Research Council and conducted in association with the *O₂* mobile network operator [18]. It set out to: understand more about how mobile phone data are used in association with health data for research; gain public views on the use of such data; and to develop recommendations for an ethically-founded framework. The researchers conducted workshops with various groups of citizens, including one with the Consumer Panel in June 2017. Following an overview of the study, members stated they did not feel well-informed about the use of phone data in health research. While they saw the value of this, they wanted to be better informed, including having the option to opt-out, and they were particularly concerned about the data being sold for profit. Their views were incorporated into articles on public views and the development of an ethically-founded framework for the use of mobile phone data for health research, as well as a detailed report for *O₂* [19, 20].

Green-blue spaces, common mental health and wellbeing

The green-blue spaces project is funded by the UK National Institute for Health Research [21]. It's exploring if there is a link between spending time in parks, woodlands, and beaches and better mental health. The researchers met with the Consumer Panel in June 2018 to discuss the development of a typology to help ensure the study is not making unreasonable assumptions on how people use these spaces. The members were asked: the purposes for which they use their garden; what would hinder them from using green-blue spaces; and their reasons for visiting green-blue spaces. It became evident that how people use and value green-blue spaces was quite variable. The Panel noted that people with mental health difficulties might be hindered from visiting green-blue spaces by reason of their condition and that they would like to see greater investment to assist people. The input from the Panel has been taken forward to inform the development of a typology characterising green-blue spaces to represent how people engage with their outdoor environment, and it will be used to inform policy making [22].

Using child-related family court records for anonymised data linkage research

The Family Justice Data Partnership has been funded by Nuffield Family Justice Observatory to

Table 1: Topics discussed by the Consumer Panel

| Date | Topic | Purpose of engagement |
|-------------|---|--|
| Jan 2017 | Sharing data across organisations | Views on the exchange of anonymised information across centres rather than the data being used or held in one centre alone |
| Jan 2017 | Influence of antipsychotics on children | Sharing information and gaining viewpoints on the findings |
| Jan 2017 | Anxiety and depression as risk factors for cardiovascular disease | Guidance to develop the proposal |
| Jan 2017 | Supporting government commissioned research | Advice on a proposed way of working |
| Apr 2017 | Reasons for anxiety among people with Multiple Sclerosis | Sharing information and gaining viewpoints on the findings |
| Apr 2017 | Using genetic data with health data for research * | Views on the issues to consider in bringing genomic data into SAIL |
| Apr 2017 | Engaging with general practitioners about providing data to SAIL | Ideas to assist with this process |
| Jun 2017 | Self-report data for the Multiple Sclerosis Register* | Advice on the frequency people Multiple Sclerosis should be asked to provide information |
| Jun 2017 | The reuse of administrative data | Opinions on a proposed way of storing and reusing data in an administrative data research centre |
| Jun 2017 | The use of mobile phone data for health research* | Engaging the Panel in a workshop to gain their views |
| Jun 2017 | Administrative Data Research Centre Wales | Reviewing the text to be used on the website |
| Sept 2017 | Blue-green spaces and mental health* | Views on mental health and the outdoors, and how to engage the general public on this study |
| Sept 2017 | MSc in health informatics | Whether the members would be willing to receive students to discuss their research proposals for the Masters |
| Sept 2017 | Prognostic tool to identify psychosis using Artificial Intelligence | Opinions on issues to be aware of due to the sensitivity of the topic |
| Sept 2017 | Research events: seminars and science festival | Informing the members they are welcome to attend research events put on by Swansea University |
| Jan 2018 | SAIL marketing plan | Views on the proposed approach to promoting SAIL |
| Jan 2018 | The use of mobile phone data for health research* | Feeding back the results from the public engagement exercises and the recommendations outlined in the project report |
| Jan 2018 | Self-report data for the Multiple Sclerosis Register* | Feeding back on the Register participants' views on questionnaire frequencies |
| Jan 2018 | Parenting tips and training | Advice on the development of a grant proposal |
| Apr 2018 | Using genetic data with health data for research* | Engaging the Panel in a workshop to gain their views |
| Apr 2018 | Welsh Government Flying Start Data Linkage Project | Discussing possible dissemination pathways for the study |
| Jun 2018 | Blue-green spaces and mental health* | Working with the Panel to develop a typology for the study |
| Jun 2018 | Individual deprivation in Wales | Advice on developing an individual level measure of deprivation based on linked data |
| Jun 2018 | Suicide Risk Estimation Support Tool | Views on using artificial intelligence to estimate suicide risk |
| Sept 2018 | Multi-morbidity and polypharmacy | Advice on priorities to include in the study protocol |
| Sept 2018 | The ACTIVE project – physical exercise among school children | Sharing information and gaining ideas on advising schools about prioritising physical exercise |
| Sept 2018 | Data-driven wearable technologies | Gaining input on terminologies to use when communicating with the public |

