About the author

Dr Ian Thomas is a Research Associate based at the Administrative Data Research Centre Wales/Wales Institute of Social & Economic Research, Data & Methods, Cardiff University
# Contents

Executive summary.................................................................................................................................................................................................4

Introduction..................................................................................................................................................................................................................................................8

Part one: Project overview......................................................................................................................................................................................................................10

Part two: Statutory homeless data in Wales........................................................................................................................................................................14

Part three: The wider 'data landscape' in Wales.......................................................................................................................................................................19

Part four: Perceptions of a new system.............................................................................................................................................................................27

Part five: Learning from others......................................................................................................................................................................................33

Part six: Designs for a new homelessness data system in Wales......................................................................................................................................52

Part seven: Way forward.................................................................................................................................................................................................................74

References..................................................................................................................................................................................................................................................75

Appendix A..............................................................................................................................................................................................................................................77

Appendix B..............................................................................................................................................................................................................................................144
Executive summary

In January 2018 a meeting was held between UK Collaborative Centre for Housing Evidence (CaCHE) Wales Hub Co-Investigators and Welsh Government officials from Knowledge and Analytical Services in order to explore opportunities for collaboration. It was agreed that a current priority in Welsh Government is the need to improve the homelessness data infrastructure.

For several years there have been discussions about making a radical shift towards individual level data recording and reporting in relation to statutory homelessness in Wales. Moreover, there have been more recent calls to explore data linkage with homelessness data, following the successful linkage of housing/homelessness data with health data in Wales (McGinn, L. et al. 2016) and Scotland (Waugh et al. 2018). However, progress on a new data agenda has been limited by a lack of resource to undertake a feasibility assessment. This project therefore had the initial purpose of contributing to the feasibility assessment of introducing a new individual level data collection, reporting, and analysis process in relation to statutory homelessness in Wales. The specific aims of the project were to:

- Assess current statutory homelessness data collection practices across local authorities
- Engage stakeholders to explore the sector’s opinions on a move to a new data collection
- Examine alternative approaches to collecting homelessness data

A range of approaches were adopted in order to achieve these aims, including: a survey with local authority Housing Options teams; a consultation with the sector on their use of published statutory data and perceptions of a new system; a stakeholder workshop; desk based review of systems; an evidence review; informal stakeholder conversations; and other public engagement such as blogging in order to promote the project activities and prompt feedback.

The outcome of the project has been a set of four options for a new homelessness data system in Wales, summaries of which are provided overleaf. However, over the course of the project it became clear that a new data system provided the opportunity to completely re-align how all homelessness data in Wales are collected and used, not just ‘statutory’ homeless data; the four options therefore reflect this expanded scope.
## Summary of options

<table>
<thead>
<tr>
<th>Option</th>
<th>Strategic fit</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Timeframe</th>
<th>Relative cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformed reporting</td>
<td>Individual level statutory data will enable pathways through the legislation to be explored, specifically repeat homelessness. Data held in a data linkage infrastructure will facilitate linkage to other public service data and Housing Support Grant recipient data. Own data can also be brought into the linkage infrastructure, for example a cohort of people taking part in an intervention, in order to engage in quasi-experiments using matched data-leading to increased evaluations by services.</td>
<td>+ Can potentially retain legacy systems in local authorities. + In the larger transformed landscape, individual technologies can be adapted (swapped, replaced, discontinued) as needs change.</td>
<td>Combined data source is not accessible for operational/daily needs. Time delay between upload and outputs/access. Large number of data flows; however, automation where possible should reduce burden.</td>
<td>1-2 years – based on concurrent development of individual level statutory system and Housing Support Grant data management solution.</td>
<td>Low – adopts proven technologies and attempts to work within current data collection and reporting practices.</td>
</tr>
<tr>
<td>Federated data: A 'data broker' acts as a co-ordinator for the system and is connected to data providing agencies making up the federation, i.e. local authorities. Requests for data are made to the broker who then 'pulls' data from agencies, after they have agreed to the data being shared. Data are de-identified prior to leaving a data provider’s system, and data from different providers can be linked as the same method of de-identification is applied.</td>
<td>Enables more responsive measurement of homelessness through the ability to query 'live' data held by organisations. Can contribute to improved evaluation by running queries of project participants across a homelessness system. + Data providers have control over who has access to their data. + Can be scaled up as the number of organisations increases. + Only de-identified data move across the federation; all personal data remains with the data provider.</td>
<td>Time lag between user making request for data and results. Data quality dependent on organisational practices. Resource required at providers to provide a data contact to process/agree data requests. Dependent on local system being capable of supporting interfaces. Data are snapshots of local systems at a point in time.</td>
<td>&gt;5 years – large timescales required in order to tender for a ‘data broker’ and have associated technologies/systems in place.</td>
<td>High – requires relatively novel approach (to homelessness sector in UK) with technological investment.</td>
<td></td>
</tr>
<tr>
<td>Option</td>
<td>Strategic fit</td>
<td>Strengths</td>
<td>Limitations</td>
<td>Timeframe</td>
<td>Relative cost</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Homelessness data warehouse</td>
<td>Provides a centralised data collection that can be used to improve measurement and research/evaluation. Use of interfaces to the warehouse could enable more service-oriented live querying to be undertaken, thereby contributing to person-centred data integration.</td>
<td>Data quality is improved through standardised data structure. Facilitates data mining and exploration. Topic-specific data can be brought together from different data sets in the warehouse to form ‘Data Marts’ which enable faster querying of data, i.e. data on rough sleepers from across different service provision.</td>
<td>Not accessible for operational/daily needs; however, interfaces for the warehouse could enable individual level querying. Data will only be as recent as the last update. Requires substantial planning, budget and technological investment.</td>
<td>2-3 years (in order to procure data warehouse and set up governance around data flows. The legal agreements to share personal data with the warehouse would require health timeframes.</td>
<td>Medium – draws on proven technologies, however new data infrastructure required through the warehouse and associated technologies to make local data systems interoperable with it</td>
</tr>
<tr>
<td>Integrated data management platform:</td>
<td>Could potentially facilitate system approach to homelessness service design through the sharing of data. Data can also be used for research and measurement, however data are for operational use rather than statistical purposes, and so quality of data may be reduced—or require continuous work to maintain.</td>
<td>Provides real-time data for use by practitioners. Potential economies of scale by combining resources.</td>
<td>Highly dependent on the building and maintenance of consensus. Difficult to sustain as the numbers of participants in the system grow; new entrants are forced to adhere to system design they did not have a say in. Legacy data collections must either be linked to the new system, or dropped in favour of the new system, in order to avoid doubling workloads. Can become outdated by the time the system is functional (if a pan-Wales approach adopted).</td>
<td>&gt;5 years (for pan-Wales approach) – long timeframe in order to tender for data management solution/provider and to develop the solution to meet sector needs</td>
<td>High – draws on proven technologies, however totally new data infrastructure required and assignment of a sector coordinator</td>
</tr>
</tbody>
</table>
Introduction

Homelessness remains an important social and political priority in Wales, as evidenced by the latest Welsh Government (2019) strategy on homelessness, which includes its vision for a Wales where homelessness is rare, brief, and un-repeated. Underpinning this vision are several policy principles (Figure 1), including:

- Earlier prevention activities, as part of a much broader shift in focus from crisis alleviation to preventing homelessness ideally before households become ‘threatened’ in the legal sense
- Person centred services that work in a ‘trauma informed’ way and are co-produced with individuals with lived experience of homelessness
- Collaborative working across service and policy areas to change public services’ response to homelessness as an ‘issues’ for and of housing

As part of the Strategy, Welsh Government acknowledge the use of data within a ‘whole system’ approach to addressing homelessness in Wales, where services are coordinated in order to provide quick, appropriate, and long-term accommodation, i.e. addressing the aspiration of homelessness in Wales that is brief and un-repeated. The strategy identifies two possible uses of data; firstly, the need to measure the issues and the response, and secondly, to evaluate services and approaches to homelessness to continually learn and improve. In the Minister for Housing and Local Government’s plenary address introducing the new strategic policy statement on homelessness1, it was further acknowledged that there were currently data quality concerns, alongside gaps in knowledge around homelessness. As part of the Strategy, a new homelessness action group was convened, with one of their activities to be to look at gaps in data.

With the Strategy and its related activities, there is a renewed political will in Wales to improve data on homelessness, and one that acknowledges that better data can potentially help measure the extent and profile of people experiencing homelessness, whilst also generating a culture of service development and commissioning which learns and develops using data. The whole systems approach being advocated in the strategy, alongside a greater emphasis on early prevention, also calls for an increased attention to homelessness ‘beyond statutory provision’, as the strategy itself indicates, ‘The duties in Part 2 of the Housing (Wales) Act 2014 should be the last line of defence—not the first’. Now is therefore an opportune time to make changes to the homelessness data infrastructure in Wales to meet the needs of stakeholders, and contribute to ending homelessness.

Figure 1: The vision for homelessness in Wales (rare, brief, un-repeated), underpinned by associated activities/priorities

<table>
<thead>
<tr>
<th>RARE</th>
<th>BRIEF</th>
<th>UN-REPEATED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced risk of becoming homeless</td>
<td>Periods of housing instability are as short as possible</td>
<td>Housing solutions are long term</td>
</tr>
<tr>
<td>Prevention</td>
<td>Effective and efficient services</td>
<td></td>
</tr>
<tr>
<td>Cross-sector collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person Centered</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Plenary address given on the 8th October 2019 [https://gov.wales/oral-statement-homelessness](https://gov.wales/oral-statement-homelessness)
One of the main sources of information relating to statutory homelessness in Wales comes from aggregate data on people assisted by local authorities under the Housing (Wales) Act. Supplementing this data is the annual rough sleeper count, which provides estimates of those people bedded down on the streets and therefore at the more extreme end of homelessness. Though statutory homeless and rough sleeper counts provide some insight into homelessness in Wales, the nature of the data being collected, i.e. aggregate figures, and the content of what is being collected, i.e. the information they present, are increasingly out of step with the needs of policy, practice, and research. Furthermore, there is an underutilisation of other routinely collected data in Wales, such as that collected under specific housing grants, but also from public services more generally, all of which could be used to greater effect in ending homelessness.

For several years there have been discussions in Wales about making a radical shift towards national individual level data reporting in relation to statutory homelessness in Wales. This shift would bring Wales in line with both Scotland and England; Scotland have collected individual level data since 2001, whilst England moved to an individual level collection in April 2018. This report provides an overview of a project aimed at exploring the feasibility of introducing a new homelessness data system in Wales.

**Report structure**

Part one of this report outlines the original objectives underpinning the project and activities that were undertaken, including a survey of local authority statutory ‘Housing Options’ teams, a workshop, consultation with the sector, and a review of international examples of data systems. Part two and three sketch out the homelessness data landscape in Wales, critically reviewing the current approach to statutory homelessness data, before going on to consider other data sources on homelessness in Wales. This decision to think about the wider homelessness data landscape was informed by several factors, addressed in Part 3, not least of which was the need to think about the bigger picture of homelessness in Wales as part of a much larger push to increase early prevention rather than crisis prevention.

Part four provides insight into the sector’s views of a new data system, primarily drawing on findings from the survey of Housing Options teams and the workshop with sector stakeholders. Part five draws together learnings from other data systems internationally, and the wider literature on data system design and operation, in order to identify design elements for a new homelessness data system. Elements cover aspects such as governance, data architecture, data access, and the purpose of a system, and are presented as a series of options of how other systems have addressed these areas. Part six outlines fours options for a new homelessness data system in Wales, bringing together the learnings from the project. The options are presented as whole functioning systems, from inputs to outputs; however, they can be thought of as a set of interchangeable parts. Part seven concludes this report by suggesting next steps in determining the future homelessness data landscape in Wales.
Part one: Project overview

This report outlines the findings from a project part funded by the UK Collaborative Centre for Housing Evidence (CaCHE) and Welsh Government, with the aims of:

- Assessing current statutory homelessness data collection practices across local authorities
- Engage stakeholders to explore the sector’s opinions on a move to a new data collection
- Examine alternative approaches to collecting homelessness data

A range of approaches were adopted in order to achieve these aims, including: a survey with local authority Housing Options teams; a consultation with the sector on their use of published statutory data and perceptions of a new system; a stakeholder workshop; desk based review of systems; an evidence review; informal stakeholder conversations; and other public engagement such as blogging in order to promote the project activities and prompt feedback. The outcome of the project has been a set of four options for a new homelessness data system in Wales—presented in Part six. This part of the report outlines the main data activities in engaging with stakeholders and learning from other systems, with these being the main sources of information upon which the options were generated.

Engaging stakeholders

A survey was conducted with statutory ‘Housing Options’ teams in order to gain insight into their current data collection practices in relation to statutory homelessness. The survey focused on several key areas that were identified from the literature as important when collecting national individual level data, namely: how data are stored and the ability to export data; the ability to link data, including the collection of personal data needed to link data sources; and prior history of sharing data within and without the local authority. At the end of the survey, respondents were asked their views of a new data collection. Survey links were provided via Welsh Local Government Association contacts, as well as the Welsh Government’s own email lists of data contacts at local authorities. Multiple individuals within each authority could complete the survey, as no single person would have been able to answer all questions; responses were therefore aggregated to each local authority. 11 of the 22 authorities in Wales responded to the survey. Responses were received from mainly urban/valleys areas, with two from mainly rural authorities. Only three of the responding authorities were from outside of the South Wales region.

In October 2018, a workshop was held at Cardiff University and was attended by over 40 delegates representing a range of stakeholders from the housing/third sector in Wales. Opening the workshop were short presentations on new data collections in Scotland and Wales, from the Centre for Homelessness Impact and the Street Homelessness Information Network (SHIN) pilot, respectively. Workshop attendees were split into three groups, with each group having at least one representative from each of the four main stakeholder types (third sector, local government, national government, and the research community). Each group was presented with a short description for a homelessness data collection system for Wales, including: the status quo, a pan-wales data management system, and individual level data reporting system. The choice of scenarios was based on an initial review of international data systems. The task for each group was to identify opportunities/benefits and challenges for their scenario. At the end of the workshop there was a group feedback session and brief discussion.
Evidence reviews: international data systems and wider literature

The bulk of the project involved a desk-based review of international administrative data systems, drawing on published materials in order to gain insight into the design of other collections. Alongside this targeted review, a further evidence review of uses of administrative data and research into the experiences of setting up new data systems was also conducted. The review of the wider administrative data literature provided insights into the benefits and issues encountered by others when initiating new data systems, in order to learn from a diverse set of experiences, particularly from data system developments in health and social care settings. Healthcare settings tend to have well developed data systems due to the routine production of administrative data, such as medical notes and medical test results. Social care settings in the UK have gone through changes in their approach to data, partly in response to high profile cases where the lack of a joined-up approach has led to death/misconduct, and which have garnered media attention and scrutiny. Health and social care therefore offered insights into different ends of the data spectrum: using data differently in a well-developed setting and changing how data are collected and shared in order to improve services for people.

The systems review primarily covered homelessness data systems, however non-homelessness collections were included in order to learn from other areas, particularly within Welsh Government. Systems were reviewed if they involved the gathering and/or the transmission of personal sensitive information, for example data about individuals or cases. Relevant systems were identified through an iterative process drawing on the knowledge of policy, sector stakeholders, and literature referencing administrative data collection and analysis. In total, 53 systems from 9 countries were included in the review. Each system was considered in terms of the following general questions:

- What are they designed to do / what is their purpose?
- What populations do they cover?
- Who is involved/responsible for collecting the data?
- What data are collected and how is this stored and accessed?
- What are their potential/stated weaknesses and strengths?

Addressing these questions provided a broad overview of the collections, which then guided thinking around how a new data system could work in Wales. Overviews of each of the systems can be found in Appendix A.

A list of Welsh and UK data systems is provided in Table 1, split by whether the system was directly related to homelessness. The international systems in Table 2 have been split into their respective countries, in order to reflect the fact that multiple systems could be in operation within a single country, covering a range of different geographies and service areas. In the United States, for example, communities could have a centralised data collection on people assisted by homeless services in their area (e.g. New York City Coalition on the Continuum of Care Homeless Management Information System); however, multiple communities within a State could also come together to form a larger data resource (e.g. CARES of NY Regional Homeless Management Information System), whilst the same data was also aggregated to feed into national data reporting on homelessness (e.g. Department of Housing and Urban Development homelessness data collections).

---

2 The United States’ Department of Housing and Urban Development homelessness data collection is included, despite its use of aggregate data, as the system itself structures how communities gather data at the local level in order to feed into this national system of reporting.
### Table 1: Welsh and other UK based systems included in review, by population covered (homeless or non-homeless)

<table>
<thead>
<tr>
<th>Homeless populations</th>
<th>Non-homeless populations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wales</strong></td>
<td></td>
</tr>
<tr>
<td>Supporting People, Wales</td>
<td>Lifelong Learning Wales Record</td>
</tr>
<tr>
<td>Street Homeless Information Network (SHIN) pilot</td>
<td>SSDA903 collection/Looked after children Census</td>
</tr>
<tr>
<td>Lifelong Learning Wales Record</td>
<td>Welsh Community Care Information Service</td>
</tr>
<tr>
<td>SSDA903 collection/Looked after children Census</td>
<td>Pupil Level Annual School Census</td>
</tr>
<tr>
<td>Welsh Community Care Information Service</td>
<td>European Structural Fund participant database</td>
</tr>
<tr>
<td>Pupil Level Annual School Census</td>
<td>Housing Stock Analytical Resource for Wales, UK Secure eResearch Platform</td>
</tr>
<tr>
<td>European Structural Fund participant database</td>
<td>Ffynnon</td>
</tr>
<tr>
<td>Housing Stock Analytical Resource for Wales, UK Secure eResearch Platform</td>
<td>Common Transfer System and 'school2school'</td>
</tr>
<tr>
<td>Ffynnon</td>
<td>Secure Anonymised Information Linkage databank</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other UK</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainstay</td>
<td>Expanded Troubled Families programme</td>
</tr>
<tr>
<td>Greater Manchester Tackling Homelessness Information Network (GM-Think/M-Think)</td>
<td>Better Futures</td>
</tr>
<tr>
<td>In-Form DataLab</td>
<td>Multi-Agency Safeguarding Hubs</td>
</tr>
<tr>
<td>Combined Homelessness and Information Network (CHAIN)</td>
<td>Dementias Platform UK Data Portal</td>
</tr>
<tr>
<td>Supporting People Client Record System and Outcomes Framework</td>
<td>Kent Integrated Dataset</td>
</tr>
<tr>
<td>Scottish statutory homelessness collections</td>
<td>Connecting Care</td>
</tr>
<tr>
<td>Homelessness Case Level Information Collection (DELTA)</td>
<td>Continuous REcording of lettings and sales (CORE)</td>
</tr>
<tr>
<td>Alert system (Client referrals)</td>
<td>NHS Scotland Corporate Data Warehouse &amp; Data Marts</td>
</tr>
<tr>
<td></td>
<td>North West London Whole Systems Integrated Care (WSIC) data warehouse and dashboards</td>
</tr>
<tr>
<td></td>
<td>Ministry of Justice DataLab</td>
</tr>
</tbody>
</table>

### Conceptual framework

Throughout this report a framework has been adopted in order to conceptualise the process via which operational data are used beyond their immediate day-to-day purposes, i.e. become administrative data that are used for performance measurement, research, and potentially feeding back into decision making. These stages in the cycle from operational to administrative data relate to the gathering of information about a persons’ case when they interact with an agency or service. Information becomes data when it is codified and recorded for storage in local data management systems—‘information gathering and data recording’. Data are often, though not always, extracted and transmitted to another organisation, for example local or national government—‘data extraction and transmission’. Finally, data can become information again when it is used to form the basis of research, statistics, and decision-making—‘data analysis and reporting’. This framework is loosely applied throughout this report, as only some of these stages may be relevant to any single system. For example, the Combined Homelessness Information Network (CHAIN), operates as a shared data platform on rough sleeping across Greater London Authorities. However, as organisations can enter data directly into the CHAIN system, extraction and transmission of data are therefore not an integral part of the systems design.
**Table 2:** International systems included in the review, split by country

<table>
<thead>
<tr>
<th>Country</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ireland</td>
<td>Pathway Accommodation &amp; Support System</td>
</tr>
<tr>
<td>Denmark</td>
<td>Client registration system database</td>
</tr>
<tr>
<td>Poland</td>
<td>Homelessness and housing exclusion (BIWM) Data Standard</td>
</tr>
<tr>
<td>Estonia</td>
<td>X-tee e-Estonia**</td>
</tr>
<tr>
<td>Australia</td>
<td>Specialist Homelessness Services Collection</td>
</tr>
<tr>
<td></td>
<td>Specialist Homelessness Information Platform</td>
</tr>
<tr>
<td>New Zealand</td>
<td>New Zealand Integrated Data Infrastructure (IDI)**</td>
</tr>
<tr>
<td>United States</td>
<td>Department of Housing and Urban Development homelessness data collections, National</td>
</tr>
<tr>
<td></td>
<td>New York City Coalition on the Continuum of Care Homeless Management Information System</td>
</tr>
<tr>
<td></td>
<td>Chicago Homeless Management Information System</td>
</tr>
<tr>
<td></td>
<td>Online Navigation and Entry System, San Francisco</td>
</tr>
<tr>
<td></td>
<td>Clarity - Nevada Statewide Community and Homeless Management Information System</td>
</tr>
<tr>
<td></td>
<td>CARES of NY Regional Homeless Management Information System, New York State</td>
</tr>
<tr>
<td></td>
<td>Ohio Human Services Data Warehouse</td>
</tr>
<tr>
<td></td>
<td>Michigan’s Statewide Homeless Assistance Data online Warehouse (SHADoW)</td>
</tr>
<tr>
<td></td>
<td>Veterans Health Administration Corporate Data Warehouse, National**</td>
</tr>
<tr>
<td></td>
<td>Virginia Longitudinal Data System**</td>
</tr>
<tr>
<td></td>
<td>North Carolina School Works**</td>
</tr>
<tr>
<td></td>
<td>Knoxville Homeless Management Information System (KnoxHMIS)</td>
</tr>
<tr>
<td></td>
<td>Stella P, National</td>
</tr>
<tr>
<td>Canada</td>
<td>Homeless Individuals and Families Information System, National</td>
</tr>
<tr>
<td></td>
<td>Calgary Homelessness Information Management System</td>
</tr>
<tr>
<td></td>
<td>Shelter Management Information System, Toronto</td>
</tr>
</tbody>
</table>

** Non-homelessness data system

Having outlined the methods used in this project and clarified some of the terminology used, Part 2 of this report addresses the first aim of the project, to assess current statutory homelessness data collection practices.
Part two: Statutory homeless data in Wales

This section of the report provides a critical overview of the current approach to gathering, recording, transmitting and reporting statutory homelessness data in Wales, drawing together findings from the survey of local authority Housing Option teams where appropriate, in order to elaborate on current practices. It outlines issues with the current approach to statutory data that were raised by local authorities and the sector generally and concludes by suggesting ways in which a new data system could address these issues. To begin with however, there is a brief introduction to the legislative framework underpinning the statutory data. In any discussion of administrative data, it is important to understand their origins from specific organisational practices, in this case, the disposal of statutory duties by local authorities in Wales.

Statutory homelessness provision and data collection in Wales

The Housing (Wales) Act 2014 introduced a fundamental reform to homelessness legislation in Wales, placing new duties on local authorities to help prevent and relieve homelessness. Though some local authorities had been undertaking preventative steps prior to the Act, the new legislation now means that more households are now offered some form of meaningful assistance if they are found to be homeless or threatened with homelessness. The Act has several parts that are monitored as part of the current statutory homelessness statistics:

- The duty to assess a case
- Provision of advice and assistance, which is the only recourse to assistance if the household is not threatened with or is not homeless
- Prevention of homelessness if the household is found to be threatened with homelessness
- Help to secure accommodation for homeless applicants
- A full duty to secure accommodation where the household belongs to an ‘at risk’ or priority need group, offered after relief efforts have failed

Local authorities are also able to provide temporary accommodation whilst permanent accommodation is being found, which includes placement in bed and breakfasts, hostels, the private rented sector, and other accommodation. Limited information on temporary accommodation placements are also collected under the statutory data returns.

Figure 1 summarises the flows of data and the main stages in the process through which information from people approaching authorities becomes statistical outputs. In brief, the current collection involves Housing Options teams completing an aggregate reporting template and submitting this to Welsh Government. Each of the stages from information gathering to analysis and reporting is discussed in the following sections. Included in Figure 2 are issues with the current collection that were raised by the sector, some of which will be elaborated upon through this section’s discussion.

---

1 Priority need categories collected in the statutory data, include: households with dependent children; where a household member is pregnant; where there is a vulnerable member of the household; if the applicant is a care leaver; is aged 16 or 17, is fleeing domestic violence, leaving the armed forces or prison; and households homeless in an emergency.
Figure 2: Data flows as part of the current statutory data system in Wales; from information about people who are homelessness, to final statistical outputs. Underneath are presented some of the issues with the current approach that were identified by the sector.

Information gathering and data recording
- Inconsistent information gathering and data recording across local authorities
- A lack of detail on who people are
- Misses prevention and relief work done by other agencies
- Focus on the Act reduces relevance to wider audience

Data extraction & transmission
- Lack of appropriate data management skills/knowledge amongst staff completing returns
- Translation of local data into aggregate reporting form can be time/resource intensive

Analysis & reporting
- Unable to measure flows of people, geographically or through the system
- Cannot measure impact on prevention of homelessness in the long term
- Reporting and analysis focuses on statistics, with limited research options due to aggregate nature, therefore limited assessment of ‘what works’

Information gathering and data recording

At a local level, Housing Options teams gather information and record data about individuals/households who approach them for advice and assistance. The survey with local authorities indicated that a range of ‘off-the-shelf’ housing software solutions were being used in order to store and manage Housing Options data, including:
- Abritas (by Civica)
- Housing Management Solution (by Northgate Public Services)
- Housing Jigsaw (by National Practitioner Support Service)
- Open Housing (by Capita One)

Several Housing Options teams reported using multiple software systems in tandem, with one authority reporting use of four different software solutions in order to record and track work done under the Act, including placements in temporary accommodation (TA).

Local authorities are provided with a data template they are required to complete, which contains a detailed explanation of the data required. However, guidance is framed in terms of completion of the data template, rather than the collection of data itself. In comparison, in both Scotland and England, and most other homelessness data collections internationally, data providers are giving clear data specifications, including variables and values, whilst ultimately being able to design their data systems as local needs require. Though the provision of guidance is only one part of improving data quality—see discussion in Part Five—it does create consistency across authorities and improves the comparability of data from differing authorities. In Wales, the lack of a common national data specification for use by local authorities when designing/commissioning their local data management systems has potentially lead to divergences in local authority recording practices; something which the sector identified as a challenge with the continued use of aggregate reporting.

The survey of Housing Options teams highlighted that some teams were going beyond what is being asked for by Welsh Government. For example, though only the main applicant’s demographic characteristics are requested in the aggregate return, 10/11 of the authorities who responded to the survey indicated that they collected information on other householders, including information on relationships between householders, as well as personal data such as name, date of birth, gender, and postcode. Personal data on all household members is important if the intention of any new system is to track all people facing homelessness, rather than just the main applicant for assistance. Being uniquely able to identify people experiencing homelessness in a household can help identify repeat homelessness, where the main applicant changes, i.e. a different partner applies for assistance, and open new areas of research and policy based on evidence derived from data linkage. From a service provision perspective, Housing Options teams indicated that having greater detail about all people in the household meant that more appropriate assistance and accommodation could be provided—10/11 authorities indicated that collecting data on more than the main applicant was a good idea for a new data system.

The data requested by Welsh Government are limited breakdowns of the characteristics of households being assisted, broad types of assistance being given, and the outputs of local authority decisions in terms of how the legal duty under the Act was discharged. Though local authorities likely collect a range of detailed information about the people they assist as part of their day-to-day operations, stakeholders expressed that the data being collected and reported at a national level do not provide enough detail on who people seeking advice and assistance are. This lack of a detailed bigger picture on homelessness is noticeable when compared with international homelessness data systems. Community Homeless Management Information Systems (HMIS) in the United States primarily collect data on people’s characteristics and the structural circumstances that may have led to their homelessness, e.g. unemployment and low income, rather than on the type of assistance received. Within HMIS, assistance provided can be inferred by the general type of agency submitting data, e.g. shelter, outreach team etc. Similarly, the new H-CLIC collection requires authorities to record data on people’s backgrounds, including employment status, nationality, benefits receipt, and accommodation at the time of seeking assistance. Having greater detail on who people are and factors leading to their homelessness is arguably an integral part of any universal prevention effort, in tracking and monitoring the antecedents of homelessness (Fitzpatrick et al. 2019).

Data extraction and transmission

In order to populate aggregate reporting forms, data from Housing Options’ data management systems are extracted and processed to generate the required counts. Counts relate to the different legal stages under the Act, for example initial assessments and those made at prevention and relief stages. Households are only included in the counts if their case has reached a conclusion under that stage. Some of the Housing Options teams and local authority respondents indicated that extracting and processing data in order to complete aggregate forms was time consuming, particularly in instances where there were no dedicated resource within the housing team for data management, and it fell to caseworkers to perform data processing tasks. Aggregate forms are submitted to Welsh Government, once completed, on a quarterly basis via a secure file transfer system (AFON).
In addition to counts of households who apply for assistance under the Act, the aggregate return also covers households housed in temporary accommodation (TA). However, TA data only relate to households accommodated at a given point in time, and only give a very broad indication of how long a household was in that accommodation at that point in time. Current data fail to provide adequate national level information on TA to enable analysis of its use, including average time spent in TA, movement between different forms of temporary accommodation, and how TA relates to assessments and outcomes under the Act. Linkage between assessments under the Act and placements in TA, and subsequent movements between different TA, would enable the appropriateness of TA use in Wales to be evaluated. Importantly, many local authorities who participated in the data collection survey indicated that they were able to link information on applications for assistance under the Act to placements in temporary accommodation (9/11).

Built into the current aggregate reporting forms are a series of validation rules that check for logical errors, such as incorrect totals in tables, and compares figures to previous reporting periods in order to check for unexplained jumps in homelessness numbers. Welsh Government also undertakes validation once the data collections have been received from local authorities, which may result in queries to the local authority if any anomalous data are submitted. However, as the methods used to gather information and record data underpinning the aggregate data varies with each Housing Options team, the lack of consistency in the ‘raw’ data complicates the ability of Welsh Government to undertake quality assurance.

Data analysis and reporting

Once the aggregate forms have been validated and cleaned, data from all authorities are combined to form the basis of statistical outputs, including statistical releases and provision of local authority figures on StatsWales. The current outputs by Welsh Government focus on the production of statistics, and, due to the aggregate nature of the data, there are limited opportunities for research using these data, something which stakeholders at the workshop picked up on—that the current collection lacked ‘flexibility’ to conduct analyses. One of the recurrent data limitations highlighted in the sector engagement activities, and previous literature, is that aggregate data returns do not permit people’s movement through the various sections of the Housing (Wales) Act to be assessed. As aggregate data are collected on households assessed at each stage under the Act, the same households may appear multiple times within the statistics if they are assessed under different sections of the Act, i.e. where prevention and relief fail, and the household is owed a full duty to be housed. Individual level data collection and national data linkage would enable people’s journeys through the housing system to be assessed by enabling their multiple assessments under the Act to be viewed as a single period of homelessness.

An aspiration for a new homelessness data system in Wales should be to engage in data linkage, which has the potential to greatly improve the sector’s understanding of what works in achieving better outcomes for people who are homeless. The sector also raised the use of data linkage as part of the engagement activities. As indicated previously, authorities who responded to the survey were already collecting the individual level data necessary to link data, i.e. name, date of birth, gender, and address. However, fundamental to any data linkage at scale, and the use of data for research generally, is having a legal and ethical basis to process data for these purposes. As part of the survey with Housing Options teams, questions were asked about the use of privacy notices and prior data sharing. Privacy notices are an important method for informing people of how their data are being used under data protection legislation. The majority of Housing Options teams had privacy notices in place outlining how client data were being used by the housing service (10/11). Less than half of authorities (5/11) indicated that their privacy notices stated that people’s data would be used for research purposes, with slightly more indicating that the data could be used for statistical purposes (8/11). In terms of sharing data, 9 out of the 11 Housing Options teams indicated that their privacy notices stated that data would be shared with others.

5 For more detail on statistical outputs relating to homelessness available on StatsWales, see: https://statswales.gov.wales/Catalogue/Housing/Homelessness
6 Chapter 14 in Downie, M. (2018) focuses on the state of homelessness data in Great Britain, including recommendations regarding setting up data linkage systems in England, Scotland, and Wales.
Local authorities command large amounts of data, being responsible for collating a number of data sources that are used by Welsh Government for national statistics, e.g. education and social care data; most authorities generally collect more data on residents in their regions than is reported to Welsh Government, in order to guide local service delivery. However, the extent to which these data are utilised for joined up delivery varies; more ‘data mature’ authorities have the capacity and organisation-wide buy-in to mobilise this data in integrated ways, whilst others use it predominantly for monitoring purposes (Wales Audit Office 2018). This variety of data maturity and integration was also found in the Housing Options survey; one authority provided evidence of how data from different systems, including housing benefits data, was being brought together; most authorities gave evidence implying that their systems were less integrated, for example using Excel to record TA placements. Linkage at a local level would enable locality specific information to be generated. However, for national policy development and evaluations of the legislation, national data linkage and sharing would be necessary.

Wales already has national data linkage infrastructures in place in the form of the Secure Anonymised Information Linkage (SAIL) databank and other offshoots from the SAIL system, as well as expertise in administrative data and data linkage at the Administrative Data Research Wales. Wales therefore could design national level data linkage into any future data system. However, though the technical and data governance infrastructure exists, data are limited to health and education data in the SAIL databank; there is limited access to central government data sources at a national level. Given that poverty is a key driver of homelessness (Bramley and Fitzpatrick 2018), and therefore an important focus for prevention efforts, social security data held by the Department for Work and Pensions (DWP) and income data held by HMRC are key data sources in understanding the structural causes driving poverty.

