Research article

Tightrope walking towards maximising secondary uses of digitised health data: A qualitative study

Ann R R Robertson
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, UK

Pam Smith
Nursing Studies, School of Health in Social Science, The University of Edinburgh, UK

Harpreet Sood
National Health Service (NHS) England, 80 London Road, London, UK

Kathrin Cresswell
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, UK

Ulugbek Nurmatov
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, UK

Aziz Sheikh
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, UK

Author address for correspondence
Ann R R Robertson
Centre of Medical Informatics, Usher Institute of Population Health Sciences and Informatics, Teviot Place, The University of Edinburgh Edinburgh, UK
Email: A.R.R.Robertson@ed.ac.uk

ABSTRACT

Background Timely progress with attaining benefits from Health Information Technology (HIT) investments requires UK policymakers and others to negotiate challenges in developing structures and processes to catalyse the trustworthy secondary uses of HIT-derived data.

Aims We aimed to uncover expert insights into perceived barriers and facilitators for maximising safe and secure secondary uses of HIT-derived data in the UK.

Methods We purposively selected individuals from a range of disciplines in the UK and abroad to participate in a thematically analysed, semi-structured interview study.

Results We identified a main theme of 'tightrope walking' from our interviews (*n* = 23), reflecting trying to balance different stakeholders' views and priorities, with sub-themes of 'a culture of caution', 'fuzzy boundaries' and 'cultivating the ground'. The public interest concept was fundamental to interviewees' support for secondary uses of HIT-derived data. Small scale and prior collaborative relationships facilitated progress. Involving commercial companies, improving data quality, achieving proportionate governance and capacity building remained challenges.

Conclusions One challenge will be scaling up data linkage successes more evident internationally with regional population datasets. Within the UK, devolved nations have the advantage that 'small scale' encompasses national datasets. Proportionate governance principles developed in Scotland could be more widely applicable, while lessons on public engagement might be learned from Western Australia. A UK policy focus now should be on expediting large-scale demonstrator projects and effectively communicating their findings and impact. Progress could be jeopardised if national data protection laws were superseded by any Europen Union-wide regulation governing personal data.

BACKGROUND

There is very considerable international interest in maximising the opportunities linked to exploiting 'big data' in healthcare and in many other contexts. Substantial funding has enabled the setting up of the UK-wide Farr Institute to support innovations 'leading to advances in preventative medicine, improvements in healthcare delivery, and better development of commercial drugs and diagnostics'.[1] The Institute aims to provide the infrastructures for collaborative working, to build health informatics capacity in the UK and to deliver patient and societal benefits from research using electronic health records, routinely collected data and population-based health datasets. Similarly, substantial investment is supporting a nationwide Administrative Data Research Network (ADRN) designed to facilitate researcher access to other datasets that are routinely collected by government departments.[2] These developments are in response to the escalating role of digital data in society and a wealth of health-related data becoming available following national and international implementations of Health Information Technology (HIT). Timely progress with realising potential benefits for patients, public health, society and the economy from the investments in HIT systems relies on policymakers and others negotiating challenges for developing secondary uses of HIT-derived data.

This interview study was the qualitative component of a larger, mixed-methods investigation into maximising the safe and secure exploitation of data held in HIT systems.[3] The interviews were designed to explore a diverse range of experts' views on the current state of the rapidly developing field of digital data and the perceived barriers to and facilitators for realising medium-to-long-term benefits from secondary uses of data recorded in HIT systems. We define secondary uses here as the use and reuse of health-related data for purposes other than for the direct, clinical care of individual patients. The current scope of secondary uses includes for conducting epidemiological and pharmacovigilance research studies, for facilitating recruitment to randomised controlled trials, for carrying out audits and benchmarking studies and for financial and services planning by healthcare organisations. The aim of the qualitative study was to provide expert, 'insider' insights into the current state and potential future of secondary uses to inform policymakers, managers and others with an interest in seeking returns on investments in UK HIT systems.

METHODS

Ethical permission

We obtained ethical approval for the interview study from The University of Edinburgh's Centre for Population Health Sciences Ethics Committee, and each interviewee provided informed consent prior to taking part.

