Welsh Parliament
Equality and Social Justice Committee

Annex 1: Summary of key themes and issues emerging from data justice evidence received by Equality and Social Justice Committee

1. Setting the context: understanding data justice

There is a need to have a clear and shared understanding and definition of data justice. A scoping report from the Committee's expert advisers provided background including several key definitions and themes:

> **Health Data Justice has been defined** as ”an orientation to the study and use of health-related data in ways that aim to redress the exclusions of structurally marginalized communities from systems of healthcare and public health, the oppressions faced by communities when participating in such systems, and the institutions responsible for governing participation.[...]”
>
> **The datafication of healthcare affects how research is delivered and care carried out**. It may change how patients access services, what treatment is on offer, and to whom. The use of algorithms to predict and calculate interactions with these services means that any errors or biases in the systems may cause significant physical and mental harm. Research has shown that whilst such systems may affect us all, 'they don't impact us all equally'. **At the heart of a health data justice approach is a focus on inequality and the impact on marginalized communities.[...]**

> …data justice debates have highlighted the extent to which the introduction of **data-driven innovation**, particularly in public services, **may shift priorities in public policy** based on computational translations of social issues that have consequences for social justice.[…]

The scoping report also highlighted the potential for technology to exacerbate data inequalities, noting that **"the starting point of any algorithm is the selection of the problem it is designed to solve"** and that:

> The decision to address a healthcare issue using a data-driven technology may become an issue of social justice when it focuses on the needs of an advantaged group at the risk of exacerbating socioeconomic, racial or gender injustices.[…]The decision to pursue the use of data-driven technologies in healthcare may result in resources being directed to specific patient groups when there is a focus on a particular aspect of the technology.

2. Public understanding of health care data

NHS Wales produces, relies on, retains and reports on significant amounts of patient data. Through the **Welsh Clinical Portal,** clinicians have access to millions of patients' digital records, which could be shared with thousands of NHS users. Much of the data is not 'visible'. The Committee heard that there can also be mixed messages about data: a strong focus on data security, data loss, risk and concerns about who is using the data should be balanced with potentially more positive messages about how data can be used for research and improving patient care. The expert adviser paper noted the Caldicott Committee in stating that:

> The preservation of privacy also relies on restricting access to data to minimize the risk of breaches. However, the importance of sharing data for the benefit of patient safety and clinical care is a well-established principle in healthcare.

The Committee heard concerns about the extent of public awareness and the need to build better public knowledge on healthcare data and its usage. Chris Carrigan of useMYdata stated that:

> There's a general lack of understanding among people about how their healthcare data is used—it's difficult for a patient to understand what is in their healthcare data because, by and large, they can't see it all—and a lack of understanding among the public about the real benefits that usage of healthcare data for the NHS, for planning and research, can bring. What the public tend to hear are largely negative messages about data loss, data hacking and risk, and therefore that engenders the perception of the public that data is something that maybe can't be

> trusted, when actually we want to move the agenda forward to something where it says, 'Data can be trusted. Use it in the right way and we can really use it to save lives'.(para.8)`

He also outlined how useMY data's involvement in workshops with patients, the public and clinical staff to examine questions around data use.

Professor Ann John from Swansea University noted the key role of public awareness of the value of healthcare data:

> […]we saw really clearly during COVID that trust is one of the fundamental issues upon which people make these decisions. Often, it's the most vulnerable, marginalised and underserved—you know, the very people we want to do this work for and with—who will be the people who opt out, which basically means they stop having a voice. (para.19)

Likewise, the expert adviser scoping paper noted the need to be clear about the use of data:

> As a taxpayer-funded health service the public feel strongly that decisions about the use of NHS data take into account the long-term potential benefits for future generations over short-term financial gain.

3. The reasons for data collection

Alongside enhanced patient and public understanding, building and ensuring trust is key in developing and using data. The evidence the Committee heard suggested that whilst there is a good degree of trust in the NHS and research institutions, patients are more hesitant about sharing data with private sector organisations.