Table 1: Cont.: Topics discussed by the Consumer Panel

| Date | Topic | Purpose of engagement |
|-----------|--|--|
| Jan 2019 | Data governance for the use of free-text clinical data | Seeking opinions of the study design |
| Jan 2019 | Fire and safety interventions and health implications | Gaining views on the development of the study |
| Jan 2019 | University of Manitoba public engagement | Learning about the Panel and the members' experiences to inform setting up a similar structure |
| Jun 2019 | Effectiveness of the Caldicott Principles | Sharing information on a study and gaining input on design |
| Jun 2019 | Online emotional wellbeing intervention | Exploring ideas on developing the website and resources |
| Jun 2019 | Built environment and adolescent mental health | Obtaining feedback on a fellowship application |
| Sept 2019 | Using child-related family court records for anonymised data linkage research | Exploring the social issues in using potentially sensitive records |
| Sept 2019 | Effect of stopping inhaled steroids in people with Chronic Obstructive Pulmonary Disease | Views on study design and any ethical issues that have not been considered |
| Sept 2019 | Website content development | Providing the Panel with a variety of images to assess suitability and preference |

These are the main topics discussed by the Consumer Panel and the purpose of the engagement during the three-year period from 2017 to 2019.

* As can be seen, researchers sometimes discuss their work with the Panel on more than one occasion through the life-cycle of the project.

effect step-change in the use of family justice administrative datasets, to ensure timely, accessible outputs are provided to the Observatory and its range of audiences [23]. The researchers met with the Panel in September 2019 to introduce the study and gain members' views on issues to consider in the use of potentially sensitive data. Members acknowledged the need for this type of research in order to protect children and wanted to be reassured the findings would inform practice. Among the topics they would like to see enabled were: earlier identification markers of children being at risk before the crisis point of reaching the courts; and linkage to crime data to inform supportive interventions to help children remain out of the criminal court and reach their potential in life. The panel also advised on the public engagement pathway for the study, including beginning with gaining advice from professionals in the field, and the importance of including both parents (not just mothers) in discussions. The advice provided is being taken forward to inform the work of the Data Partnership [24, 25].

Survey

We gained viewpoints from members on their experience of being on the Consumer Panel and their perceptions on the work of SAIL. Due to the COVID-19 crisis we were unable to hold a meeting (in April 2020) and so this work was conducted as an email survey (mid-April to mid-May 2020). Members were given the option of returning their completed survey to one team member to be passed anonymously to the originator for analysis. We received eight responses from among the 16 members. They were initially asked to write about their own experience of being part of the Panel. The main positive points arising from this part of the exercise were: feeling welcomed; receiving clear explanations; hearing about research studies; receiving follow-up information on studies; and being able to

contribute to the work. Some member comments are given below:

'I was warmly welcomed and have found my experience of being a member of the panel both very informative and my comments at meetings valued.'

'It's been interesting to see how researchers do appear to take account of what members of the panel have to say and that we have been able to make a difference.'