Sampling

We planned to interview approximately 20–25 individuals with diverse expertise and involvement in health-related digital data. Potential informants were selected based on their current activities related to secondary uses of data held in HIT systems in the UK, with additional, international participants being invited from regions with a world-wide reputation for work involving HIT-derived data. The sampling frame for recruiting interviewees was constructed to access a broad spectrum of expert opinions from a range of stakeholders in secondary uses (Box 1). Potential participants were identified from a systematic scoping review of the literature,[4] from our research team's contacts and through 'snowballing', whereby one interviewee suggests a colleague. The purposively sampled[5] interviewees included policymakers, health professionals, data scientists, social scientists, academics, researchers and representatives of the pharmaceutical industry, the legal profession and the third sector.

Box 1 Interview sampling frame

- Policymakers
- NHS National Services Information Services Division/ Electronic Data Research and Innovation Service (eDRIS) staff
- Clinicians (secondary and primary care)
- Academic researchers
- Data scientists
- Social scientists
- Legal professionals
- Ethics experts
- Farr Institute/Scottish Collaboration for Public Health Research and Policy
- Commercial sector (pharmaceutical industry)
- Third/voluntary sector
- International experts in secondary uses of HIT-derived data

---

Data collection

We conducted semi-structured, in-depth, one-to-one interviews face-to-face or by telephone. We explored interviewees' experiences and views of the current state of using digitised health-related data, how they thought secondary uses of HIT-derived data would develop in the future and the principal facilitators and barriers they perceived to achieving this. We used a brief topic guide (Appendix 1) as a foundation for the interviews, and the researcher adapted and developed questions in response to the individual participant's role and interests. Interviews were digitally audio recorded and professionally transcribed verbatim before being cleansed of personal identifying information; the researcher made contemporaneous notes of two interviews where the interviewees requested no audio recording. Interviews lasted between 30 and 60 minutes, mostly lasting approximately 45 minutes.

Data analysis

Qualitative data collection and analysis were iterative. This allowed the interviewer to explore emerging themes further in subsequent interviews and to seek alternative viewpoints from interviewees from different disciplines. The qualitative data were analysed thematically[6] supported by the qualitative

software package NVivo 10. Our thematic analysis was informed by a range of theoretical approaches employed in previous work[7–9] and we discussed initial and refined coding categories and themes at regular team meetings throughout the data collection and analysis process. Towards the end of data collection, we convened two workshops for team members specifically to discuss the interview transcripts and analysis combining inductive and deductive approaches.

RESULTS

We approached 28 potential interviewees (declined, $n = 1$; no response or subsequently could not be contacted, $n = 4$), leading to 23 interviews with participants throughout the UK and in Australia ($n = 1$), Canada ($n = 2$) and the USA ($n = 1$). One interviewee (Scottish Government) subsequently withdrew consent, reporting new workplace regulations against giving interviews, and that audio file and transcript were subsequently deleted from our dataset.

The main theme generated by the qualitative analysis was 'tightrope walking'. This overarching theme of negotiating pathways through multiple, sometimes conflicting, considerations included three sub-themes of 'a culture of caution', 'fuzzy boundaries' and 'cultivating the ground'. These are shown in Table 1 with their associated main coding categories identified from the interview data. Selected key findings are summarised below, supported by illustrative quotes.

Tightrope walking

Metaphors such as 'balancing act', 'juggling' and 'tightrope walking' were frequently used in many of our interviewees' accounts of working towards maximising the safe and secure use and reuse of clinical data. These metaphors referred in particular to four main areas where participants spoke of challenges they felt required continual, careful negotiation:

i.   balancing protections for individual patient privacy and using available health-related data for the 'public good' and 'in the public interest';
ii. fostering public trust in expanding the secondary uses of HIT-derived data, and engaging patients and the public with the processes;
iii. achieving proportionate governance in secondary uses for dataset linkage research for trustworthy but also faster studies;
iv. efforts to balance perceived costs and rewards among different groups, for example between hospital staff involved in collecting data for clinical care and academics using these data for research purposes.

Support for secondary uses among our sampled interviewees would be expected given each of our participants was actively involved in secondary uses of HIT-derived data in some capacity. Advocacy for pursuing benefits from secondary uses was, however, consistently related to the concept of the public interest and, in the UK context of publicly funded National Health Services (NHS), to ideas of a social contract and reciprocation.