Summary

The collection of individual level statutory data could address several of the issues with the current system at the analysis and reporting stage, which mostly relate to the limited research options available with aggregate data and the inflexibility of aggregate data to respond to ad hoc policy questions. Improving the evidence base and ability to conduct analysis and evaluation of what works in Wales (and what is not working), could have positive impacts on policy, practice, and service commissioning. That some Housing Options teams had shared data, and with a range of different organisations, was a good indicator that they were open to the idea mobilising data for social good. Furthermore, the collection of personal data on all household members, which a majority of Housing Options teams were undertaking, means that should a new data system require individual level data to be reported, this will not cause a dramatic shift in data recording practices—for the responding authorities at least. However, some authorities in Wales still perceive data sharing as risky, and there is a lack of clarity on what data can be shared, as has been identified by the Wales Audit Office (2018), and which may hamper a new data system based on the sharing of personal data. There was a general sense that the current data lack detail on who the homeless are, what experiences they have faced, and where they have come from (geographically, and in terms of tenure). Arguably, these are the most important pieces of the data jigsaw that a new system should address, in order to design programmes and services that work toward improved prevention and targeting people ‘at-risk’.  

7 A development from the SAIL databank has been the UK Secure eResearch Platform, a data platform for the secure linkage, storage and use of data for research purposes. The SAIL databank has become a user of UKSeRP. For more information on the SAIL databank, see https://saildatabank.com/about-us/overview/
Part three: The wider ‘data landscape’ in Wales

The project was originally intended to explore a new approach to statutory homelessness data in Wales. However, the scope of the project was expanded in response to several drivers that emerged from the literature and sector engagement activities, including:

- Data on ‘statutory homelessness’ relates to a specific experience of homelessness at a particular point in time. There are a range of other homelessness data sources in Wales that relate to other experiences; combining these data can provide policy and practice with a greater understanding pathway into and out of homelessness, and therefore the ability to better evaluate and design effective services, programmes, and policy;

- Homelessness is more than a ‘problem of housing’, and therefore more effective services should take a cross-sector response, itself fostered by data sharing.

- Early prevention efforts require ways of assessing population level housing (in)stability, in order to identify groups at-risk of becoming homeless, prior to them having to access ‘crisis’ prevention/relief—statutory data are only part of the picture;

- There are new data and policy developments in Wales that will potentially change the data needs/requirements of the sector, including: the piloting of the Street Homeless Information Network, the ending of the Supporting People grant and its amalgamation into the Housing Support Grant, and the shift to Housing First.

The following discussion draws these drivers together to explore the ‘data landscape’, being the range of data sources and flows of data between different agencies in Wales. The homelessness data landscape is mapped out in order to illustrate the rich data sources that already exist, albeit in isolation, that could be combined under a new approach to data. Secondly, the discussion considers how data from other public bodies and the wider housing (non-homeless) sector could contribute to early prevention efforts. The third element of this part of the report explores future changes to policy and funding that will impact on the data landscape in Wales by influencing how the sector operates.

The national homelessness data landscape in Wales

Figure 3 is an intentionally simplified representation of the current data landscape relating directly to homelessness, illustrating how, in this instance, data collected by local authorities and third sector organisations (far left hand side) feed into a set of information sources and outputs (far right hand side). It should be noted that although there has been some move to undertake research and evaluation using linked data (McGinn et al. 2016), most outputs within the current landscape are aligned with the measurement of homelessness and activities undertaken to assist people. Research and evaluation studies are largely ad hoc, under reported, and local; ideally, we would want a landscape where research and evaluation were routine, well reported so that everyone can learn from its findings (good and bad), and national in scale (where appropriate). Aside from the statutory data, there are two main formal data sources that provide insight into the experience of homelessness in Wales at a national level: Supporting People and rough sleeper counts.
**Supporting People outcomes data**

The Supporting People (SP) Programme provides grant funding to local authorities to enable them to either deliver directly or commission housing related support services. Examples of provision under SP includes street outreach and older people’s services such as home modifications. Welsh Government and other agencies acknowledge that SP is vital in supporting homelessness prevention efforts and in tackling poverty (Wales Audit Office 2017a). Funded services are contractually obliged to provide data on client outcomes, as well as contribute data to an evaluation of the Supporting People programme being conducted through data linkage. Welsh Government do not publish any statistics in relation to SP; however, some local authorities produce top-line figures of units/people supported by under SP projects commission/run in their area. Local authority reporting on people supported under SP is inconsistent across authorities, and only infrequently are data reported as stand-alone statistical/research outputs. To summarise, very little is publicly known on the people being assisted under SP funded schemes, despite the fund being an integral part of homelessness prevention activity in Wales.

**Rough sleeper street counts**

Local authorities in Wales in partnership with other agencies undertake a data collection exercise in order to monitor the extent of rough sleeping across Wales. For the purposes of the collection, rough sleepers are defined as people who are sleeping overnight in the open air or in buildings/other places that are not meant for habitation. There are two phases to the collection: in October local authorities engage in a two-week information gathering exercise, which is followed by a count of people seen sleeping rough on a single night in...
November. The two-week exercise involves local authorities conducting a questionnaire-based survey among relevant local agencies and services who work with rough sleepers. Data from these surveys have been used in previous rough sleeper count exercises to target where to conduct the single night count. Questionnaires are completed for each person who attended a service for rough sleepers and are analysed by the Local Authority Housing Network rather than Welsh Government. Authorities were encouraged to use a unique identifier in order to prevent double counting of surveyed people. The results of the questionnaires and counts are summarised in statistical reports produced by both the Welsh Government and Local Authority Housing Network.

Across SP, rough sleeper, and statutory data, there is potentially a high degree of overlap in data collection due to the nature of funded programmes having contractual obligations to provide data, and because agencies engage in other data collections. For example, The Wallich is a homelessness charity in Wales that provides outreach work with rough sleepers, with this activity having been primarily funded by SP. As such, data from the Wallich on rough sleepers could conceivably be included in the SP collection, as is their contractual obligation under the SP funding rules, and any statistics on people sleeping rough that the Wallich contribute to e.g. the Street Count. Harmonizing data from different sources and rationalizing flows of data through a new data system could reduce burden on data collecting organisations and duplication of effort in reporting into multiple data systems, as well as providing a way of assessing the number of unique people experiencing homelessness in Wales.

Data coverage and gaps

Table 3 summarises the extent to which the current homelessness data landscape can be used to enumerate different experiences of homelessness, as defined by the ETHOS definitions of homelessness and housing precarity (Amore et al. 2011). There are several caveats attached to each data source under each homelessness experience, compounded by inability to bring these sources together to generate unduplicated counts of people experiencing homelessness.

Statutory data provides some insight into all types of homelessness, only for those individuals who seek assistance from statutory services. Though there is some international evidence that people accessing certain services will access others 'at some point', for example in Denmark where rough sleepers overlapped with shelter users to an extent that the two groups could be considered the 'same' (Benjaminsen and Andrade 2015), qualitative insight from client facing workers that suggests that some individuals (with high needs) may disengage from mainstream services, and engage with low-threshold services (Russell and Thomas 2019). In this regard, SP funded services could provide information on people who access third sector services directly—and who may be disengaged from statutory provision.

In Wales there is a conspicuous lack of data on people accessing hostels and other emergency accommodation, except where placed in those accommodation types by an authority, which is in stark contrast to many of the international homelessness administrative data systems where these provider types are far more prevalent, e.g. Denmark, United States, Canada—discussed in detail in Part five. People's use of hostels and other emergency accommodation may be captured under SP data. However, given that hostels receive funds from a mixture of sources, including donations and housing benefits, then reliance on SP for insight into hostel and emergency accommodation use, where this is not linked to statutory provision, may always be under-estimated.

One gap in data that is particularly intractable using only administrative sources will be the ‘hidden’ homeless, or ‘Homeless people living in conventional housing with family and friends’ in the ETHOS definition. Crisis adopt a definition of hidden homelessness as including concealed households, sharing households, and overcrowded households (Fitzpatrick et al. 2017). Generally, a person who is hidden homeless would not normally be accessing housing services, nor would they be visible as part of street counts. At present, figures for hidden homelessness in Wales are estimations based on national sample surveys, predominantly for the UK/England, or from the Census—in the case of overcrowding data. Hidden homelessness may be partly captured in statutory data where the reason for homelessness was stated as ‘being asked to leave by family and friends’. However, this figure is likely an under-estimate as it represents only those individual/households who seek assistant, and therefore make themselves visible. How to collect ‘better’ data on hidden homelessness in Wales is discussed in Part Six, however in summary, we should not rely solely on administrative data to assess the scale of homelessness in Wales, but choose multiple methods and data sources appropriate to the populations we wish to measure.

---

8 For statistical releases on rough sleeper numbers and greater detail on the methodology of the rough sleeper count in Wales, see [https://gov.wales/national-rough-sleeper-count](https://gov.wales/national-rough-sleeper-count)
Table 3: Summary of types of homelessness, as defined by the European Typology on Homelessness and Housing Exclusion (ETHOS), and coverage in the main data sources in Wales

<table>
<thead>
<tr>
<th>ETHOS definition of homelessness</th>
<th>Statutory data</th>
<th>Supporting People</th>
<th>Rough sleeper count</th>
</tr>
</thead>
<tbody>
<tr>
<td>People living rough</td>
<td>Only where a person indicates this as the reason for their homelessness/threat of homelessness</td>
<td>Only people attending/receiving assistance from funded services</td>
<td>Only where the person was ‘bedded down’ during the data collection time-period</td>
</tr>
<tr>
<td>People in emergency accommodation</td>
<td>Only where a household is placed in temporary accommodation by the authority</td>
<td>Only people attending/receiving assistance from SP-funded services</td>
<td>Only a count of the number of units of bed-space, rather than who is occupying that bed-space</td>
</tr>
<tr>
<td>People living in accommodation for the homeless</td>
<td>Only where a household is placed in temporary accommodation by the authority</td>
<td>Only people attending/receiving assistance from funded services</td>
<td></td>
</tr>
<tr>
<td>People living in institutions</td>
<td>Only where a person indicates this as the reason for their homelessness/threat of homelessness</td>
<td>Only people attending/receiving assistance from funded services where spend category is ‘People with Criminal Offending History’</td>
<td></td>
</tr>
<tr>
<td>People living in non-conventional dwelling</td>
<td>Only where a person indicates this as the reason for their homelessness/threat of homelessness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homeless people living in conventional housing with family and friends</td>
<td>Only where a person indicates this as the reason for their homelessness/threat of homelessness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Beyond homelessness: Integration of data across different policy areas

In their ‘SHARE’ framework, the Centre for Homelessness Impact indicate that a person-centred ecosystem of services, alongside homelessness as a shared priority cutting across policy and practice areas, can offer the best chance of ending homelessness⁹. One way to facilitate these outcomes would be to encourage sharing of data across homeless services, whether through direct data sharing about cases, or through integrated data systems that can help to disrupt siloing of policy through linked data research. However, from a prevention perspective, data from homeless services potentially reflect points after which early intervention is effective: it may therefore be beneficial for a future system to at least incorporate the possibility of linking to and sharing data with other public services. A fictitious housing/homelessness pathway is presented in Figure 4, in order to illustrate the interactions a person may have with various services as they move from housing instability into homelessness. Though someone may interact with several services over the period leading up to and during their homelessness, housing providers such as social landlords, and public services such as General Practitioners and the Police, are potential points of contact with the person during ‘at risk’ periods.

To a certain extent, the Housing (Wales) Act encourages co-operation and cross-sector working to prevent/ alleviate homelessness, particularly between social services and local housing officers (Section 95 of the Act). A new data system could provide a mechanism to formalise and standardise the integration of data—such as social care, social landlords, healthcare, and criminal justice system—as part of early prevention efforts to assist people be before they need to access ‘crisis’ services. The integration of data could take place at a national level through data linkage infrastructures that enable actionable intelligence leading to policy development (Fantuzzo and Culhane 2015) or it could occur through the use of live systems and local data sharing platforms that assist the identification of people ‘at-risk’. For example, missed payments to a social landlord could be an early warning sign that the household is facing difficulties, and therefore an opportunity for the landlord and the Housing Options team/housing support service to offer advice and assistance, or preventative work.

Figure 4: Illustrative example of a person’s interactions with different services over the course of their housing/homelessness pathway

<table>
<thead>
<tr>
<th>Person’s housing pathway</th>
<th>At risk of homelessness</th>
<th>Loss of home</th>
<th>Homelessness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Street outreach</td>
<td></td>
<td>Seen bedded down</td>
<td></td>
</tr>
<tr>
<td>Housing Options</td>
<td>Assessment</td>
<td>Legal Duty</td>
<td>Outcome</td>
</tr>
<tr>
<td>Temporary accommodation</td>
<td>Entry</td>
<td>Hostel</td>
<td>Exit</td>
</tr>
<tr>
<td>Housing support service</td>
<td>Start</td>
<td>Support</td>
<td>End</td>
</tr>
<tr>
<td>General Practitioner</td>
<td>Appointment &amp; referral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental health community team</td>
<td>Visit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Police</td>
<td>Call-out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social landlord</td>
<td>Notice: missed payment</td>
<td>Eviction</td>
<td></td>
</tr>
</tbody>
</table>

⁹ For more information relating to the SHARE framework, see: https://www.homelessnessimpact.org/share
From the review of systems, several examples were found where data were integrated from services with remits outside of homelessness, e.g. other public services and housing providers. In the United States, regulations associated with receipt of government funding for the formation of Continuum of Care outline rules around the formation of Homeless Management Information Systems servicing the continuum (Department of Housing and Urban Development 2011). These rules permit homeless and non-homeless projects not funded by state funds to participate in the HMIS. The rationale for increasing the number of participating organisations is that it increases the number of data points on people and can therefore help the HMIS identify returns to homelessness with greater accuracy, e.g. a person being seen on the streets by outreach services not being funded by the state can be identified. Another example of multiple services contributing data is in Scotland, where in 2014 the Scottish Government expanded the PREVENT1 data collection on preventative assistance to enable registered social landlords (RSLs) to voluntarily submit data. RSLs were asked to collect the person’s name in order to enable linkage to the mainstream homelessness collection (H1). The decision to expand PREVENT1 acknowledged the fact that agencies other than local authorities are undertaking important prevention and relief work\textsuperscript{10}.

The Geelong Project in Australia demonstrates the power of thinking outside of sector silos. Geelong used a whole school-screening assessment tool to identify young people at risk of experiencing homelessness and school disengagement. If a young person was identified as at risk, then an offer of support to the young person and family was made. In the original Geelong study, the pre-trial baseline number of adolescents entering the Specialist Homeless Services system in the study area (230 entries) decline by 40 per cent compared to the post-trial baseline (100 entries) (McKenzie 2018). This approach to homelessness prevention in schools has since been piloted in Canada (The Upstream Project\textsuperscript{11}), and Wales will start its own pilot in partnership with End Youth Homelessness Cymru.

On the horizon in Wales

In addition to making more efficient use of the large amounts of data already collected on the experience of homelessness in Wales, there are several ‘on the horizon’ factors that will influence the data landscape in Wales, namely: (1) Housing First, (2) the Street Homeless Information Network, (3) changes in the Supporting People programme through its amalgamation with other funding streams related to housing/homelessness alleviation—the Housing Support Grant (HSG), and (4) new technologies and approaches to homelessness prevention/alleviation reducing the detail of data collection.

**Housing First**

A ‘Housing First’ (HF) approach was originally developed in New York City in the early 1990’s as part of Pathways to Home, whereby people were offered settled accommodation alongside person-centred supported recovery services (Tsemberis 2010). Some of the general principles behind the original Pathways project, namely the provision of housing alongside support, have been adapted by various countries under the label of ‘Housing First’. The Welsh Government have endorsed the principles of HF and for increased use of this approach to tackling homelessness in Wales, particularly rough sleeping (Welsh Government 2018a). However, HF, as originally envisaged, calls for a change in the whole system of service delivery, not just programme level change to provide increased support. As a high intensity service, HF is designed to be targeted to those people with the greatest need/vulnerability (Watson et al. 2013), and is therefore facilitated by coordination of services and standardised assessments of needs across ‘intake’ organisations (Turner 2014). It is this whole system approach to quickly assessing and referring people that is currently missing from Welsh Government guidance on the adoption of HF. The design of any new data system may therefore want to take into account the need for coordination, standardisation of intake data in order to rapidly assess people who will benefit from HF, and for streamlined referrals across different service providers—if HF is to truly become embedded in service provision. The Street Homeless Information Network (SHIN) pilot may provide a testbed for a single technological solution for common assessment of need with people who are most likely to be suitable for HF programme, i.e., rough sleepers and others with street-based lifestyles.

\textsuperscript{10} A similar issue was raised by the sector in Wales and England, that there is work being undertaken by agencies other than Housing Options teams, which was missing from statutory data.

\textsuperscript{11} For more details on Upstream, see the project website at: https://www.theupstreamprojectcanada.org/
Street Homeless Information Network (SHIN) pilot

In 2018 the Welsh Government released its *Rough Sleeper Action Plan*, outlining its determination to tackle rough sleeping in Wales (Welsh Government 2018b). One of the subsequent developments under this rough sleeper plan has been the piloting of a new approach to centralise data on rough sleeping in Wales: the Street Homeless Information Network (SHIN) pilot. The aspiration of SHIN is to provide a platform to enable sharing of data between organizations working with rough sleepers across Wales, including third sector outreach teams and local authorities. Data from SHIN could also be used to provide continuous counts of rough sleepers, potentially replacing the rough sleeper counts—much like the CHAIN system in Greater London. As of 2019, SHIN is undergoing a pilot study with the Wallich in order to determine the viability of a national data platform for sharing rough sleeper data, as well as finalising the governance structure and data flows across the system. SHIN could potentially become an important tool to facilitate cross-organisational working and will also provide a unique data source of individual level data on rough sleeping in Wales. The SHIN pilot will also provide an ideal opportunity to explore and learn from the development of a pan-Wales homelessness data system.

SHIN has not been incorporated into the design options presented in Part six of this report as the scope and functionality of SHIN is liable to change based on the findings from the pilot. However, if SHIN remains as a data sharing platform for services interacting with rough sleepers and people with a street-based lifestyle, then the options may change in order to reduce duplication of data capture. For example, there is potentially a high degree of cross-over between Supporting People funded outreach and a future national version of SHIN. There could therefore be some value in expanding SHIN’s remit to encompass all funded non-statutory services, and unfunded services on a voluntary basis, e.g., Specialist Homeless Information Platform (SHIP) in Australia and Homeless Individuals and Families Information System (HIFIS) in Canada.

Housing Support Grant

In 2017 the Welsh Government proposed to combine ten different funding schemes into one combined ‘Early Intervention, Prevention and Support Grant’; this included Supporting People Programme funding. The proposal met with opposition from the housing sector, who argued that to combine the schemes would potentially result in a reduction in money going to ‘Supporting People’ type services—as the ring-fence around the SP funding would be removed—and would dilute the focus on homelessness, to the detriment of people experiencing homelessness12. Instead of a single large grant scheme, a ‘housing focused’ grant was proposed and has been approved: the Housing Support Grant (HSG). Though the structure and detail of the grant have not been finalised, the HSG merges Supporting People, Homelessness Prevention and Rent Smart Wales Enforcement, and will potentially lead to new reporting arrangements in order to monitor the funding of services.

---
In Welsh Government’s 2019 homelessness strategy it was suggested that the HSG will be made available on the basis of a single costed strategy at a local level; this follows, to an extent, the model of funding services in the United States (discussed later in this report), where funded communities are required to form a continuum of care. Though the details of how the HSG will be monitored have not been published, they will likely follow a similar approach to measuring outcomes and activity under the SP fund. However, given that the data being generated to monitor SP were found to be lacking in quality for evaluative and commissioning (Wales Audit Office 2017a), the new grant scheme provides the ideal opportunity to ‘re-vamp’ reporting arrangements. Specifically, a new HSG reporting system should incorporate and enable data linkage, broadly to public services data to enable the impacts of SP on other services, but more importantly to statutory data and future rough sleeper data if the SHIN pilot proves functional. It is important that future evaluations of HSG can examine its effects on housing stability/homelessness reduction, in order to demonstrate whether the fund is having the desired outcome on housing stability/homelessness prevention.

New technologies and the changing response to homelessness

As part of the increased focus on early prevention under the new homelessness strategy there may be changes in how Housing Options and other services operate, including the increased use of technology. However, new forms of technological intervention may not align themselves with current approaches to ‘administrative data’, or current thinking within policy and practice circles about what constitutes ‘data’ of value. Early prevention and advice services that take place virtually or via telephone may not be appropriate venues for the generation of detailed data collections currently possible when undertaking face-to-face intake interviews. Shelter Scotland, for example, have trialled the use of ‘chat-bots’ to provide automated advice, whilst there were several examples from the international literature of the use of telephone hotlines, e.g. OneLink in Australia acts as a single point of advice and contact for human services across the Australian Capital Territory. Some authorities in Wales are already using first-contact services: Carmarthenshire County Council for example manages demand for homelessness services through a telephone call handling service (Wales Audit Office 2017b). If new technologies are more widely adopted, the administrative data being collected may provide less detail on people’s homelessness/risk factors of homelessness, particularly those at risk who are more likely to have been assisted by low threshold advice and assistance. There may therefore need to be a move towards increased use of ‘transactional information’ in order to provide measures of housing precarity, such as website views and call rates.

Concluding thoughts on the data landscape in Wales

Policy and practice are constantly changing, and the current data landscape in Wales is a product of different funding streams converging, diverging, and sometimes disappearing, alongside changes in legislation and shifting policy foci—at present Welsh Government have strategies on people leaving the secure estate, rough sleeping, young people, and prison leavers. A new homelessness data system for Wales must therefore strike a balance between flexibility and consistency: it should evolve as the needs of the sector do, as they have done and no doubt will continue to do so, but also generate a cohesive picture of the experience of homelessness in Wales, regardless of who is funding a service and how that funding changes. Having a legal duty to assist homeless people means that a large majority of people who face precarious housing in Wales will come into contact with a Housing Options team, and, combined with the Housing Support Grant funding and the monitoring associated with this, alongside the SHIN rough sleeping pilot, Wales has the potential to be able to enumerate homelessness in its broadest definition at a national scale—without having to resort to estimation techniques based on smaller scale data collection exercises or point-in-time counts. A pool of rich data exists in every service provider, and probably in far greater detail than is requested for official statistics and reporting purposes. The data simply needs harmonising, rationalising to reduce duplication of effort and to be mobilised to create the most impact.

13 For more details, see Shelter Scotland’s blog relating to the chat-bot: [https://blog.shelter.org.uk/can-a-chatbot-help-your-charity/](https://blog.shelter.org.uk/can-a-chatbot-help-your-charity/)
Part four: Perceptions of a new system

This part of the report provides an overview of the sector’s perspectives on a new homelessness data system, drawing primarily from the survey with local authority Housing Options teams and the workshop with the sector. Due to the way in which these two forms of engagement were structured, together they offer the opportunity to gain insight into the sectors opinions generally, and in relation to specific formats of a new system; namely, a pan-Wales data management data system and new reporting arrangements. Synthesising these sources generated a series of opportunities and common areas of concern for the sector. Each of these areas is discussed in turn, with reference to the wider literature on data systems in order to provide insight into how to either maximise the opportunities or address concerns.

Opportunities of a new system

The following discussion explores some of the main opportunities of a new homelessness data system identified by the sector (Table 4). In addition to practical opportunities to reform data collection, reporting, and use to correct current deficiencies, a final set of opportunities identified were those around efficiencies and savings. Though this latter group are tangentially related to outcomes focused on ending homelessness, efficiencies and savings are particularly important as they reflect the more tangible opportunities of a new system—whereas many of the other opportunities require cultural change in order to fully maximise their potential and may not be measurable in an easily quantifiable sense.

**Improvements in data**

One of the main opportunities relating to data raised in the workshops was for improvement in data consistency and therefore quality, as at present authorities record statutory data differently and this can pose issues when trying to analyse data from across Wales and when sharing data. A new data system could afford the opportunity to improve the quality and consistency of data, whilst the submission of individual level data may increase data providers’ data quality assurance practices given that the data would be directly scrutinised.

Outside of quality, the configuration of a new data system was viewed as generating different opportunities. In the case of a centralised data management system, where all parties access the same data source, workshop participants highlighted that the timeliness of data would be improved, which in turn could lead to more responsive policy and services, e.g. live ‘needs’ mapping and person-centred care. However, there were benefits of a new reporting arrangement which would be undermined by a pan-Wales data management approach: under a new reporting arrangement, data collection processes could be determined by local need—with improved guidance and input being provided by Government to improve consistency—whereas a pan-Wales data management system would either necessitate that local authorities and agencies move entirely to the new system, or use local systems and periodically update the pan-Wales data management system—thereby potentially generating additional work.
Table 4: Summary of sector-identified opportunities of a new homelessness data system

<table>
<thead>
<tr>
<th>Opportunity</th>
<th>Specific areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improvements in data (quality/access/scope)</td>
<td>Improved quality of data through standardisation</td>
</tr>
<tr>
<td></td>
<td>Collect different forms of data more suited to the goal of ending homelessness in Wales, i.e. needs, structural causes of homelessness, prior housing situation</td>
</tr>
<tr>
<td></td>
<td>Timeliness of data enabling more responsive services/policy and live needs mapping</td>
</tr>
<tr>
<td></td>
<td>Incorporation of other services into data system, e.g. Supporting People/Housing Support Grant funded services.</td>
</tr>
<tr>
<td>Increased capability to conduct research and evaluation leading to policy/commissioning change</td>
<td>Ability to conduct research &amp; analysis, leading to policy change, specifically:</td>
</tr>
<tr>
<td></td>
<td>• data linkage</td>
</tr>
<tr>
<td></td>
<td>• evaluation of what works</td>
</tr>
<tr>
<td></td>
<td>• understanding indicators of homelessness</td>
</tr>
<tr>
<td></td>
<td>• better outcome measures, specifically returns to homelessness</td>
</tr>
<tr>
<td></td>
<td>Develop a more nuanced picture of homelessness in Wales through sub-group analysis</td>
</tr>
<tr>
<td>Joined up working and service improvement</td>
<td>Improvement in experience of services through continuity of care and seamless referrals across the system—no one 'falling through the cracks'</td>
</tr>
<tr>
<td></td>
<td>Potential to generate collaboration across service providers</td>
</tr>
<tr>
<td></td>
<td>Customer experience may be standardised with processes and procedures brought into alignment</td>
</tr>
<tr>
<td></td>
<td>Ability to draw on data from other services in order to speed up assessments</td>
</tr>
<tr>
<td>Addressing power imbalance</td>
<td>Reconfigure how data are generated through person-centred data collection enabling service users to access their own case notes</td>
</tr>
<tr>
<td>Efficiencies and savings</td>
<td>Collect once, use multiple times is more cost effective</td>
</tr>
<tr>
<td></td>
<td>Cost savings, if software solutions provided by Welsh Government</td>
</tr>
<tr>
<td></td>
<td>Information gathering and data collected can be led by local need for certain system designs</td>
</tr>
<tr>
<td></td>
<td>Local authorities will potentially be able to utilise their existing systems of information gathering and data generation under certain designs</td>
</tr>
<tr>
<td></td>
<td>Upload of data without need to process could potentially save authorities time/resource</td>
</tr>
</tbody>
</table>

In addition to enabling improvements in data quality, the creation of a new data system could also provide the opportunity to revise what data are recorded about people’s experiences of homelessness. As part of the engagement activity, stakeholders were asked to indicate what improvements they would like to statutory statistics. Responses included more information on the circumstances of people experiencing homelessness, including where they have come from (e.g. geographically and in terms of tenure), their household and socio-economic status (e.g. income), and if they have experienced adverse childhood experiences. There was also an interest in greater detail on temporary and supported accommodation (both statutory and non-statutory).

**Increased capability to conduct research and evaluation**

Linked to improvements in data quality and types of data collected, there was a general consensus from across the various engagement activities that a new data system could increase the sectors’ capability to measure the experience of homelessness in Wales, and to conduct research and evaluation of policies and practice. Though research and evaluation were referenced as opportunities in their own right, they were also valued as leading to policy development and service improvements. Evaluations could be used to identify what works in ending and/or preventing homelessness, with this being very much in line with the evidence driven approach to ending homelessness advocated by the Centre for Homelessness Impact (2019) and the new homelessness strategy for Wales. Some of the proposed analyses/statistical opportunities of a new data source identified by the sector included: longitudinal analyses, measures of repeat homelessness, data linkage to devolved and non-devolved service data, sub-group analysis, and conducting an administrative data census of homelessness in Wales.
However, in order to conduct research and evaluations, data needs to be accessible and presented in several formats so that it can be used by sector actors with different levels of resource and specialisation in data analysis. Identifying what format of data will be of greatest use to the sector was considered when creating the system options in Part Six. Each design enables slightly different forms of access, some of which may be preferential, though all options have data linkage designed into them, as this was something strongly advocated by the sector.

**Joined up working and service improvement**

In addition to the possibility of ‘better’ data and therefore new research and evaluation capabilities leading to improved policymaking, the sector also identified opportunities for day-to-day service delivery through a new data system. Improving service user experience was one of the top opportunities of a design based around a pan-Wales shared data management solution. If multiple services were able to use the same data platform, stakeholders suggested that this could speed up assessments, as data from other services, such as healthcare, could be used when considering suitability of temporary accommodation and ‘risk’. Furthermore, where a household/person moved authorities, having a single platform accessible across multiple geographies could reduce the time needed to undertake an assessment in other authorities, as a person’s historic data would be recorded on the system and accessible to other agencies. However, the sector identified several challenges with this type of data sharing platform, which will be discussed in the following section.

**Addressing power imbalance**

In their options review for a new street homelessness data system in Scotland, the Centre for Homelessness Impact identified that people with experience of homelessness sometimes felt a sense of disempowerment when accessing services (Russel and Thomas 2019). Not only are service providers in a position of power in determining access, data generation can be an extractive one-way process over which the person seeking assistance may perceive they have little control. Data protection laws provide some means of redressing these power imbalances, for example the rights to access, rectify, or erase data\(^\text{14}\). Organisations have a legal duty to inform people about their data rights and how their data are used. However, there are more active ways in which data systems can be designed to foster a sense of agency. As part of the workshop’s group discussion, the opportunity for people to have access to notes about themselves was raised, with the suggestion that this could enable improvements in data quality as well as transparency. This approach to ‘co-creating’ data has been adopted by Greater Manchester Tackling Homelessness Information Network (GM-Think), for example, which encourages people accessing services to view their data along with their case worker. Whether the co-creation of data leads to improvements in data quality is uncertain, however, being able to view and control their own data could potentially make services more person-centred.

**Efficiencies and savings**

Workshop participants highlighted that under new reporting arrangements, where individual level statutory data are extracted and transmitted to Welsh Government, this could contribute to efficiency savings by collecting data once and using it multiple times. This approach to data fits within recent developments within Welsh Government through the formation of the Administrative Data Research Unit, whose remit is to increase the use of routinely collected data held by Welsh Government. Furthermore, a local authority respondent indicated that if data were uploaded, with no additional requirement to process the data beforehand—as is currently the case—then this could potentially reduce the burden of providing data. However, despite the sector identifying that efficiencies were possible, a Nesta study of data use in local authorities in England indicates that making an ‘investment to save’ business case may not be possible, and involves a certain amount of experimentation and ‘a leap of faith’ that the insights gained from new projects would be worth the investment (Symons 2016).

---

Challenges of a new system

The following discussion explores some of the main challenges highlighted by the sector, including additional resource implications, data protection, gaining and maintaining consensus, and investment in legacy systems. Misuse of data and potential harms form certain forms of collective data were also a challenge of a new system, though only within the context of data management software being used by multiple organisations. Finally, though improved data were seen as an opportunity, there were concerns raised about the processes required to generate this data, with these being grouped as a residual theme of ‘data related issues’.

Resource and practical implications

A pressing concern from the perspective of already stretched services was the challenge of resourcing a new data system, with this being raised for both new reporting arrangements and designs based around shared data management systems. In addition to cost implications of any new IT infrastructure, strain on already pressured staff was also cited as an area of concern. However, the cost/resources required in setting up and running a new system will vary drastically dependent on the system design chosen. Options that simply require new reporting requirements/arrangements, i.e. replacing aggregate returns with individual level, will necessitate some alteration of local data management systems if they cannot output the required data. However, Welsh Government could adopt a similar approach to MHCLG during the introduction of the H-CLIC system and provide grant funding for use by local authorities in order to upgrade their IT infrastructure. Alternatively, as is currently the case with the SHIN pilot study, and was the case with the Ffynnon software, Welsh Government could fund the provision of a software solution for use by multiple service providers.

Table 5: Summary of sector-identified challenges of a new homelessness data system

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Specific areas of concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource implications</td>
<td>Technical barriers &amp; resource requirements (including staffing, infrastructure, and funding)</td>
</tr>
<tr>
<td></td>
<td>Lack of relevant data management skills amongst some local authorities makes the task of preparing data resource intensive—additional burden under a new system may not be practical</td>
</tr>
<tr>
<td>Data protection and privacy</td>
<td>Data protection legislation—specifically the GDPR—and gaining consent for use of data</td>
</tr>
<tr>
<td></td>
<td>Determining who will be able to access data &amp; data security generally</td>
</tr>
<tr>
<td></td>
<td>Public trust and confidence in data use</td>
</tr>
<tr>
<td>Investment in legacy systems</td>
<td>Authorities have invested heavily in their own systems, therefore centralised systems represent wasted resource</td>
</tr>
<tr>
<td></td>
<td>Relatedly, there may be duplication of data on different systems, in cases where agencies continue to use their own systems</td>
</tr>
<tr>
<td>Harm to people experiencing homelessness</td>
<td>Misuse of data, for example the punitive use of information to withhold services</td>
</tr>
<tr>
<td></td>
<td>Permanency of records of service use can come to affect people after the fact and for many years after the encounter—‘black marks’ on people</td>
</tr>
<tr>
<td>Gaining &amp; maintaining consensus</td>
<td>Deciding on common data items to be collected</td>
</tr>
<tr>
<td></td>
<td>Determining who will be able to access data</td>
</tr>
<tr>
<td></td>
<td>Trying to meet conflicting stakeholder agendas and needs</td>
</tr>
<tr>
<td></td>
<td>Initial momentum building, and maintaining this momentum</td>
</tr>
<tr>
<td>Data related issues</td>
<td>Poor quality data in ‘live’ data sharing platforms</td>
</tr>
<tr>
<td></td>
<td>Responsibility for data quality monitoring and auditing</td>
</tr>
</tbody>
</table>
In addition to added resources implications from a change in systems, local authority stakeholders raised the concern that depending on the type of system, the practice of service provision could be altered by the need to collect different forms of data at specific points in the process of assisting people. For example, an inflexible data management system that has not been designed with the practice of service provision in mind will be burdensome on both the case worker and the person seeking assistance, e.g. being asked a series of unnecessary questions as ‘the system’ requires the case worker to do so. Alternatively, having some form of consistency in what is being asked and how it was asked was highlighted by one of the local authorities as a possible benefit of the system, leading to the standardisation of customer experience. The design of any new system should account for the work being done by those entering data into the system, and fit managerial and policy expectations around these data, i.e. take frontline staff as experts in what is feasible/ethical to record in a ‘usual’ interaction—rather than a top-down approach to design that forces service interactions to conform to certain measurement practices, possibly leading to data omissions, poor quality data when it is entered, and ultimately data that are not useful for management, policy, or practice.