'The Panel has provided a platform to engage with researchers, keep informed about developing policy around the use of linked data and hopefully we the panel members have contributed to improving patient health outcomes in some small way.'

The main learning points were: some members feeling apprehensive on joining the Panel; topics having to be cut short due to time constraints; discussions losing focus; not always having the opportunity to make a point; and possibly increasing the frequency of the meetings.

'When I first joined the SAIL Panel I was not sure how it all worked and how I would fit in and make a difference.'

'Sometimes, I did feel that the Group was a little too large for me to be able to give as much feedback as I would have liked.'

'Perhaps the frequency of meetings should be increased to six a year rather than just four and would allow for unforeseen cancellations of scheduled meetings.'

Members were then asked to provide their views on strengths, weaknesses, opportunities and threats (SWOT) for SAIL. The main themes arising from the responses are provided in Table 2. These and the views on being a Panel member have been provided to management so they can be taken forward to inform the work of SAIL and how we conduct our engagement.

Table 2: Summary of the SWOT analysis with the Consumer Panel

| Strengths: What does SAIL do well? | Weaknesses: What could SAIL do better? |
|--|---|
| Wide range of datasets and data linkage Reputation with the health/social care service and research community Enabling access to anonymised data Public engagement | Greater awareness raising among local authorities Costs for data access and time to data provision Lack of data from some general practices Unclear how SAIL engages with the wider public |
| Opportunities: What's on the horizon for SAIL? | Threats: What dangers does SAIL face? |
| Expanding the types of dataset held Working across national boundaries Issues arising out of the current [Covid-19] pandemic Dissemination of outputs and impact to show the benefits of linked data research | Continuation of funding if the Covid-19 crisis results in cut-backs Commercialisation of data and presumed consent Competitors with better datasets Maintaining standards: security, anonymity and faith of the public |

This table shows the main themes in the responses to the SWOT analysis.

Wider public involvement & engagement

In addition to working with the Consumer Panel, we engage with many wider public groups at bespoke meetings or existing events such as science festivals. These include citizens of particular age groups, or from different sectors of society, or with particular health conditions and areas of interest. We provide some examples as illustrations.

Multiple sclerosis register The United Kingdom Multiple Sclerosis (MS) Register is funded by the MS Society with the aim of adding new knowledge about MS to inform policy and practice. It combines clinical data from neurology clinics with a wealth of self-reported data (via Patient-Reported Outcome Measures (PROMS)) from people with MS and routinely-collected data, as available [26]. As well as carrying out research based on the data, the Register has a programme of public engagement to ensure the views of people with MS are incorporated into developments [27]. This includes regularly attending events to meet with people with MS and a range of engagement activities. The Register team recognises that continued co-development with people with MS is pivotal to continued provision of data and to meet their needs as participants. As a result, usability changes have been made to the Register website and the option to download a copy of their own PROMS results has been added. Further work is underway, including a study to explore the acceptability of digital technologies in self-monitoring [28].

Adolescent mental health data platform Engagement with young people is at the heart of the Adolescent Mental Health Data Platform (ADP) [29]. The views and opinions of young people steer our research to make sure we are answering the right questions in the right way. The 'Be Heard' initiative was developed as part of the ADP website to engage young people affected by mental health issues either personally or through family and friends. As part of this, the Big Data and Mental Health Research Survey collected the views of almost 3000 young people aged 16–24 on the way in which the data should be used for research. Participants expressed views about how data should be protected, who should have access to it

and what kinds of research it should be used for. The results of this survey will be used to guide our research going forward and how we communicate this with participants and the public. The 'Be Heard to Be Seen' project funded by Medical Research Council is an arts-led participatory approach to understanding 'big data' in mental health and use of images by young people. The aim of this project is to bring together researchers from across disciplines of arts, data science and medicine with young people to explore the issues around the generation and use of 'big data', including health records [30, 31].