'Part of my view would be to say if you're an NHS patient there are rights and there are responsibilities and part of your responsibility of being treated within a state healthcare system is for your information to be used for the benefit of you and other members of society'.
(16: medical professional)

'The NHS is still a socialised system here [in Scotland], you know, it's retained that much more so than in

Table 1 The main theme and sub-themes generated by the interview data and their main coding categories

Main theme *Tightrope walking*	Sub-theme *A culture of caution*	Sub-theme *'Fuzzy' boundaries*	Sub-theme *Cultivating the ground*
Coding categories	Coding categories	Coding categories	Coding categories
Public interest	Collaboration	Terminology	Innovations in secondary uses
Legal complexities	Incremental steps	Types of organisation	Data quality
Proportionate governance	Data hugging	Conceptual	Different models
Pharma	Ethics		Workforce capacity
Who pays/gains			Precision medicine
			Relationships
			Funding
			Public engagement
			Translational capacity

England, and I think that makes a difference in terms of making an argument for health research for patient benefit when the overall boundary around that is a socialised healthcare system'. (15: social scientist)

'There is a level of concern if commercial entities are involved, we know that, but it's much more subtle than that. It's not just public – good, and private – bad'. (03: legal academic)

Interviewees acknowledged likely benefits, both economic- and research-related, from attracting, for example pharmaceutical companies to the UK – providing appropriate safeguards were in place and public and health professionals' trust was not jeopardised. A media outcry over the selling of datasets[10,11] in England (care.data) had reflected genuine, widespread distaste about commercial exploitation of data, according to participants. They believed such events had negatively influenced public perceptions more widely, even although the systems put in place in the devolved NHS Scotland and elsewhere were quite different from those that were operating in England.

'In Scotland we don't have one single large database which is where the Care database in England got itself in trouble. Here all the datasets are all left with the original suppliers of the data, so no-one can get into some big database and get access to all the data'. (05: Farr Institute)

While Scotland had not followed the open access initiative that had been instigated in England, some interviewees still recognised that there could nonetheless be advantages to big data open access:

'People can do analyses, they can hold public sector to account if they can get access to the data. These all seem like good things. Breaches of individual privacy aren't a good thing so, again, we're balancing advantages and disadvantages from making data available'. (05: Farr Institute)

An NHS National Services Scotland (NSS) interviewee suggested that for the most part, he believed people were very happy with the idea of NHS staff conducting research on NHS-derived data, and slightly less so if it was academics carrying out research on those data unless the academics were closely aligned to the NHS. The involvement of commercial bodies and particularly of the pharmaceutical industry, he said, 'made people more nervy'. Personally he had no bias against working collaboratively with pharmaceutical companies because he believed they did positive work and needed to do research in order to do more of it.

'We always ensure that we have control over publication, so we maintain a certain level of independence from the pharmaceutical industry … And that so far has been fine, but perhaps that's mostly because this work has been relatively low scale, small numbers'. (06: epidemiologist)

Another interviewee likened managing the sometimes conflicting priorities encountered in a step-by-step approach to maximising benefits from HIT-derived data to 'walking a tightrope', whereby he believed protecting the NHS should always remain the first priority.

'It (the data) is collected for patient care, firstly, and then for running the health service, so we need to be able to plan how many hospitals, how many doctors, we need to be able to look at quality of care, we need to be able to do those things. Anything which leads to the public withdrawing confidence has the potential to bring down the health service'. (05: Farr Institute)

An exemplar in governance for data linkage studies, highlighted by study participants both from the UK and internationally during interviews, was the principles-based proportionate data governance framework developed collaboratively in Scotland.[12] This framework had four elements: an account of the principles and instances of best practice, information on who was a data controller and in what circumstances, a model of proportionate governance and a training element. Combining safeguards with the flexibility of a principles-based approach was considered a model for others in the UK and abroad who also hoped better to balance researchers' needs for reasonably fast approvals to access data and the recognised need to protect privacy, confidentiality and data security. The framework had not been designed to be specific to Scotland or to be specific to health-related research.

'We deliberately designed it in a generic fashion so it could be picked up by anybody in any sectors actually considering what needs to be taken into account in data linkage'. (03: legal academic)

The idea of 'proportionate' governance in this framework comprised consideration of data anonymisation, consent for using the data and if a proposed data linkage would be in the public interest. These three considerations were underpinned by the governance framework's triad of 'safe people' (accredited researchers), 'safe environments' (for example accessing research data only from a "safe haven") and 'safe data', which covered technological capabilities such as for anonymising, and zipping and unzipping research datasets.