Work undertaken by Understanding Patient Data has echoed this, as did Chris Carrigan (useMYdata) in evidence to the Committee:

> […]the further a person gets away from their treating clinician, the less trust they have in who's using their data. So, high trust in GPs and then hospitals, and then academia, universities, and then out to commercial companies. (para.28)

Similarly, Professor Ann John set out that:

> People trust the NHS, they trust the people involved in their direct care, they trust academics. There can be a bit more mistrust with industry. But I think it goes back to that active involvement, transparency and engagement at all levels, and communication with people. (para.63)

Rhidian Hurle, Executive Medical Director from Digital Health and Care Wales (DHCW) stressed that:

> It's the patient's choice to divulge their clinical history, their medical history, their personal history, and it's entirely their choice. The relationship between a healthcare professional and an individual is based on trust. (para.219)

This sense of patient ownership and understanding of the reasons for data collection were also touched on in evidence from Professor Ronan Lyons from the SAIL Databank at Swansea University. Professor Lyons talked about the tangible benefits of data collection relating to the COVID-19 vaccine and in smaller clinical trials:

> […]our ability to bring data together from many different sources, that allowed us to understand who was getting infected and why they might be getting infected. It allowed us to look at the context of infections, whether that was in the general community, in healthcare settings, in care homes, and in school populations as well. And then, it also allowed us to look at the effectiveness of interventions. The most obvious one of those would be in the vaccine roll-out: how well did the vaccines work at protecting people[…] (para.250)

> And often, clinical trials are run in groups of people who actually have just got one condition, or even a mild version of it, because it's easier to show whether the treatment works or not. But then, those drugs and vaccines are used in the entire population, and you also need to look at what they call the real-world evidence data, as to how well they work, or not. (para.250)

In their written evidence **BMA Cymru Wales** stated that:

> […]sharing patient level data is going to be vital in ensuring a data led planning process for NHS and Social Care in the future. Developments such as the National Data Resource (NDR) and the SAIL databank are welcome in acting as a repository to enable efficient and accurate planning and research to take place. This can only function when confidential patient data is shared and linked to other data sets in Primary Care, Secondary care, and Social Care.

> This requires data governance and safety to be engineered into the process from the outset and an honest debate with the public held so they are aware their medical records will be used for this purpose with safeguards and individual confidentiality maintained.

4. The quality and completeness of health data

The Committee heard concerns about this, including the potential under-reporting of data amongst minority ethnic and socially disadvantaged groups, as well as people who have limited contact with health services. This could lead to their needs not being reflected in the planning, development and delivery of services. In addition, there were questions about whether all available data was of sufficient quality to be used effectively, and significant amounts of research being based around either very large or very small population bases.

The expert advisers set out that:

> The exclusion of women from some areas of medical research and differences in the health care of female patients is well documented and shows up in the data.[…] If this type of exclusion is carried over into algorithm design, as this example shows, it can cause severe harm to women.

And that:

> The historic lack of ethnicity recording on death certificates failed to give the authorities a clear picture of the uneven impact COVID had on minority ethnic populations and those working in face-to-face occupations in the early stages of the pandemic. This delayed the implementation of protective measures which may have saved lives.

Professor Ann John noted the need to build trust around data collection “at a population level, but also very specific work with underserved populations” (para.19):

> What we've found with certain populations, and we've seen it a lot during COVID, is that the poor quality recording of ethnicity data was a real issue. (para.22)

Professor Ronan Lyons from SAIL noted that:

> I know that collection of information on ethnicity has been mandatory for many years in the NHS right across the UK, but it's still not brilliant, so it is something that does need to be improved.[…]I wouldn't be confident that the NHS measurement of ethnicity is that good. (para.307)

Professor John noted that “it's really important to understand the resources it takes to reach underserved groups, but also work out the ways to make sure they have a voice in the data” and the need for consistent funding to undertake that work and “keep people engaged” (para.81).