Data protection and privacy
Across all forms of sector engagement, from presentations to surveys, it was rare for the prospect of a new data system to be discussed without mention of data protection, specifically the General Data Protection Regulation (GDPR). Data protection legislation is often cited and perceived as a barrier to organisations sharing data (Symons 2016), and this study was no different in this finding. However, the extent to which this perception is true in practice has been called into question. In a scoping exercise of public bodies and their data sharing practices, conducted by the Law Commission (2014), a lack of clarity about what was permissible under data protection legislation was a challenge when sharing data. Similarly, as identified by the Wales Audit Office (2017), local authority staff in Wales were often unclear on what data they could share, with whom, and for what reasons. Though there have been several efforts by the Information Commissioners Office to ‘de-mystify’ data protection, the perception that legislation prevents sharing persists. Any new data system should therefore incorporate training in data protection legislation conducted by a third party in order to maintain Welsh Government impartiality.

Investment in legacy systems
From the survey of local authority Housing Options teams, several respondents indicated they have made significant investment into their data management systems; a similar situation could also be imagined with other service providers who have purchased ‘off the shelf’ software. One local authority respondent reported that they had a system that dealt with housing registers, temporary accommodation, statutory homeless assessments, and other housing services; a change in the data collection system could therefore negatively impact on current data infrastructure investments if these had to be altered or replaced with another solution. There are additional resource implications if, for example, local authorities and services choose to use their bespoke data management systems but are also required to submit data to a pan-Wales data platform. Though Application programming interfaces (API’s) can be used to increase the interoperability of local bespoke systems and pan-Wales platforms, for example by automating data sharing with these platforms, unless there is a benefit to sharing data as part of the pan-Wales platform it would represent a wasted investment. For example, there could conceivably be a situation where only a small number of services utilise a pan-Wale platform directly, whilst the majority simply input data via API’s or periodically, such that the pan-Wales platform merely becomes a convoluted data collection mechanism.
Harm to people experiencing homelessness
Where 'live' data are shared with other organisations, the sector indicated that this may have the potential to result in harm to the person seeking assistance. Forms of harm that were identified by the sector included the misuse of data to withhold services from someone, and ‘black marks’ placed against individuals based on prior service interactions. Memorandum of understanding and service agreements can be used to outline 'permissible uses' of data systems and discourage potentially harmful uses. As is the case in many HMIS in the United States, having a full audit of requests to access case files can also reduce ‘misuse’ of data when making case decisions, and limit access to data on a need to know basis—with the requester having to stipulate why access to particular data are needed. Furthermore, through the use of consent and other forms of user involvement—see the previous discussion on addressing power imbalance—the potential uses of data that lead to harm to the person can be mitigated, i.e. service providers can only see information when granted permission by the person receiving assistance.

Gaining consensus
Several concerns raised by the sector fell under a common theme of consensus: either the sector indicated that an aspect of a new system would require consensus, or that a common purpose and drive was necessary to the uptake of the system. In terms of sustainability, consensus can lead to a shared sense of value being placed on the data, and therefore increase the sustainability of a new system (Lane 2018). The variables to be collected and who would have access to data would require consensus, with this being an issue affecting both new reporting arrangements and data management options for a new system. Competing interests and different key performance indicators were identified as challenges to consensus. If consensus on what data are collected is achieved, a further challenge identified was then the applicability of this data to others. For example, under new statutory homelessness reporting arrangements, consensus may be achieved amongst local authorities and Welsh Government on data to be collected, however this data should also be useful to other sector stakeholders—therefore there needs to be some outward focus on the wider use of data.

Data related issues
A final set of challenges with a new data system related specifically to data, including concerns around data quality and responsibility for monitoring. There were specific concerns about the low quality of 'live' data, i.e. data in a shared data management system. A possible solution to this issue, as is the case with CHAIN in Greater London, would be to make data quality monitoring and data management one of the functions of the provider of the data platform, e.g. merging records belonging to the same person and generating data quality reports for providers. However, data quality will always be the responsibility of the data provider who will ultimately have to correct any errors. Several systems in the international review attempted to encourage data providers to be more pro-active in maintaining their data quality, rather than waiting until data were required for upload/sharing with a central collection. Having a strong data governance function, the provision of data quality reports, and setting data quality targets were some of the mechanisms through which data standards could be monitored and maintained. Across all the systems in the international system review, having managerial and front-line buy-in of the importance of data meant that data quality improved as the relevance of the data being collected increased.

Summary
Having considered information gathered from the different engagement activities, the sector appears to be in favour of a new data system in Wales, in theory, and with some reservations. Stakeholders saw the value of having new data in order to drive policy and practice decisions but have understandable concerns about the resource implications and impacts on services that a new data system could create. The identified challenges will require involvement of the sector and people with experience of homelessness in order to generate solutions that are appropriate to the working practices and lives of the people ‘behind’ the data. The sector could identify few opportunities in maintaining the status quo: where they did, opportunities identified related primarily to the familiarity of the sector with the current statutory data approach and maintaining consistency with the past. A new data system brings with it the opportunity to give Wales a future-facing direction in addressing homelessness.
Part five: Learning from others

The following discussion synthesises findings from the review of international data systems, encompassing 51 systems from over 9 countries. Rather than attempt to create typologies of whole systems, this section’s discussion is structured around eight crosscutting elements that emerged as important design considerations, summarised in Table 6. Design elements cover different points in the flow of data, from its recording by organisations to its use to generate statistics and research outputs. Possible choices for each element are presented based on examples from the review e.g. how data governance has been approached in other systems. It is these design choices that make up the system designs presented in Part Six of the report. The conclusion to this part of the report takes the learnings from the review—the design elements—and applies them to the challenges/concerns raised by the sector in Part Four of the report. However, before discussing individual elements, the first part of this discussion compares several national approaches to homelessness administrative data systems in order to illustrate the breadth of approaches and their distinctive origins.

Table 6: Summary of design elements for a new data system and the options for addressing these elements identified from the review

<table>
<thead>
<tr>
<th>Design element</th>
<th>Options based on review</th>
</tr>
</thead>
<tbody>
<tr>
<td>System aim &amp; new ways of working</td>
<td>Collective Information, Research &amp; evaluation, Measuring the problem, and the sector’s response</td>
</tr>
<tr>
<td>Data architecture model</td>
<td>Centralised, bringing together data into a single dataset/system, e.g. data warehouses/lakes or integrated systems used by multiple agencies Federated, where data remain with data owner(s) and are brought together when required, e.g. an interlinked network or a hub-and-spoke model where all systems link to a single ‘data broker’</td>
</tr>
<tr>
<td>Supporting change (Procurement/financial aspects)</td>
<td>Grant funding, Provision of technologies</td>
</tr>
<tr>
<td>Governance</td>
<td>Delegated to a single governing body, whether internal or external to the system, Collective governance where each stakeholder is represented</td>
</tr>
<tr>
<td>Data quality</td>
<td>Data standardisation and having data standards, Monitoring by another body, (Semi)automated validation</td>
</tr>
<tr>
<td>Ethical &amp; legal</td>
<td>Consent to share and process personal/sensitive data from the person, Using legal means to share/process data, e.g. drawing on specific legislation as enablers, Sharing based on contractual obligation</td>
</tr>
<tr>
<td>Data sharing mechanisms</td>
<td>Processing prior to leaving data providing organisation (e.g. aggregation), De-identification of individual/case level data prior to sharing, Sharing personal information, with higher levels of data security, e.g. Trusted Third Party and split file processes</td>
</tr>
<tr>
<td>Data access &amp; accessibility</td>
<td>Portals, dashboards, and open data, Data downloads, Data labs (either submit own data or work within a ‘safe setting’)</td>
</tr>
</tbody>
</table>
International examples of homelessness administrative data systems

Table 7 summarises 7 national homelessness administrative data systems that focus on the collection of individual/household level data—though aggregate level data are reported to government in the United States. Administrative data are a by-product of day-to-day operations of organisations (Hand 2018), predominantly fulfilling the role of monitoring and performance measurement. The national approach to homelessness provision therefore influences what organisations and services operate, and therefore what administrative data is being generated. The United Kingdom is unique in the extent to which there is a legal right to housing assistance, with the administrative data being generated closely aligned to monitoring outputs under legislation, i.e. decisions on the legal status of households. Though other nations have legal rights to emergency accommodation, i.e. Denmark and Ireland, and others have legal rights that are infrequently enforced at a local level, e.g. France, in the international arena homelessness provision predominantly takes the form of state funding of homelessness services. The United States, Australia, Canada, Ireland and Denmark all fund and therefore monitor homelessness services. However, as funds can be provided to a range of services in United States, Australia, and Canada, data collections in these countries accommodate differences in their data providers, e.g. gathering data in low threshold services such as outreach where interactions may be brief alongside shelters, where access is naturally more structured around start and end dates. The ability to measure repeat homelessness in nations that collect data from different services is enhanced as it enables the re-appearance of individuals in other services to be identified. Pooling data also leads to the generation of ‘system wide’ performance measures that enable communities of services to assess how well they—as a community—are tackling homelessness; rather than focusing on how individual service types are performing.

Unsurprisingly, with its focus on decisions under legislation, Wales’ main national data collection (i.e. statutory data) is most like Scotland and England. However, there is still some variation between these three nations, with the greatest difference being between England/Wales and Scotland than between England and Wales, driven partly by Scotland’s removal of priority need which significantly simplifies its statutory approach—and therefore data. In comparison to Scotland’s single decision stage, Wales and England have three (prevention, relief, and full duty). However, England and Wales benefit from a formalised statutory prevention stage, with data collection on prevention falling under the main statutory system, whilst in Scotland prevention sits outside of the statutory process, with activities collected under the remit of prevention being gathered under a separate collection to statutory decisions—known as the PREVENT1 collection.

In 2018, England began a new system of individual statutory homelessness data collection—the Homelessness Case Level Information Collection (H-CLIC). Unique to H-CLIC, compared to the other national systems, is its requirement to submit personal identifiable data to government, with all other nations either choosing to anonymise e.g. via aggregation (United States), or de-identify data (Scotland and Denmark). However, the decision to transmit personal identifiable data has been problematic, with data sharing agreements between local authorities and government not being in place a year after the collection started (in April 2018), leading to delays in personal data transfer, and therefore undermining the ability of H-CLIC to link data between authorities. It should also be noted that the HL1, HL3 and PREVENT1 data collections on statutory homelessness and prevention work in Scotland only share de-identified data with Scottish Government; however, as part of a national data linkage study, Scottish Government undertook a time consuming exercise in gaining consent for and then organising the transmission of personal identifier data into a data linkage infrastructure.
Though all the national data collections will, to some extent, require data providing organisations to work in particular ways in order to generate individual level/de-duplicated data, there were several examples of governments/funders specifying the working practices of funded services. HUD in the United States sets rules around how communities in receipt of funding must structure and organise themselves in a specific way—that of ‘Continuum of Care’—whilst also mandating certain types of centralised data collection—through the formation of HMIS. These requirements have evolved over time in response to learning from best practice on how funded communities have handled their data. Along similar lines, Canada, under their Reaching Home strategy stipulate that services/communities form collective ways of work, and use a free data management software solution, the Homeless Individuals and Families Information System (HIFIS), to manage their data. Where a community chooses not to use the HIFIS software, they cannot use funds they receive to create a data system that would undertake the same functions as HIFIS—thereby ensuring that more communities adopt HIFIS and therefore collect data in a more standardised way.

Alongside administrative data collections, many countries also use other methodologies such as point-in-time counts, in order to generate insights into populations that may not be interacting with public services, and therefore not ‘visible’ within administrative data. In the United States, point-in-time counts of people sleeping rough are collected alongside data from HMIS systems in order to generate data on sheltered and unsheltered persons. Alternatively, in Denmark, shelter data are supplemented with a survey of rough sleepers, which is sent to a sample of rough sleeper service providers. Australia and Ireland do not have national rough sleeper counts, though the different states/regions do undertake their own rough sleeper count, e.g. the City of Melbourne conducts biannual street counts, whilst Dublin have a twice-yearly street count in Spring and Winter. Scotland are unusual in not collecting data on rough sleeping as a separate count exercise: instead, information on rough sleeping is gained from the statutory data collection (HL1), which asks whether a person has slept rough in the past 3 months or was sleeping rough the night prior to seeking assistance. It should be noted however that Scotland are contemplating a move to a separate collection on rough sleeping (Russell and Thomas 2019).

Having provided some insight into how other nations ‘do’ homeless data, the following discussion explores design elements that emerged from the review of all data system systems, with many of the data systems in the country comparison being returned to throughout the discussion.
### Table 7: Country profiles of national homelessness data systems

<table>
<thead>
<tr>
<th>Country</th>
<th>Homelessness provision</th>
<th>Administrative data collections/system(s)</th>
<th>Main data provider types</th>
<th>Level and sensitivity of data</th>
<th>Data provision</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scotland</td>
<td>Legal right to housing</td>
<td>HL1 (assessments) HL3 (temporary accommodation) PREVENT1 (prevention) also known as Housing Options in Scotland Various local authority data management systems ProcXed (as the data validation and transfer system)</td>
<td>Statutory services provided by local authorities, and non-statutory prevention work they undertake Some Registered Social Landlords can provide PREVENT1 data on a voluntary basis Persevering</td>
<td>Case level, with some individual level data on main applicants under the HL1/PREVENT1 De-identified through encryption of National Insurance Numbers (NINO) &amp; use of local authority case identifier</td>
<td>Mandatory collection of Housing Options (prevention) data, and from 2016 HL3 data on temporary accommodation was made mandatory</td>
<td>Scotland’s case level data collection has been running since 2001. However, it has undergone various changes over time in line with policy developments and data needs; namely the removal of priority need, and increasing prevention work. There are several individual collections covering statutory assessments, placements in temporary accommodation, and prevention work. Data providers are responsible for structuring their local data management systems, with data being extracted into standardised templates/formats. Data are then either emailed or uploaded to Scottish Government using a data upload and validation portal (ProcXed).</td>
</tr>
<tr>
<td>England</td>
<td>Legal right to housing</td>
<td>Homelessness Case Level Information Collection (H-CLIC) DELTA (as the data validation and transfer system) Various local authority data management systems</td>
<td>Statutory services provided by local authorities</td>
<td>Individual level Contains personal/sensitive data including unencrypted NINO, and personal data in separate file</td>
<td>Voluntary</td>
<td>In 2018 a new individual level data collection was introduced in England to coincide with changes in homelessness legislation. As with Scotland, local authorities are able to design their own local data management systems to suit their needs/capacities. The H-CLIC collection the name of the file format that individual level data must be structured before being uploaded Ministry of Housing, Communities &amp; Local Government (MHCLG) via the web-based upload and validation portal DELTA. For smaller authorities, with a lower caseload, data can be entered directly into DELTA; larger authorities upload files of structured data.</td>
</tr>
<tr>
<td>Country</td>
<td>Data provision</td>
<td>Main data provider types</td>
<td>Administrative data collections/systems</td>
<td>Level and sensitivity of data</td>
<td>Overview</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>Primarily through federal funding, including the Homeless Emergency Assistance and Rapid Transition to Housing (HEARTH) Act</td>
<td>Longitudinal Systems (LDS)</td>
<td><strong>Longitudinal Systems (LDS)</strong> to HUD; Organisations providing data to LDS include, among others, Safe havens, Supportive housing, Street outreach, and prevention activities.</td>
<td>Aggregate data provided to LDS; Individual level personal/sensitive data accessible via some HMIS systems to enable sharing of personal data needed to identify the same clients across services.</td>
<td>The HEARTH Act brought into law the requirement for communities receiving funding from HUD to form Continuum of Care (CoC) and placed certain requirements to form HMIS systems. The purpose of HMIS systems is to bring together data across a community, in order to generate de-duplicated aggregate counts of people being assisted by the CoC. HMIS systems are also used to assess performance. BHPS can be designed to fit each community's needs, and are usually overseen by a lead agency responsible for coordinating the HMIS system and maintaining data quality.</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>Primarily through funding of the National Housing and Homelessness Agreement (NHHA)</td>
<td>Specialist Homelessness Services Collection (SHSC)</td>
<td><strong>Specialist Homelessness Services Collection (SHSC)</strong> to AIHW, including Data Exchange 2.0 (the transfer system).</td>
<td>Individual level De-identified data provided to SHSC through the creation of a Statistical Linkage Key (SLK) based on personal data. SLK enables linkage across different data sources.</td>
<td>The SHSC is the name for the individual level data set relating to people assisted by services funded under the national funding agreement. The SHSC can be used to provide data on various aspects of homelessness, including the number of people assisted, the services they received, and their needs. SHSC data can be used for research and planning purposes, and is managed by the Australian Institute of Health and Welfare (AIHW). There are two mechanisms via which data can be recorded and reported: data can be recorded and reported directly to the SHSC, or through the SHIP, which is used by funded agencies to manage their own data. Data extracts provided directly to the SHSC can be used for research and planning purposes, and are managed by AIHW.</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Homelessness provision</td>
<td>Administrative data collections/system(s)</td>
<td>Main data provider types</td>
<td>Level and sensitivity of data</td>
<td>Data provision</td>
<td>Overview</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
<td>------------------------------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>----------</td>
</tr>
</tbody>
</table>
| Canada  | Primarily through funding of homelessness services under the Reaching Home strategy, which came into effect in 2019. | National Homelessness Information System (NHIS) Homeless Individuals and Families Information System (HIFIS) Some regions have their own bespoke Homeless Management Information Systems (e.g. Calgary) | Organisation eligible for funding under Reaching Home can include:  
- Housing services, e.g. shelters, transitional housing, permanent supportive housing, outreach  
- Prevention and shelter diversion services  
- Client support services, e.g. skills development  
- Treatment services | Individual level De-identified data provided via HIFIS system to NHIS HMIS systems and HIFIS when used for community level data systems would enable sharing of personal data needed to identify the same client across services | Mandatory for funded services | HIFIS is a data management/collection solution provided free of charge to organisations/communities funded by the state and has been developed and supported by Employment and Social Development Canada. HIFIS has evolved over time from a downloadable software, to a web-based platform, whose capabilities have also evolved in response to sector needs; for example, it can act as a local data management solution and can be used to pool data across a community (e.g. Saskatoon city). Data collected via HIFIS are periodically extracted and transmitted to government as part of the NHIS, the federal data monitoring initiative. However, some areas use their own bespoke homelessness data management systems (e.g. Calgary), with data being extracted from these and reported as part of the NHIS. Under the ‘Reaching Home’ strategy, communities will have to adopt HIFIS if they do not have a suitable HMIS system in place and cannot use funds to create their own version of HIFIS. |
<p>| Ireland | Legislation does not set out a statutory duty to assist but does empower local authorities to meet the needs of people experiencing homelessness, including working with NGOs to house people in an emergency. Funding of homelessness services is also made under legislation, for example Housing Assistance Payment (HAP) scheme, which is a social housing support payment. | Pathway Accommodation and Support System (PASS) | | | | In January 2011, PASS went live nationally across nine regions in Ireland, with the aim of providing a shared database of clients and accommodation. PASS automates contact between clients and services and is used to allocate emergency bed spaces. Within each region the lead local authorities for homelessness provide monthly reports on homelessness to the Department of Housing, Planning and Local Government, which identify the number of people utilising State-funded emergency accommodation on a regional and county basis. Dublin conducts a rough sleeper count exercise in both spring and winter to augment the administrative data on shelter users. |
| | | | | | | |</p>
<table>
<thead>
<tr>
<th>Country</th>
<th>Homelessness provision</th>
<th>Administrative data collections/system(s)</th>
<th>Main data provider types</th>
<th>Level and sensitivity of data</th>
<th>Data provision</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>Homelessness interventions are largely integrated into social services provision, with Section 110 in the Social Assistance Act requiring municipalities to provide temporary accommodation for people experiencing homelessness; however, there is no ‘right to housing’, outside of emergency provision.</td>
<td>Client registration system database</td>
<td>Shelters funded under Section 110 in the Social Security Act</td>
<td>Individual level When enrolling in a shelter, the individual must register their unique Central Personal Register (CPR) number. Through the CPR, exact matching can be used to records within the shelter data and also other public services data.</td>
<td>Mandatory</td>
<td>Data are collected by Statistics Denmark in order to monitor state provision of temporary accommodation. With data being the responsibility of the statistical agency, which also has a well-developed data linkage framework, the scope of the data collection is reduced as information could be obtained through data linkage to other sources held by Statistics Denmark. In addition to the shelter data, a national street count is undertaken roughly every 2 years, and involves surveys sent to local authorities and social services, who then fill out questionnaires about each individual homeless person they work with/encounter.</td>
</tr>
</tbody>
</table>
System aim & new ways of working

Having a clearly articulated aim is an important first step in the design of any data system. Data protection legislation measures the proportionality of data sharing and processing against these aims, i.e. that the volume of data and type of processing are appropriate and warranted. Furthermore, being able to clearly articulate the purpose of an innovation—in this case a new data system—and how it differs from other ways of working can help in generating a sense of ‘coherence’ or shared understanding, which itself is the first stage of embedding its use in practice (Murray et al. 2010). A data system that replicates current data management practices and has little relevance to front-line staff will have difficulties in sustaining its consistent use. From the review of systems, and the wider literature, three very broad uses of administrative data can be defined, relating to measurement, research, and integrated operational purposes (Kumar 2015).

Starting with the latter of these, measurement is mainly related to the activities of counting the homeless and the actions of services, often within the context of performance measurement. The current statutory data collections in each of the devolved UK countries fall primarily into this purpose: their aim is to measure homelessness and local authority decisions under housing legislation. However, H-CLIC in England is making some headway towards routinising research use through research and evaluation of the Homelessness Education Act. The Ministry of Housing, Communities & Local Government (MHCLG), who oversee the H-CLIC collection, have expressed an interest in using the new individual level data to conduct data linkage studies to evaluate the wider issues and impacts of statutory provision (MHCLG 2018).

Though there were other examples of administrative data mainly for measurement being used for research, e.g. Denmark’s shelter data (Benjaminsen 2016a, 2016b; Benjaminsen and Andrade 2015; Nielson et al. 2011), from the international literature the Virginia Longitudinal Data System (VLDS) is an example of a data system intended purely for research and evaluative purpose. The VLDS brings together education data across the breadth of schooling, enabling learner pathways to be explored; researchers can apply to access these data. The aim of the VLDS is to improve student outcomes through enabling research of linked education data. The distinction between research and measurement is therefore a difference between describing homelessness and related activities on the one hand and understanding the mechanisms and causal impacts on the other.

A final purpose for administrative data identified in the reviewed systems related to their direct use in determining the day-to-day operations of services through data sharing and integration. The intention of this data integration is to extend the information available to organisations beyond their own data, and therefore entails combining administrative data from multiple organisations/policy areas. For example, the North West London Whole Systems Integrated Care data warehouse brings together administrative health data—inpatient, outpatient, and primary care—at an individual level and visualises this data in order to reconstruct a person’s prior use of health services. There were however examples of platforms purely for the purposes of sharing data. The Alert system in the UK provides a mechanism for public services to make referrals to local authority housing teams; whilst school2school in Wales enables the transfer of pupil data between local authorities/schools whenever a child moves school. However, platforms that bring together data mainly enhance multidisciplinary working within a service or sector, rather than being the cause of it. As the example of single assessment systems in adult social care in England illustrates, there still needs to be cultural change within a sector to break down structural divisions for collaborative practice and shared data to have impact (Abendstern et al. 2008).
A new homelessness data system in Wales should conceivably address all these different aims in order to improve early prevention and the effectiveness of homelessness services. Table 8 summarises some of the potential new ways of working that a new data system could open, under each aim, and links these ways of working to the homelessness strategy for Wales’ rare, brief, and un-repeated, and person-centred services. Measuring the issues and impacts of responses to homelessness is important in making services accountable to both the public and service commissioners (Culhane et al. 1997), as well as providing insight into the issues faced by people experiencing homelessness, which can then help guide service delivery. Research and evaluation have an important role in identifying what works in ending/relieving homelessness, as well as undertaking population level analysis to create measures of risk and therefore prevalence of homelessness. Finally, collective information directly impacts service delivery and decision-making by front line service providers, either directly or through streamlining collaboration and referral.

Table 8: What a new data system for Wales could enable under the three main purposes of a data system (observing/testing/doing) and how this contributes to ending homelessness in Wales.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Capabilities</th>
<th>Strategic contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Prevention</td>
</tr>
<tr>
<td>Integrated information</td>
<td>Sharing of data between organisations</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Case management based on historical service data</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Streamlined referrals and single point of entry into the housing system</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Identification of people ‘at risk’ of homelessness based on shared data—thereby enabling early intervention</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Not having to recount prior service interactions</td>
<td>x</td>
</tr>
<tr>
<td>Research &amp; Evaluation</td>
<td>National data coverage enabling research to identify population segments at-risk and predictors of homelessness</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Evaluation of policies/programmes &amp; the Act</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Evaluation of individual services in order to determine what works</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Provide a sampling frame for research, leading to national studies of homelessness and user voice surveys15</td>
<td></td>
</tr>
<tr>
<td>Measuring the issues</td>
<td>New indicators of the prevalence of homelessness in population</td>
<td>x</td>
</tr>
<tr>
<td>&amp; the response</td>
<td>Measures of repeat homelessness service use, or ‘re-entry’ into homelessness</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Measures of pathways through homelessness &amp; the homeless service sector</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Improved demographics on people experiencing homelessness, including needs profiles</td>
<td>x</td>
</tr>
</tbody>
</table>

15 As an example of research with homeless people, the Journeys Home survey in Australia was a longitudinal study of 1,700 people who were either homeless or at risk of homelessness, and were identified via Centrelink, the social security payment department of the Australian Government. See the following link for more details on the Journeys Home survey: [https://melboumainstitute.unimelb.edu.au/journeys_home](https://melboumainstitute.unimelb.edu.au/journeys_home)
Data architecture model

The term ‘data architecture model’ is used here to refer to the structure of a data system, broadly classified as either centralised or federated architecture—see Table 9 for an overview of the different architecture and their sub-types. In a centralised model, data exist in a central data repository, which is created either by: physically extracting data from one system and depositing it in another (e.g. data warehouses such as the Kent Integrated Data set), or by simply extracting and providing that extract to another organisation (e.g. the current reporting practices for national homelessness data in England and Australia), or several different organisations interface with the same central repository (e.g. integrated systems such as CHAIN or the SHIN pilot). Most of the systems reviewed adopted this centralised approach, given that it generates large centralised data sets/sources that can then be easily accessed/queried. However, there were a limited number of systems adopting an alternative architecture, ‘federated’ approach, mainly in the United States.

Within a federated data system, data largely remain the (legal) responsibility of the organisation that collects it. Upon request, data are automatically extracted, or ‘pulled’, from local systems and can be combined with other data sources. A single organisation, or ‘data broker’ usually manages data requests from organisation, and is responsible for requesting data from data owners, and creating the datasets made available to the requester. However, data are for single use only, i.e. for the use by the requester, and as such, data within a federated model is not stored outside of the participating organisations’ systems in a permanent repository. The X-tee system in Estonia is an interesting example of a completely automated federated system that enables ‘live’ querying of other agency databases—and forms the backbone of Estonia’s ‘e-Government’.

The federated model has been adopted by certain data systems in the United States that have been created to enable research using education data across the education pathway—known as ‘P-20W state longitudinal data system’. An example of such a federated model is the Virginia Longitudinal Data System, which enables research access to de-identified data. As State legislation on processing personal/sensitive data varies the federated model has been adopted to enable research access and linkage between multiple data sources, without exchanging personal data or processing data outside of its original host organisation, thereby working within the confines of local data legislation.
Table 9: Summary of data models with examples from the review

<table>
<thead>
<tr>
<th>Model</th>
<th>Sub types</th>
<th>Examples</th>
<th>Benefits</th>
<th>Limitations¹⁶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centralised:</td>
<td>Data are ‘pushed’ to a central location, forming a permanent data ‘pool’</td>
<td>Organisational data: Data are reported to a central organisation (e.g. government or state) who then merge data to create a data set. Data are quality assured ‘in transit’ between provider and central organisation</td>
<td>Data are historical and are therefore unaffected by changes in the operational data Approach is commonly used and organisations are familiar with the concept Data quality of the centralised data can be monitored and standardised Having a single data set/source reduces access time</td>
<td>Centralised source is a single point of weakness for potential data breaches Timeliness of data where data are uploaded to warehouses or to another organisation, however, automated data transfers can be used for nightly uploads</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H-CLIC Specialist Homelessness Services Collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kent Integrated Data set Michigan’s Statewide Homelessness Data online Warehouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined Homelessness Information Network Homeless Individuals and Families Information System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warehouse:</td>
<td>Data are pooled together in a specifically designed IT infrastructure that automates the processing of incoming data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management system:</td>
<td>Data are recorded/accessed simultaneously by different users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Federated:</td>
<td>Data are ‘pulled’ from data collecting organisations on demand, and are for single use only</td>
<td>Live federation: Organisations are able to query one another’s databases in-real-time</td>
<td>Queries relate to live data therefore can help with urgent time sensitive issues Data remain the legal responsibility of the data provider, therefore reduces the need to reach consensus on who can view what Technologically, more easily scalable as simply requires the creation of a new interface to the federation Federation works with available data, rather than trying to harmonise across numerous systems</td>
<td>Technical fault with data management solution, and or data broker can lead to system outage, i.e. single point of technological failure Data quality the responsibility of the data provider, and managing different quality standards may become resource intensive Less commonly used approach; novelty may to a concern to some organisations Data are for single use and affected by changes in organisational data—no two queries may return exactly the same results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X-tee (Estonia’s e-government platform)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Virginia Longitudinal Data System North Carolina School Works</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹⁶ Adapted partly from National Centre for Education Statistics (2012)
Supporting system change (Procurement/funding arrangements)

Not all new data systems will require changes in local data management solutions, for example where a change in the level at which data are reported may still be compatible with how local authorities store data. However, where change is required there were found to be two general approaches to achieving a system change either through grant funding or the provision of technology.

The approach adopted by England in the introduction of the H-CLIC collection was to provide funding to local authorities that would enable them to upgrade their local data management systems in order to store and output the required individual level data. Roughly £9,200 was made available to each authority, however it can be assumed that not all authorities needed to make use of this fund, and that some authorities needed significantly more. In the United States, state-wide educational research platforms can apply for funds under the Statewide Longitudinal Data Systems (SLDS) Grant Program, provided by the National Center for Education Statistics (NCES). Grants are very large; with the smallest grants being on the order of several million dollars. In addition to the SLDS grant, states also fund their longitudinal systems through local grants and contributions from partners. Grant funding to enable system change can enable local data management solutions to be procured that fit community needs—in the case of the SLDS in the United States, to contend with highly restrictive state data sharing laws. However, as was the case in England, services may tend to gravitate towards standard software solutions on the market, such as Housing Jigsaw, negating the element of choice. Furthermore, as services/communities can design their systems in isolation, this can lead inconsistent practice, particularly when limited guidance is given to organisations on what the end ‘product’ or data outputs should be.

An alternative approach to organisations/communities procuring their own data systems is for there to be a standard data management solution made available free of use. The Homeless Individuals and Families Information System (HIFIS) in Canada, the Specialist Homelessness Information Platform (SHIP) in Australia, and Alert client referral system in the United Kingdom are examples of the provisioning of software that enable data sharing/management. Smaller organisations without their own data management solution are likely to benefit the most from this approach. However, if use of provisioned software is a requirement, then this may represent a loss of functionality for more ‘data mature’ organisations/communities who have their own approach—as was implied by authorities in Wales during engagement activities. A compromise to provisioning and mandating software solutions is to create a functional but cost-effective data management system, and include clauses in any funding that if an organisation should choose to use their own systems, grant funding cannot be used to supplement these systems—this being the case in Canada, where communities either use HIFIS or their own data solutions.

Data Governance

Data governance refers to the management of different aspects of data creation and use within an organisation, including its quality, availability, retention, and security. Data governance is usually overseen by a governing body, composed of several sub-committees with overview of different aspects of data processes. From the review of systems, data governance committees were found to be composed in two main ways: either through a collective/consensual model or through delegation. In addition to these two basic structures, who is a member of this governing body can also vary, and include academic institutions, government departments, or sector service providers. Taking the first of these data governance models, collective governance relates to situations where several different stakeholder organisations are part of the data governance framework for a data system. The P-20W state longitudinal education data systems included in the international review used collective governance models, whereby different government departments providing data to the systems participate in the governing body, i.e. education departments and those related to welfare and work.

---

17 See the Statewide Longitudinal Data Systems Research homepage for more details: [http://slds.rhaskell.org/](http://slds.rhaskell.org/)
Collective governance enables all data providers and stakeholders in a data system to have equal say in determining how their data should be collected and used. For example, the Virginia Longitudinal Data System in America uses a voting system to approve any system changes. However, the limitation of this approach to data governance is that where a system includes many participating bodies, then the size of the governing body can become impractical when trying to meet the competing needs of different sector actors. Furthermore, collective governance can prevent timely access to data if agreement to use data is required from each contributing organisation—this is the case for the Looked After Children data collection in Wales, where permission to use data held by Welsh Government needs to be sought from each of the 22 local authorities.