'This is about robust research use. That implies that there are appropriate ethical checks and balances, that there is suitable anonymisation, where that's possible and practical for the research… There's consent, there's anonymisation and a third avenue which is authorisation – the idea that you can actually have, for example, ethics bodies that can authorise the linkage of identifiable data in the public interest so long as certain types of criteria are met'. (03: legal academic)

The primary legislation controlling how personal data in the UK could be used currently was the amended 1998 Data Protection Act (DPA).[13] The legal and regulatory context in which personal data were either shared or linked was described as one of the 'labyrinthine' complexities. The DPA

itself was reported to be hard to understand and often mis-understood. There were different legal systems within the UK and further legal complexities should secondary uses also pertain to continental Europe or elsewhere internationally. As a European Directive, DPA legislation had been implemented differently across European Union (EU) member states. Now, however, it was proposed to replace the various national data protection laws with a single, uniform EU Regulation.[14] Uniformity was intended to introduce standardised, personal data protection legislation across the EU and facilitate data movement as well as save costs to businesses, but some interviewees feared that the advent of a rigid regulation could instead be 'potentially a huge threat' to health-related secondary uses research. Part of that concern lays in fears that any recognition of health research being conducted in the public interest could be overwhelmed by lobbyists for privacy protections with respect to the commercial exploitation of personal data. An interviewee described research funders' concerns as getting:

> '…drowned out by a lot of extreme views about what should happen in terms of consumer privacy, rather than just patient privacy'. (03: legal academic)

We now in turn consider each of the three sub-themes subsumed under the 'tightrope walking' meta-theme.

A culture of caution

Misconceptions about the contents of the current DPA controlling personal data, according to UK interviewees, was a contributing factor to inconsistent attitudes among data controllers to sharing different types of personal data and to the phenomenon of "data hugging" – an overly cautious approach to data sharing – by some. It was believed that professional attitudes could also be a problem, and some perceived general practitioners (GPs) in particular to be unnecessarily cautious about making primary care data available for secondary uses research. Better information and education could challenge data hugging.

> 'One of the ways to address the culture of caution is to raise awareness among the custodians, and also researchers who want access to data, that fundamentally the law is there as much to protect privacy as it is to facilitate the responsible sharing of data'. (03: legal academic)

More positively, a culture of caution was deemed fruitful in the context of multidisciplinary collaborators taking incremental steps to lay solid foundations for expanding secondary uses of HIT-derived data. In this context, a past record of personal relationships with multidisciplinary colleagues and scale were both identified as important facilitators. The smaller size of a Canadian province, an Australian state or a large healthcare organisation in the more fragmented and commercialised health systems of the USA meant that individuals working in various disciplines relevant to secondary uses were all likely to know of one another and could 'pick up the phone' to each other. The small UK

nation of Scotland had the advantage that 'small-scale' encompassed national datasets. International interviewees described Scotland's position in regard to developing secondary uses as enviable.

Fuzzy boundaries

An array of secondary uses-related fuzzy boundaries was identified by interview participants from lack of clarity or consensus over terminology, for instance whether secondary uses was an appropriate term and how it should best be defined, to widespread conceptual fuzziness about data 'ownership', being a data controller, and the differences among sharing data, linking identifiable data and linking data using anonymised, aggregated datasets. Interviewees also spoke of the fuzzy boundaries of hybrid organisations where it was no longer necessarily always clear if an organisation could clearly be classed as belonging to the public, private or voluntary sector.

> 'Let's take the NHS. It has research functions within it and a lot of their researchers conduct it in conjunction with universities, and there are joint posts, so you can't put a boundary around that system very easily. And outside, that boundary is even more fluid because the health system itself is really increasingly a combination of public and private'. (15: social scientist)

Fuzzy boundaries could also be construed as advantageous in so far that they offered possibilities for opening up debate and discussions between diverse stakeholder groups. For example even when research participants had given consent at the start of a study, unforeseen ethical dilemmas could arise as further information came to light over time, such as in genetics research:

> 'The plan is to actually look at some real cases where this does occur and actually see what the individual concerned would like to have done, because the real problem is the balance between alarming people unnecessarily and not rescuing people from situations they need to be rescued from. Unfortunately the reality is, even when people in studies have explicitly said they do not want to have any data back from the study, a lot of them still believe that they would be contacted if a life-threatening thing was found and that's simply not true'. (18: academic/commercial)