Dr Robert French from useMYdata noted that for researchers running and using large data sets “if you're looking for average effects, then it probably doesn't matter if some people are missing from the data”, which does however emphasise the importance of looking more actively for those smaller groups (para.108). The expert advisers echoed this:

> The drive to centralize the collection of health data in repositories such as the National Data Resource and SAIL databank in Wales reflects the need for large quantities of data to safely train and test algorithms. Even if efforts are made to ensure data is representative it can result in biased outcomes for a patient if the design process is not carefully monitored.

Professor Lyons noted that a lots of research was completely reliant on “how good our healthcare data is” (para.250) but struck a slightly more positive note to the Committee:

> One of the things that we use as a population denominator for our work is, essentially, the NHS registration system, which everybody in Wales is associated with, and then we've also looked at service use across many, many different factors, and it's surprising how small a group is missing in total. But it does not, by any means, show the whole picture, and so, particularly, you will find that young men tend to be missing from lots of things. (para.299)
>
> And so we have a piece of collaborative work now with Public Health Wales that looks at vaccine, the ethics of vaccination and the uptake by different groups, and the census data allows us to demonstrate that particular groups actually are at higher risk of some diseases, and also uptake in some of those groups isn't nearly as good as possible. (para.300)

Ifor Evans, Executive Director of Strategy at DHCW argued that:

> I can't think that there is any reason to think that there are groups that have been excluded from that data that is used for strategic planning within the health boards in terms of the form and shape of the services that are available. I think that we do have some examples of where we've used the information in the data very effectively in order to target services at groups in need or who are vulnerable. (para.183)

Rhidian Hurle from DHCW highlighted the importance of ensuring a joined-up patient record:

> In terms of joining up the care, the delivery of a health and care record should be available wherever the patient presents. It's not a context-specific thing. It's not a GP surgery; it's not secondary care, tertiary or some community care. It should be available. (para.151)

BMA Cymru Wales, whilst noting that “the GP record is widely acknowledged to be the most consistent and life-long patient record in the health sector”, also stressed that efficient and accurate planning needed links across primary, secondary and social care.

5. The drivers for data collection

NHS Wales produces, relies on, retains and reports on significant amounts of patient data, which is driven directly as result of patient contact with health services.

However, the Committee heard evidence that the key drivers for some data collection – especially linked to research and funding – could very largely be availability of resource or funding linked to specific projects or programmes. This had the potential to skew priorities, but also shift the focus away from under-represented groups and their specific needs were more under-represented.

Professor Ronan Lyons from SAIL acknowledged the partnership working with commercial companies and the reliance of SAIL on funding from research grants provided by organisations outside the NHS and university sector:

> Essentially, the research that's conducted with us is carried out by academics and, also, Government social researchers, as part of an Economic and Social Research Council grant. We don't allow companies to have direct access to data, but what happens is that companies often partner with a university and fund the research through a university—that's how that works. But I think this is an area that's going to grow considerably over the years.(para.272)
>
> We look at the competitions that are out there for funding,[...]and apply in a very competitive sense to answer those questions. We have been very successful in years, and we've brought many millions of pounds of money into Wales and employed quite a number of people in doing that. So, a lot of what we do, basically, is determined by the questions that the research funders put up, because we need to have people to be able to answer the questions. (para.291)

Professor Ann John noted that “there's evidence that shows that, where research is very active, and data research as well, that improves the care for people” (para.87). Rhidian Hurle also argued for the benefits of a robust shared medicines record, stating “the ability to provide one source of the truth, which is updated, will bring massive benefits in terms of patient safety” (para.204).

BMA Cymru Wales noted however that:

> SAIL has demonstrated itself to be a trusted research environment with effective safeguards and processes allowing safe access to data, leading to high levels of GP practice engagement. However, GPs have seen little tangible use of SAIL to benefit their populations directly given that level of engagement to date. This should be addressed by Welsh Government (as the funding body) to mandate usage for NHS planning purposes where necessary.

6. Data complexity

Data is often found in a wide range of systems and formats, which can make its better integration and use more difficult. The NHS in Wales has a mix of paper and digital records, and these are spread across a range of systems and organisations including hospitals, community care settings, and independent primary care contractors. This poses challenges for seeing the whole patient journey and the whole population.