An alternative approach to collective data governance is through delegation of authority to another body. The HL1 and PREVENT data in Scotland could be said to have a delegated structure in that the Scottish Executive are able to provide access to anonymised data for research purposes, without having to contact all 32 local authorities in Scotland. The benefit of delegation is therefore the gains in efficiency and time if one party, representing others, makes decisions about data. The downside of delegation is potentially a lack of control over data, where organisations join a system at a later date and are required to submit to governance structures they were not a part of agreeing.

Data quality

Maintaining good quality data is key to accurate evidence-based decision-making, particularly if the aspiration of a new case/individual level homelessness data collection is to form the basis for national statistics on statutory homelessness in Wales. In general, data quality within some local authorities in Wales is an on-going issue, as evidenced by the Wales Audit Office (2018) in their report on data use and in responses to engagement activities as part of the current project. From the review of systems, there were various mechanisms that were employed to improve quality, including: data standardisation and data standards, monitoring data quality (either by a lead organisation or an external body), and (semi)automated validation. Rather than either/or, all these strategies could be employed to ensure high quality and consistent data. The collection processes underpinning the statutory homelessness data in Wales are currently not consistent to any great extent, due in part to the lack of a national data specification for the collection processes of the data. Furthermore, though individual level Supporting People data are collected via a standard data template, the data are not reported on as a ‘national statistic’ and are therefore do not have to be subject to the same rigorous assessments as the statutory data. Taking the example of HMIS in the United States, standardisation of data could involve all services who have a homelessness remit collecting the same basic details regarding the people they assist. Standardisation streamlines the process of combining data from different providers, and therefore reduces the time required to clean and analyse data. As standardisation requires organisations to record data in the same way, this necessitates a set of data standards being agreed upon—a standard collection is proposed in Part Six.

However, creation of data standards is a passive approach to ensuring data quality. The onus is placed on organisations to adhere to those standards, which may wane over time and create issues when data are combined and errors in data have already occurred. A more active approach to data quality is through monitoring of data standards. For example, the provision of reports on data quality can help highlight data issues prior to any reporting deadlines, providing time to correct these. Having regular reports can also illustrate areas where data quality is consistently low, and therefore areas of a data system that require further examination—as they may be inappropriately designed—and potential areas where further training with data entering staff needs to be undertaken.
Across all the data systems reviewed there were varying levels of automated validation. In the case of H-CLIC, when individual level data are uploaded to the DELTA system, data undergo a series of comprehensive checks, ranging from formatting of input data to ensuring that data are consistent within the person. In the Kent Integrated Dataset, data being deposited within the data warehouse undergo a fairly simple set of validation checks that are automated to the extent that aggregate statistics are generated from the deposited data and compared (presumably manually) to published data to determine if any data errors have occurred. The current software used to upload the aggregate statutory homelessness data to Welsh Government (Afon) has the capacity to handle individual/case level data and undertake validation of that data, e.g. SSDA903 collection/Looked after children data.

Ethical and legal

The biggest perceived barrier to collecting and sharing individual level data relates to data protection legislation. However, the sheer range of data systems that exist in the UK—and internationally in far more ‘data restrictive’ environments—demonstrates that it is possible to work with individual level administrative data in a legal and ethical way. Drawing on the review of systems and data protection legislation, three main ‘gateways’ for the sharing and processing of individual level personal data exist: (1) consent, (2) through legislation, and (3) obligation. Where data were being used in an integrated way or was shared via a common platform, consent for use of data was often requested e.g. CHAIN in Greater London and Online Navigation and Entry System in Chicago. However, within the UK (and EU) data protection context, consent is valid when freely given, which roughly equates to agreement to use of data without fear of repercussion or coercion. It has therefore been argued by the Information Commissioners Office (ICO) that public authorities should avoid consent where other means can be used, given that people using public services may feel coerced to provide data in order to gain ‘better’ services. A counter argument to this would be that although the validity of consent may be questioned in certain contexts, it is still an important ethical practice in engaging people in how their data are being used and addressing power imbalances in service provision. However, where consent is asked, use of other legal means to continue to use data undermines the practice of gaining consent.

For consent to be meaningful there must also be the ability for people to withdraw consent, which can undermine the quality of data and cause additional burden on data governance structures in having to accommodate procedures and processes for assessing the adequacy of consent and removing those who withdraw their consent. A nuanced approach to consent could address some of the conflicts between requiring data for statistical and research purposes and wanting to empower people using services. For example, HMIS systems in the United States offer people using services the ability to hide their data from other users of integrated data platforms, to differing extents, whilst having a ‘core data’ element that is used for reporting requirements.

Aside from consent, the other main gateway to enable use of data was through legal gateways, particularly in instances where the purpose of a data system was statistical and/or research related. As an example, the Data Protection Act (2018) has provisions for data sharing to undertake a ‘public task’, defined the processing of personal data necessary for:

- the administration of justice;
- parliamentary functions;
- statutory functions;
- governmental functions; or
- activities that support or promote democratic engagement.

---

18 This argument has been laid out in the guidance document for the H-CLIC system, see Data Protection Guidance document: https://gss.civilservice.gov.uk/user-facing-pages/mhclg-homelessness-statistics-user-forum/
However, when drawing on legislation to legitimate the sharing and processing of personal information, there are still obligations to the data subject in making the processing fair and transparent, and informing them that their data are being used in certain ways. In the case of the H-CLIC system, privacy notices were issued in order to meet obligations to be transparent in how data are used. The ICO advised MHCLG to use a layered approach to issuing privacy notices through posters in public places or an information leaflet, and electronically placing the notice on the authority website.

A final mechanism through which data can be shared is through obligation, usually to a funder. An example of such an obligation would be the Supporting People fund in Wales, where funding recipients are required to submit data on people they assist as a term of receiving the fund. Similarly, though potentially not contractual in a strict legal sense, grantees in receipt of United States Department of Housing and Urban Development (HUD) funds are required to provide HUD with unduplicated counts of people assisted by funded projects, along with a series of performance outcome metrics. However, funding of services in the United States falls under specific legislation—the HEARTH Act—with the provision of data being a part of this legislation, therefore falling under contractual and legal gateways.

**Data sharing and privacy preservation**

In order to share data, many data collection systems both nationally and internationally require data sharing agreements in place between data providers and processors. In both the English and Scottish systems, data sharing agreements have been set up in order to enable the transfer of personal data. In Scotland, agreements were single use enabling the transfer of data for Scottish Government’s data linkage project. In England, a series of data sharing agreements are being set up between MHCLG and local authorities enabling them to share personal identifiers under H-CLIC—the scope of these agreements will presumably cover research uses, as one of the intended applications for the individual level data is for data linkage research to explore the effects of the Homelessness Reduction Act.

Some of the approaches to ensuring the security of data whilst sharing data were to either (1) for the data provider to process their data, i.e. aggregation, (2) anonymise or de-identify data before sharing, or (3) to share personal data with added measures to ensure that disclosure risks were minimised. Rather than any one of these approaches being 'better', measures taken to ensure the privacy and security of data vary depending on the purpose of the system. For example, where data are part of a data-sharing platforms used for operational purposes, then this necessitates that people are identifiable given that the purpose of such a platform is to use the data to make decisions about the case. Where the purpose of a data system is mainly statistical, as is the case with the current data collection on statutory homelessness in Wales, then personal identifiers are not necessary, and therefore data can be aggregated or de-identified prior to leaving the data providing organisation. In cases where data were for statistical or research purposes, the capacity to link data was the deciding factor in what method of privacy protection was used. For example, where data were to be used for longitudinal forms of analysis, where knowing the same person over multiple years was integral, then de-identification of data renders data effectively anonymous yet linkable.

A possible mechanism for sharing data without disclosing people’s identities is through de-identifying data prior to it leaving the data providers’ system. De-identification means that data are effectively anonymised, whilst retaining the ability to link together data relating to the same person—whereas anonymisation would render data un-linkable. The same person receives the same unique number throughout the data source, and preferentially across data sources. There are various examples of how data can be de-identified, such as: creating a unique number based on personal data (e.g. Statistical Linkage Key in the Specialist Homelessness Service Collection (SHSC)/Specialist Homelessness Services National Minimum Data Set in Australia); assigning the person a unique number at random and retaining lookup tables to enable the same person to be assigned the same number if they re-enter a service (Local Authority Child Identifiers used in the Looked After Children Census in Wales); and ‘hashing/encrypting personal identifiers, or already existing unique identifiers (e.g. National Insurance numbers), by converting data into strings of characters—encryption offers the ability to de-encrypt once the data reach the data processor/requester.
Scotland's statutory homelessness collection uses National Insurance Numbers that have been hashed, and therefore rendered de-identified, in order to attempt to link people seeking assistance in different authorities. However, an issue with using hashing is that slight differences in the underlying data can lead to data linkage errors; for example, a single incorrect number in a National Insurance Number will lead to entirely different hashed identifier, and therefore a missed match between two records that belong to the same person. In addition to hashed National Insurance Numbers, local authorities in Scotland are required to provide individuals seeking assistance with an identifier number that is unique to that person in that authority, thereby enabling the same person to be identified if they appear within the same authority's data. A similar approach is used in the case of child social care data in Wales: each child has a unique identifier for that child in that authority. Multiple methods to link between records are often adopted (Harron et al. 2017), in order to maximise the future use of data.

De-identification at the source overcomes the risk-averseness of organisations around sharing personal data. However, unless the same method of de-identification is shared with all other data providers participating in a project, the de-identified data cannot be linked to other individual level data. However, in each of the GB regions there are options available for national data linkage that enable de-identification in a consistent way, many of which adopt a ‘split file’ method. In a split file approach, personal identifiers (name, date of birth, gender, and postcode) are split from data relating to service histories, or ‘payload data’. A trusted third party is in receipt of the personal identifiers, who then de-identifies the data by linking to a ‘population spine’—being a population level set of unique identifiers for each person. For example, the SAIL databank uses the NHS Wales Informatics Service as the trusted party who themselves use the Welsh Demographic Service as their population spine, which contains NHS Wales National Health Numbers (Jones et al. 2016); Scotland use the National Records of Scotland Indexing Service and adopt Community Health Index (CHI) as their population spine, which takes the CHI number as the unique number (Waugh et al. 2018). Once the data subject has been linked to the population spine using the personal identifiers, their unique identifier is extracted and encrypted, with this encrypted number enabling consistent data linkage across multiple sources. Unfortunately, this method of de-identifying data does not overcome the initial hurdle of organisations being risk averse when sharing personal data, as the personal data still need to be shared with the trusted third party.

A final approach to sharing personal data is for the recipient organisation, for example the data processor, to assume some legal responsibility for that data, and to ensure that there are practices in place to ensure the security of the data. The H-CLIC system adopts this approach, with MHCLG being in receipt of personal identifiers so that the data can be de-identified in house. In order to reduce risks that data are disclosed and can be linked back to a living person if data are mistakenly or maliciously intercepted in transit, H-CLIC also adopts a split file process in order to separate the personal data from the payload data, i.e. there are separate files of personal details and anonymised H-CLIC data sent at separate times.
Data accessibility and access

For the generation of data to have value, the information based on the data needs to be made accessible as information to a wide range of audiences with varying levels of data literacy, whilst there also need to be means via which data can be accessed. One common finding across reviewed systems was that the use of visualisation significantly improved the accessibility of data: and this held true across data at different units of analysis, i.e. individual level up to aggregate data. There were several examples of area level dashboards relating to homelessness in the United States, drawing on individual level data being collected as part of regional HMIS systems, e.g. the community dashboards associated with the Knoxville HMIS\(^\text{19}\) and Los Angeles Homeless Services Authority\(^\text{20}\). When directly linked to local data systems, well designed dashboards can provide easily interpretable ‘live’ insight into homelessness in an area, which can enable communities to track the impacts of their activities on, for example, returns to homelessness. Along similar lines is the Stella P software provided by HUD in the United States, which enables communities to input data from their HMIS systems, which is then automatically analysed to generate data visualisations of stock and flow through their homeless systems. Operating at a different level were the North West London Whole Systems Integrated Care dashboards that visualise information about individuals in order to reconstruct their interactions with health and social care services; visualisation can then be used by people involved in a persons’ care to make future decisions. In addition to increasing the accessibility of data by presenting pre-analysed ‘bits’ of information, automating analysis and outputs can also shorten the feedback loop between data, information, and action, with this being the driving force behind many campaigns to end homelessness, such as Community Solutions and the Build for Zero campaigns in the United States.

For sensitive data where the unit of presentation was at the individual or household level, modes of accessing data were understandably more constrained, with only de-identified or completely anonymised data being made available in most cases. This ‘granular’ data could either be downloaded, e.g. Virginia Longitudinal Data System, or, more often, were accessible within a secure data environment, e.g. New Zealand Integrated Data Infrastructure. The benefits of enabling access to data within a secure environment are that use of the data can be controlled, for example by having in place statistical disclosure controls where outputs must be checked by the data owner/controller before being released to the analyst. Secure data environments can range from physically secure settings, such as the Virtual Microdata Laboratory in operation by the ONS for access to sensitive data such as the Census; alternatively, a secure environment could be a virtual workspace within which all research is conducted that is accessed remotely, such as the Secure Anonymised Information Linkage databank ‘Gateway’. Data that can only be accessed from a physical setting may act as a barrier to use as it requires travel to the safe setting, which might preclude non-funded researchers, i.e. use of data by third sector organisations wanting to evaluate their services; remotely accessible secure environments therefore reduce this access barrier.

Meta-data are essential for ensuring that ‘granular’ data are accessible to researchers and analysts. Metadata are data about data: when properly constructed, they should outline the variables contained in a data set, the values the variables take, and provide a summary of how a data set should be structured for use. However, despite the importance of meta-data in making sense of data, there is a widely acknowledged lack of meta-data which can be a barrier to its access (van Panhuis et al. 2014), significantly frustrating any use of the data in a timely manner. In addition to making research and analysis a difficult task, a lack of meta-data can also complicate sharing of data between organisations—which speaks back to the need for data standards (themselves a form of meta-data), to build a common ‘data language’.

---

\(^{19}\) Example of the Knoxville HMIS data dashboard can be found at: [http://knoxxhmis.sworpswebapp.sworps.utk.edu/dashboard/](http://knoxxhmis.sworpswebapp.sworps.utk.edu/dashboard/)

\(^{20}\) Example of the Los Angeles Homeless Services Authority data dashboard can be found at: [https://www.lahsa.org/data](https://www.lahsa.org/data)
To close this section, the concept of ‘data labs’ offer an interesting example of making data accessible to service providers who may not have the capacity or capability to conduct primary research with individual level data. Though the definition of a ‘data lab’ is used to describe a number of data access options, the term is used here to refer to an organisation that works with other people’s data, and outputs analysis for them—rather than simply offering virtual or real secure environment where the organisation/person accesses and conduct analysis themselves (e.g. In-from DataLab in the UK). The Ministry of Justice DataLab (MOJ DataLab)\(^1\), for example, enables third sector organisations the opportunity to explore the possible associations between their services on people’s recidivism—measured as reconviction rates. Organisations submit the personal identifiers (i.e. name, date of birth, address) of the individuals taking part in the programme/services they wish to explore to the MOJ DataLab. The Lab then links those individuals to their criminal justice data—primarily prison data, but also police data—and generate a comparison group of similar characteristics. The Lab output a set of standard reports that compare the recidivism rates between the people receiving the programme/service, and the comparison group, and return this report to the third sector organisation.

The benefits of opening data are that it increases the accountability of services, by making data everyone’s business. Enabling services to link their data to a national data resource could also potentially improve the homeless sector in Wales’ capability to undertake evaluations and therefore identify what works in ending/alleviating homelessness, with increased evaluation being one of the activities called for under the new homelessness strategy in Wales. Collectively, better evaluation has the potential to improve the commissioning of services through evidence-based decision-making. Finally, opening access to data also represents a ‘free’ research resource: data can be used to undertake research without the need to commission the research in the first place.

Summary

This part of the report has primarily dealt with the practical aspects of a new system. Table 10 combines the areas of design identified in the proceeding discussion with learnings from the wider literature on system design, in order to directly address the challenges of a new homelessness data system highlighted by the sector. The aim of Table 10 is to illustrate potential ways through which the challenges of a new system can be addressed, at least in part, through design. In the following part of this report (Part 6), the different elements of design are combined in order to construct options for a new data system in Wales that address the needs for improved data and smarter use of data, and which acknowledges the sectors’ concerns/challenges raised across the various engagement activities.

---

\(^1\) For more information on how organisations access the Ministry of Justice DataLab, and previous analyses conducted by the lab, see https://www.gov.uk/government/publications/justice-data-lab
Table 10: Summary of how design choices could potentially be used to address sector concerns/challenges raised

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Possible mechanism for redress</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resource implications</strong></td>
<td>In order to mitigate the financial implications of any system changes, grant funding could be made available to improve local IT infrastructure (e.g. H-CLIC in England and HMIS systems in the United States), or the provision of free software licences to a standard software solution (e.g. HIFIS in Canada). Streamlining data flows through automation and centralisations could potentially reduce the burden of collecting/reporting more data on already overburdened front-line staff.</td>
</tr>
<tr>
<td><strong>Data protection and privacy</strong></td>
<td>Consent, where data are to be shared with others for service delivery purposes (e.g. Online Navigation and Entry System), whilst legal gateways can be used for sharing for research and statistical purposes (e.g. H-CLIC). Notices can indicate which parts of the data being provided are subject to which legal gateway, i.e. client notes subject to consent and basic ‘core data’ subject to legal gateways for research and statistical purposes. Data sharing agreements will need to be in place, regardless of which legal gateway is used.</td>
</tr>
<tr>
<td><strong>Investment in legacy systems</strong></td>
<td>Use of API technologies and/or a federated model of data sharing (e.g. North Carolina School Works) would enable local authorities and other service providers to retain their legacy systems whilst still enabling access to their data. Interfacing technologies were used in a situation where a pan-Wales data management system were adopt, if a large enough number of local authorities/services choose to retain their current systems, the cost of a software solution may outweigh actual use. Adequate provision of grant funding may also mitigate the lost investment when moving to another system—however highly bespoke systems may be difficult to replace. Alternatively, basing a system around data centralisation, such as a data warehouse where data are transformed when in the warehouse to create a 'standard’ data set, could allow the sector to retain their legacy systems.</td>
</tr>
<tr>
<td><strong>Harm to people experiencing homelessness</strong></td>
<td>Harm due to data breach is most likely to occur when data are in transit. Data encrypted prior to exiting the local data system reduces harm caused by privacy breaches, or, alternatively, adoption of split file mechanism where personal data have to be shared.</td>
</tr>
<tr>
<td></td>
<td>In order to reduce misuse of data by service providers, memorandum of understanding and service agreements signed by those accessing data can be used outline ‘appropriate’ uses of data. Furthermore, for systems that share data between organisations, having an electronic audit can help discourage ‘improper use’ and provide a documentation trail if a review of a case is called.</td>
</tr>
<tr>
<td></td>
<td>Deletion of data from integrated systems can be a mechanism for preventing ‘black marks’ being placed against people based on their prior interactions. For example, data in shared data management systems can be purged if the person has not re-entered the system after a given period of time. Furthermore, mechanisms could be used to enable services to only access records of another service when consent is given.</td>
</tr>
<tr>
<td><strong>Gaining consensus</strong></td>
<td>Research into the embedding of innovations in organisational processes (Murray et al. 2010) highlights that designing a system whose outputs are relevant and therefore valued by those collecting the data, and/or that fits into current service delivery practices, have the greatest potential to improve uptake—in this case in a new data system. In the design of a system, sector consultation will be important, in order to agree on what data will have the maximum value to a range of sector actors.</td>
</tr>
<tr>
<td></td>
<td>Depending on the scope of the system, different forms of governance may be required. As the sector encompasses a large number of different services, reaching a sector wide consensus in a way that is timely and practical for operational purposes may be difficult, therefore a combination of consensus and delegated governance could be adopted, i.e. representatives from Regional Collaborative Committees of the Supporting People/Housing Support Grant fund, rather than include all grant funded service providers.</td>
</tr>
<tr>
<td><strong>Data (quality) related issues</strong></td>
<td>The change in level of data being transmitted to Welsh Government, from aggregate to individual/case, may naturally generate shifts in data quality practices as it increases scrutiny of underlying data. Having a data governance presence with a remit over data quality will provide a mechanism for maintaining data standards and a platform to organise changes in data as the sector evolves.</td>
</tr>
</tbody>
</table>
Part six: Designs for a new homelessness data system in Wales

This part of the report brings together the learning's from the sector engagement and evidence reviews to outline four options for a new homelessness data system in Wales. Each option is briefly summarised before providing detail on its technical configuration (e.g. broadly addressing data recording, extraction, and transmission), followed by a discussion on how the system fits within the homelessness strategy for Wales. In order to aid assessment of each option a summary of the strengths, weaknesses, opportunities, and challenges of each is provided. Included with each option is a summary diagram of the main data flows across the system, with a more detailed technical system diagram being provided at Appendix B. Across the systems there are some common elements that are discussed separately in order to limit repetition, these being: (1) scope and standardisation of data, (2) access to data in a secure data infrastructure for data linkage, (3) measuring hidden homelessness.

Data scope & standardisation
All options generate individual level data on household members being assisted. Collecting data on only the main applicant would limit the ability of Welsh Government to generate statistics on repeat homelessness where the main applicant changes and reduced the reliability of statistics and research if household membership has to be inferred from other administrative data. As the sector have raised data linkage as an aspiration of new data collection (Downie 2018) data on all individuals being assisted is justifiable in order to achieve this aim. In terms of statutory homelessness all of the system options include the collection of data relating to placements in temporary accommodation at the placement level, i.e. start and end date, with the ability to link between placements and decisions. Being able to link between placements in temporary accommodation and decisions under statutory legislation will vastly improve understanding of the use of temporary accommodation, something which the sector has indicated is an area of concern (Golten 2015).

In order to be useful as a basis for decision-making, data need to be interpretable to people outside the original context of its collection. Standardisation of data collection can therefore facilitate the creation of a common data language. Having a data ‘standard’, or set of pre-defined data items to collect, covering assistance provided to individuals will also enable a broad picture of homelessness in Wales to be constructed when integrating data—if only for the purposes of research and statistics. As an example of what the data standard could look like for the homelessness service sector in Wales, Table 11 is a synthesis of elements from the Specialist Homeless Service Collection in Australia, the recommendations for core variables by the Mutual Progress on Homelessness through Advancing and Strengthening Information Systems (MPHASIS) group (MPHASIS 2008), and the structure of the BIWM Data Collection Standard in Poland (Wygnanska 2015). A core set of data items would be data enabling clients to be linked across services, including name, date of birth, gender, and postcode, and if possible, unique national identifiers such as the National Insurance Number. Non-core data refers to elements that may not be applicable to all people assisted across the sector and would include details on needs and what services were provided, which can be tailored to suit sectors and collected as and when necessary. The core data elements address the need to understand who the homeless are, what services they are accessing, and whether they are still homeless after leaving those services.
During a seminar given as part of the engagement activities for this project where options for a new system were being outlined, one of the attendees working closely with the sector suggested that data requests should not be burdensome on initial interactions between staff and people accessing services. Furthermore, in their options appraisal for a new rough sleeper data system in Scotland, the Centre for Homelessness Impact found that some of the people accessing ‘low threshold’ services were doing so specifically because of the reduced requirement to provide personal details (Russell and Thomas 2019). Though the ‘core element’ of the data standard requests personal data that may not be possible to collect on initial interaction, for example when someone is bedded down on the street, this data could be built up over a timescale that fits within the normal working practices and populations of people being assisted. Furthermore, as indicated in the discussion in Part three, changes in the use of technology and how services interact with people may require greater emphasis being placed on ‘transactional data’, such as telephone call counts, which although they lack depth, are indicators of service use.

Table 11: Example data ‘standard’ for a sector wide data collection framework

<table>
<thead>
<tr>
<th>Core</th>
<th>Clients</th>
<th>Services</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collected by all services on all clients, again where feasible in the process of assisting the person</td>
<td>Date linkage variables: Name, or nickname</td>
<td>Service provider</td>
<td>Housing status after leaving service (including tenure)</td>
</tr>
<tr>
<td></td>
<td>Date of birth</td>
<td>Dates of entry into service, or data of outreach</td>
<td>Date of exiting service, where service has an ‘exit’</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Postcode of most recent permanent address</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>National Insurance Number (if known)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Housing status (including tenure and type to assess hidden homelessness)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-core</th>
<th>Clients</th>
<th>Services</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collected by services where relevant to the working practices of services</td>
<td>Other demographic characteristics</td>
<td>Assistance provided (e.g. Housing Act relevant assessments)</td>
<td>Employment status</td>
</tr>
<tr>
<td></td>
<td>Household composition</td>
<td></td>
<td>Unmet needs</td>
</tr>
<tr>
<td></td>
<td>Employment status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Client needs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data accessibility
All options enable access to data outside of Government, including access via interfaces, dashboards, and through integrated data infrastructures. The one common method of accessing data is through a UKSeRP (Jones et al. 2016), which would provide a secure e-research platform enabling access to (linked) de-identified administrative data. Examples of UKSeRP’s included in the review are the SAIL databank, the Housing Stock Analytical Resource (HSAR) which is leased by Welsh Government in order to hold housing data, and the Dementias Platform UK as an example of a UKSeRP enabling access to dementia research internationally. Through a specifically designated UKSeRP for homelessness data in Wales, approved users would be able to access data securely from their own computer terminals. Users of this ‘homelessness SeRP’ would be able to conduct longitudinal analysis using statutory data, engage in data linkage to other public service data e.g. health and education data, and would also be able to bring data from their own cohorts of individuals into the SeRP in order to examine the impacts of interventions. Setting up or ‘leasing’ a new UKSeRP is preferential to adopting an already existent UKSeRP, such as HSAR, as it will provide the opportunity for sector specific data governance arrangements to be made.
Measuring hidden homelessness
As highlighted by the discussion in Part Three, the hidden homeless would not normally be accessing housing services nor would they be recorded as part of rough sleeper counts, and therefore are largely obscured from the homelessness data landscape. However, information on the extent of hidden homelessness could be achieved, in part, if public services such as General Practitioners or healthcare professionals routinely recorded the housing status of people they encounter—though there are ethical and practical barriers of requiring already overburdened staff to ask and record additional personal information.

At present, people who are of ‘No Fixed Abode’ (NFA) are identifiable in hospital data as having a postcode recorded as ‘ZZ99 3VZ’. NFA is an ill-defined term that could conceivably refer to anything from rough sleeping to sofa surfing. Furthermore, the recording of NFA as a specific postcode field is incompatible with the current secure data infrastructure in Wales (SAIL databank) which anonymises postcodes. There are however ‘social determinant of health’ codes available within both primary and secondary health data that could be used to identify the homeless. However, in both the case of diagnoses codes and the use of postcodes to record NFA, the reliability and consistency with which these codes are applied is not known—the Administrative Data Research Centre Wales have a programme of work exploring the extent to which ‘homelessness’ can be identified in health data.

Alternatively, where someone accesses a homelessness specific service, prior histories of hidden homelessness could be recorded. The H-CLIC system in England collects data on accommodation type which includes living with friends and family. However, the use of administrative data to count hidden homelessness is limited to retrospective assessments of housing situations amongst those who become ‘visible’ at some point. A more rigorous method of enumerating hidden homelessness across the population would be to amend the National Survey for Wales to include a set of questions to assess the extent of concealed, ‘doubled up’, and overcrowded households, and to ask these questions of all households—there are examples from Los Angeles (Agans et al. 2014) and Vancouver (Eberie et al. 2009) where sample surveys have been used to enumerate the hidden homeless through surveys. The remainder of this part of the report discusses each of the options in detail.

Option 1: Transformed reporting
(Individual level data)

Nearest example(s) from the review

- H-CLIC in England
- Scotland’s HL1 and PREVNT1
- Homeless Individuals and Families Information System in Canada
- Combined Homelessness and Information Network (CHAIN)

---

22 The terminology used in the United States to refer to multiple co-habiting households.
Overview

This option would result in the creation of a new individual level data set that could be used for research and statistical purposes; initially covering statutory homelessness, the scope of this data set could be expanded to cover HSG recipient data. De-identified individual level data would be reported to Welsh Government via Afon—a data transfer and validation tool currently in use by Welsh Government. For HSG data, a centralised data management solution is proposed, which could either be downloaded to act as a local area data management solution (e.g. HIFIS in Canada), or through a web-based data management solution (e.g. CHAIN). The HSG data management solution would streamline reporting, data transfers, and provide data management facilities for smaller organisations who lack their own IT solutions.

Figure 5: Transformation of data flows to reduce provider burden through technologies (Transformed)

How will it work?

Welsh Government would provide data specifications for statutory homeless data outlining what variables to collect, what values the variables should take, and at what stage in the Housing Options process data should be collected. The guidance provided to Housing Options teams can be used to (re)design local data management systems to harmonise data across Wales and improve data quality generally by improving consistency. Where current systems can output the required individual level data, then legacy systems can continue to be used. When England moved to the H-CLIC data collection, funds were made available to local authorities in order to improve their local data management systems to meet the new data requirements; a similar approach could be adopted in Wales.

Housing Options teams would extract individual level data from their local systems and upload (transmit these) to a secure web-platform, which would undertake validation of the data using a set of pre-defined validation rules. The Housing Options team would have the opportunity to amend any data errors and re-upload the data, before Welsh Government download it onto their local database. The AFON system used for the current statutory homelessness data submission could be adapted to process the new data—AFON currently manages individual level data for the SSDA903/Looked After Children Census. Alternatively, a separate software solution could be procured specifically for homelessness/housing data, for example DEWi for Schools’ data (see Pupil Level Annual School Census).
It is proposed that rather than Welsh Government being in receipt of personal identifiers, that these be sent to a secure data infrastructure via a trusted third party for de-identification for data linkage studies. In order for Welsh Government to identify and report on people experiencing repeat homelessness, personal identifiers can be hashed at the local authority side before transmission, or by the upload platform itself through an integrated function. Alternatively, hashed National Insurance Numbers could be used to achieve de-identified linkage, with a similar approach being adopted by Scotland’s HL1 collection to identify repeat homelessness nationally, and when linking between PREVENT1 entries provided by registered social landlords and HL1 data. Welsh Government would be responsible for any final cleaning of the statutory homelessness data before depositing this the secure data infrastructure.

**Housing Support Grant (HSG) addition:**
A more ambitious transformation of the reporting of homeless data in Wales could be undertaken to change the reporting of the future HSG. It is envisaged that a separate system using a software solution that is either downloaded and acts as a local data management system that periodically uploads data to Welsh Government (e.g. Homeless Individuals and Families Information System in Canada or Ffynnon in Wales), or a web-based data management solution used by multiple agencies (e.g. CHAIN in Greater London). The addition of the HSG platform could potentially help alleviate the burden on councils to process data prior to sending it to both Welsh Government and the secure data infrastructure; it also provides smaller HSG funded organisations with a data management functionality they may lack, and opens up new ways of working—particularly if a web-based data management model is adopted that enables data sharing.

**What outputs will it generate?**

The main output from the system is an individual level data set relating to statutory homelessness provision in Wales, which has unique identifiers that enable the extent of repeat homelessness to be evaluated. The data will be accessible via a secure data infrastructure, which, in addition to enabling access to the homelessness data in a secure environment, can also be used for data linkage studies, particularly to other public sector data, even without the HSG addition. The statutory data could still be used on an ad hoc basis for data linkage studies with non-statutory data. For example, a housing service provider could bring their data into the secure infrastructure for use on a specific project they were conducting/commissioning, and link this to the statutory data to explore the impact of their services on statutory homelessness. Alternatively, a ‘Data Lab’ could be funded whereby services provide their personal data to the lab, who then link this to the centralised homelessness collection, and generate standard outputs that provide data information on returns to homelessness services23.

As data are first processed by Welsh Government before being deposited in the secure data infrastructure, Welsh Government will be able to produce national statistics prior to data being publicly available. However, because of this processing, there may be a delay between submission of data and it being publicly available for research purposes. It is envisaged that quarterly aggregate statistical outputs, for example on StatsWales, would fill the sector’s general information requirements, with individual level being deposited on an annual basis in the secure data infrastructure.

Having national level statutory homelessness data would enable research and evaluation, though this would be limited to statutory homeless outcomes. However, having national level data will improve decision-making in relation to the Act and future homelessness related policy by providing an evidence base not currently available. If the data being collected is expanded to provide more detail on who homeless people are and their housing situations after leaving statutory services, rather than only recording ‘outcomes’ based on how their case was legally discharged, then this will meet the sectors needs for more detailed information on homelessness in Wales. By recording more information relating to people experiencing homelessness, rather than legal cases going through the homelessness system, an expanded data collection could also hopefully improve the accessibility of the statistics and information being produced by the new collection by avoiding excessive technical/

---

23 See for example the Ministry of Justice ‘Justice Data Lab’, which links third sector data to offender records, in order to provide information on reoffending of service users: https://www.gov.uk/government/publications/justice-data-lab
An expanded data collection can also contribute to homelessness prevention, when linked to population level data, by helping understand issues faced by segments within the (homeless) population.

**Opportunities from a HSG addition:**

Having a separate software system for HSG data management could provide a mechanism for person-centred progress tracking for those engaged with grant-funded services. As an example of a person-centre outcomes, the GM-Think system enables people accessing services to update their goals and achievements. Providing the ability for the person being assisted to control their own assessments and progress reporting can potentially address issues raised about the Supporting People grant Outcomes Framework, that outcomes were potentially being recorded without the person’s input (Wales Audit Office 2017a).