Cultivating the ground

Interviewees in our study highlighted developments in the secondary uses of HIT-derived data, particularly the growth of dataset linkage studies and the introduction of additional health-related datasets, such as genomics and other biotechnology data, primary care data, medical images and laboratory results. Large datasets linked to individual patients would advance developments in precision (also known as personalised) medicine, in which healthcare is individually customised. An example of progress with such work came from North America, where a national network of genetics research had been funded by the National Human Genome

Research Institute to support genomic medicine.[15] In addition to the potential to deliver faster, cheaper research and to enhance medical knowledge and drug safety, interviewees suggested a more holistic approach to health and health care would develop through a growing numbers of studies that linked health-related datasets with each other and with datasets from other sectors. Combining HIT-derived data with education, housing and justice datasets, for example, would increasingly generate evidence to support public health initiatives and evidence-based policy making beyond specifically health policy, as well as to potentially support more robust policy evaluations.

Relationships between different sectors still had to be fully worked out. UK and international interviewees reported this particularly to be the case with respect to establishing mutually satisfactory working relationships with pharmaceutical companies, such that for-profit, private businesses saw benefits from participating in secondary uses research while simultaneously clearly safeguarding the public interest principle and retaining public support. Most interviewees acknowledged that collaborative working with commercial companies, including the pharmaceutical industry, would be an important component of achieving any aspiration to create economic wealth through secondary uses of digital data. Job creation could be a measurable, medium-to-longer term return on investment in HIT systems. Value might also accrue, however, which would be harder to define, such as some future overall improvement in public health as a consequence of evidence-based policies that had reduced environmental damage. That would be according to an interviewee:

'… a very different vision of wealth creation'. *(15: social scientist)*

Natural language processing, according to some interviewees from the UK, should soon allow new datasets derived from unstructured information in electronic health records to become available for research. A further, potential resource for secondary uses research was patient reported measures.

'It's another layer that will come … and could be in many ways invaluable and wonderful extra data. It would just require another level of thinking about'. *(06: epidemiologist)*

Patient-entered data in records would have to be flagged in order for researchers to understand if and how to incorporate those data in research. More generally, data quality in clinical records, and especially the quality of data following dataset linkages, was raised as still being a challenge to conducting robust secondary uses research, by both UK and international interviewees.

'Data quality is certainly an issue. The primary goal, I think, has probably got to be interoperability, and I would say patient access to those integrated records as well… In order to have interoperability you've got to have a standard and the standards are a very technical thing. Getting everyone to agree that that's enforced and

getting the people who are funding the development to understand that it's important, all of these things are all sort of steps along the way'. (22: GP)

'Data quality is a huge issue with clinical datasets. … So a measure of weight for instance will depend in part on what the measurement of weight was but also on somebody correctly entering that into the database and there will be typographical errors in that without question. If you're entering enough data there will be data entry errors, inevitably'. *(06: epidemiologist)*

In very large epidemiological studies, which were now possible using population datasets, the influence of some data errors should be attenuated by the number of data items. Another issue was how well researchers understood the variables in the datasets with which they were working. An interviewee referred to the publication initially of misleading findings from a study comparing lengths of hospital stays in two Canadian provinces; in this instance, that research team had been unaware that in one of the province's databases, acute patients were recorded as being discharged after a given period of time and then recorded as readmitted as a different type of patient.

'The provinces are very independent and so making sure you're making real comparisons, rather than apples and oranges comparisons and assuming that they're the same thing, is challenging and there are real examples of disasters which have been made across provinces'. (08, Manitoba Centre for Health Policy)

Most participants raised the need for stable and adequate funding to continue supporting organisations that were advancing the secondary uses of HIT-derived data and help address outstanding data quality issues.