Genomic data integration The genomic data integration study Jedi was funded by the Medical Research Council to explore ethical, legal and social issues in the use of genomic data linked to phenotypic health data for research. The study design was multi-faceted and included: a review of legislative and regulatory requirements; example studies using genomic and phenotypic data in combination; data governance arrangements in published research; case studies of organisations working with genomic and phenotypic data; and public views and expectations. The engagement activities took the form of eight public workshops across a range of demographics and focused particularly on the acceptability of differing models of data use [32]. These viewpoints along with the other findings of the study were used to develop a set of recommendations towards a framework for the use of genomic and phenotypic data for research, particularly in data safe havens [33].

Discussion

What this study adds

This paper has provided a summary of our PI/E policy which is based on published standards. It has set out our rationale, scope and practical approaches to PI/E for SAIL and some of our related Population Data Science initiatives. It has summarised the work of the Consumer Panel, including showing the array of topics discussed over a three-year period with examples of researcher engagement and feedback. It has illustrated some of our wider engagement activities and

how citizen input has been taken forward. Via the survey, it has provided insights into people's experiences of being part of the Consumer Panel, and perceptions of the strengths and weaknesses in the work of SAIL, plus opportunities and threats. Although there will be jurisdictional, cultural and organisational differences, we believe that the material covered in this paper will be of interest to other data focused enterprises across the world.

Lessons learned

From our experience with the Consumer Panel over the years and from the recent survey, we've noted some members of the public may feel apprehensive joining an established group focused on what can be a highly technical field of work. People need to feel welcomed, to be encouraged to ask questions and be provided with digestible information. There is a need in all PI/E activities to manage group dynamics and work hard to ensure everyone has the chance to provide their views. It is important that people are given enough time to assimilate information and consider their response, particularly on topics where presenters are experts and might not appreciate the steepness of the learning curve for people new to the topic. We have been encouraged by the keenness of citizens to engage and make a contribution to research and developments.

We have found that PI/E activities can be challenging and rewarding in various measures; this is how it should be. If it was not challenging, there would be limited opportunities to learn; if it was not rewarding then it would be harder to see the benefit. PI/E may reveal novel ideas, new perspectives on existing ideas, fundamental questions on issues taken for granted, views on related societal events and their impact on individuals, amongst many anticipated and unanticipated opinions. Particularly for enterprises such as SAIL and related Population Data Science initiatives that don't naturally meet face-to-face with research participants, having a variety of PI/E activities provides the opportunity for citizen input during the course of studies and developments.

In terms of modes of engagement, we have found that can be a challenge to find the optimum balance between informing citizens and properly eliciting their views, and along the continuum of providing a blank sheet to a *fait accompli*. Activities usually need to begin with some information but without leading or coercing the viewpoints. We believe there is a journey towards awareness and understanding and it is important to work with people along the way, neither expecting an expert level of knowledge nor assuming a lack of ability to grasp the issues under discussion. It is important to afford all individuals due respect, and also to encourage everyone to think not only about issues that pertain to themselves, but to include considerations for wider society as part of social responsibility. We seek to take a pragmatic approach, prioritising the communication modes that allow maximum public input commensurate with the purpose of the activity.

Limitations

We acknowledge that there are limitations in our PI/E activities. Due to time pressure to produce the PI/E policy, we were unable to work with the Consumer Panel. However, since the policy necessarily had to align with national standard and the SAIL operating model, the scope for input would have

been limited. There will be greater scope in the development of the strategy to enact the policy. The Consumer Panel is not representative of Welsh population but is skewed towards over 50s age groups. We have not yet had opportunity to consider and act on the outcomes of the Panel survey, but we have provided a summary in this paper to indicate their perceptions and possible learning for others. However, since PI/E is an ongoing area of work, we will always have some work in progress, rather than all completed. In our wider activities we acknowledge that, despite our best efforts, we might not achieve full representation of all sectors of the public, or to be certain the views we elicit are a true, stable reflection of public opinion. This is not specific to our work, but is a challenge common to achieving meaningful PI/E since the results will be dependent on complex factors such as personal worldviews, cultural values and recent social or political events and their relative prominence.