Ifan Evans from DHCW highlighted the problem:

> [...] there are years before we will have a comprehensive, fully accurate and all the history and timely digital health records. The NHS is data-rich and information-poor from a digital perspective. There is a lot that is still on paper, and there's a lot that is on paper before it reaches the digital record, because digital information has to be inputted by someone somewhere. (para.133)

Rhidian Hurle from DHCW added that the patient record “should be available wherever patients present, regardless of the healthcare context” (para.191) but also that:

> [...]there's a complexity of digital systems across the healthcare context, and there will be data held in those systems that, at the present time, isn't available within the national record. It is our ambition to line up the systems so that we collate that data as we go along. (para.177)

Chris Carrigan of useMYdata noted these issues stating “the more systems, the more technology and computer systems are involved in the process, the more difficult it is to bring that data together” (para.39) but also highlighted that the ability for something like e-prescribing to span primary care, secondary care and community has been beneficial to patients (para.95). BMA Cymru Wales expressed caution about the roll-out of enhanced patient access to records, noting that “much of the data which would be available to patients is extremely complex”.

The relationship with social care data adds in another layer of complexity. Progress on the Welsh Community Care Information System – meant to link up health and social care data in the community - has been patchy and Audit Wales has identified areas of concern.

7. Data security

Safeguards, security, and effective information governance remain critical in the collection and management of data, and patients can be understandably concerned about who is accessing, using and sharing data. Appropriate systems and frameworks need to be in place. Particular sensitivities were noted around areas such as mental health services. At the same time, the Committee heard reassurances around how information security is managed both in terms of NHS and research data.

The expert advisers highlighted the information governance and security issues facing healthcare:

> The health sector is a highly regulated space focussed on the protection from harm of both patients and those that treat them. With new techniques and products being developed at speed their introduction into healthcare throws up new ethical and legal challenges which those responsible for information governance and legal responses struggle to keep up with.

They also noted twin dilemmas associated with ”the delicate balance between the protection of citizens’ rights and the desire to support innovators who may approach the use of health data with different aims and sensibilities” and the desire to build trust in data sharing which has led to “a focus on privacy and security measures”.

The NHS Confederation in a blog post from June 2022 on harnessing data to improve patient outcomes said:

> Data governance teams must record a rationale for their decision at the same time they grant a user access to data. Likewise, data owners must record a rationale when they approve the use of a data set for a purpose.

Another NHS Confederation blog from September 2022 argues for improved data in community health services but which understands the whole patient journey in real time.

Effective data governance is also a key issue for health data in Wales. Written evidence submitted by DHCW to a joint Health and Social Care/Public Accounts and Public Administration scrutiny session noted:

> There are however, varying programme governance arrangements for national digital programmes, [...]Consistency and simplification of programme delivery would help to ensure clear accountability.

Chris Carrigan (useMYdata) acknowledged the need for data safeguards and the central role of the patient in choosing what's in their health data and who can use it but posed the query:

> [...] where does the data sit, who controls it, who determines what research is important or what gets done and what are the safeguards? And that's all about transparency and including patients and the public in that conversation. So, I think e-prescribing is an interesting example, where it does span multiple traditionally separate sectors of primary care, secondary care and community for real benefits to patients, and actually there can be real benefits to drive research as well. (para.95)

Professor Ann John highlighted the concerns relating to security around potentially more sensitive patient information:

> There is a sense that people feel that mental health data is more sensitive. Some of that is to do with stigma; some of that is to do with the fact that when you're taking a mental health history, you're asking much more personal details than sometimes you would ask in a physical health history. (para.63)

Mr Hurle (DHCW) stressed that "the information governance programme across Wales is strong", "you're actively being watched when you've been given permission to use the record" (para.225) and there are things that can be done to ensure only the people who need to see the record are actually able to see the record. However, Darren Lloyd from DHCW highlighted some of the challenges in ensuring data security across care settings:

> we (*DHCW*) ensure that the access controls within those systems are adequate to the needs of those health professionals and those individuals accessing the data, so that it is appropriate at the time it is required. And that requires quite a number of different factors and a certain element of intelligence within those systems and those services to allow that to happen, but that means then that we have a fully comprehensive record that is available at the point of care under different circumstances, because those controls are within those systems and services. (para.195)

BMA Cymru Wales noted that "under GDPR regulations, GP partners (who hold GMS contracts) are considered as the data controllers for any data that they hold about their patients". It is critical that GPs understand the need for robust security of patient data and ensure that such security is in place.