**Table 12:** Summary of technical elements for transformed individual level data option

<table>
<thead>
<tr>
<th>Design element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose and new ways of working</strong></td>
<td>Offers a new source of data on statutory homelessness which can feed into improved research and evaluation, and therefore policy and service delivery. Having improved data quality/standardised data can potentially lead to data sharing through shared ‘data language’. The HSG addition could potentially lead to increased cross-sector working.</td>
</tr>
<tr>
<td><strong>Data architecture model</strong></td>
<td>Data are extracted by agencies from their local data management systems and transmitted to Welsh Government who centrally combine and clean the data. Web-based data upload software used to manage data transmission process &amp; validate data. The HSG data management platform would have a centralised architecture, where data can be viewed by other organisations, with consent from the person accessing the service.</td>
</tr>
<tr>
<td><strong>Supporting change</strong></td>
<td>Grant funding to be made available to local authorities in order to improve their local data collection system to record/and output individual level data and temporary accommodation data. HSG data management/sharing platform could either be a downloadable software for use by services, or a web-based platform. Having a web-based platform may enable larger organisation who do not wish to use the system as their local data management solution to still interface with it to upload data in batches.</td>
</tr>
<tr>
<td><strong>Governance</strong></td>
<td>Collective governance of statutory data in the data infrastructure in order to maintain sector control over its own data and to keep informed of uses of their data. Additional governance structure will be required if the HSG data management solution is adopted.</td>
</tr>
<tr>
<td><strong>Data quality</strong></td>
<td>Semi-automated validation of data as it is being uploaded, and further cleaning by Welsh Government of the combined data. HSG data quality to be overseen by the governance body for the data management system and/or the service provider.</td>
</tr>
<tr>
<td><strong>Ethical and legal</strong></td>
<td>Uses legal means to share/process data as part of a national collection as data are for statistical and research purposes, rather than being for data management, and will therefore not directly impact a person’s provision of care.</td>
</tr>
<tr>
<td><strong>Data sharing</strong></td>
<td>De-identification/encryption of data being provided to Welsh Government, will enable same individuals to be identified however. Normal process of de-identification by a trusted third party in order to transfer data into the secure data infrastructure.</td>
</tr>
<tr>
<td><strong>Data access</strong></td>
<td>Statutory homelessness data to be deposited by Welsh Government in the secure data infrastructure. Access to the homelessness data to be directed to the infrastructure, rather than to KAS (although KAS will be involved in the governance function of the data infrastructure, and/or may have delegated control over aspects of the data). Statistical outputs produced by Welsh Government. HSG data have various destinations, dependent on needs, i.e. individual level data may be required by commissioners of services at local authorities, Welsh Government receiving aggregated reports in order to monitor provision, whilst individual level data (including personal data) sent to the secure data infrastructure.</td>
</tr>
</tbody>
</table>

24 Shaded and italicised text refers to additional options for each system
Table 13: Strengths, Weaknesses, Opportunities, and Challenges (SWOT) analysis of transformed individual level data (Option 1)

<table>
<thead>
<tr>
<th>Strengths:</th>
<th>Weaknesses:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Having a centralised dataset on statutory homelessness will support improved research &amp; evaluation of statutory services, and measuring the issues &amp; response to homelessness in Wales. If local systems can extract the required data then no additional IT infrastructure are necessary.</td>
<td>The combined data are not accessible for day-to-day case decision-making.</td>
</tr>
<tr>
<td>Provision of ‘core data’ guidance can improve the consistency of information being gathered by services. Data upload portal requires less input from sector as it simply manages data flows, therefore procurement and development should be fairly quick; alternatively data upload can be managed by adapting current technologies used by Welsh Government—such as AFON—thereby reducing setup time/costs. Can maintain current statistical outputs, given that the data being provided to Welsh Government will be the raw data that currently form the basis for the statutory outputs. Communities can organise their data systems to meet local data needs, therefore does not preclude the use of local data warehouses or shared case systems.</td>
<td>There are a large number of data flows between different agencies, particularly if HSG continues to be collected in the same manner as SP data, without a centralised collection software, however, transfers could be automated to reduce burden on data providers. Delay between submitting data and actionable intelligence as data have to be processed and go through data cleaning and linkage by Welsh Government, and through statistical release procedures, before entering into secure data infrastructure.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunities:</th>
<th>Threats/challenges:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition of HSG funded service data increases the scope for data linkage. Formation of a ‘DataLab’ for homelessness research and evaluation in order to generate actionable intelligence, as has been suggested using H-CLIC data in England. Creating reciprocal data exchanges, e.g. reporting back to authorities and data providers, gives added value to sharing data, particularly if the analysis is based on data that the data provider themselves would not normally have access to. Both the increase in consistency/quality of data across local authorities (and Housing Support Grant recipients) opens up possibility of improved data sharing (through a common data language) and for more centralised data management systems in the future. Continuum of Care systems in the United States have similarly evolved from centralised data, to shared data management.</td>
<td>Future legislative/funding changes may impact on data needs, however having a data upload that is structured around interconnected modules will improve the adaptability of the system, e.g. Lifelong Learning Wales Record and H-CLIC. Local authority perceptions around data protection legislation and their ability to share data will have to be overcome. Where local databases cannot meet the required data extract format, this will create an initial burden on the authority in order to improve local systems. Need to find a consensus amongst local authorities/sector on what data items should be collected and included in guidance. However, data being requested can largely draw upon data currently being provided by Housing Options and Supporting People programmes. Capacity issues at Welsh Government in order to process data. Requires grant funding from Welsh Government in order to improve local data management system to output required data extracts (where not possible to do so).</td>
</tr>
</tbody>
</table>
Option 2: Federated model of data access

Nearest example(s) from the review:

- Virginia Longitudinal Data System
- North Carolina School Works (NCSW)

Overview

This option would improve access to sector data for research and statistical purposes by providing a mechanism to access sector data directly from data providers, rather than centralising data. Requests for data would be submitted to a data broker, who would then be responsible for seeking approval from the data provider, e.g. a Housing Option team. The broker extracts the relevant data from the agency’s database if they agree to allow their data to be shared. Prior to exiting the data provider’s system, data are de-identified before being combined with other data from the federation (if requested) and made available for download by the organisation/person requesting the data. There is the potential for this system to evolve to enable organisations forming part of the federation to query one another’s databases in real time (e.g. X-tee in Estonia).

Figure 6: Federated model of data sharing across the sector (Federated)
How will it work?

This option adopts a ‘federated’ rather than a ‘centralised’ data architecture; rather than agencies pushing their data on a regular basis, i.e. through data uploads to Welsh Government, data remain in the agencies database until they are requested, and are then automatically ‘pulled’ from their databases. The option relies on a hub-and-spoke model for the data federation, with data requests and flows of data being processed by a single organisation: the ‘data broker’. The data broker can be a single organisation or group of organisations, that are either created specifically to undertake the role of broker or use existent departments with similar functions. For example, the Virginia Longitudinal Data System is managed by several state departments with involvement from Universities and other organisations in the development of the infrastructure of the system. The data broker for the North Carolina School Works system is Government Data Analytics Center located within the State’s Department of Information Technology, who specialise in mobilising data and business intelligence. The data broker acts as an interface between the different organisations participating in the system, and between the federation and external researchers/data requesters. Flows of data through the federation would be as follows:

- A proposal/request for data is made to the data broker; included in this proposal would be what variables are required, why, and what uses the requested data are for. In addition to the proposal, there may be other clauses attached to data access. For example, in order to conduct research using data from the Virginia Longitudinal Data System, researchers having to work with the data providing organisation to produce outputs that are mutually beneficial; similarly, access to a number of other central UK Government data sources, e.g. the HMRC datalab, require analysts/people using the data to be associated with a UK based academic institution, a not-for profit organisation, or government department (either directly or sponsored by)—mainly to limit commercial use of data and data being accessed outside of the UK;

- The data broker reviews the request, and submits individual requests for data to the relevant data providers in the federation. For example, if the data request required Housing Options data for Wales, the broker would forward requests to all 22 statutory housing teams in Wales;

- The data provider agencies review the request, and either approve or decline it;

- Following approval, the data broker would ‘pull’ data from the relevant data provider’s systems. In the process of being extracted, data are de-identified. Where data from multiple providers is requested, all data as part of the same request undergo the same de-identification, which enables the same person to be identified across the different data providers. To continue the Housing Options example, the same person who made applications to other authorities would be identifiable within the provided data extract.

- Following extraction, the data broker would prepare a final data set, or series of data sets, which are reviewed by the data providers a final time, before being made available for download by the person/agency who made the original request. Any identifiable data in the data sets made available for download would be encrypted, such that none of the unique identifiable information could be linked back to ‘real world’ data, e.g. any unique ID numbers cannot be linked back to the original data provider’s system.

Data providing participants in the federation will largely be able to operate their own data management practices as they currently do; the federated system merely improves the ability for different systems to talk to one another via ‘adaptors’ that connect them to the data broker. Data providers can potentially benefit from the ‘pull’ rather than ‘push’ nature of the system as the time/resource needed to process data submission will be reduced—the ‘adaptors’ and data broker are responsible for extracting and processing data.

The provision of data specifications for Housing Options and HSG fund recipients will increase the general standard and consistency of data being collected, thereby reducing the workload of the data broker in understanding the different data types—it also has the potential to improve the ability of people outside the housing sector to understand and use the data they request. There may need to be some changes in how data are stored within local systems to improve the ability of the data broker to access data; for example, creating a partitioned off area of local data management systems that is accessible to the data broker.
Initially, data providers within the data federation would be limited to Housing Options teams and HSG funded organisations as a proof of concept—and as there is a demonstrable need for data access to these data sources in order to maintain current funding obligations and the production of national statistics on homelessness. However, over time, the federation could be expanded to include other housing service providers, such as social landlords, and other public services. One of the main barriers to participating in the federation would however be the ‘data maturity’ of the agency wishing to take part. If an organisation does not have a sufficiently high enough technical capacity, then achieving a connection may require development of the local IT infrastructure. Data maturity may be a particular issue for smaller HSG funded agencies, therefore it is proposed that data under the HSG funding be collected in a centralised data management software, similar to Option 1, that can then act as the interface the collective HSG data.

It is envisaged that Welsh Government becomes a ‘super-user’ of the federated system, with access rights to certain data, for example HSG funded recipient data and statutory homeless data for monitoring purposes. Local authorities could similarly be super-users of HSG data, if they are to retain responsibility for commissioning services—as is the current arrangement under the Supporting People programme fund.

**Networked federated system adaptation:**

Developing the federated approach further, direct networking between organisational data management systems without outside of the broker would enable agencies to query one another’s data management system in real time. A similar approach is used in Estonia’s X-tee. This method would potentially only be viable for larger public sector organisations where there were clear statutory gateways enabling data access for case management purposes.

**What outputs will it generate?**

The outputs from this system would be improved access to sector wide data, and the ability to link together data from different agencies, whilst limiting privacy related issues—given that data are de-identified. Specifically, this system would result in annual anonymised extracts of both Housing Options data and HSG data being held by Welsh Government, or another agency, and which form a permanent historical record accessible to the sector. A request for combined Housing Options/HSG data would provide data on individuals using statutory and non-statutory data within the year. However, one of the challenges that government may face with this option is the cultural shift in access to data that it creates; Though Welsh Government would presumably still wish to continue generating national statistics, requests for data can be made at any time via the broker.

The unique de-identification of data prior to leaving the data provider’s system provides an added level of security as it limits the potential for disclosure of identities through combining different data requests. However, this method of de-identification will mean that annual extracts of data will not be linkable over time, which will limit the ability to explore repeat homelessness in national statistics. For example, if Welsh Government request data from Housing Options teams on a quarterly basis, the extent to which repeat homelessness occurs within the quarter can be reported on, however the data will not support the measurement of repeat homelessness over the year. However, the limited ability to monitor repeat homelessness could be overcome by (1) requesting data over a longer time period, (2) asking authorities to monitor repeat homelessness themselves and flag this within their data systems, (3) depositing data in a secure data infrastructure, as this will be individual level and linkable over multiple years and data sources.
As with all options, individual level data are deposited in a secure data infrastructure to enable data linkage to data not held by members of the federation, for example healthcare data or education data. Depositing data in the infrastructure would be a separate exercise involving each federation member separating their data into the person identifiers and ‘payload’, before sending these to the trusted third party and secure infrastructure, respectively. As the data in the infrastructure are de-identified but linkable, then longitudinal data analysis could be undertaken, specifically repeat homelessness over the life course of the data collection. The rational for having broker access and a secure data infrastructure is that the broker access system would generate live data—being extracted directly from systems—that may therefore improve the sector’s ability to respond to real world events as they occur, because of their live nature. Whereas data in the data infrastructure would be updated on an ad hoc basis, and therefore be suited to research of historic data. The quality of the data deposited in the data infrastructure may suffer due to the lack of centralisation, which makes having a strong data governance function, an essential aspect of maintaining data quality in this system design.

**Opportunities from a networked federated system:**
The outputs possible from a networked federation would impact on the provision of services to people. In effect, a federated network would operate as a shared data management system across the nation, distributed over several organisations. It is claimed that X-tee in Estonia has led to substantial reductions in working time needed to process and make data requests to other agencies.

**Table 14: Summary of technical elements for the federated model of data access option**

<table>
<thead>
<tr>
<th>Design element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose and new ways of working</strong></td>
<td>The main outputs from this system will be improved access to timely data for use by sector actors for statistical and research purposes. <em>The inclusion of direct networking of data provider systems would alter how agencies work together by increasing collaboration and referrals through data sharing.</em></td>
</tr>
<tr>
<td><strong>Data architecture model</strong></td>
<td>Data remain in local data management systems and are ‘pulled’ from these when a request is made. Requires the creation/allocation of a data broker to manage the flows of data and requests across the federation. ‘Adaptors’ in each federation member will manage automated data extraction by enabling the broker to interact with the local system directly. <em>Development of the federated approach would be the direct networking of systems outside of the data broker, to enable live querying of systems and access to identifiable data.</em></td>
</tr>
<tr>
<td><strong>Supporting change</strong></td>
<td>Large Welsh Government grant to be managed by a lead authority/organisation, with the federated approach being ‘rolled out’ incrementally, much like the WCCIS healthcare system.</td>
</tr>
<tr>
<td><strong>Governance</strong></td>
<td>Use of data is ultimately under the governance structures of the individual data providers, delegated governance to the data broker will increase the ability of the broker to act on behalf of its members. Collective governance of data in the data infrastructure would help maintain sector control over its own data and to keep informed of uses of their data. Additional governance structure will be required for the HSG data management solution.</td>
</tr>
<tr>
<td><strong>Data quality</strong></td>
<td>Data quality is the responsibility of the data provider, however the provision of guidance to Housing Options/HSG providers will improve consistency across the same provider type. The sector will generally be encouraged to sign up to the ‘data standard’ so that consistent data are available across the housing sector—the new HESA data infrastructure has adopted a similar approach to encourage data providers to take ownership and responsibility for the quality of their data. An external organisation, to maintain data quality across the member organisations could be set up, e.g. Workforce Data Quality Initiative in the United States.</td>
</tr>
<tr>
<td><strong>Ethical and legal</strong></td>
<td>Privacy notices and legal gateways can be used to enable sharing and processing of data. A memorandum of understanding would be required outlining what is required of participating data providers in the federation.</td>
</tr>
<tr>
<td><strong>Data sharing</strong></td>
<td>Data are de-identified prior to leaving the data provider’s data system. Where data are to be sent to the secure data infrastructure, the normal process of de-identification by a trusted third party will be used.</td>
</tr>
<tr>
<td><strong>Data access</strong></td>
<td>The main data access point for data by both the sector and other stakeholders would be the data broker, or the secure data infrastructure for data linkage research. Welsh Government would be an access point for annual de-identified data; however this could be delegated to another organisation on their behalf.</td>
</tr>
</tbody>
</table>
### Table 15: SWOT analysis of federated model of data access (Option 2)

<table>
<thead>
<tr>
<th><strong>Strengths:</strong></th>
<th><strong>Weaknesses:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data that leaves a data-provider’s system is de-identified; may appeal to services as it addresses concerns around sharing personal data.</td>
<td>Though requests are based on ‘live’ data, there will still be a time lag between the request being made and the data being provided—given that the request needs to be approved and processed before being released. However, by retaining ‘official’ data collection periods, where authorities clean and ready their data for access by Welsh Government, i.e. quarterly, and by using technology to automate data requests, any impact on the production of official statistics can be mitigated. The system would merely automate a process that is currently manually instigated.</td>
</tr>
<tr>
<td>Supports research and reporting needs, with data requests being compiled from ‘live’ databases, therefore up-to-date analysis is available to the sector.</td>
<td>As data being accessed by the broker are de-identified before extraction, this means that a separate data upload into the secure infrastructure would be required—if linkage to data ‘outside’ the federation is ever required.</td>
</tr>
<tr>
<td>Ownership of data remains with the data-providing organisation, and access to it is controlled by the agency itself, which can reduce ‘turf battles’ over access to data.</td>
<td>Quality of data is determined by local processes, and low quality data could negatively affect data linkage research; this could be addressed, to a certain extent, through the data standard driving harmonisation of data across the sector, and/or an external moderator.</td>
</tr>
<tr>
<td>System is scalable, to an extent, to multiple data providers, as any new federation members are simply set up with a connection to the data broker; however, it requires that the data provider be sufficiently ‘data mature’ to enable interfacing of their data management systems.</td>
<td>Outputs are likely to vary based on changes in the underlying live data, i.e. queries of the same data source of the same time period may result in slightly different data as operational system data can be changed. If Welsh Government retain permanent copies of annual extracts of data, then these can become a stable historic record.</td>
</tr>
<tr>
<td>Removal of data providers from the federation is easy as it simply requires the removal/de-activation of their connection device.</td>
<td>Data broker is a single point of technological failure as the ‘hub’ of the system. Furthermore, if a connection to a federation member is lost, then this may impact data quality/completeness of returned data.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Opportunities:</strong></th>
<th><strong>Threats/challenges:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Could develop into a live federated sharing of data, much like the X-tee in Estonia where agencies are able to query one another’s systems for individual level data; this would therefore support collective information and service delivery.</td>
<td>High levels of coordination and sustained engagement would be needed to set-up and run the federation.</td>
</tr>
<tr>
<td>Federation could extend outside of Wales, to bordering local authorities; people moving across borders, and having access to data about these people, was raised by a local authority in the survey.</td>
<td>Contracting of a data broker will be a challenge given the novelty of the approach.</td>
</tr>
<tr>
<td>Could potentially reduce the number of FOI queries that are being answered as these data requests can be directed to the broker; it might reduce the time spent answering FOIs as the broker is responsible for extracting the data.</td>
<td>Staff resource is required at the data provider in order to process requests for data and maintain data access.</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>High levels of coordination and sustained engagement would be needed to set-up and run the federation.</td>
<td>Requires substantial financial input in order to fund ‘adaptors’ and also the formation/running of the data broker.</td>
</tr>
<tr>
<td>Contracting of a data broker will be a challenge given the novelty of the approach.</td>
<td>Technologically complicated, which may not be compatible with smaller agencies, i.e. those funded by HSG; having a shared system for HSG funded agencies, with its own governance structure should enable these agencies to contribute data to the federation without being overly burdensome to them.</td>
</tr>
<tr>
<td>Staff resource is required at the data provider in order to process requests for data and maintain data access.</td>
<td>Requires clearly defined Information Governance structure for the federation and HSG data management solution.</td>
</tr>
<tr>
<td>Requires clearly defined Information Governance structure for the federation and HSG data management solution.</td>
<td>Requires multiple data-sharing agreements and memoranda of understanding to be in place, which can be time consuming to set up.</td>
</tr>
<tr>
<td>Requires multiple data-sharing agreements and memoranda of understanding to be in place, which can be time consuming to set up.</td>
<td>As the number of agencies increases, the knowledge needed by the hub to manage the data increases.</td>
</tr>
</tbody>
</table>
Option 3: Data warehousing

Nearest example(s) from the review:
- Ohio Human Services Data Warehouse
- Veteran’s Health Administration Corporate Data Warehouse in the United States
- North West London Whole Systems Integrated Care
- Kent Integrated Dataset

Overview

This system operates via a single centralised data warehouse for homelessness data in Wales that pools data from across statutory and non-statutory services. Personal identifiable data would be uploaded to the warehouse in order to enable linkage across data sources, and to enable greater flexibility in the warehouses’ capabilities over time. Initially, interfaces with the warehouse would enable the sector to query the data in order to return aggregate data reports to guide service commissioning and policy making. ‘Data marts’ would also bring together data on specific topics, such as regions or rough sleeping. The long-term goal of this system is to enable the warehouse to be queried in order to return outputs that relate to individuals to guide care decisions (e.g. Veteran’s Health Administration Corporate Data Warehouse and North West London Whole Systems Integrated Care).

Figure 7: Central warehousing of data to inform policy and service design (Data warehouse/lake)
How will it work?

The system is configured around the periodic extraction and transmission of data from data providers to a centralised national data warehouse. Initially, local authorities would contribute data to the warehouse, with this expanding to include Housing Support Grant funded services, before finally opening up to other services. However, as the process of reporting data to a warehouse may be burdensome for small HSG recipients, some of whom may lack a data management platform, local authorities could submit data on their behalf, as is currently the case with Supporting People data submitted to the SAIL databank.

Data can either be manually extracted and uploaded into the warehouse (e.g. Kent Integrated Dataset), or Automated Programming Interfaces (APIs) can be used in order to automate the extraction and transfer—and also achieve a higher frequency of data upload, i.e. nightly. The warehouse would hold individual level which would include personal identifiers in order to link data within the warehouse, and for later linkage studies when deposited in the secure data infrastructure. In order to ensure that the warehouse runs optimally and to simplify data governance arrangements across the system, data being uploaded should follow a given format. The format of the data can vary by the uploading service type, e.g. Housing Options or HSG funded service, if each ‘type’ is consistent in using the same format. However, a ‘core data’ standard will help ensure a minimum level of consistency across all service types.

After data are extracted from the data providers’ systems they enter a ‘staging area’, where they undergo any additional transformation, cleaning, and validation. The staging area could also be used to stagger data which is being uploaded at different frequencies. For example, if one authority uploads Housing Options data on a nightly basis and another on a weekly basis, then the staging area could be used to temporarily store data so that it all enters the warehouse on a weekly basis.

In order to make searches of the data more efficient, data could be collected into thematic ‘data marts’. Examples of data marts could include local area data marts, which combine individual level data on people resident in local authorities into a data mart for that authority. Alternatively, a data mart could be used to combine information on rough sleepers from across different agency uploads. A rough sleeper data mart would be particularly useful if data depositing in the data mart was opened to other public services. An example use case would be enabling healthcare professionals or the police to flag a person as a rough sleeper, with this data then being automatically routed to the rough sleeper data mart for use in planning service delivery by rough sleeper outreach teams.

Data would be accessible via interfaces that enable the data marts/warehouse to be queried directly. Access to this query function would be limited either to those agencies who are providing data to the warehouse, and/or those who have been approved as users. General users would be able to access data for research purposes via the secure data infrastructure. In order to get from the warehouse to the secure data infrastructure, personal data would undergo the usual split file process—being sent to a trusted third party for de-identification. However, the main part of the administrative data, i.e. the ‘payload’ data, could either go to Welsh Government for further cleaning, and to produce national statistics. Alternatively, data could go directly to the secure infrastructure from the warehouse service provider, under the oversight of a data governance structure.
What outputs will it generate?

The system would generate a new source of data on homelessness in Wales by pooling together information from different agencies. The warehouse, through the data marts and querying interfaces, would also improve the sectors access to data, without having to rely on Welsh Government to generate outputs. The Michigan’s Statewide Homelessness Data online Warehouse (SHADoW) enables querying of data that combines multiple Continuum of Care systems, enabling service providers to gain insight into homelessness across the state. The types of queries that could be undertaken would all return counts of individuals; in order to maintain data security, queries with less than a certain number of returns would be suppressed, e.g. less than 5 people in the case of SHADoW.

In order to increase the accessibility of data beyond those who are permitted to query the warehouse, area level data dashboards could be developed that draw on ‘live’ data. As an example, the Knoxville HMIS Community Dashboard brings together data on homelessness in the region, including the reasons for the cause of homelessness, system performance measures, change in housing status of the people the community assists, and bed utilisation data. Data dashboards could combine—through APIs—other open data sources, such as DWP benefits data and HMRC data, in order to have a cross-sectoral perspective on the determinants of homelessness. The bringing together of both structural factors on the causes of homelessness, placed alongside the experience of homelessness itself, could potentially lead to new insights into early prevention.

The benefit of a data warehouse is that in having a data resource that is not simply a data set held by government—as was the case in Option 1—and is not limited to de-identified data—as is the case with the secure data infrastructure—then this enables software applications to be developed that can then interface directly with the data. As an example, the Veteran’s Health Administration Corporate Data Warehouse has enabled new data exploration tools to be developed drawing on warehouse data (Rajeevan et al. 2017), including a tool enabling clinicians to identify patients for recruitment into a study about medications in old age.

Individual level data outputs & dashboards:
A future adaptation of this option would be to enable the system to generate outputs based on individual level data and to provide access to these for use by frontline staff. The North West London Whole Systems Integrated Care provides person level data visualisations of priori contacts with health and social care systems, with the aim of bringing data together to generate information summaries for care providers. The challenge of adopting this approach would be the ethical and privacy concerns around making data available at a person level; though these could be address by seeking consent from the person to sharing data with other warehouse use members, as is the case with the NW London system.
Table 16: Summary of technical elements

<table>
<thead>
<tr>
<th>Design element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose and new ways of working</td>
<td>Offers a new source of homelessness/housing sector wide data to improve measuring the issue and response and improve sector access to data. Development of the system to output individual level data to guide decision-making can lead to person-centred service delivery.</td>
</tr>
<tr>
<td>Data architecture model</td>
<td>Individual level data (including identifiers) would be extracted from local data management systems and uploaded into a centralised data warehouse. Data can either be manually extracted, or automatically through APIs. Data uploaded to the warehouse would first enter a 'staging area' where data area temporarily stored and transformed; the staging area also means that data which are input with different frequencies can enter the warehouse at the same time. Thematic sub-sets of data, or 'data marts', can be created that group data for easier querying, e.g. create local authority data marts cover geographic areas, or rough sleeper data mart combining data on rough sleepers.</td>
</tr>
<tr>
<td>Supporting change</td>
<td>Welsh Government to procure the warehouse, or assign a lead organisation to oversee procurement. Stakeholders can fund their own interfaces to the system once warehouse populated with data and governance function fully established.</td>
</tr>
<tr>
<td>Governance</td>
<td>A shared governance structure is proposed covering both the warehouse and data infrastructure, providing consistent oversight of the whole system of data storage and use.</td>
</tr>
<tr>
<td>Data quality</td>
<td>The staging area can be used to undertake validation and cleaning of the data. However, general improvements in the quality of data should be implemented through the use of the 'core data' standard.</td>
</tr>
<tr>
<td>Ethical and legal</td>
<td>Legal gateways could be used to share and process data if the warehouse were to remain for statistical and research purposes. However, if individual level data are to be made accessible, or the system used to flag individuals ‘in-real-time’, then consent would be appropriate as the person’s care is affected directly.</td>
</tr>
<tr>
<td>Data sharing</td>
<td>Split file process could be used in transferring personal data and other data from providers to the warehouse in order to minimise disclosure from a data breach. Data encryption and decryption could also be adopted in transit between the provider and warehouse. A split file process would also be used in order to share data with data infrastructure. Welsh Government would not be provided with personal data, but instead use a universal system ID designated to each person with a warehouse entry. Data queries could return only aggregate data, with rules on the suppression of low cell counts where queried data were lower than some given rule, i.e. cells less than 5.</td>
</tr>
<tr>
<td>Data access</td>
<td>Access points to the data would be via the ability to query the warehouse directly, and the secure data infrastructure for data linkage research and general access to the homelessness data. Queries made to the warehouse would return aggregate reports and output anonymised data analysis. However, taking the example of North West London Whole Systems Integrated Care, data relating to a person’s pathway of service use could be made available in order to inform service delivery.</td>
</tr>
</tbody>
</table>
Table 17: SWOT analysis of data warehousing (Option 3)

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supports reporting and research through the ability to query historical data relating to homelessness/housing related support.</td>
<td>Data within the warehouse will lack context and therefore may be difficult to interpret by those outside of the original data collection agency; the drive towards improved data standards should help develop a shared 'data language'</td>
</tr>
<tr>
<td>Data uploads can be tailored to suit data providing organisations capabilities, e.g. can be quarterly updates that are manually extracted from local systems and uploaded, or automation could be used at a far higher frequency of upload.</td>
<td>Data in the warehouse are only as recent as the latest update; however, APIs can be used to automatically upload new data into the warehouse every night, as is the case with the VHA Corporate Data Warehouse.</td>
</tr>
<tr>
<td>Data quality improvements by focusing data reporting on a 'core data' standard.</td>
<td>Initially not available for case-level decision making; however, add on tools can be used to improve access for case-based decision making, as is the case in the North West London Whole Systems Integrated Care dashboards.</td>
</tr>
<tr>
<td>Data marts bring together topic specific data to enable faster querying of data.</td>
<td></td>
</tr>
<tr>
<td>Legacy systems can be retained and linked to warehouse via automated programme interfaces or other means.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Threats/challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional information tools that enable exploration of the data in the warehouse can be added over time—as has been the case with the VHA Corporate Data Warehouse, where developers have created tools that enable ‘at risk’ cases to be flagged and reported to medical professionals.</td>
<td>Data quality monitoring will need to improve at the data providers end; however, a ‘staging’ area to the warehouse would enable some limited cleaning and processing of the data prior to depositing in the warehouse. Furthermore, a function of data governance structure could include data quality.</td>
</tr>
<tr>
<td>Long term plan for the system is to enable the warehouse to generate individual level summary dashboards to guide person-centred care.</td>
<td>Finding a balance between the level of detail of data and the scope to conduct useful analysis.</td>
</tr>
<tr>
<td>Warehouse could be expanded to include other public service data; social security data would be particularly beneficial.</td>
<td>Legal/ethical requirements surrounding use of data may be complicated if individual level data querying is chosen; may require use of consent rather than legal gateways as the data would be used to actively determine care.</td>
</tr>
<tr>
<td></td>
<td>An agreement with all data providers will need to be reached in order to share personal identifiers with the warehouse; providers are already risk averse.</td>
</tr>
<tr>
<td></td>
<td>If the option to enable access to individual level data is adopted, misuse of data possible as decisions made based on data cannot fully be controlled; a memorandum of understanding on how data can be used would help guide ethical decision making.</td>
</tr>
<tr>
<td></td>
<td>Furthermore, having an electronic audit of what queries have been run by users can aid in auditing decision-making if required.</td>
</tr>
<tr>
<td></td>
<td>Decision on who will host the data warehouse could be politically charged and will require input from all data providers.</td>
</tr>
</tbody>
</table>
Option 4: Centralised integrated data management

Draws on elements from:

- Online Navigation and Entry System in San Francisco
- Combined Homelessness and Information Network in Greater London
- Pathway Accommodation and Support System in Ireland

Overview

The purpose of this system option is to increase sector collaboration by enabling agencies assisting people who are homeless to share data and coordinate services through a common data management platform. The platform would be built around a single ‘housing record’ for each person, being a historical record of their different service interactions. The setup of this system could vary in terms of scale, i.e. having a single system covering Wales (e.g. Street Homeless Information Network pilot), or systems covering communities that are commissioned independently, but are ‘inter-operable’ (e.g. HMIS’s across the United States). However, if areas are not mandated to form community data management systems, then this option simply becomes a variation on Option 1, as community systems would need to be situated within a much larger national framework in order to harmonize data. A ‘core data’ element from each housing record would form the basis for Welsh Government statistics and research in the secure data infrastructure.

Figure 8: Integrated data management systems for the sector (Integrated)
How will it work?

The system is configured around a web-based data management solution that can be accessed by approved housing service providers. An example of this approach to data management and collection is Street Homeless Information Network (SHIN) in Wales and Combined Homelessness and Information Network (CHAIN) in Greater London. Within the system, a person has a centralised record that is built up over their various interactions with different housing service providers. Following similar developments in health and social care, for example the Summary Care Record in England and the Welsh GP Record, having a centralised care record means that the person receiving assistance will not have to repeat their information. Furthermore, like care records, for an agency to be able to read someone’s full housing record, they must have permission from the individual.

This system is designed for use by multiple agencies working across a range of service provision types, therefore the data being collected as part of the housing record would be under two parts: (1) the ‘core data’ standard covering universal information to be completed by all agencies, and which provide more detail on the homeless persons’ circumstances and their housing outcomes; (2) agency specific data, such as decisions under the Act for Housing Options teams or HSG Outcomes Framework data, thereby enabling the system to be used for day-to-day data management purposes. Neither of these data elements would be visible to other users of the system unless permission were given by the person whose record it is. The only data from the housing record that can be accessed without permission would be basic details needed to verify a person’s identity in order to assign a record to them, e.g. name, date of birth, last settled address, National Insurance Number. Following the example of CHAIN, when a ‘new’ person interacts with a service, the caseworker would search for the person on the system using their personal details.

Data would be entered onto the system through service specific interfaces, which would be able to accommodate any specific functionality required, e.g. fields for recording assessment under the Act. The goal would be for all organisations to use the system to record data about their interactions with a person; however, some local authority Housing Options teams may have well developed data management systems and may not want to use the centralised system as their main data management tool. In these instances, API interfaces could be used to extract data from the Housing Options system into the centralised system.

Rather than having an extract of the entire system, which would be inherently complex given the number of organisations using the system, Welsh Government would be provided with de-identified ‘core data’ elements; though some of the service specific data elements could be shared, such as decisions under the Act, to enable the legislation to continue to be monitored. Welsh Government would be responsible for cleaning the data, before depositing it in the secure data infrastructure. Personal identifiers would undergo the trusted third-party process before being deposited in the secure data infrastructure to enabling linkage to data outside of the integrated system. The AFON system current employed by Welsh Government can handle secure file transfers and can be adapted to handle the upload/download of the ‘core data’.

As many organisations would be contributing data then validation and maintenance of data quality would need to be an on-going function—rather than something which occurs in the extraction and transfer stage. In the example of CHAIN, and in the United States, where HMIS’ adopt a data platform approach, the provider for the software is also contracted to manage data quality. This could be achieved through (1) setting mandatory fields, (2) having a standard for missing data, e.g. 90 per cent complete, and (3) the generation of data quality reports fed back to data providers.
The functioning of the actual data management system, i.e. how data are ultimately stored, whether cloud based or if a data warehouse is used, will be left to the vendor to determine in order to make the procurement process competitive and value for money—rather than dictating unnecessary IT infrastructures. Rather than having a single procurement exercise covering Wales, as was the case for the Welsh Community Care Information Service, a series of community level integrated data management systems could be procured, rather than a ‘one-size-fits-all’ system. This latter approach is adopted by the United States Department of Housing and Urban Development and the various instances of community level homelessness information management systems operated by service they fund. Enabling local areas in Wales, e.g. local authorities or groups of authorities, to procure their own data management systems would enable them to design systems to meet their local/collective needs, whilst also sharing any resource implications. ‘Localist’ arguments around data collection and use suggest that the quality and relevance of data decays with distance from the source (Cornford et al. 2013), therefore area specific systems can potentially preserve this relevance.