'I just think it should be viewed as unacceptable to spend many millions on running a cohort and then not be willing to spend £1,00,000 to ensure that if you now do a co-analysis with another cohort that the variables that you're going to use actually do mean the same thing in the two cohorts'. (18: academic/commercial)

'We are looking at incorporating genomic data and kind of treating this as a mechanism for reuse of omic-related data. We don't have funding yet and we're working on that'. (09: PopData)

In Canada, healthcare is organised by province and British Columba holds longitudinal data for the whole of the province's population at a research support facility known as PopData.[16] A repository of Manitoba's range of research datasets is housed in the Manitoba Centre for Health Policy.[17] These are different models from England where the Health and Social Care Information Centre collects data nationally.[18] In Scotland, where datasets are not held centrally, researchers seeking to work with NHS data apply through the electronic Data Research and Innovation Service;[19] if a proposal is approved, researchers then access the appropriate, indexed and linked data for their study from a secure

'safe haven'. In the UK, Wales has also developed a remote access system for researchers called the Safe Anonymous Information Linkage.[20,21]

Interviewees from abroad and from the UK identified the need for funding for staff development and capacity building in order to grow the available workforce with the range of required skills for further secondary uses development, such as in building HIT infrastructures, mathematics, designing new research methodologies and analysing and interpreting the findings from large dataset studies. Interviewees advocated increased training opportunities in health informatics and data science across a range of existing academic disciplines to foster multidisciplinary collaborations, with stable funding to support new PhD studentships for years to come. In the UK context, the recent substantial investments in the ADRN and the Farr Institute,[1,2] with just such organisational aims as these, were seen by interviewees as a vitally important step to support continued progress.

Insufficient funding was, however, suggested by one UK interviewee as an explanation for the UK's failure to date to implement a national public education and public engagement campaign for secondary uses work with HIT-derived data. Others highlighted a reported example of highly successful public engagement from Western Australia (WA), where WA Data Linkage System[22] research projects used state-wide population data and where there was no patient consent to opt in or out. Every WA Data Linkage System project had a consumer advisory group attached to it, and every study had to produce a lay summary of the research carried out, its findings and its impact in clear, everyday language. Over the years, the state's community had become the main 'champions' for data linkage studies using unconsented data, according to a staff interviewee.

> '20 years ago they would have opposed it because they would have been worried about privacy. But we've got such a fantastic track record of doing wonderful research which has benefited the community and we get out there and tell everyone about it, so I think that's an important aspect of it… What our consumers said was that we know you've got data on us, if you don't use it to improve the health system and you don't use it to avoid harm we will sue'. (14: WA Data Linkage System)

An important factor in accruing positive support for data linkage research in WA was translational capacity, which reflected the close working relationships between WA Data Linkage System staff and local policymakers, as well as a history of significant public involvement at all levels and stages of the research process.

> 'The translation starts with involving the policy people and even the practitioners as well as the consumers actually at the beginning of the project'. (14: WA Data Linkage System)

DISCUSSION

While our qualitative study was mainly concerned with secondary uses of HIT-derived data in the UK context, aiming to generate potentially helpful findings to UK policymakers in particular, we included an international dimension. We conducted a total of 23 interviews to sample a diverse range of expert views on current and future challenges for maximising the safe and secure exploitation of HIT-derived data. The interview data analysis generated a main theme of 'tightrope walking', with three sub-themes of 'a culture of caution', 'fuzzy boundaries' and 'cultivating the ground'. Tightrope walking was one of a number of metaphors interviewees used for the perceived, multiple balancing acts recognised as underpinning the advances to date in exploiting secondary uses – particularly evident in the small, devolved UK nation of Scotland – and still seen as necessary for making further progress. Reported instances of 'tightrope walking' included finding the balance between carrying out potential dataset linkage research that would be in the public interest and safeguarding the privacy of individuals and data security, and attracting private, commercial interests while maintaining proportionate data governance, transparency and public support. Small-scale environments and prior, multidisciplinary, collaborative relationships could facilitate this balancing of competing views and priorities.

There is a marked variation within the UK due the contrasting population sizes of its constituent nations, its devolved national health services and the quite different approaches taken in the past towards exploiting HIT-derived data, notably between the large nation of England, where it was decided to hold population datasets centrally, and elsewhere in the UK. Nevertheless, the recently established Farr Institute, with a remit that includes delivering patient and societal benefits from HIT-derived datasets, is a UK-wide organisation.[1] The advent of the Farr Institute and its sister organisation, ADRN,[2] were welcomed as important facilitators for the UK to derive benefits from secondary uses of digital data. Hoped-for benefits in the medium to longer term included increased translational research as well as the wider societal benefits of job and wealth creation.

The main short-term scope for research innovation was universally held to lie in greater exploitation of dataset linkage studies. The potential now for the Farr Institute and the ADRN, and other related UK-wide endeavours such as the Alan Turing Institute,[23] to work closely together is an important step towards increasing cross-sectoral dataset linkage studies in the UK, with the promise of building an evidence base to underpin policy decisions and policy evaluations beyond specifically healthcare policy.