Conclusion

We have set out our policy and multifaceted approach to PI/E, with examples as illustrations. SAIL has a longstanding Consumer Panel to provide advice on the work of SAIL and related Population Data Science initiatives. In addition, we have a programme of PI/E with the wider population of Wales, and with specific sectors and interest groups as befits the topic under study. Although there will be jurisdictional, cultural and organizational differences, we believe that the material covered in this paper will be of interest to other data focused enterprises across the world.

Acknowledgements

We express our grateful thanks to the members of the Consumer Panel and all citizens who take part in our PI/E activities. Funding for the SAIL Databank as an infrastructure support service is provided by HCRW (grant references are not used by this funder), the ESRC (grant ref: ES/S007393/1) and the MRC plus consortium of HDRUK funders (grant ref: NIWA1).

Ethics statement

Ethical approval was not required for the work of this paper since it is based on our approach to PI/E. The Consumer Panel members were not engaged as research participants, but as advisors to our work.

Conflict of interest

The authors declare they have no conflicts of interest.

References

1. Jones KH, Ford DV, Thompson S and Lyons RA (2019) A Profile of the SAIL Databank on the UK Secure Research Platform. IJPDS, 2:4, doi: <https://doi.org/10.23889/ijpds.v4i2.1134>

2. McGrail KM, Jones KH, Akbari A *et al* (2018) A Position Statement on Population Data Science: The Science of Data about People, IJPDS, 3:4 doi: <https://doi.org/10.23889/ijpds.v3i1.415>
3. Swansea University (2020) Population Data Science <https://popdatasci.swan.ac.uk/>
4. National Institute for Health Research School for Primary Care Research (2020) <https://www.spcr.nihr.ac.uk/PPI/what-is-patient-and-public-involvement-and-engagement>
5. INVOLVE <https://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/>
6. Arnstein, SR (1969) A Ladder Of Citizen Participation, Journal of the American Planning Association, 35: 4, 216-224 doi: <http://dx.doi.org/10.1080/01944366908977225>
7. UK Research and Innovation (2015) Concordat for engaging the public with research <https://www.ukri.org/files/legacy/scisoc/concordatforengagingthepublicwithresearch-pdf/>
8. Health and Care Research Wales (2020) <https://www.healthandcareresearch.gov.wales/public/>
9. EU General Data Protection Regulation (2016) <https://gdpr-info.eu/>
10. HM Government (2018) UK Data Protection Act <http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted>
11. Priest S, Goodwin J and Dahlstrom MF (Editors) (2018) Ethics and Practice in Science Communication, University of Chicago Press
12. Understanding Patient Data (2020) <https://understandinpatientdata.org.uk/>
13. Aitken M, Tully, MP, Porteous C, *et al* (2019) Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research. IJPDS, 4:1:06 doi: <https://doi.org/10.23889/ijpds.v4i1.586>
14. UK Public Involvement Standards Development Partnership (2019) UK Standards for Public Involvement <https://drive.google.com/file/d/1U-IJNJCfFepaAOruEhzz1TdLvAcHTt2Q/view>
15. SAIL Databank (2020) Public engagement <https://saildatabank.com/about-us/public-engagement/>
16. Jones KH, McNerney CL and Ford DV (2014) Involving consumers in the work of a data linkage research unit. International Journal of Consumer Studies, January 2014, 38:1:45–51, doi: <https://doi.org/10.1111/ijcs.12062>
17. SAIL Databank (2020) Information Governance Review Panel <https://saildatabank.com/saildata/data-privacy-security/#secure-access>