What are the outputs from this system?

Whereas other system options lead to changes in service delivery indirectly, through improving evidence, this system directly seeks to impact how services are delivered, by integrating data from a wide number of services into the day-to-day activities of client-facing staff. Increased access to data can potentially lead to increased collaboration in the delivery of care, and also, as a consequence, person-centred provision of care. The opportunity exists for other public services involved in a person’s care to be granted read/write permission in order to improve service delivery. For example, the Housing (Wales) Act makes provision for sharing of information between social-care and housing services; having a shared data management system would be a mechanism for achieving that aim in a secure environment by granting social workers, for example, permission to view data—where the person consents to this level of access.

The collection of data from several different organisations within the housing sector will increase the coverage of data on homelessness in Wales and thereby improve understanding of changes in the number of people experiencing homelessness and reasons for homelessness. In order to open up data access and exploration to organisations outside of the system, including the general public and policy-makers, data dashboards based on aggregate live data (e.g. Knoxville HMIS Community Dashboard) could provide summaries of homelessness across different geographies and for different topics/policy concerns—such as rough sleeping, or youth homelessness.
### Table 18: Summary of technical elements of a centralised integrated data management option

<table>
<thead>
<tr>
<th>Design element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose and new ways of working</strong></td>
<td>This option enables increased collaboration by facilitating the sharing of information across providers, whilst putting the person in control of their data. If local areas are empowered to create their own shared data management systems, then this has the potential to lead to greater investment in the system.</td>
</tr>
<tr>
<td><strong>Data architecture model</strong></td>
<td>Web-based data management solution shared by different agencies that can be used for their data management purposes. Data are structured around a ‘housing record’ that relates to all interactions between the person and housing services. Data would be downloaded by Welsh Government and cleaned before being deposited in the data infrastructure.</td>
</tr>
<tr>
<td><strong>Supporting change</strong></td>
<td>For a pan-Wales approach, Welsh Government would either procure a software provider or seek to appoint a lead organisation to procure on their behalf. For area level data management systems, a grant mechanism could be used much like the United States funding of Continuum of Care, stipulating what the ‘core-data’ element should be and enabling communities to design their own systems around that—however there is no guarantee that communities will form area wide data management systems unless this is mandated.</td>
</tr>
<tr>
<td><strong>Governance</strong></td>
<td>A collective governance structure would have oversight of both the data management solution, and the data infrastructure, as with the data warehouse.</td>
</tr>
<tr>
<td><strong>Data quality</strong></td>
<td>Quality of data would be checked on entering it into the system. Part of the data governance remit would also be to maintain data quality across the system. Prior to the data being made publicly available via the secure data infrastructure, Welsh Government would download and undertake cleaning of ‘core data’.</td>
</tr>
<tr>
<td><strong>Ethical and legal</strong></td>
<td>As the data are meant to be accessible across the system, then consent for data sharing would be required. However, a two-teared approach is suggested: (1) consent for detailed access to case file, (2) legal gateways to enable processing and use of ‘core data’ for statistical and research purposes.</td>
</tr>
<tr>
<td><strong>Data sharing</strong></td>
<td>Permission to view data must be granted by the person whose housing record it is. The only data visible without permission would be basic details that indicated that the person had an entry on the system.</td>
</tr>
<tr>
<td><strong>Data access</strong></td>
<td>Use of a single system can reduce the need for data exchanges between organisations. For general access to the data for research purposes, data are available in the secure data infrastructure. National statistics would be generated by Welsh Government, in the process of preparing the data for depositing in the secure data infrastructure. Data dashboard would provide ‘live’ or near live updates of homelessness in Wales.</td>
</tr>
</tbody>
</table>
Table 19: SWOT analysis of centralised integrated data management (Option 4)

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>People have control over who can see their data, thereby shifting the balance of power from agencies to the person. Validation would be incorporated into the data management system, including error messages when invalid information is entered, and ongoing validation/quality assurance provided by the vendor for the system. Reduced duplication of data entry if the same records can be accessed by other services, and reduced resources required to prepare referrals as these could be made through the system. Data dashboard allows the sector to visualise and explore data on homelessness in real-time.</td>
<td>Data management data are complex, meaning that a simple download form the system will provide an excess amount of information. A ‘core data set’ of key data from across the different organisation types would need to be created and in the context of service delivery. Heavily reliant on continued organisational investment in the system, given that only HSG funded services have an obligation to provide data.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Threats/challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Could potentially increase coordination across the sector if all agencies involved in the care of a person work from the same data/information. Supports data management functions if there is a national move toward Housing First model. Potential to increase the scope of the system to include tracking of emergency accommodation for the homeless, e.g. shelter spaces or transitional housing spaces. Having data management systems covering different geographies would enable them to be designed to meet local needs.</td>
<td>The lack of a mandate to use such a system threatens the integrity of national data on homelessness, i.e. authorities/organisations could choose to collect data in other ways, or simply not participate, yet more systems to collect their data. In effect, funding area data management solutions could become Option 1, where local authorities are at liberty to design their own local systems but provide consistent data to Welsh Government. Pan-Wales approach forces collaborative cultural change on the sector which may not lead to sustainable uptake. There may be a long lead in time before the system becomes functional due to the need for consultation and procurement before the solution is developed and tested; could potentially mean that the end product is based on out-dated sector requirements. If the software solution is provided under contract via a ‘for-profit’ supplier, then there is likely to be a degree of inflexibility in the end product whereby additions to the system will incur further costs. For a pan-Wales approach, finding a consensus amongst all agencies using the system on the functionality of the interface will be time consuming. New users to the system may not be able to gain the same benefits as early adopters, as they will have to adopt a system that they were not part of designing. Local authorities and other agencies may be tied to IT providers, or have developed bespoke IT solutions that meet their local needs and achieve fairly similar goals.</td>
</tr>
</tbody>
</table>
Part seven: Way forward

Data are most effective when embedded in national homelessness strategies as it implies a commitment to the use of data within evidence-based policy and practice (Edgar et al. 2007). Though the new homelessness strategy for Wales states that data are important in understanding the scope and response to homelessness in Wales, a more concrete plan for the role of data linked to the strategy should be the next steps in determining the homelessness data future for Wales. In terms of the actions laid out in the homelessness strategy for Wales, a new data system could conceivably:

- Generate population level data as the basis for early prevention efforts
- Enable data linkage in order to quantify the impact of housing/homelessness on other policy areas, and vice-versa, and thereby increase the scope of homelessness prevention as more than an ‘issue’ of housing
- Provide mechanisms for accessing and engaging in research using individual level data, in order to foster a culture of learning amongst policy-makers, commissioners, and client-facing services
- Facilitate data sharing, which itself can contribute to collaboration, either through increased harmonisation of data and/or by providing technological solutions that enable data sharing/shared data
- Integrate data from across a wide range of services at the person level in order to provide person centred case management, whilst also reducing the need to repeat service histories

Each option realises these capabilities to different extents. For example, Option 1 (individual level statutory data reporting) may not be a platform for data sharing, however it certainly does not prevent this practice from happening—if organisations and services have the willingness to do so. Harmonizing data collected by statutory services and the homelessness sector can in fact facilitate data sharing through a common data language, and so Option 1 should certainly not be considered less meaningful as a mechanism for collaboration than Option 4, say (a shared data management system). The system design options should therefore be taken as starting points for a conversation on how data can facilitate different aspects of the task of ending homelessness in Wales.
References


Appendix A: Data system summaries

<table>
<thead>
<tr>
<th>Data system:</th>
<th>Lifelong Learning Wales Record (LLWR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>The LLWR contains individual level data relating to post-16 learning provision in Wales, including: Further Education, Work Based Learning, Community Learning, Higher Education at Further Education, and Welsh Language learning. Similar data collections are undertaken in England under the Individualised Learner Record, however unlike English ILR data, the LLWR has been comparatively under-utilised for research and linkage purposes. The lack of research uses of the LLWR may change in the future as data have been deposited in SAIL for linkage and research studies. The LLWR is a long running data collection at Welsh Government, with data being available from 2003/04 onwards.</td>
</tr>
<tr>
<td>Purpose:</td>
<td>Used for planning, funding, monitoring and quality assurance of post-16 provision.</td>
</tr>
<tr>
<td>Geographical coverage:</td>
<td>Wales</td>
</tr>
<tr>
<td>Population:</td>
<td>Learners in Wales enrolled on learning activities/programmes funded by Welsh Government.</td>
</tr>
<tr>
<td>Update frequency:</td>
<td>Continuous, however in reality data are submitted in batches dependent on reporting periods.</td>
</tr>
<tr>
<td>Data collection method/storage:</td>
<td>Data are submitted to Welsh Government in XML format.</td>
</tr>
<tr>
<td>Data presentation/sharing:</td>
<td>Primarily data are used for internal funding purposes, however data also form the basis of statistical outputs on learner numbers and characteristic in Wales. Reconciliation reports are provided to data providers in order to maintain data quality and ensure data submitted at the end of the financial year are timely and accurate. Data have been deposited in SAIL for research purposes. Privacy notice provided to learners indicates that data are used for linkage and research.</td>
</tr>
<tr>
<td>Information collected:</td>
<td>Personal information are collected, including:</td>
</tr>
<tr>
<td></td>
<td>- Name</td>
</tr>
<tr>
<td></td>
<td>- Address</td>
</tr>
<tr>
<td></td>
<td>- Telephone number</td>
</tr>
<tr>
<td></td>
<td>- National Insurance Number</td>
</tr>
<tr>
<td></td>
<td>- Gender</td>
</tr>
<tr>
<td></td>
<td>- Surname at 16</td>
</tr>
<tr>
<td></td>
<td>- Date of birth</td>
</tr>
<tr>
<td></td>
<td>- National identity</td>
</tr>
<tr>
<td></td>
<td>- The last school attended</td>
</tr>
<tr>
<td></td>
<td>- Year left school</td>
</tr>
<tr>
<td></td>
<td>- Unique Learner Number that enables linkage to other education data sources. In addition to personal data, data relating to the learning activities undertaken by the learner are collected. Data are split into 4 main areas:</td>
</tr>
<tr>
<td></td>
<td>- the Learner (LN) dataset which includes information about the learner such as name, date of birth, ethnic origin and gender;</td>
</tr>
<tr>
<td></td>
<td>- the Learning Programme (LP) dataset which gives information about the current programme of learning being undertaken by the learner and any characteristics which may change over time;</td>
</tr>
<tr>
<td></td>
<td>- the Learning Activity (LA) dataset which collects data on the individual activities or courses undertaken by the learner on his/her programme of learning, and</td>
</tr>
<tr>
<td></td>
<td>- the Award (AW) dataset, which provides information on the awards for which the learner is entered and those achieved. Each learner has a single entry in the LN dataset, but may have multiple programmes, activities and awards associated with them.</td>
</tr>
<tr>
<td>Strengths:</td>
<td>As the data are collected for operational and funding purposes, the data are of good quality and there are well developed validation procedures in place. Modular nature of the data collection has meant that new data collections have been added over time.</td>
</tr>
<tr>
<td>Weaknesses:</td>
<td>Data are deleted after 10 years in the case of Further Education and Community Learning, and 5 years for Welsh for Adults. Data structure is complex, which has grown in complexity over time as additional data sets have been added. Resource intensive in maintaining database, validating data, and reporting.</td>
</tr>
<tr>
<td>Reference material:</td>
<td><strong>Main webpage:</strong> <a href="https://beta.gov.wales/further-higher-education-data-collections">https://beta.gov.wales/further-higher-education-data-collections</a></td>
</tr>
</tbody>
</table>

Data system: SSDA903 collection, superseded by Looked After Children (LAC) census

Summary

Under Social Services and Well-being (Wales) Act 2014 local authorities are required to assess children in need and their families. From 2016-17 onwards individual level data in relation to looked after children are collected during the year through the Looked after children census. This replaces the existing datasets Looked After Children (SSDA903), Adoptions of Looked after children, Educational qualifications of care leavers and Care leavers on their 19th birthday. As part monitoring social services functions under the Social Services and Well-being (Wales) Act 2014, several variables are used from the LAC Census in order to form the basis of performance measures.
| **Purpose:** | Provide Welsh Government with the necessary information to evaluate the outcome of policy initiatives and to monitor objectives on looked after children, both during their time in care, and on reaching adulthood. |
| **Geographical coverage:** | Wales. |
| **Population:** | Children who are looked after or in receipt of care from local authorities. |
| **Update frequency:** | Annual. |
| **Data inputting/collecting organisation(s):** | Local Authorities. |
| **Data collection method/storage:** | Data are uploaded in Excel format to Welsh Government via web-based secure uploading site, Afon. Afon undertakes validation of data, for example looking for missing data and also incorrect/impossible responses. Welsh Government carry out further validation using SAS software in order to check that the episodes of care are complete. Afon generates data quality reports that enable authorities to address issues prior to the data being sent to Welsh Government via the system. |
| **Data presentation/sharing:** | Data form the basis of statistical outputs. Access to the data has been granted for research purposes in the past. Welsh Government provide a privacy notice which outlines how and what data are shared between local authorities and Welsh Government. |
| **Information collected:** | Data within the SSD403/LAC relate to episodes of care, and are broadly split between variables relating to the child and those relating to their periods in care. Personal data relating to the child includes:  
- Child identifier unique to that child in that authority  
- Sex  
- Date of Birth  
- Ethnic origin  
- Disability code  
- Child's home postcode i.e. the address from which they entered the care system  
Data which relates to the episode of care includes:  
- Dates episode commenced and ended  
- Reason for episode  
- Legal status  
- Child in Need code (category of need)  
- Type of placement (foster care, residential care, placement with parents, etc.)  
- Reason episode ceased  
Each episode represents a period of being looked after under the same legal status and in the same placement. A new episode is started when either the legal status or the child's placement changes. |
| **Strengths:** | Data collection is based around episodes therefore enabling longitudinal analyses. Furthermore, as episodes are the building blocks for periods in care, analysis can be tailored to suit different research purposes. Data feed into performance measures for Social Services. |
| **Weaknesses:** | Child identifiers are only unique within the same authority therefore movement of children across authority boundaries cannot be accounted for. Confusing situation in relation to ownership of the data held by Welsh Government as permission needs to be sought from local authorities before Welsh Government will agree to provide access to data they hold. |

**Reference material:**

**Research/data use examples:**
**Data system:** Welsh Community Care Information System (WCCIS)

**Summary:**
Social care services are often provided by a range of organisations which each use their own, sometime incompatible ways of collecting data, for example using paper-based forms. The new WCCIS integrates information in a single national system that makes it possible - on a need to know basis - for information to be shared securely between health and social care services. The Welsh Government provided £6.7m in capital funding for the initial set up costs for the WCCIS and has made a further £4m available to support its implementation via the Integrated Care Fund. The procurement was managed by a Joint Procurement Board consisting of representatives of NHS Wales and local authorities from the Wales System Consortium. Bridgend County Borough Council led the process.

**Purpose:**
Support care in the community, helping to reduce unnecessary hospital stays. Carers will have easy access to all the important information they need and the individual won't have to repeat their details time and again to different carers, leading to better consistency of care. Sharing of information meets the required by the Integrated Health and Social Care, Social Services and Well-being (Wales) Act 2014.

**Geographical coverage:**
Wales.

**Population:**
People in receipt of care and social services in Wales.

**Update frequency:**
Live data.

**Data inputting/collecting organisation(s):**
Range of data inputting organisations/people, including those related to NHS Wales and local authority social care providers:
- Adult’s services
- Children and Family services
- Financial services
- NHS mental health services
- NHS community health practitioners
- Various therapy providers, e.g. occupational therapy
- Integrated teams across health and social care

**Data collection method/storage:**
The WCCIS platform was created by Microsoft, so its appearance will already look familiar to professionals in the Health and Social Care sector.
The system benefits from a user-friendly menu structure, detailed search functions and electronic client records that are easy to navigate.
Users can make referrals, manage their caseloads and share information.
In compiling the specification of requirements there was a high degree of commonality between processes in health and social care services and more than 70% of the requirements were common to both areas.

**Data presentation/sharing:**
Frontline staff can access and record information ‘on the go’ using mobile devices such as tablets and smart phones.
There is the option for data for an organisation to be presented as a dashboard for quick interpretation.

**Information collected:**
- Clinician details
- Patient identifier
- Test results
- Notes and images
- Referrals
- Assessments
- Care planning
- Case and contact management
- Diaries/scheduling

**Strengths:**
It will enable professionals across health and social care to work together within their local area, the region and across the whole of Wales.
Make better informed decisions as all relevant information is in one place
WCCIS will support service change already underway to reshape social services and health care in Wales
Reduce duplication in data capture and enable easier information sharing by the creation of a single citizen record for both health and social care

**Weaknesses:**
Cost and lengthy period associated with procurement and transfer onto new system.

**Reference material:**
Main webpage: [http://www.wales.nhs.uk/nwis/page/66175](http://www.wales.nhs.uk/nwis/page/66175)
Main webpage for developer: [https://www.careworks.co.uk/portfolio-item/wccis](https://www.careworks.co.uk/portfolio-item/wccis)
**Data system:** Supporting People, Wales

**Summary**
Launched in 2003, the Supporting People (SP) programme is a Welsh Government programme that helps approximately 60,000 people each year to live independently. SP is an early intervention programme which aims to prevent people from becoming homeless. There are two key elements to support provided by Supporting People: long-term and short-term. Long-term support is designed to help people retain or gain independence and avoid the need for more costly interventions such as entering care. Short-term is intended to be more preventative of housing crisis and avoiding homelessness. Overall, the approach of the SP programme is in keeping with the aims of The Housing (Wales) Act 2014 to provide preventative action to avoid homelessness.

**Purpose:**
The collection of client data is aimed at monitoring the activities funded as part of SP, and in contributing to the longitudinal evaluation of SP through data linkage.

**Geographical coverage:** Wales.

**Population:** People in receipt of SP funded services, with specific service groups being:
- substance misuse (drug/alcohol)
- domestic violence
- families
- generic floating support
- mental health issues
- people over 55 years
- young people, 16-24 years

**Update frequency:** Annually.

**Data inputting/collecting organisation(s):** Local Authorities.

**Data collection method/storage:** Data are collected by local authorities. Excel spreadsheet is used in order to collect data relating to SP outcomes.

**Data presentation/sharing:** Data collection spreadsheet automatically creates summary tables and summary graphs. Individual level data have been shared by local authorities with SAIL in order to form part of the feasibility study in the use of data linkage to evaluate SP, which subsequently led to the full evaluation. Local Authorities are required to provide a disclaimer relating to how data will be used in order to evaluate the programme.

**Information collected:** Outcomes framework collects information on:
- Length of service and type of service
- Service user needs (3 needs to be identified, lead need)
- Dates support started and ended, and the date of any reviews
- Achievement against a number of areas: feeling safe; contributing to safety; managing accommodation; feeling part of the community; managing money; engagement in education/employment; health (including mental health)
- Reason support was ended

Personal data specified for data linkage include:
- NHS number (where available)
- First name
- Last name
- Address
- Postcode
- Date of Birth
- Gender

**Strengths:**
Outcomes framework could provide a measure of programme success in terms of improvements in client circumstances. Collection of individual level data has meant that Welsh Government have been able to engage in a wider data linkage evaluation of the SP programme to explore its effects on public services use.

**Weaknesses:**
Numerous stakeholders found the outcomes framework to be insufficient as a means of assessing the impacts of the SP fund (Public Accounts Committee 2018). Supporting People outcomes data is not being provided to SAIL so will not be available for analysis; current research uses of the data are therefore limited to the entries and exits from SP support, rather than the being able to account for the person-level transformative nature of that support, and the effects of this on wider service usage.

**Reference material:**


**Research/data use examples:**
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>Pupil Level Annual School Census (PLASC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>PLASC is mandatory for maintained schools in all sectors and is a census of pupils taken in January each year. Schools are required to report on school level and pupil level data to provide an invaluable research base to better inform educational policy and practice in Wales. The PLASC is usually combined with attainment data in order to form the National Pupil Database, now known as the National Data Collection. In order to link the PLASC to other data collections, for example the Children in Need Census, the Unique Pupil Number is used. PLASC data are available from 2003/04 onwards.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>Monitor learner numbers and characteristics in Wales.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Wales</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>All pupils registered on the census day.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Annual</td>
</tr>
<tr>
<td><strong>Data inputting/collecting organisation(s):</strong></td>
<td>Local Authorities submit to Welsh Government; Schools report to Local Authorities.</td>
</tr>
<tr>
<td><strong>Data collection method/storage:</strong></td>
<td>Data uploaded to Welsh Government via web-based secure uploading site, Data Exchange Wales initiative (DEWi). Files are uploaded to DEWi by schools and are validated by the system according to the collection type. Validation looks at the format of the file, what data is present, and checks that the data conforms to defined sets of values. LEAs then access the system to assist in the resolution of any issues found in the files from schools within their authority. Further validation of data is undertaken at Welsh Government by comparing submission to previous years’ in order to query changes which appear greater than historical trends.</td>
</tr>
</tbody>
</table>
| **Data presentation/sharing:** | Data provided by the PLASC are used for:  
- the fair & accurate distribution of funding to school through the Welsh Local Government Revenue Settlement,  
- the update of the National Pupil Database (NPD) which links pupil’s records to assessment and allows schools, LEAs and central government to quickly and accurately carry out previously complex statistical analyses,  
- ensuring a greater accuracy of pupil level data to help schools to adhere to the principles of the Data Protection Act (DPA).  
Data form the basis of routine statistical outputs and aggregate data are publicly available on StatsWales. Access to the data has been granted for research purposes in the past, with data having been deposited in SAIL for linkage studies. PLASC data provides much of the contextual data that is used in the All Wales Core Data Sets (AWCDS). The AWCDS are delivered to schools (primary, middle and secondary) and LAs as useful contextual analysis tools for consideration of accountability and planning within a school’s individual circumstances. The AWCDS are also used by Estyn as part of its inspection materials. It is important that the PLASC data is accurately reported by schools to ensure that these tools are useful and appropriate to the school and LA. |
| **Information collected:** | PLASC data covers school, class and pupil level, including:  
- School identifiers  
- School characteristics (type, language, number of free school meal children)  
- Classes data (name, number of teachers, level of Welsh teaching in class)  
- Personal identifiers, including Unique Pupil Number  
- Pupil characteristics (Local Authority care, free school meal eligibility, ethnicity, first language)  
- Special Educational Needs |
| **Strengths:** | Validation of data when inputting into DEWi allows Schools to directly address as many data errors as possible. Data are fed back to Schools as part of the All Wales Core Data Sets. This therefore provides Schools with information/analysis which they may not have otherwise been able to produce themselves. |
| **Weaknesses:** | Being a census, then school moves and other within-year changes to the child’s circumstances are only registered in the following year. |

**Reference material:**


**Research/data use examples:**

### Data system
Common Transfer System and 'school2school'

### Summary
The Common Transfer System (CTS) and school2school system form part of the pupil transfer/referral system in England in Wales. The school2school system enables maintained and independents schools to post information about pupils being transferred, so that the recipient school/local authority can retrieve the data. In Wales, there is a statutory obligation to transfer information using the CTS via school2school. The Common Transfer File is a standardised file format containing information about the pupil, including assessment results and attendance data, which enables sharing across different school management information systems. School management information systems, or local authority systems, should be able to generate the required transfer files by automatically drawing down data from the local system—rather than manual input.

### Purpose
Enable the secure transfer of information about children when a child moves schools.

### Geographical coverage
Wales (and England)

### Population
Children in maintained and independent schools

### Update frequency
Ad hoc usage when child transferred

### Data inputting/collecting organisation(s)
Schools or local authority administrators where

### Data collection method/storage
- Guidance documents suggest that school management information systems be designed to output the transfer file.
- Data uploaded to secure school2school website.
- Files can be sent to the new school within the site by quoting the Local Education Area and school establishment code.
- Where the pupil's destination school is not known, the school2school system has a separate secure area, the 'Lost Pupil Database':
- Data can be stored on the school2school system, for example in instances where the pupil moves outside of the area able to receive the common transfer file. Files can be retained, should the pupil return to a maintained school later.

### Data presentation/sharing
- Data are integrated into the recipient school's management information system.
- The Lost Pupil Database can be searched in order to locate 'missing' pupils, if they are transferred and no CTF is received by the recipient school. If the pupil is located, their file can be downloaded.
- Only when the recipient school have verified that the pupil is enrolled at their school should they then download the data from school2school.
- Reports can be run from the school2school system so that data providers can monitor the number of files uploaded to the system.
- Welsh Government undertake data linkage between pupils in the Lost Pupil Database with data from the PLASC in order to identify whether pupils are enrolled at another school in Wales.
**Information collected:**

Items in the transfer file include:

**School Details -**
- Sending School's LEA number.
- Sending School's Establishment Number.
- Receiving School's LEA Number.
- Receiving School's Establishment Number.
- Pupil School history (if available).
- Academic year.

**Pupil Details -**
- Unique Pupil Reference Number.
- Surname.
- Forename.
- Date of Birth.
- Gender.
- Ethnic Code.
- National Identity.
- Welsh Language Ability.
- SEN Provision (status, primary and secondary need).
- Eligibility for Free School Meals.
- Free Milk Eligibility.
- In Care Indicator.
- EAL Acquisition.
- The address where the pupil normally resides.
- Surname for at least one contact person and details of their relationship to pupil for.
- An indicator that medical information exists that may be relevant to the pupil's new school.

**Attendance Data -**
- School Year.
- Number of sessions possible (half days).
- Number of sessions attended.
- Number of authorised absences.
- Number of unauthorised absences.

**Strengths:**

- Enables the secure transfer of files.
- School2school can be used to message others within the network of users.
- Lost Pupil Database can be used to track pupils missing from the education system.

**Reference material:**

Main webpage(s) for Common Transfer System in Wales:


### Data system:

| Summary | The Ffynnon system was commissioned by Welsh Government and designed to be a pan-Wales performance management software solution for use by local authorities, fire and rescue services, and national park authorities. The software went live in 2007/8, and cost roughly £4 million in software development and ongoing support over a 4-year period. The defining feature of the Ffynnon system was that it was not compulsory to use, and services could adapt the system to suit their purposes—organisations were assisted by the developer to customise the system. The funding for Ffynnon ended in 2014. The development of the system was guided by a project board, bringing together stakeholders from a range of public sector users. |
| Purpose: | Provide a software solution to create a performance management culture amongst public service in Wales. |
| Geographical coverage: | Variable; local authority level or cross local authority in the case of national parks and fire services. |
| Population: | Variable: entities could cover individuals (such as South Wales Police crime data), or aggregate data (school level reports) |
| Update frequency: | Variable |
| Data inputting/collecting organisation(s): | Local authorities, fire and rescue services, and national park authorities |
| Data collection method/storage: | Ffynnon was designed to be web-based, facilitating remote access and data upload. The systems are hosted within a collaborative environment. |
| Data presentation/sharing: | System was designed to enable flexible presentation of data, such as performance dashboards. There was the ability to transfer data within an organisation and to others. Data entered into the system was managed locally. |
| Information collected: | Information collected using the Ffynnon system varied, given the high degree of flexibility in the system. As the system had a performance focus, the system could potentially include indicators, risks, projects and milestones, actions. Example uses of the system included a pilot to transfer crime data from their administrative system into Ffynnon in order to populate dashboards. |
| Strengths: | Users of the system reported that it reduced time and staffing resources needed to collect and collate performance reports. Dashboards made the data easier to interpret. There was a general sense that the system led to a step change in performance management practices—where the system was utilised productively. System improved collaborative working for some organisations. |
| Weaknesses: | Cost-benefit of system may have been equal when compared to organisations developing their own performance management system; Welsh Government managing the contract was one area of savings, whilst the flexibility of the system and ability to have additional functions was value for money. Flexibility/developmental nature of the system may have been a disadvantage. Uncertainty surrounding the continued funding of the Ffynnon system may have led to users purchasing their own system, over which they would have greater control and certainty. |

### Reference material:

### Data system:
Street Homeless Information Network (SHIN) pilot

### Summary:
SHIN takes its cue from the CHAIN database and has the aim of establishing a network of organisations that support and record information on rough sleepers in a consistent manner. CHAIN may provide more reliable estimates of the number of rough sleepers in Wales, whilst also making it easier to analyse the problems faced by rough sleepers and drive service improvement. SHIN is in the initial stages of its development, and is currently being piloted within the Wallich. As the system is still being constructed it is difficult to identify positive/negative aspects and therefore those identified are possible issues, given current information on what the system will potentially do and what data it will collect.

### Purpose:
Establish a network of organisations that support and record information on rough sleepers.

### Geographical coverage:
Wales.

### Population:
People who have been seen rough sleeping by outreach teams, and those who have visited services reporting that they are a rough sleeper. There may also be some inclusion of those involved in 'street culture' activities such as drinking and begging.

### Update frequency:
Continuous.

### Data inputting/collecting organisation(s):
Voluntary and statutory homelessness service providers who have contact with rough sleepers.

### Data collection method/storage:
Web based information sharing platform.

### Data presentation/sharing:
- Client data is accessible via a central database.
- Dashboards for data inputting organisations enabling real-time analysis of their own service user population.
- Potential for view only access to the data.
- Detailed information about interactions with services are hidden from general view.
- Reports can be created from data, however limited to own organisations case-load.

### Information collected:
- Basic identifying and demographic data
- Support needs (for example mental health problems or substance use)
- Circumstances prior to rough sleeping
- Contacts made with outreach workers on the street (i.e. instances of people being seen rough sleeping)
- Key outcomes, actions and events (including arrivals at and move on from accommodation)

### Strengths:
- Potentially enable more valid counts of rough sleepers in Wales to be undertaken, replacing or adding to the single night/week counts.
- Bring together detailed information on the support needs and issues faced by rough sleepers in Wales.
- Potentially provide an electronic case management system for smaller organisations who may not have had resources to create their own.

### Weaknesses:
- Data collection by outreach workers will fluctuate with resource availability.
- Non-participation of some organisations may affect data coverage when attempting to provide national statistics.
- Concerns raised by the Wallich in the wording of the National Assembly for Wales enquiry into rough sleeping that SHIN could be used to verify a rough sleeper's story and prove they were a rough sleeper – this relates to the potential punitive use of these forms of centralised systems.

### Reference material:
### Data system
- **European Structural Fund participant database**

### Summary
The European Structural Funds (ESF) provide funding for key labour market interventions aimed at developing human capital in order to promote employability and increase employment skills in Wales. The fund itself is overseen by the Welsh European Funding Office. ESF funding has been made available in 2 tranches: firstly between 2007 and 2013 under the Convergence and Competitiveness programmes, and from 2014 to 2020. The ESF 2014-2020 funding has 3 priority areas: Poverty, Skills for Growth, and Youth Employment and Attainment. Each funding area of Wales (e.g. West Wales and the Valleys, East Wales) have specific objectives under each of these priority areas (or axis).

Each beneficiary of ESF-funded project/programme is responsible for creating a Monitoring and Evaluation plan (M&E plan). The M&E plan should set out who is responsible for collecting data, to whom and how often it will be reported, how it will be used in the management of the operation and how data quality will be managed and assured. In addition to data needed for reporting purposes, beneficiaries are also encouraged to collect other management data in order to inform their own operations.

In addition to monitoring data collected when participants leave a programme, a survey of former participants has been conducted: the ESF Leavers’ Survey. The surveys have been administered via interviews, and were conducted in 2009, 2010, 2011, 2012 and 2013. The Leavers’ Survey has since been replace by the ESF participants’ Survey for the 2014-2020 funding period.

### Purpose
- **Organisations delivering EU-funded projects are required to provide data to enable the progress of projects, including performance against key targets, to be monitored.**

### Geographical coverage: Wales.

### Population:
- Persons benefiting directly from an ESF-funded intervention. Only those persons who can be identified and have been asked for their personal data and for whom specific expenditure is earmarked are reported as a participant.

### Update frequency:
- Suggested annually, however varies with each beneficiary.
- Frequency of reporting is agreed between the beneficiary and WEFO and is set out in the beneficiaries Monitoring and Evaluation plan.

### Data inputting/collecting organisation(s):
- ESF-funded service providers, also known as beneficiaries.

### Data collection method/storage:
- Beneficiaries are provided with an Excel data spreadsheet for completion.

### Data presentation/sharing:
- Participants are issued with a privacy notice that explains why personal individual level data are being collected, what it is being used for and who will have access to the data.
- The beneficiary of funding can decide to provide participants with the option to consent providing certain information.
- Data from ESF-funded project participants for the period 2007 to 2013 are in the process of being deposited within SAIL.
- ESF 2007-2013 data have been used in order to undertake the ESF Leavers Survey.
**Information collected:**

Data are collected at two main points, once upon entry and again upon exit from an ESF-funded project/programme. If a person is accessing multiple interventions from the same funded project/operation, then only one entry is recorded against that person.

Personal identifiers that could be used to allow data linkage are collected, including:

- National Insurance Number
- Unique Learner Number
- Name
- Address (incl. postcode)
- Date of birth

Personal non-sensitive data must be recorded on the following for an ESF participant for them to be assessed against ESF targets:

- Gender
- Employment status
- Age
- Education level
- Household situation

Ethnicity, migrant status and disability are also requested, however as these a personal sensitive data then ESF participants are given the option as to whether they provide these data.

Data relating to the service provision is limited, but includes:

- Course/programme title
- Start and end date
- Whether the participant left the programme early
- The objective of the activity (aligned with specific objective of the project)
- The beneficiary (i.e. who provided the programme)

On completion of the ESF support, participants are assessed on the following in order to measure the impacts of the ESF participation:

- Improved Labour Market situation upon leaving
- Job searching upon leaving
- Entering education/training upon leaving
- Qualification gained upon leaving
- Work relevant certification upon leaving
- Entered employment, including self-employment, upon leaving
- Completing work experience placement
- Completing volunteering opportunity
- Training in maths, Science Engineering and Technology
- Increase in hours worked upon leaving
- Returning to work after a period of absence
- At reduced risk of becoming NEET
- In employment, including self-employment six months after leaving
- Improved Labour Market situation six months after leaving

**Strengths:**

Guidance has been provided to those involved in implementing a Structural Funds operation in relation to the effects of the GDPR on data collection.