Our study identified several, common 'barriers' or challenges, including improving data controllers' understanding of the current laws governing personal data protection and how these should be applied to sharing HIT-derived data; increasing understanding of and public support for research exploiting digital data and dataset linkage studies; secondary uses workforce development and capacity building; and improving the quality of HIT-derived data, including the quality of linked data.

While funding, the Farr Institute and the ADRN, and a flexible, incremental approach to problem solving, drawing on multidisciplinary collaborative relationships, were seen as general but important facilitators for realising benefits from HIT systems, two specific examples of addressing challenges – proportionate governance and public engagement – also emerged from our study, which could serve as lessons for others elsewhere.

Much collaborative work had gone into building processes and infrastructure by the Scottish Informatics Programme[24] before the 2013 creation of its successor, the UK-wide Farr Institute. This earlier work had included developing an innovative, principles-based, proportionate approach to data linkage governance, which interviewees in the UK and abroad believed could be widely applicable.[25,26] The flexibility of a principles-based approach to proportionate governance, rather than a prescriptive one, could allow it to be used by health and non-health sectors in various jurisdictions. Its wider use outside Scotland could, for example, be helpful in addressing the challenge of widely acceptable governance where pharmaceutical companies sought to conduct research on HIT-derived data.

WA gave an example of successful public engagement.[22] There attitudes to using HIT-derived data without consent to opt in (or out) had reportedly shifted from negative to positive over an estimated period of some 20 years. The main factors believed to be responsible for the shift were having patient and public involvement at all levels and every stage of all data linkage projects, combined with government policymakers setting much of the research agenda. There was then wide dissemination of the research findings, explicitly linked to any associated policy and service changes. Working very closely with policymakers and the public could improve the visibility of data linkage study findings and their impact in a virtuous circle of translational research and public engagement. To follow this example in the UK, the main focus next for policymakers might be on facilitating increasing the number – and public knowledge – of completed data linkage studies and maximising awareness of the benefits these had brought.

Strengths and limitations

A strength of this study is the geographical spread and diversity of interviewees from different disciplines and geographical areas that we sampled in order to address a focused research question. Nonetheless, a clear overarching theme emerged from our interview data. Purposive sampling is prone to researcher bias but is a viable sampling technique for expert elicitation, which was the purpose of this work. The interviewee sampling frame was constructed by an experienced research team, not by an individual, and while we would not make claims for data saturation, the team ended further interviews once we had elicited useful 'insider' insights

into challenges and facilitators, as we had sought to do. The combination of inductive and deductive analysis approaches has we believe helped to identify important, potentially transferable lessons beyond the UK. It is noted that this study did not include interviews with non-experts in this field, that is with patients, carers and members of the general public, whose perspectives may be quite different from the experts' perspectives reported here.

CONCLUSIONS

Substantial investments made in HIT are producing a wealth of digital data with the potential to benefit our knowledge of diseases, drug safety, service delivery and public health. One still unresolved challenge for better exploiting secondary uses of HIT-derived data will be scaling up data linkage research successes more evident internationally at the smaller scale of province or state. Scotland has an advantage in that 'small scale' encompasses national datasets. Proportionate governance principles developed in Scotland are likely to be applicable elsewhere, while lessons on public engagement might be learned from Western Australia. A UK policy priority now might be to increase the numbers and visibility of completed outputs from data linkage studies. To date, there has already been significant, if uneven, progress with developing secondary uses in the UK, with that progress attributable in part to flexibility and an incremental, collaborative approach to working in a field characterised by 'fuzzy' boundaries. The success of a flexible approach could be jeopardised by rigidity if national personal data protection laws were to be superseded by a single, uniform EU-wide regulation governing personal data.