18. Jones KH, Daniels H, Heys S and Ford DV (2018) Challenges and Potential Opportunities of Mobile Phone Call Detail Records in Health Research: Review. JMIR Mhealth Uhealth 2018;6(7):e161. DOI: [10.2196/mhealth.9974](https://doi.org/10.2196/mhealth.9974)
19. Jones KH, Daniels H, Heys SM and Ford DV (2019) Public views on using mobile phone Call Detail Records in health research, JMIR Mhealth Uhealth 7(1):e11730 DOI: [10.2196/11730](https://doi.org/10.2196/11730)
20. Jones KH, Daniels H, Heys SM and Ford DV (2019) Towards an ethically-founded framework for the use of mobile phone CDRs in health research, JMIR Mhealth Uhealth 7(3):e11969 DOI: [10.2196/11969](https://doi.org/10.2196/11969)
21. Song J, Fry R, Mizen A, *et al* (2018) Association between blue and green space availability with mental health and wellbeing, IJPDS 3:4, Conference Proceedings for International Population Data Linkage Conference 2018.
22. Mizen A, Fry R, Wheeler B and Rodgers SE (2019) Co-producing a typology for Green and Blue spaces for a longitudinal, national dataset of Green and Blue spaces. IJPDS, 4:3, Conference Proceedings for International Conference on Administrative Data Research 2019
23. Nuffield Family Justice Observatory for England & Wales (2019) Improving intelligence for the family justice system – A Nuffield Family Justice Observatory Data Partnership <https://www.nuffieldfjo.org.uk/app/nuffield/files-module/local/documents/NFJO%20DP%20Summary%20Document.pdf>
24. Johnson RD, Ford DV, Broadhurst K, *et al* (2020) Data Resource: population level family justice administrative data with opportunities for data linkage, IJPDS, [in press]
25. Bedston S, Pearson R, Jay MA, *et al* (2020) Data Resource: Children and Family Court Advisory and Support Service (Cafcass) public family law administrative records in England, IJPDS, 5:1. DOI: <https://doi.org/10.23889/ijpds.v5i1.1159>
26. Ford DV, Jones KH, Middleton RM *et al* (2012) The feasibility of collecting information from people with Multiple Sclerosis for the UK MS Register via a web portal: characterising a cohort of people with MS. BMC Medical Informatics and Decision Making, 2012,12(1):73 doi: [10.1186/1472-6947-12-73](https://doi.org/10.1186/1472-6947-12-73)
27. Middleton RM, Allen-Philbey K, Schmierer K, *et al* (2019) A model for participant engagement in research: the MS register JNNP 90: 12:175 <http://dx.doi.org/10.1136/jnnp-2019-ABN-2.155>
28. Allen-Philbey K, Middleton RM, Tuite-Dalton K, *et al* (2020) Can we improve the monitoring of people with multiple sclerosis using simple tools, data sharing and patient engagement? doi: [10.3389/fneur.2020.00464](https://doi.org/10.3389/fneur.2020.00464)
29. Adolescent Mental Health Data Platform (2020) <https://www.adolescentmentalhealth.uk/>
30. Marchant A, Turner S, Balbuena L, *et al* (2019) Self-harm presentation across healthcare settings by sex in young people: an e-cohort study using routinely collected

linked healthcare data in Wales, UK. Archives of disease in childhood. Archdischild-2019-317248. <http://dx.doi.org/10.1136/archdischild-2019-317248>

31. Ingham K, John A (2017) Virtual Embodiments <http://www.kareningham.org.uk/virtual-embodiments/>.
32. Jones KH, Daniels H, Squires EL and Ford DV (2019) Public views on models for accessing genomic and health

data for research: a mixed methods study. J Med Internet Res 21:8. <http://dx.doi.org/10.2196/14384>

33. Jones KH, Daniels H, Heys SM, *et al* (2020) Towards a risk-utility data governance framework for research using genomic and phenotypic data in safe havens. J Med Internet Res, 22(5):e16346, DOI: <http://dx.doi.org/10.2196/16346>