Roles of data controller are clearly defined.

ESF-funded project guidance includes a document relating to monitoring and evaluation.

Not consenting to certain data does not prejudice access to services.

**Weaknesses:**

ESF Leavers’ Survey/Participants’ Survey and administrative data from the participant database have not yet been linked together.

Recording one entry where a person is accessing multiple interventions from the same project/operation may not give a complete picture of the interventions provided to a person.

**Reference material:**


**Research/data use examples:**

### Data system

Housing Stock Analytical Resource for Wales, UK Secure eResearch Platform (UKSeRP)

### Summary

The aim of the Housing Stock Analytical Resource (HSAR) is to bring together a range of data on the characteristics, fabric, condition, and energy efficiency of the housing stock in Wales. Welsh Government are exploring long term options for hosting data about individual properties contained by HSAR and are considering using a UK Secure eResearch Platform (UKSeRP). UKSeRP is a high powered, safe and secure e-research platform, developed by Swansea University, which delivers several benefits including considerable processing power, a version control facility, and the advantage of utilising the existing skills and experience of their data linking team.

### Purpose

Hosting of HSAR within a dedicated UKSeRP will enable the different elements of HSAR to be retained permanently in a linkable format for research purposes.

### Geographical coverage

Wales.

### Population

Predominantly properties within Wales; however may expand to include data on persons within properties (data acquisition depending)

### Update frequency

Ad hoc, as and when data become available. Ideally, data will be updated every 6 to 12 months.

### Data inputting/collecting organisation(s)

Data originate from various sources, including in house data sources (e.g. Rent Smart Wales) and large external data providers such as Valuation Office Agency, Land Registry, and HMRC data.

### Data collection method/storage

Data will be at the property level within the UKSeRP, however person level data (e.g. DWP/HMRC) will be held within SAIL.

### Data presentation/sharing

It is hoped that the Digital Economy Act 2017 will provide the legal gateway for data sharing between public bodies and Welsh Government. As the resource is to be at the residence level, linkage between data sources will be made through the Unique Property Reference Number (UPRN).

Welsh Government will undertake in-house analysis to inform policy. The intention is for there to be a facility for external users to access aggregate data and some non-sensitive property level data through the UKSeRP.

Data within UKSeRP can be linked to other data already contained within SAIL databank, thereby allowing linkage between properties and the people living in those properties.

### Information collected

Information from various sources will provide a range of data at the property level, including:
- Tenure
- Property type
- Property age
- Number of rooms (bedrooms, bathrooms)
- Price paid
- Dates of property transactions
- Heating intervention schemes run by Welsh Government

Person level data will potentially include:
- Benefits and income data (DWP)
- Personal income (HMRC)

Some of these data are speculative acquisitions, particularly those requiring DWP/HMRC data and data from the Office of National Statistics.

### Strengths

Placing HSAR within a UKSeRP will provide a permanent housing analytical resource, retaining this in an anonymised format, which therefore reduces privacy concerns.

Data linkage to other sources will enable novel policy questions to be addressed, for example the links between tenure/property conditions and educational outcomes.

### Weaknesses

As data are from disparate data sources, they will vary in terms of population coverage and quality. Access restrictions to different data sources will apply, particularly data which is being acquired through the Digital Economies Act as this legal gateway only applies to the use of data by other public bodies.

### Research/data use examples

### Data system:
Scottish statutory homelessness collections (HL1, PREVENT1, HL3)

### Summary
Paragraph 4.22 of the Code of Guidance on Homelessness states that where an application is made under the Homeless Persons legislation, this should be recorded through the electronic case-based HL1 returns to the Scottish Executive Housing Statistics branch. The HL3 system went live in 2001, with major revisions being made to the collection in 2007 following a consultation with local authorities, voluntary bodies and other parties. Revisions considered changes in the legislation and other emerging information needs. The collection has undergone minor changes considering further legislative amendments to remove priority need.

PREVENT1 was added based on the recommendation that local authorities, and Scottish Government, should develop a measurement for homelessness prevention work. The decision to have a separate collection rather than expand the HL1 was based on the fact that prevention work was happening ‘upstream’ of the HL3 collection, and the expansion of the current collection could potentially disrupt the HL1.

### Purpose:
Monitor the Housing (Scotland) Act 1987, including households housed in temporary accommodation.

### Geographical coverage:
Scotland.

### Population:
All applications to local authorities in Scotland which fall under section 28(1) of the Housing (Scotland) Act 1987 are collected via HL1 and PREVENT1.

All placements within temporary accommodation are reported under the HL3.

### Update frequency:
Quarterly.

### Data inputting/collecting organisation(s):
Local authorities.

### Data collection method/storage:
Data collection systems are not universal, i.e., each authority has its own collection system.

Upload data to Scottish Government, minus identifier data, via a secure file upload/data validation platform (ProcXed).

A detailed data specification is provided to local authorities which outlines the structure of authority collections, including possible values for different fields within the collection.

For HL3 (temporary accommodation) collect, there is a ‘stand-alone’ version of the data collection tool that can be used in lieu of a local IT solution. Stand-alone collection can be downloaded for use by authorities in order to produce summary analyses of input data and produce output reports that can then be send to Scottish Government.

### Data presentation/sharing:
Data form the basis of quarterly and annual statistical releases published by Scottish Government.

Data held at Scottish Government have been used to undertake ad-hoc analyses, some of which have been published on the Scottish Government website.

Data linkage studies have been conducted using HL1 and PREVENT1 data – see research examples below.

Personal data are not routinely shared with Scottish Government. Data linkage studies have required sharing of personal data with National Records Scotland for linkage to health numbers (much like the SAIL process).

Scottish Government indicate that they use National Insurance numbers to examine repeat homelessness. Homeless applications are considered repeat if the circumstances and family members recorded for the case are the same.

### Information collected:

#### Data collected under HL1 and PREVENT1:
- Applicant characteristics, including National Insurance number. NI number is one-way hashed at the point of submission to Scottish Government.
- Information about the households
- Reasons for approach
- Prevention activity/assistance
- Decisions
- Outcomes

#### Data collected under HL3 (Temporary accommodation):
- Date of offer
- Type of accommodation offered
- Take-up of offer
- Dates on entry/exit

Repeat homelessness is assessed by linking to prior applications for assistance. Previous cases where a homeless household has applied to a local authority can be associated with a current case. Encrypted national insurance numbers can also be used to link repeat persons across authorities.

### Strengths:
- Personal data are not shared with Scottish Government which reduces privacy risks.
- Guidance on data structure increases the consistency of collections across authorities.
- PREVENT1 collection has been extended to Registered Social Landlords to enable them to input data, thereby expanding the potential data points for understanding wider community prevention efforts.
### Weaknesses:

Repeat homelessness is not used more prominently as a success measure of the legislation. Lack of routine sharing of personal identifiers means that data linkage is time consuming as data have had to be re-acquired from all 32 authorities by Scottish Government.

### Reference material:

- Main webpage: [https://www.gov.scot/Topics/Statistics/Browse/Housing-Regeneration/RefTables](https://www.gov.scot/Topics/Statistics/Browse/Housing-Regeneration/RefTables)

### Research/data use examples:

- Author unknown (2015) Health and homelessness data linkage study - Fife
**Summary:**
In 2018 the Homelessness Reduction Act introduced new statutory duties for local authorities in England, specifically around the provision of prevention work and the relief of homelessness. These new duties meant that the previous data collection for statutory homelessness, the P1E collection based on the return of aggregate data, was no longer deemed fit for purpose. The design of H-CLIC has been shaped by the Homelessness Reduction Act and collects data that enable the effects of the Act to be monitored. When introducing the new system, £3 million was made available to support IT upgrades across England, or to otherwise support the transition from P1E to H-CLIC - £9,202 per local authority. Local case management systems are required to output data which fits the H-CLIC structure (See diagram below). These data are then uploaded to an online file transfer/data validation tool: DELTA.

**Data system:**
Homelessness Case Level Information Collection (H-CLIC) & DELTA.

**Purpose:**

**Geographical coverage:**
England.

**Population:**
People seeking assistance from statutory housing services in England where the applicant is believed to be homeless. Households who contact the local housing authority about their housing who are not homeless or threatened with homelessness and so no application is taken, are not to be recorded on H-CLIC.

**Update frequency:**
Quarterly collection, however, cases can be uploaded continuously.

**Data inputting/collecting organisation(s):**
Local authorities.

**Data collection method/storage:**
Each local authority designs its own data management system.
Upload to Department for Communities and Local Government (DCLG) using online secure transfer system and validation system known as DELTA.
Smaller authorities (fewer than 50 households per quarter) can input data directly into system via online version of the collection form in DELTA.
Data are uploaded in XML format in order to account for the relational nature of the data sets.

**Data presentation/sharing:**
Statistical reports and aggregate data published by DCLG.
Ministry of Housing, Communities and Local Government (MHCLG) are engaged in a data sharing exercise in order to obtain local authority homelessness data. The intention of the project will be to link H-CLIC to benefits receipt data, police crime data, and children’s educational attainment. Intention to link to children’s educational attainment warrants the collection for children’s personal details.
Personal data and main H-CLIC data are sent to separate teams within MHCLG. Personal data go through a process of de-identification by creating a unique identification number.

**Information collected:**
Data upload is structured around several linked dataset:
- Main application details, e.g., reasons for homelessness.
- Demographics of all household members.
- Support needs of household members.
- Actions taken under housing legislation, including prevention, relief, final duty.
- Placement in temporary accommodation.
- Personal data (name, date of birth, etc.) for each household member collected.

**Strengths:**
Splitting of personal data from main data provides an additional level of data security.
Online version of H-CLIC/DELTA for smaller authorities potentially saves them having to invest in complex systems.
Increased opportunities to undertake ad hoc analysis into the needs of people receiving statutory homelessness assistance.

**Weaknesses:**
Data structure is complex.
Each authority has had to develop its own approach to local data management; cost implication of providing funds to authorities to re-design their systems to be compliant.
Potential re-alignment of local data collection practices to conform to the new structure.
Provision of data is not legally required, nor is any funding affected by non-receipt of data, therefore authorities can choose not to provide data. Data imputation necessary where data missing,
Use of data by researchers outside of MHCLG is not permissible under current privacy notices/sharing agreements.
Reference material:


Examples of a software solution for use by local authorities:

https://www.housingjigsaw.co.uk/products/prah/

| **Data system:** | Alert system. |
| **Summary:** | The Alert system is a free tool for use by agencies in England, Northern Ireland, and Wales to support partnership working by providing a platform for making referrals between organisations. The system was originally designed to meet the Duty to Refer under the Homelessness Reduction Act, which applies to several public bodies, including social services, healthcare providers, probation services, and jobcentres. However, on recognition that referrals were an important part of service provision, other referrals can be made through the Alert system, such as modern slavery referrals and child safeguarding concerns. Alert is provided by Housing Jigsaw, itself a product of the National Practitioner Support Service and Housing Partners. Users can pay for an upgraded version with greater functionality. |
| **Purpose:** | Provide a secure platform for making inter-agency referrals. |
| **Geographical coverage:** | England, Northern Ireland, and Wales. |
| **Population:** | Enables referrals of different types. |
| **Update frequency:** | System is live so referral updates are real-time |
| **Data inputting/collecting organisation(s):** | Referrals can be made between specified public bodies where there is a statutory duty to refer, and also wider agencies. |
| **Data collection method/storage:** | System can be accessed via web-portal, meaning there are no user licences. Alert is also a module within housing Jigsaw products. |
| **Data presentation/sharing:** | Enables users to make referrals through sharing consistent information. Referring agency completes a consent to share declaration before submitting referral. Provides an audit of referrals; each one is automatically assigned a unique referral code. There is the option, when submitting the referral, so that the person who is being referred also receives a copy of the referral notice. The referring person has a record of all referrals than have been made by them. |
| **Information collected:** | When making referrals, the type of referral is chosen from a drop-down menu. For housing referrals, additional information is required/can be included, relating to the type of referral (parts of the Act), nationality, ethnicity, whether the applicant has formal identification, and free text to outline agency involvement already in place. Basic information collected for the person being referred includes:  - Name (First and Last)  - Date of birth  - Gender  - Nickname/maiden name  - NHS number  - Address (Postcode)  - Contact details |
| **Strengths:** | Simple and user-friendly interface increases uptake of the system. Updates to referring agency mean that they are kept up to date without having to check in with the other organisation. |

**Reference material:**
Main webpage: [https://www.housingjigsaw.co.uk/our-products/alert/](https://www.housingjigsaw.co.uk/our-products/alert/)
**Data system:** Combined Homelessness Information Network (CHAIN)

**Summary**
CHAIN is a multi-agency system for recording data about rough sleepers and the wider street population in London. CHAIN is funded by the Greater London Authority and managed by St Mungo's Broadway and represents a unique source of detailed data about rough sleeping. As the system is designed for sharing information, data protection is an important issue; before a client's information is entered CHAIN, they sign a data protection statement which explains what CHAIN is and how their data might be shared. Reports from the system are used at an operational level by commissioning bodies to monitor the effectiveness of their services, and at a more strategic level by policy makers to gather intelligence about trends within the rough sleeper population and to identify emerging needs. People are counted as rough sleeping when they are bedded down, either in the street or in areas not meant for human habitation.

**Purpose:**
Sharing of information about rough sleeper needs, ensuring that the person receives the most appropriate support and that efforts are not duplicated.

**Geographical coverage:** Greater London Authority.

**Population:**
Rough sleepers and wider street population, for example those engaged in street drinking. Rough sleepers are included on the system where they are 'verified', meaning that they have been seen bedded down.

**Update frequency:**
Live system; based on contacts with person.

**Data inputting/collecting organisation(s):**
Voluntary sector partners, including:
- Outreach teams
- Rough sleeper assessment and reconnection services, such as No Second Night Out
- Accommodation projects, including hostels, second-stage accommodation and supported housing projects
- Day centres
- Other specialist services

**Data collection method/storage:**
Web based information sharing platform that forms a by-person list of rough sleepers in the GLA area.

**Data presentation/sharing:**
Dashboards for data inputting organisations enabling real-time analysis.
Researchers and public bodies in the GLA area can use CHAIN data to improve their work.
Statistical reports published by GLA, produced by CHAIN team.
Data have been linked to health records (see Brodie et al. 2013).

**Information collected:**
Mandatory fields include:
- Name
- Date of birth
- Gender
- Ethnicity
- Nationality
- Immigration status
- Support needs
- Institutional history
- Last settled base
- Reason for leaving last settled base

**Strengths:**
Enables continuous count of rough sleeper observations and the co-ordination of services. Can be used to speed up interactions with people who are homeless as it can be used as a reference guide for that individual's needs.
Creation of CHAIN lead on to St Mungo's developing Real Systems as a social enterprise for helping other third sector organisations to improve their data monitoring systems.

**Weaknesses:**
Data are not comparable with rough sleeper count in rest of England and therefore may be comparatively under-referenced.
Data collection by outreach workers will fluctuate with resource availability, i.e. changes in initiatives which increase street outreach may increase contacts.
Non-participation of some organisations may lead to biases/omissions in the data.
Potential for data to be used punitively.

**Reference material:**
Main webpage: [https://www.mungos.org/work-with-us/chain/](https://www.mungos.org/work-with-us/chain/)

**Research/data use examples:**
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>In-Form DataLab</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>The DataLab is an offshoot of the Homeless Link initiative and draws data directly from a range of organisations using the In-Form case management system. Over 150 providers across the UK currently use the In-Form software. Together, the pooled data provides a source of information on homeless people using homelessness services.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>The aim of the DataLab is to be a central source of information about homeless people in England and provide a more comprehensive picture of the effectiveness of interventions in ending homelessness.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>England/expanding to other In-Form clients</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>People accessing homelessness support services.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Ad hoc, as-and-when new data are necessary.</td>
</tr>
<tr>
<td><strong>Data inputting/collection organisation(s):</strong></td>
<td>Homelessness service providers who currently use an In-Form system can choose to participate in the scheme, and have their data deposited within the DataLab. Service providers will vary, but could including: - Outreach teams - Rough sleeper assessment and reconnection services, such as No Second Night Out - Accommodation projects, including hostels, second-stage accommodation and supported housing projects - Day centres - Other specialist services supporting homelessness</td>
</tr>
<tr>
<td><strong>Data collection method/storage:</strong></td>
<td>In-Form is based on the Salesforce software. Data is transferred from individual In-Form platforms, into a centralised data base, using a 'Salesforce to Salesforce' connection. In the process of creating the data base, data are de-duplicated. The data stored within the Lab's data base is compliant with GDPR and UK Data Protection laws. Participating organisations must sign data sharing agreements with the In-Form DataLab. Client consent is not necessary, however, as good practice, the In-Form DataLab suggest that organisations provide information to service users about the potential use of their data. Data within each participating organisations In-Form system is mapped to a In-Form DataLab standard framework. The In-Form DataLab team will then test the connection and mapping.</td>
</tr>
<tr>
<td><strong>Data presentation/sharing:</strong></td>
<td>Data are available via an online database which can be queried. New service being provided by the DataLab is to provide participating organisations with insight reports and analytics about their data, and the ability to benchmark themselves against the DataLab database itself. Provides some incentive for participating given lack of sector benchmarking.</td>
</tr>
<tr>
<td><strong>Information collected:</strong></td>
<td>Every instance of In-Form has a set of standard fields – In-Form ‘objects’. These objects can be imported into the single data base which forms the DataLab. Standard fields/objects are those with drop down menus, which minimises the need to recode variable values in order to create continuity across the In-Form instances. Personal data are hashed in the process of transferring data from each In-Form instance to the Lab and enable de-duplication of people.</td>
</tr>
<tr>
<td><strong>Strengths:</strong></td>
<td>Access to a large pool of people who are homeless has the potential to generate innovative insights into homelessness. Participation in the system has reciprocal benefits for the data provider as they receive additional analysis. Organisations can choose which of their services data contribute to the DataLab.</td>
</tr>
<tr>
<td><strong>Weaknesses:</strong></td>
<td>Data are not universal, being from geographically dispersed services. Personal data quality will vary with service type and homelessness population. Data are not accessible to researchers who are not data providers.</td>
</tr>
<tr>
<td><strong>Reference material:</strong></td>
<td>Main webpage: <a href="http://in-form.org.uk/faq">http://in-form.org.uk/faq</a></td>
</tr>
</tbody>
</table>
**Data system:** Supporting People Client Record System and Outcomes Framework

**Summary**
The Supporting People (SP) programme in England was launched in 2003 with the main objective of helping end social exclusion by preventing housing crisis, enabling vulnerable people to live independently. SP provides a single funding stream for the delivery of housing related support and is fully devolved to local authorities. The Department for Communities and Local Government (DCLG) devised a framework for assessing the work being done under the SP grant, which included a Client Record System and the National Outcomes Framework. The Client Records Office, Centre for Housing Research (CHR), University of St Andrews was responsible for organising the collection and analysis of the SP programme data collection. From April 2011 the Department for Communities and Local Government ceased collection of supporting people client record and outcome data, with this becoming the responsibility for individual authorities. However, until 2014/15 CHR continued to collect client outcomes data on behalf of subscribing authorities.

**Purpose:**
Assess SP service quality and gathering management information.

**Geographical coverage:**
England.

**Population:**
People being assisted under the SP programme grant fund. The primary client groups include:
- Older people with support needs
- Older people with mental health problems
- Frail elderly
- Mental health problems
- Learning disabilities
- Physical or sensory disability
- Single homeless with support needs
- Alcohol problems
- Drug problems
- Offenders or at risk of offending
- Mentally disordered offenders
- Young people at risk
- Young people leaving care
- People with HIV/AIDS
- Homeless families with support needs
- Refugees
- Teenage parents
- Rough sleeper
- Traveller
- People at risk of domestic violence
- Generic/Complex needs

**Update frequency:**
Monthly updates of new client records.

**Data inputting/collecting organisation(s):**
Service providers of SP funded services report to local authorities.

**Data collection method/storage:**
Paper forms were available for office use; however an Access-based programme was available (SP Digital) free for download from the Client Record Office. The Client Record On-Line Submission System (CROSS) allowed Client Record data to be entered via the internet. For providers with their own in-house data management systems, an Electronic Data Transfer (EDT) system was available.

**Data presentation/sharing:**
SP Digital had pre-designed templates so that providers could produce summary information. Data quality was checked by CHR upon receipt; after this validation the data then became part of the Client Record database. An exercise was conducted by CHR in order to link Client Record forms to Outcomes. In order to achieve a link, several fields were linkable, namely: (1) the person involved (National Insurance (NI) number or the client/tenant code assigned by service); (2) the service with whom the contract is with SP service ID and NCP ID; and (3) the date on which the support started (support start date).
### Information collected:
Client records were completed each time a client entered a housing related support service funded by SP. The client record form gathered information about the client and how they came to receive housing-related support, but the client remained anonymous.

Outcomes forms for both long- and short-term services collect characteristics and outcomes achieved. Outcomes for long-term services were collected for a sample of clients in services after completion of their annual support plan. In 2009/2010 the sample was 10% for older people services and 50% for all other services.

Outcomes measured included:
- Maximising income, including being in receipt of the correct benefits
- Reducing overall debt
- Obtain paid work/participate in paid work
- Participate in chosen training/education and where applicable, achieve a desired qualification
- Participate in chosen leisure/cultural/faith/informal learning activities
- Participate in chosen work like/voluntary/unpaid work activities
- Establish contact with external services/family/friends
- Management of physical and mental health
- Management of substance misuse
- Management of independent living as a result of assistive technology/aids etc.
- Secure/obtain settled accommodation (new for 2009/10)
- Comply with statutory orders and processes (in relation to offending behaviour)
- Better manage self-harm, avoid causing harm to others, minimise harm/risk of harm from others
- Greater choice and/or involvement and/or control at service level and within the wider community

### Strengths:
Different ways in which organisations could provide CHR with data fits in with local data collection practices across local authorities in England.

Outcomes framework is client focused, in comparison to a focus on numbers assisted.

### Weaknesses:
Client data collected by CHR were anonymous which prevent linkage to other data sources. Lack of a clear ID number complicated the process of linking between client forms and outcomes.

### Reference material:

### Research/data use examples:
**Data system:** Expanded Troubled Families Programme.

**Summary:** Troubled Families is a programme of targeted interventions for families with multiple problems, including crime, anti-social behaviour, truancy, unemployment, mental health problems and domestic abuse. Local authorities identify ‘troubled families’ in their area and usually assign a key worker to act as a single point of contact. Central Government pays local authorities by results for each family that meet set criteria or move into continuous employment. The Troubled Families programme is administered by the Department for Communities and Local Government (DCLG) and covers England only. Expanded collection forms part of a comprehensive data linkage evaluation.

**Purpose:** Monitor the Troubled Families programme. Payment is by results; therefore local authorities are required to evidence that they have achieved a positive outcome for the families.

**Geographical coverage:** England.

**Population:** Families in receipt of assistance under the Troubled Families programme. The administrative definition of a ‘troubled family’ was a household that was:
1. Involved in crime and anti-social behaviour
2. Have children not in school
3. Have an adult on out of work benefits
4. Cause high cost to the public purse
In order to be included in the programme, the local authority had to satisfy themselves that a family meet all the core criteria (1-3), or two of the core criteria and the 4th criteria of high cost. The definition of ‘high cost’ was up to the local authority to determine.

**Update frequency:** 6 monthly uploads.

**Data inputting/collection organisation(s):** Local authorities.

**Data collection method/storage:** Data collection for the programme has change from its initial phase, to place a greater emphasis on individual outcomes and data linkage. Data collected for the Expanded Troubled Families Programme, fall under 3 broad areas:
1. Basic information on individual family members in order to conduct the National Impact Study (NIS) using linked administrative data,
2. Family level outcomes data on important issues not available in administrative data which are collected and tracked locally as Family Progress Data (FPD);
3. Information relating to the costs of providing programmes at the local level, which are used as part of a Cost Savings Calculator.
FPD is submitted to DCLG via an online information system, whilst NIS data are sent to ONS via secure file transfer (e.g. Egress).

**Data presentation/sharing:** Data sharing agreements are in place between authorities, DCLG, and ONS enabling the transfer of programme data for data processing. ONS act as the trusted third party for data linkage as part of evidencing/evaluating the programme. All personal identifiers contained within the data provided by local authorities is encrypted using a one-way cryptographic hash. This enables linked data sets to be retained in an anonymised format for use by DCLG.

**Information collected:** Essential data items collected under the NIS includes:
- Identifiers (name, date of birth, gender, postcode)
- Relational data
- Eligibility for the programme
- Date screened for the programme
- Intervention start and end date
Desirable items within the NIS are unique identifiers that enable exact matching to administrative sources, such as National Insurance Number, Unique Pupil Numbers, and NHS numbers. Under the FPD data collection sent to DCLG, indicators of relevance to a range of family members are collected.

**Strengths:** Encryption ensures the privacy of the individual while allowing ONS to link the data to other information sources if required.
Personal data files are opened in the ONS Secure Data Linkage Facility. Designing data collection to enable data linkage from the outset enables saving in evaluations, as well as reduced burden on data providers to collect outcomes data.
**Weaknesses:** Data agreements dictate the time period that the final linked pseudo-anonymised data set can be retained by DCLG, until 2025. Coordination of data collection for the programme was limited in less ‘data mature’ local authorities, requiring investment in more advanced software. Limited/no opportunities for academic research of data set; only certain named individuals within DCLG can access the data.

**Reference material:**

| Ministry of Housing, Communities & Local Government (2018) Supporting disadvantaged families |
| Annual report of the Troubled Families Programme 2017-18. London: Ministry of Housing, Communities & Local Government |

**Research/data use examples:**

**Data system:** MainStay

**Summary:** MainStay is an online system that allows housing and support services to access a common assessment and referral facility. The system uses the information inputted by assessors to ensure that persons can be quickly prioritised and suitably accommodated for their needs. The main benefit is one comprehensive assessment for access to all services linked via MainStay. The aims of the system are to prevent homelessness, reduce repeat homelessness, and reduce evictions and abandonments.

**Purpose:** To provide improved access to short-term Housing Related Support services for single persons (and couples in some circumstances) within the region of Liverpool City Council.

**Geographical coverage:** Liverpool.

**Population:** Single people who are homeless (and couples in some circumstances).

**Update frequency:** Continuous.

**Data inputting/collecting organisation(s):** Voluntary and statutory homelessness services partners.

**Data collection method/storage:** Information management system with a common assessment tool.

**Data presentation/sharing:** Clients are searchable across the database to avoid creating duplicate records. Can access data on clients not assessed by own service; system alerts case owner that data has been accessed.

**Information collected:** Referral form collects initial information needed to assess client, including:

- Reason for referral
- Name
- Date of birth
- National Insurance number
- Current accommodation and last settled accommodation
- Disability status
- Economic status
- Ethnicity
- Immigration status
- Sexual orientation

As part of the processing of the application, further details are obtained about the individual, including a risk and needs assessment of the client which can be used to set a "risk/needs level" and prioritise high risk/need people for assistance. Alerts can then be added to the person's file.

**Strengths:**

- Avoids multiple assessments for service users and accommodation providers.
- Consent to share data is single page and easy to read.

**Weaknesses:**

- Non-participation of some organisations may bias usability of data for statistical purposes and referrals.
- 'Freely given consent' may be questionable, given it may affect service access if information is not shared.

**Reference material:**

### Data system: Better Futures toolkit – Housing Support Enabling Unit

#### Summary: ‘Better Futures’ is a toolkit for promoting a person-centred and outcomes-focused approach to the assessment and review of housing support. At the centre of the Better Futures system is a person-centred Outcomes Framework to promote and measure change. This framework enables individuals to collaborate with their key worker to identify their own priorities, plan their support and chart their progress towards achieving a better quality of life and sustaining independent living. The framework is accompanied by a web-based IT ‘outcomes measurement’ system. This can be used to chart and visualise individuals’ progress towards their own goals. It can also be used to generate reports to help inform the service delivery and evidence outcomes the services are helping to achieve. Better Futures is not a case management system or a statistical reporting tool. The web-tool is managed by the Housing Support Enabling Unit (HSEU).

#### Purpose: IT tool and outcomes framework designed to enable housing support service providers to record an individual’s support needs over a period.

#### Geographical coverage: Scotland.

#### Population: People receiving support from housing services.

#### Update frequency: Continuous.

#### Data inputting/collecting organisation(s): Housing support providers.

#### Data collection method/storage: On-line tool.

#### Data presentation/sharing: Reports can be run for individuals or organisations in order to explore effectiveness in reaching outcomes. Outcomes star visualises progress.

#### Information collected:
- Name
- Date of birth
- Gender
- Ethnicity
- Housing situation
- Local authority where a person lives
- Goals and elements of support
- Scoring on goals

#### Strengths: Enables people using services to set personal outcomes and monitor their own progress. Strong support amongst stakeholders for involving service users and enabling individuals to shape their own goals and support requirements.

#### Weaknesses: Limited functionality as a method for enumerating homeless households.

#### Reference material:
Data system: Greater Manchester Tackling Homelessness Information Network (GM-Think, formerly M-Think)

Summary: M-Think/GM-Think is a data platform that enables services to share information quickly and safely in the Greater Manchester area. The system is part of the Inspiring Change Manchester programme, led by Shelter England and funded by the Big Lottery. The aim of the programme is to transform the support received by people with multiple needs. M-Think originally covered the City of Manchester area, however this was expanded to cover the Great Manchester area under the GM-Think system. The design of M-Think (now GM-Think) involved several stakeholders, including people with experience of homelessness, in order to design a system that was responsive to their experiences of service use (good and bad).

Purpose: At the core of GM-THINK is the principle of ‘telling your story once’ – consensual information-sharing between organisations so the individual can get access to the right services, at the right time.

Geographical coverage: Greater Manchester area.

Population: People with multiple complex needs, includes problem drug or alcohol use, mental health or emotional well-being, homelessness and offending.

Update frequency: Continuous as system is ‘live’.

Data inputting/collecting organisation(s): Voluntary and public sector organisations providing services to the Council under the Homeless Prevention Grant recipients were contracted to input data into the M-Think database.

Data collection method/storage: Web-based multi-agency data base that enables sharing of information by organisations working within the Greater Manchester area.

Training in the system has been an important part of making case workers aware of the importance of sharing information, and the impacts of the language they use to record data.

Data presentation/sharing: Data are structured around a ‘front page’ and associated case notes for the person that are input by the different services they may interact with.

The person seeking assistance can view data about themselves with their case worker. Notes created by other organisations can also be viewed alongside individuals from those organisations.

Sharing of client information across the system is possible, thereby avoiding people having to retell their story every time they approach a different service.

Information collected: Information collected under the M-Think/GM-Think system includes:

- Basic demographics and assistance profile which are used to create a profile page.
- Support plan/action plan.
- Plans for the person in terms of accessing support and keeping safe.

Some elements of the collection is designed from the person's perspective, rather than the case workers’. For example, asking about the persons' ideas and ambitions. The language of the system has also been modified in order to reduce the stigmatisation of people accessing services, for example ‘safety plans’ rather than ‘risk plans’.

Strengths: Person seeking assistance is involved in their own support planning and updating personal goals.

Use of jargon free language makes data accessible to the person receiving assistance.

System is aimed at bringing together services to support people.

Reference material:
Main webpage: https://inspiringchangemanchester.shelter.org.uk/
More information on Inspiring Change Manchester:
https://www.manchestercommunitycentral.org/news/inspiring-change-manchester-expressions-interest
Shelter England (Date not known) Introducing M-Think. Available at: https://england.shelter.org.uk/_data/assets/pdf_file/0007/1149717/7472_M-Think_Leaflet_v3_PRINT.pdf
**Summary**
MASH is a model of data sharing amongst public bodies, usually local authority services. Though there are numerous ways in which MASHs’ can operate, for example as informal networks and groups, technological solutions for data sharing and case management have been developed in some cases.

**Purpose:**
Improve the safeguarding response for children and vulnerable adults through better information sharing and timely safeguarding responses.

**Geographical coverage:**
Local authority areas.

**Population:**
Children and vulnerable adults entering public services within local authorities.

**Update frequency:**
Varies, the ideal being continuous real-time sharing.

**Data inputting/collection organisation(s):**
Public bodies, for example police and social services.

**Data collection method/storage:**
Varies with hub set-up, either data are stored in a dedicated IT solution for a hub or there is linking between agency systems which already exist.

**Data presentation/sharing:**
MASHs’ are designed around information sharing between partner organisations; sharing is however on a case-by-case basis. Data are not for monitoring and statistical purposes but for active case management and decision making.

**Information collected:**
Enter of vulnerable people into MASHs is usually through some form of referral, part of which involves the collection of screening data in order to establish the risk rating for a case. Specific data collected varies with hub, however as a minimum:
- personal details relating to the child or vulnerable adult
- information relating to their service interactions

**Strengths:**
MASH models lead to more accurate assessment of risk and need and more thorough case management.
Better understanding between partner organisations.
Data sharing has a strong basis in legislation, for example child care, for example Section 10 of the Children’s Act 2004 encourages cooperation between an authority’s services in order to provide care services for children.

**Weaknesses:**
MASH system/teams handle relatively small number of at risk cases, therefore this model of operation may not scale up.
Primarily MASH models are about the co-location of people in a hub team, there are therefore time/staffing implications in order to engage with and have a presence in the hub.