However, this does also need to be set out in the context of moves towards increased patient access to and control over their own medical records and data. The evidence also touched on the ability of patients to opt out of their confidential patient information being used for research and planning. NHS England has a National data opt-out policy in place, albeit after a previously troubled policy had to be abandoned. In response to a question on whether and why he thought patients in Wales should be able to opt out of certain types of data sharing, Mr Hurle replied:

> 'Yes' is the short answer. The question is, in which context. For example, in the context of, say, a research trial, an informed consent of a patient who has capacity provides the option for that individual to remove their consent at any point of the journey. [...]So, there are mechanisms you can actively use as an individual. Once information is collated and pseudonymised, that is your identity has disappeared from the data, then by the fact that your identity has disappeared from the data, it's difficult to remove you from the data. (para.170)

BMA Cymru Wales set out that they are “wholly aligned with the theoretical concept of safe and legal access for patients to their own medical information”. Nonetheless, they also noted that “plans for full record access would probably have to be on an ‘OPT-IN’ and informed consent-based model, itself probably highly labour-intensive”, and called for an effective governance framework, a staged approach to patient data access, and appropriate redaction of any third-party data within records.

In relation to data security in research, Professor Lyons set out that:

> The provision of data to SAIL is a voluntary activity; it's not required by anybody. We have demonstrated, over the years, that the system that we've designed is robust and safe. We have to have accreditations from multiple organisations. (para.278)

According to Health Data Research UK the SAIL Databank is a flagship for the robust secure storage and use of anonymised person-based data for research to improve health, well-being and services.

8. Developments in data collection and use

The Committee heard of the potential of advances in digital and data technology such as the NHS Wales App, which could make for improved patient access to data around services and their own health record. At the same time, there was caution about the need to manage expectations about the pace of change. Note was also made of the impact of the pandemic on data collection, which had enabled some improvements in the available data on groups who may previously have been under-represented in data.

Mr Rhidian Hurle from DHCW set out that:

> I've seen a progressive increase in the volume and availability of patient data for the delivery of care to the point that now, I can see every single citizen's GP record, with their consent, and I can see every single blood test everywhere in Wales. I can see every single x-ray done everywhere in Wales from a laptop, either remotely or within the hospital. That has delivered massive benefits in terms of reducing duplication, starting conversations with patients on what matters to them rather than trying to validate their history, accurate medicines reconciliation, reducing harm by being able to see things. (para.117)

Ifan Evans of DHCW also highlighted the way in which the COVID-19 pandemic had driven change and the development of new tools and approaches:

> Contact tracing was a new thing and it was a very simple standard configuration across Wales, delivered from the cloud, with a joint data controller agreement behind it. It did not require anybody to let go of anything, not really, although local authorities did have their own contact tracing arrangements. And I think that helped us to deliver, compared to other parts of the UK, a very effective vaccine delivery programme and contact tracing programme. (para.238)

At the same time he acknowledged and reiterated several times the significant difficulties that digital change poses for DHCW, noting that “Digital is never going to be ‘done’” (para.237) and that:

> […]the biggest challenge is that driving significant change in a complex digital system is very difficult to do, and there is no shortage of digital transformation programmes that have taken a long time to introduce, and moving people onto those new systems and training them is a challenge. (para.205)

BMA Cymru Wales expressed some caution around the NHS Wales App, setting out that:

> We are reassured that the NHS Wales app […] will only proceed with summary record access once a GP practice has enabled the functionality. We strongly advise that only summary and coded data including vaccinations, would be in scope for the NHS Wales app until all these issues are addressed.