REFERENCES

1. The Farr Institute. Available at: http://www.farrinstitute.org/1_about.html. Accessed 20 October 2015.
2. The Administrative Data Research Network (ADRN). Available at: http://adrn.ac.uk. Accessed 20 October 2015.
3. Sheikh A, Smith P, Nurmatov U, Robertson A and Cresswell K. Developing a UK framework for facilitating health policy deliberations on maximising secondary uses of healthcare data: a mixed-methods study. Available at: http://www.cso.scot.nhs.uk/outputs/focus-on-research-summaries/focus-on-research-generic-health-relevance. Accessed 20 October 2015.
4. Robertson A, Nurmatov U, Sood H, Cresswell K, Smith P and Sheikh A. A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data. *Submitted to Journal of Innovation in Health Informatics.*
5. Patton, M. Q. Qualitative Research and Evaluation Methods, 3rd edition. Thousand Oaks, CA: Sage, 2002.
6. Braun V and Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;3(2):77–101. http://dx.doi.org/10.1191/1478088706qp063oa.
7. Cresswell K, Worth A and Sheikh A. Implementing and adopting electronic health record systems: how Actor-Network Theory can support evaluation. *Clinical Governance: An International Journal* 2011;16:320–336. http://dx.doi.org/10.1108/14777271111175369.
8. Laurie G and Sethi N. Information governance of use of health-related data in medical research in Scotland: current practices and future scenarios. Available at: http://www.law.ed.ac.uk/file_download/publications/2_188_informationgovernanceofuseofhealthrelate.pdf. Accessed 20 October 2015.
9. Nutley S, Walter I and Davies H. From knowing to doing: a framework for understanding the evidence-into-practice agenda. *Evaluation* 2003;9:125–148. http://dx.doi.org/10.1177/1356389003009002002.
10. Davies M. Private companies set for access to patient data for just £1. *Pulse*. Available at: http://www.pulsetoday.co.uk/your-practice/practice-topics/it/private-companies-set-for-access-to-patient-data-for-just-1/20003879.fullarticle. Accessed 10 November 2015).
11. Caldicott F. Care.data hangs on engagement. *digitahealth.net*. Available at: http://www.digitalhealth.net/analytics/46830/caldicott:-care.data-hangs-on-engagement. Accessed 10 November 2015.
12. Scottish Government. *Guiding Principles for Data Linkage* 2012. Available at http://www.gov.scot/Resource/0040/00407739.pdf. Accessed 20 October 2015.
13. HM Government. Data Protection Act 1998. Available at: http://www.legislation.gov.uk/ukpga/1998/29/contents. Accessed 20 October 2015.
14. European Commission. Reform of the data protection legal framework in the EU. Available at: http://ec.europa.eu/justice/data-protection/reform/index_en.htm. Accessed 20 October 2015.
15. Electronic Medical Records and Genomics (eMERGE). Available at: https://emerge.mc.vanderbilt.edu/organization. Accessed 20 October 2015.
16. About Population Data BC (PopData). Available at: https://www.popdata.bc.ca/aboutus. Accessed 20 October 2015.
17. Manitoba Centre for Health Policy. Available at: http://umanitoba.ca/faculties/health_sciences/medicine/units/community_health_sciences/departmental_units/mchp. Accessed 20 October 2015.
18. Health and Social Care Information Centre. Available at: http://www.hscic.gov.uk. Accessed 20 October 2015.
19. Electronic Data Research and Innovation Service (eDRIS). Available at: http://www.isdscotland.org/Products-and-Services/eDRIS. Accessed 20 October 2015.
20. Safe Anonymous Information Linkage (SAIL). Available at: http://www.saildatabank.com. Accessed 20 October 2015.
21. Jones K, Ford D and Jones C. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *Journal of Biomedical Informatics* 2014; 50(100): 196–204.
22. Data Linkage Western Australia. Available at: http://www.data-linkage-wa.org.au. Accessed 20 October 2015.
23. The Alan Turing Institute for data science. Available at: https://turing.ac.uk. Accessed 12 November 2015.
24. Scottish Informatics Programme. Available at: http://www.scot-ship.ac.uk. Accessed 20 October 2015.
25. Laurie G and Sethi N. Towards principles-ased approaches to governance of health-related research using personal data. *European Journal of Risk Regulation* 2013 Jan 1;4(1):43–57.
26. Sethi N and Laurie G. Delivering proportionate governance in the era of eHealth: making linkage and privacy work together. *Medical Law International* 2013 Jun;13(2–3):168–204.

APPENDIX 1

Brief topic guide for interviews

- Interviewee's background, current position and specific role/interest in digitised health and related data
- Views on current state and future potential for secondary uses, with a focus on innovations in secondary uses of health-related data
- Views on main barriers to and facilitators for advancing secondary uses/maximising returns on investment in HIT systems