As data-driven technologies become an increasing part of care services, new types of data – and data harm – may emerge. Chris Carrigan (useMYdata) cautioned about what he saw as a potentially exponential growth in health data:

> […]we're now on the cusp where, if we look at our children and younger people in this country, they're generating data about themselves, whether it's on fitbits or apps, or whatever. They've got a digital footprint that's going to be there that we've never had. Now, the data is going to get huge, massive, massive—beyond all recognition. (para.113)

Similarly, there will be real challenges in securing better data integration between health and social care.

However, Chris Carrigan noted what he saw as the more limited priority being afforded to improving patient access to their record:

> What I don't yet see in the equation is where an individual fits within that. So, all the money is going into research, technology, big research environments. Very little is actually going into being able to facilitate me to see what is in my full health record. (para.41)

Likewise, Chris Carrigan (para.59) and Professor Ann John talked about the need to address digital exclusion; not just access to devices, but access to w-fi and thereby the broader range of digital services and programmes. (para.63)

9. Patient engagement

There was consistent acknowledgement in the evidence that patient engagement and the patient voice was still playing a limited role in the development and management of data and data systems, with the focus historically often being based primarily around service providers research funders and clinical staff. It was acknowledged that this still represented a significant challenge in developing more patient-focused and patient-led systems and data

Work by the Kings Fund has stressed the critical importance of engaging with patients and the public in building trust on data; ensuring accessibility, transparency and co-development in data use, and providing patient control.

The scoping report by the expert advisers noted that:

> Data justice work has especially drawn attention to the need to centre impacted communities in decisions about the generation, collection and use of data.[...] In some instances, this has informed approaches to design and an engagement with the conditions within which data infrastructures emerge, calling for more participatory design practices that emphasise the involvement of communities.
>
> The NHS has a long history of public engagement and consultation over the provision of health services. This culture has been extended to the field of data-driven innovation with respect to involving patients and citizens in decision-making around the use of health data beyond their direct care. Research has shown that people want a say in how their data is used because it is generated by their interactions with the health service.

Professor Ann John stated that “we need to involve the public in the meaningful design and use of data. So, they need to be involved in all of the bodies that are approving data for research and use” (para.19). Chris Carrigan set out that one of the key aims of useMYdata was to secure uncreased public and patient involvement in streamlining access to data and ensuring greater data transparency.

Mr Hurle from DHCW noted that their approach on data was “to build trust with our key partners, particularly patients” (para.146):

> The app, though, is our main engagement tool for patients. Historically, for the NHS Wales Informatics Service and DHCW in its first few years, our user base is clinicians and health organisations and people who deliver health services, not, historically, direct services to patients. But that is changing and, as we're moving into those areas, we're engaging more with patients. (para.166)

However, Ifan Evans noted that their engagement on the National Data Resource - a new national data platform that brings together data about health and social care services from across Wales – has been users of that data who are clinicians and developers, not the public. (para.164)

He also noted that:

> […]so we, as Digital Health and Care Wales, have not engaged very much directly with patients historically, because all of our services are directed at clinicians and they generally, historically, have been used in secondary care.(para.222)

Turning to SAIL, Professor Lyons described their engagement with patients and the public; a consumer panel, lay members on the independent panel considering project applications, public involvement in research teams and studies, use of expert panels involving lay members looking at which research algorithms are used. (paras.253,257)

> We have as many public engagement activities as we can. I think we've two full-time members of staff and this is what they do altogether. It's difficult to do, I think, on an abstract basis. So, it tends to be done more about a particular question. So, let's say, if we were doing some work on a disease like multiple sclerosis, we would always engage with the Multiple Sclerosis Society UK, and use their patient advocacy groups to get involved with, rather than try to recruit somebody just from the general public with that particular disease at that point in time.(para.287)

Arguing for a stronger patient voice in data development, Professor Ann John provided a direct example of where patient interest, effective care and efficient data management interlink:

> I do lots of self-harm and suicide prevention work, and that concept of people in distress having to tell the same story over and over again is something where these sorts of united data systems can help us improve the quality of care in relation to that telling of stories over and over again, but also the sharing of information from one area to another. (para.52)