**Reference material:**
For an example of the governance and data sharing agreements in place around a MASH-type system, see Merton Multi Agency Safeguarding Hub: [https://www2.merton.gov.uk/health-social-care/children-family-health-social-care/safeguardingchildren/multi-agency-safeguarding-hub.htm](https://www2.merton.gov.uk/health-social-care/children-family-health-social-care/safeguardingchildren/multi-agency-safeguarding-hub.htm)
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>Dementias Platform UK (DPUK) Data Portal</th>
</tr>
</thead>
</table>
| **Summary**      | DPUK is a coordinated and integrated way of conducting neurodegenerative disease/dementia research, and will maximise the potential of UK cohort studies, both disease-specific and population. The integrated research environment has three core activities:  
  - data sharing via an electronic data portal with over 35 existing UK population studies encompassing 2 million participants; this enables new ideas to be tested thoroughly and quickly;  
  - a clinical studies register, to recruit and match volunteers from UK population studies with clinical trials that will fast-track the journey towards new treatments for dementia;  
  - a programme of experimental medicine studies to find out what works and why. This activity is supported by state-of-the-art research networks in brain imaging, stem cells and informatics to create and coordinate the best possible science. |
| **Purpose:**     | The DPUK Data Portal's aim is to provide a central access mechanism for researchers to access the diverse datasets/cohorts in a secure research platform which is delivered using the UK Secure e-Research Platform (UKSeRP). |
| **Geographical coverage:** | International. |
| **Population:** | Varies with study, however all cohorts are individuals involved in dementias research. |
| **Update frequency:** | Ad hoc, as and when new data collections and studies are available/choose to participate. |
| **Data inputting/collecting organisation(s):** | Principle investigators for participating cohort studies are the data owners. Inputting organisations are predominantly university of health intuitions engaging in dementias research. |
| **Data collection method/storage:** | Researchers are able to bring their own data into the Analysis Environment, provided it is anonymised. The Data Portal offers a Data Linkage service which uses, deterministic and probabilistic matching technologies to create de-identified, analysis-ready data.  
  Cohort data are transferred to the DPUK UK Secure eResearch Platform, hosted by Swansea University, to be prepared for access by the DPUK research community. Data preparation are conducted by permitted persons nominated by the cohort upon signature of the DPUK Data Deposit Agreement.  
  DPUK receive cohort data 'as is', and curate the data to according to a DPUK common data model. Data curation will include liaison with the cohort research team to ensure the sense of the data is retained. |
| **Data presentation/sharing:** | Data held within the DPUK is anonymised.  
  UKSeRP will mean that ownership remains with the data owner and the governance structure puts them in the centre of the process and ensures that the data is only used in projects that the data owners are comfortable with.  
  Researchers identify which cohorts are relevant to them, apply for access to the data and then analyse it in a secure, remote environment with a complete data linkage and analysis package. |
| **Information collected:** | Information available from different dementias studies/cohorts varies, however each contains demographic data and diagnosis information, alongside other physical/clinical examinations such as cognitive functioning tests. Some studies include brain scans, blood work analysis, and other forms of biometric data. |
| **Strengths:** | Provides a secure way for national researchers to access data on Dementias research securely.  
  Virtual desktop allows for remote access to data.  
  Data owners retain control of their data.  
  Governance structure can be designed to suit the purposes of the data collection, rather than falling under the remit of SAIL governance. |
<p>| <strong>Weaknesses:</strong> | Data are anonymised, therefore the system cannot be used for operational, case level assessments. |
| <strong>Reference material:</strong> | Main webpage: <a href="https://portal.dementiasplatform.uk/">https://portal.dementiasplatform.uk/</a> |</p>
<table>
<thead>
<tr>
<th>Data system:</th>
<th>Continuous Recording of lettings and sales (CORE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>CORE provides information on social letting, sales, tenants and buyers in England, and is overseen by the Ministry of Housing, Communities and Local Government (MHCLG). CORE was initially set up in 1989 to record information on private registered social housing providers; being extended in 2004 to cover local authority housing providers. Subsequent changes have occurred to the collection over time as the social rented sector has evolved. It is a regulatory requirement for providers registered with the Homes and Communities Agency to supply the data.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>Information contained within CORE is used to inform government policy and funding decision making relating to social lettings in England.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>England</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>Tenancies and sales of social rented accommodation</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Ongoing; updated when there is a new tenancy</td>
</tr>
<tr>
<td><strong>Data inputting/collecting organisation(s):</strong></td>
<td>Social landlords, including local authority (LA) and private registered social housing providers (PRPs)</td>
</tr>
<tr>
<td><strong>Data collection method/storage:</strong></td>
<td>CORE is collected via a web-based interface, which allows manual input and batch upload. Providers of data are requested to input new records into CORE as close to the start/sale of the tenancy. Validation of data occurs through the CORE system, and providers are monitored in terms of the timeliness of their data entry and their query resolution. The CORE web-platform has been designed in order to minimise the amount of time required by providers to input/upload their data.</td>
</tr>
<tr>
<td><strong>Data presentation/sharing:</strong></td>
<td>Both quarterly and annual statistics are published by MHCLG, however the quarterly statistics are based on “live” data and are therefore revised in the annual publication. Data are available for research via the UK Data Service. Different levels of access are available dependent on the disclosivity of the data.</td>
</tr>
<tr>
<td><strong>Information collected:</strong></td>
<td>Information collected covers details of the tenancy, demographics of the household, including age and sex of each household member, and more detailed information on income and benefits receipt. However, there have been quality concerns regarding the income/benefits data. The unit of measurement within the data set is the household, however details are collected about all individuals within the household.</td>
</tr>
<tr>
<td><strong>Strengths:</strong></td>
<td>Data have been made available for research purposes. Where a landlord is not registered with the Homes and Communities Agency they are still able to supply data to CORE. MHCLG collect information on how CORE licence holders were using the CORE data.</td>
</tr>
<tr>
<td><strong>Weaknesses:</strong></td>
<td>CORE data are anonymised, therefore data linkage not possible.</td>
</tr>
</tbody>
</table>

**Reference material:**
- Main webpage: https://core.communities.gov.uk/public/GuidesAndManuals.html
- Main webpage for CORE data via the UK Data Service: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7604#!/details
**Data system:** Kent Integrated Dataset (KID)

**Summary**
Though health and social care records are available for research purposes, only a limited number of data are usually available. The KID system was designed to fill this data gap by providing system-wide information on health and care for the Kent and Medway area. Initially, the project was funded through the NHS England’s ‘Long Term Conditions Year of Care’ programme between 2013 and 2016. After this programme ended, Kent County Council and the local NHS health board decided to continue to fund the resource, in order to develop a longer-term objective of ‘place-based commissioning’.

**Purpose:**
The purpose of KID is to provide health and social care planners in Kent and Medway with insight into population health and use of services in their area.

**Geographical coverage:** Kent and Medway.

**Population:** General population; those accessing health/social care services.

**Update frequency:** Monthly.

**Data inputting/collecting organisation(s):**
Data providers cover a range of health and social care services, including:
- Primary care
- Community health
- Mental health
- Acute hospitals
- Adult social care
- Palliative care

**Data collection method/storage:**
Data originate from different organisational administrative data sources. Data owners are responsible for the validation and quality assurance of their data. However, the KID team also run a series of basic validations, including a check that all providers have provided data, whether key variables are included, and coding appears consistent across providers.

KID is governed collectively by the organisations who contribute data to the system. Underneath the steering group for the collection, sub-groups deal with different aspects, such as development of the data set and applications for use of the data.

Patients can opt-out of having their data included in KID by informing the relevant provider in writing.

**Data presentation/sharing:**
Data can be accessed under the condition that research is likely to have some benefit to Kent and Medway health and care services. Individual level data can only be accessed on Kent County Council's computer system, either physically within the authority or via a secure remote desktop.

**Information collected:**
Data included in the relate to different services, including:
- Diagnoses and clinical information from GPs
- Social care data such as the client category and care package
- Secondary care received in hospitals
- Mental health services
- Demographic information (e.g. age, sex, LSOA of residence, pseudonymised unique property reference number)

In addition to the health and social care event data, the KID system also includes an estimate of the cost of each episode of care. Costs of episodes of care differ between services, drawing on different compendia of healthcare costs available from the NHS/other sources.

As the data primarily relate to health and social care services, the NHS number forms the linkage key. However, the NHS number is encrypted to protect privacy further. Data are further “coarsened” to provide an additional layer of privacy protection, for example by replacing postcodes with LSOA codes, and replacing exact birthdays with the week of birth.

**Strengths:**
Population coverage means that analysis can drill down into sub-groups without losing power. Cost data enable economic modelling of care. The KID system is exploring the inclusion of other non-health/social care data, increasing the scope for evaluative studies.

**Reference material:**

**Research/data use examples:**
### Data system:
Connecting Care

### Summary
Connecting Care is a data-sharing platform for the Bristol, South Gloucester and North Somerset area, enabling secure access to health and social care data across the region. Health and social care professionals have used Connecting Care since 2013.

### Purpose:
By enabling sharing of data across the region and across health and social care providers, the aim of Connecting Care is to enable a joined-up approach to working, and faster more effective service delivery.

### Geographical coverage:
Bristol, North Somerset and South Gloucester

### Population:
General population; specifically, those who interact with certain health/social care services

### Update frequency:
Continuous

### Data inputting/collecting organisation(s):
Data contained within connecting care includes certain elements of information from:
- General Practices
- Hospital departments
- Community service
- Mental health service
- Local authorities

### Data collection method/storage:
Data from individual care systems is shared securely using an encrypted connection to the Connecting Care clinical database.
- Connect Care builds upon the Orion Health platform, enabling users to view data through a portal interface.
- Local systems for a service can connect to the Connecting Care system via application programming interfaces and/or through a specific technology that enables data exchange from multiple systems.

### Data presentation/sharing:
Only those directly involved in a person’s care are authorised to view their data within the system. Users are required to stipulate a reason for access to a record, which include:
- An on-going care relationship
- Consent
- Vital interests
- Safeguarding concerns
- To check the appropriateness of a referral

There are different levels of visibility within the system, such that the data seen by a care professional varies by their role in a person’s care.

Though a person can opt-out of having their data on the Connecting Care system, as the data would be shared regardless in order to care for a person, opting-out would simply mean care providers use telephone, fax and letters to organise care. Opting out means that a person’s data is no longer visible – for safeguarding reasons the data still needs to be on the Connecting Care system for review.

### Information collected:
The Connecting Care system creates an integrated digital care record, ‘Connecting Care Local Record’, combining data from various sources. Data items contained within a care record may include:
- Medications prescribed
- Contact details for those involved in a person’s care
- Medical conditions and allergies
- Radiology and pathology reports
- Care plans and other documents
- Details of operations and procedures
- Hospital, GP and social care referrals/appointments
- Safeguarding alerts

Personal data are shared with the system in order to link data and enable people to be searched in the system.

Data are only visible within Connecting Care for as long as the data is held by the collecting organisation.

### Strengths:
The Connecting Care system has a strong legal basis for sharing data, and this legal basis is made explicit to the public.
The system is being developed to include other data sources.
System can support a large volume of users.

### Weaknesses:
Research uses of data have not been explicitly designed into the system from the start.

### Reference material:
Main webpage: https://www.connectingcarebnssg.co.uk
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>NHS Scotland Corporate Data Warehouse (CDW) and Data marts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>NHS National Services Scotland have developed a central repository for data from several healthcare settings across Scotland. The repository, or warehouse, is structured around a set of data collections that are gathered into thematic areas, or data marts. The data marts enable quicker, topic specific queries than querying a central repository. Additional interfaces or informational tools have also been developed that enable performance/reporting from the warehouse, enabling data to be drilled down to person-level.</td>
</tr>
<tr>
<td>** Purpose:**</td>
<td>To support information management and decision making.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Scotland.</td>
</tr>
</tbody>
</table>
| **Population:** | Covers interactions/events under the remit of health services in Scotland, including:  
  - Accident and emergency  
  - Cancer deaths  
  - Antimicrobial management  
  - Outpatients  
  - Prescriptions |
| **Update frequency:** | Refresh of data varies by data source, for example daily or monthly. |
| **Data inputting/collecting organisation(s):** | Healthcare providers in Scotland. |
| **Data collection method/storage:** | Structured data, i.e. data in the form of standardised data sets, are uploaded to the data warehouse from other systems.  
The warehouse is dimensional, with data marts containing data for a specific theme or dimension.  
Data for some sources are split into identifiable and anonymised records and held separately in the CDW.  
Data uploaded are a mixture of aggregate and patient/event level. |
| **Data presentation/sharing:** | Data marts bring together data on a specific topic/data set, for easier querying.  
The Discovery information tool brings together metrics in order to support performance and quality improvement. Data can be drilled down into in order to provide person-centred information—depending on what permissions have been granted.  
Requests for access to the different data marts is managed by a central online system, the 'User Access System'. |
| **Information collected:** | The different data marts within the system vary in scope. The majority draw down information from a single data set/collection, for example the Accident and Emergency (A&E) data mart contains data from June 2007 to the present day on patient attendances at Emergency Departments, Minor Injuries Units and Community Hospital A&Es across NHS Scotland. |
| **Strengths:** | The ability to drill down to person-level has the potential to enable data in warehouse to inform care/and or auditing of care.  
Data marts can combine data sources in order to enable topic specific querying.  
Access to the data is managed by a central system that potential enables audits of access requests and processing. |

**Reference material:**
Main webpage: [https://www.isdscotland.org/Products-and-Services/Datamarts/](https://www.isdscotland.org/Products-and-Services/Datamarts/)

**Research/data use examples:**
### Data system:
North West London Whole Systems Integrated Care (WSIC) data warehouse and dashboards

### Summary
In order to create a person-centred approach to health and care provision, data from all medical professionals involved in a person’s care are brought together into an integrated care record. This care record is held within a data warehouse. The data dashboards are a means of visualising data and provide a suite of tools for clinicians and health professionals that can guide patient care.

### Purpose:
Coordinate person-centred approach to care for North West London patients through the creation of integrated care records and dashboards for data visualisation.

### Geographical coverage:
North West London area.

### Population:
Patients using certain healthcare providers in North West London; 95% coverage of patient population.

### Update frequency:
Varies with the data set being provided, however ranges from weekly to monthly updates

### Data inputting/collecting organisation(s):
A range of healthcare providers provide data, including:
- mental health and community Trusts
- social care data from eight boroughs
- 380 GP practices

### Data collection method/storage:
Data are stored in a data warehouse, whilst ‘dashboards’ provide a data visualisation layer to provide whole patient/population information.

### Data presentation/sharing:
Data are available in a range of outputs/tools, including:

- **Identifiable data outputs:**
  - Integrated patient summaries, which can be used to inform case meetings and care pathways
  - Case finding tools to identify patients who may benefit from proactive intervention
  - Long term care management tool

- **De-identified data outputs:**
  - Population health overviews
  - Benchmarking services tools
  - De-identified datasets for researchers

Access to identifiable data is only provided for health and care professionals involved in a person’s care, whilst access to de-identified data is made available to authorised researchers. To access the de-identified data, researchers must complete an application form outlining their intended project and required data. Data are accessed and analysed on a separate workspace. Researcher’s own data can be brought into the WSIC; for example, data on people undergoing a particular intervention can be flagged and have their data extracted from the warehouse.

Patients have the option to opt out of having their data included in the electronic patient record, which means that data are not shared with the warehouse.

### Information collected:
The dashboards bring together a range of data sources about a person’s care to form a care record, including test results, medications, allergies and social, health and wellbeing information.

Sensitive information, for example HIV and sexual health diagnosis are not shared with the warehouse.

‘Costed’ version of certain dashboards are available that provide additional information on patient spend by using sector cost data.

### Weaknesses:
Access to de-identified data is time limited (6 months) and requires re-submission of forms to extend access.

De-identified data requires knowledge of SQL to undertake analysis.

### Strengths:
Use of dashboard interface makes interpreting information easier and accessible to different audiences of users.

Flexibility in the design of the system has meant that a number of different dashboards have been created that focus on specific population health issues.

Patients are able to refuse to have an electronic record created for them.

### Reference material:
Main webpage: https://www.healthiernorthwestlondon.nhs.uk/news-resources/information-sharing
**Data system:** Secure Anonymised Information Linkage (SAIL) databank

**Summary**
The SAIL databank operates as a safe haven for individual level de-identified data, enabling linkage between different data sources. Data held by SAIL covers predominantly health records, however there has been an expansion of the databank in recent years to include more detailed education data (from early years up to Higher Education), along with some limited child social care data. Other data can be brought into SAIL and linked on a project-by-project basis, for example where organisations have conducted surveys or studies and have participant’s identifier details.

When bringing data into SAIL, a split file process is used, whereby the personal identifiers (name date of birth, gender, postcode) are sent the NHS Wales Informatics Service (NWIS) for de-identification, whilst the main bulk of the data (‘payload data’) is sent directly to SAIL. The de-identification process undertaken by NWIS generates a unique national identifier for the person for use in SAIL, the Anonymised Linkage Field (ALF). All data in SAIL have an ALF assigned, thereby enabling linkage across different data source.

**Purpose:** The purpose of SAIL is to facilitate data linkage research.

**Geographical coverage:** Varies with individual data sources; largest geographic area covered is Wales

**Population:** Varies with data source; largest coverage is population of Wales

**Update frequency:** Varies; mainly annual

**Data inputting/collecting organisation(s):** Varies, but includes Welsh Government, individual General Practitioner, Local Authorities, and NHS Wales

**Data collection method/storage:** Data being imported into SAIL are manually extracted from local databases and undergo the split file process described above. Once in SAIL, data are stored in a de-identified format, meaning they are effectively anonymised but are linkable across data sources.

**Data presentation/sharing:** Linked data can be accessed via the SAIL ‘gateway’, which is a remote virtual desktop. All outputs from the gateway undergo disclosure control checks, i.e. that no individuals are identifiable within the data.

**Information collected:** Information collected varies dependent on the data source. However, data available for linkage covers:

- Primary and secondary healthcare
- Education data, from early years up to Higher Education
- Specialised health and social care data set – for example substance misuse data

**Weaknesses:** As the de-identification happens outside of SAIL, the information provided to researchers on match quality and the quality of the underlying identifiers is limited. Use of trusted third party (NWIS) means that SAIL is now reliant on NWIS for its continued operation—unless all data in the databank are to be re-acquired.

**Strengths:** Having a common approach to de-identification enables data linkage on a national scale.

**Reference material:**
Main webpage: [https://saildatabank.com/](https://saildatabank.com/)
## Data system:
Ministry of Justice DataLab (DataLab)

## Summary
The DataLab offers an analytical service to organisations which can provide insight into the possible effects of their interventions on reoffending. Organisations can submit personal identifiable data for people engaged in their service to the DataLab, who then link this data to criminal records held by the ministry. The DataLab then create a comparison group of offenders of the same ‘type’, before comparing the re-offending rates between the intervention group and the comparison group. Organisations are then provided with a report that summarise the potential impact of their intervention/service.

## Purpose:
Improve understanding about the effectiveness of services.

## Geographical coverage:
National (United Kingdom)

## Population:
Offenders

## Update frequency:
Ad hoc, when analysis requested

## Data inputting/collecting organisation(s):
Ministry of Justice operate the lab, however organisations who work with offenders can use the lab, for example charities, public/private sector organisations, and educational institutions.

## Data collection method/storage:
Personal identifier data are emailed to the Ministry of Justice, either from a secure government email address, or by using the Criminal Justice Secure email service—whereby the person applies for an email account.
Data are uploaded using a standard template.

## Data presentation/sharing:
Organisations who submit data to the DataLab receive a report which compares reoffending rates between their intervention group and a matched comparison group.
At no time is the organisation given access to individual level offending data.
Results of the analysis are also published by the DataLab online, in order to increase transparency; publication online is a term of use.

## Information collected:
The DataLab template includes personal identifiers (name, date of birth, gender), unique identifiers (prisoner number and or police national number), start and end dates of intervention, and any other information relevant to the intervention. The template also includes an area where the purpose of the intervention can be outlined, including the nature of the work being conducted.

## Weaknesses:
Reports/analyses are limited and standardised as the purpose of the Lab is to generate routine outputs easily.

## Strengths:
Reports visualise data clearly and are written for a generalist audience.
Improves sector knowledge of the possible impacts of interventions.
Maintains strict security of data by only outputting analytical outputs, rather than enabling access to data.

## Reference material:
Main webpage: https://www.gov.uk/government/publications/justice-data-lab
### Data system:
Pathway Accommodation and Support System (PASS), Dublin

### Summary:
Official homelessness figures for Ireland are produced by local authorities through PASS. In January 2011, PASS went live nationally across nine regions: Dublin, Mid-East, Midlands, Mid-West, North East, North-West, South-East, South-West and West. The aim of PASS is to provide a shared database of clients and accommodation, and therefore enable quicker allocation of accommodation. PASS also automates contact between clients and services, and is used to allocate emergency bed spaces. There are over 2000 staff using the application nationally and 600 projects. Within each region the lead local authorities for homelessness provide monthly reports on homelessness to the Department of Housing, Planning and Local Government, which identify the number of people utilising State-funded emergency accommodation on a regional and county basis. Dublin represents over 70 percent of homelessness cases in Ireland reported on PASS. Accordingly, the Dublin Region Homelessness Executive (DRHE) also houses the national PASS office. In 2018, DRHE tendered for a new provider for the PASS system.

### Purpose:
Provides local government, State agencies and the non-governmental sector with a shared client database which facilitates client assessment, support planning and interagency case management of people experiencing homelessness.

### Geographical coverage:
9 regional PASS centres covering Ireland.

### Population:
Primarily people placed in emergency accommodation and assessments by local authorities.

### Update frequency:
Real-time, continuous data entry.

### Data inputting/collecting organisation(s):
Statutory and voluntary partners receiving government funding.

### Data collection method/storage:
Online shared system providing real-time information on homelessness presentations and emergency bed spaces.

### Data presentation/sharing:
Each of the nine PASS regional centres has a lead local authority whose role is to ensure high quality data are entered into PASS, and to produce monthly anonymised reports to central government. Clients are searchable within the database. The DRHE regularly use the PASS data in order to undertake analysis (see: [https://www.homelessdublin.ie/info/publications](https://www.homelessdublin.ie/info/publications)).

Government statistics are limited to monthly reports based on the aggregate returns from PASS system regions.

Reports can be run on regional and national data, dependent on the researcher’s status, i.e. whether they are a system researcher or a regional researcher.

Client histories can be viewed, and filters applied to identify specific records.

### Information collected
- Basic identifying and demographic information, including Personal Public Service – like National Insurance number.
- Holistic Needs Assessment.
- Support Plan.
- Key-worker recording.
- Initial Assessment & Assessment of Housing Need.
- Family composition.
- Information about placement on local authority waiting lists.
- Reasons for homelessness and type of previous accommodation.
- Data relating to outcomes.

PASS is compliant with data protection legislation. Clients are referred to using their unique PASS ID in any communications.

### Strengths:
- Actively used to plan and manage services such as bed allocation.
- Data are used for research purposes, including follow on surveys of families and some evaluative studies of specific interventions.
- National data system formed out of a priority action from the government’s homelessness strategy, therefore strong policy driver and links to overall approach to homelessness.

### Weaknesses:
- Data are purged from PASS if the person does not have contact with a service within 2 years, meaning that some episodically homeless households will be missing.
- Issues have arisen with the automated aspects of the system over time which have negatively impacted on user experiences, i.e. automated telephone calls being sent to clients resulting in them losing their place within the bed allocation system.
- National statistics based on aggregate reporting, therefore limited research without acquiring data from individual regions.
Reference material:
Main webpage: https://www.homelessdublin.ie/info/pass

Research/data use examples:
**Data system:** Client registration system database – Statistics Denmark

**Summary:** Data are collected by Statistics Denmark in order to monitor state provision of temporary accommodation. When enrolling in a shelter, the individual must register their unique Central Personal Register (CPR) number. Through the CPR, exact matching can be used and temporary accommodation use tracked over time. Data on shelter use is available from 1999 onwards. Statistics were previously published by the Social Appeals Board, however as of 2016, they were transferred to Statistics Denmark. This transfer meant that the scope of the data collection could be reduced, as information could be obtained through data linkage to other sources held by Statistics Denmark.

**Purpose:** Monitor Section 110 in the Social Assistance Act which mandates that municipalities must provide temporary accommodation for homeless people.

**Geographical coverage:** Denmark.

**Population:** People who are offered accommodation overnight; does not include drop in centres or other services that do not provide residential stays.

- Certain forms of accommodation are exempt, for example hostels for women suffering violence or the threat of violence.

**Update frequency:** Quarterly updates to Statistics Denmark.

**Data inputting/collecting organisation(s):** Homeless shelters.

**Data collection method/storage:** Upload data to Statistics Denmark using secure internet transfer system.

**Data presentation/sharing:** Statistical reports are published by Statistics Denmark on an annual basis.

**Information collected:**
- Central Personal Register (CPR) number which is the unique number for that individual.
- Periods of entering and exiting accommodation.
- Referring agency.
- Exiting information, i.e. where the person leaves to go to.
- Limited demographics; collection simplified once the data were collected by Statistics Denmark.

**Strengths:**
- Data has been collected continuously since 1999 and are not deleted enabling long term patterns of shelter use to be explored – including episodic homelessness.
- Researchers can apply to use the data through the Statistics Denmark micro-data access service.
- Mandatory collection of national unique ID means that exact matching between datasets can be undertaken.

**Weaknesses:**
- Collection only covers those using shelters.
- Data collection is not explicitly shared between shelters.
- The high level of data sharing and the maturity of the data infrastructure in Denmark is unique; benefits of data linkage are therefore not transferable to other, less ‘data mature’ countries.

**Reference material:**
General overview of accessing data held by Statistics Denmark: https://www.dst.dk/en/TilSalg/Forskningsservice

**Research/data use examples:**
Data system: Specialist Homelessness Service Collection (SHSC)/Specialist Homelessness Services National Minimum Data Set (NMDS)

Summary: The SHSC is a collection undertaken by the Australian Institute of Health and Welfare (AIHW) to gather information about people seeking services from agencies that receive funding under the National Affordable Housing Agreement (NAHA) or the National Partnership Agreement on Homelessness (NPAN) to provide specialist homelessness services. The SHSC contains data from 201 onwards. The NMDS collects information on people who are referred to, or seek assistance from, specialist homelessness services (SHS) agencies. There are two parts to the SHSC: (1) a Client collection which is collected at the person level and, (2) the Unassisted Person collection which is a simple count from the service provider, and is used to gain information about adults and children whose request for service is not able to be met by a SHS agency.

Purpose: Monitor assistance provided by organisations funded by certain government funding streams.

Geographical coverage: Australia.

Population: People who present at SHS requesting assistance. People are considered homeless if they are in ‘non-conventional’ accommodation, rough sleeping, or in short-term emergency accommodation.

Update frequency: Monthly.

Data inputting/collecting organisation(s): Special Homelessness Services agencies which is an organisation which receives government funding to deliver homelessness services to a client, and can be either not-for-profit and for profit agencies. Services provided can include:

- Accommodation services include short-term or emergency accommodation, medium-term/transitional housing, assistance to obtain long term housing, assistance to sustain tenancy or prevent tenancy failure or eviction and assistance to prevent foreclosures or for mortgage arrears.
- Other support services refer to the assistance, other than accommodation services, provided to a client. They include domestic/family violence services, mental health services, family/relationship assistance, disability services, drug/alcohol counselling, legal/financial services, immigration/cultural services, other specialist services and general assistance and support.

Data collection method/storage: Upload data into a validation platform (Validata) either directly, or into a web-version of the system (see Specialist Homelessness Information Platform below). Validata is a validation tool built by the AIHW and introduced to the SHS collection in September 2018, replacing the Specialist Homelessness Online Reporting (SHOR).

Data presentation/sharing: Statistical reports are published by Australian Institute of Health and Welfare. Data cubes and other data are available online. Can be linked to other data through the creation of a Statistical Linkage Key (SLK). The SLK is based on the client’s date of birth, sex and selected letters of the person’s given name and family names.

Information collected: The base unit for the SHSC NMDS is person who presents to a specialist homelessness service. Data relates to periods of support that are structured around episodes of assistance. Information collected includes:

- Basic socio-demographic information including name and date of birth.
- Services required by and provided to each client.
- ‘Sensitive’ data, including mental health issues, Aboriginal status, and migration status (Client has the option not to consent to these data items being shared).
- Details about accompanying children.
- Client circumstances before, during and after receiving support. Data are collected every month during service provision and again at the end of the support period.
- Outcomes including, housing status, employment/education, repeat homelessness, and pathways to housing.

Clients can report collectively, in which case information is collected about each individual, but the group are referred to as a ‘Presenting unit’.

Strengths: Personal data are not shared with the AIHW when generating statistical reports, thereby reducing disclosure risks. Use of SLK reduces processing by AIHW. Data collection on people cannot show how many times a person was unable to receive assistance.

Weaknesses: Limited to those who seek assistance at participating organisations, could bias homelessness estimates.
Reference material:
Main webpage: http://meteor.aihw.gov.au/content/index.phtml/itemId/683255

Research/data use examples:
For an overview article of Australian homelessness research, see:
For research using SHSC data:
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>Specialist Homelessness Information Platform (SHIP).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary:</strong></td>
<td>SHIP supports agencies to report for the Specialist Homelessness Service Collection (SHSC) and allows them to record information about support provided to clients and instances where a person received no immediate services. SHIP is a web based platform hosted by Infoxchange Australia, a not-for-profit social enterprise. The infrastructure externally hosted by Infoxchange, and is maintained in a secure environment which meets or exceeds the Australian Government Protective Security Protocols.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>Free platform for any Special Homelessness Service agency to use and allows agency workers to record client information, case notes, case plans, and client goals.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Australia.</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>People who present at Special Homelessness Services requesting assistance. People are considered homeless if they are in ‘non-conventional’ accommodation, rough sleeping, or in short-term emergency accommodation.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Continuous data entry; monthly upload to SHSC.</td>
</tr>
<tr>
<td><strong>Data inputting-collecting organisation(s):</strong></td>
<td>Special Homelessness Services.</td>
</tr>
</tbody>
</table>
| **Data collection method/storage:** | Organisation has access to its own version of the web-based platform which acts as the local case management system.  
Monthly upload of data into a validation platform (Validata) for use in SHSC. |
| **Data presentation/sharing:** | Statistical reports are published by Australian Institute of Health and Welfare.  
Data cubes and other data are available online.  
Can be linked to other data through the creation of a Statistical Linkage Key (SLK). The SLK is based on the client’s date of birth, sex and selected letters of the person’s given name and family names.  
AIHW do not have access to individual agency databases. |
| **Information collected:** | - Basic socio-demographic information including name and date of birth.  
- Services required by and provided to each client.  
- Details about accompanying children.  
- Client circumstances before, during and after receiving support, including housing tenure and employment status.  
- System allows for documents to be added to case files. |
| **Strengths:** | Data can be exported in order to generate outputs needed for the Specialist Homelessness Service Collection; SHIP is therefore both a case management and reporting tool.  
Reduces duplicate counting through client search function. |
| **Weaknesses:** | Limited to those who seek assistance at participating organisations, could bias homelessness estimates.  
Not a platform for sharing data between organisations. |

**Reference material:**

Main webpage:  

**Data system:** New Zealand Integrated Data Infrastructure (IDI)

**Summary:**
The New Zealand IDI is a collection of whole-population data sources that are generated by public services, including both administrative data and survey/Census data. The IDI falls under the remit of the Statistics New Zealand (Stats NZ).

The unique aspect of the IDI is that data can be linked and are made accessible for statistical and research purposes. As of 2018, the IDI holds over 166 billion pieces of information.

**Purpose:**
Provide an infrastructure for integrated data research on de-identified data.

**Geographical coverage:**
New Zealand

**Population:**
Whole population; specific populations covered vary with data source

**Update frequency:**
IDI refresh of data occurs up-to four times per year

**Data inputting/collecting organisation(s):**
Varies, however primarily public services

**Data collection method/storage:**
Data are stored as separate data tables in an SQL database.

Data are stored in a de-identified format, thereby enabling linkage studies and cross-sector analyses.

Data are linked to a population spine created using tax, birth and death records, and visa

**Data presentation/sharing:**
Data can be accessed via ‘safe settings’

**Information collected:**
Data within the IDI are used in linkage studies; therefore some standard data are required in order to facilitate linkage, including:

- Unique national identifiers from various public services, i.e. the New Zealand Inland Revenue Department number (IRD)
- Unique personal identifiers, i.e. name, gender, date of birth

These data are used to de-identify records whilst assigning a unique person identifier—the snz_uid

Data available in the IDI cover:

- Health
- Social services, i.e. benefits, youth services, and child services
- Education, i.e. primary, tertiary, and secondary education
- Justice and regulation, i.e. court charges, victims of crime, offender data
- Housing, i.e. social housing register
- Economy and businesses, i.e. business register, household panel surveys and labour force surveys
- People and communities, i.e. births, deaths, Census

**Strengths:**
The breadth of data and the permanent storage of data in a de-identified but linkable format greatly reduces the time required to engage in complex data linkage studies.

IDI is designed to enable access to data to a range of audience, not just government analysts.

Other data can be brought into the IDI in order to engage in ad-hoc linkages—see examples of data uses below

**Weaknesses:**
IDI cannot be used for case-management purposes.

Managing data quality is constant as the IDI gathers together data, the quality of data is still the primary responsibility of the data collecting organisation—given the number of depositing organisations, this means that data quality could vary if there are no provider level quality standards in place.

**Reference material:**


**Research/data use examples:**

Pierse, N. et al. (2019) Service usage by a New Zealand Housing First cohort prior to being housed. Population Health. 8:1-10
Data system: Department of Housing and Urban Development (HUD) homelessness data collections/Longitudinal Systems Analysis (LSA)

Summary: In 2001, Congress charged HUD to work with communities to develop unduplicated counts of homeless persons assisted, analyse the patterns of service use by homeless clients, and evaluate the effectiveness of programs locally - part of the Homeless Emergency Assistance and Rapid Transition to Housing (HEARTH) Act. To accomplish this mandate, communities (known as Continuums of Care, or CoC) were required to collect consistent, longitudinal data through what had become known as Homeless Management Information Systems (HMIS). Each CoC has a lead agency responsible for the administration of the HMIS. CoC are required to establish a centralised or coordinated assessment system; though this does not require that the assessment must be centralised and coordinated through the HMIS, some CoC have adopted this approach.

HUD draws on core data sets for to evaluate homelessness service provision and the scale of homelessness in the United States, including: Point in Time (PIT) counts of sheltered and unsheltered persons, Housing Inventory Count (HIC), and HMIS data (now known as Longitudinal Systems Analysis data).

Purpose: Congressional mandate to monitor homelessness in the United States.

Geographical coverage: United States.

Population: People who present at homelessness-related services that forms part of communities (known as Continuum's of Care, or CoC) and also those who are unsheltered. People experiencing domestic violence are exempt from having their data recorded on HMIS systems.

Update frequency: Annual update to HUD.

Data inputting/collecting organisation(s): CoC in the United Stated (399 in 2017).

Data collection method/storage: HMIS are designed to output a series of files known as the Longitudinal Systems Analysis files, which are then reported to HUD.

LSA data are uploaded to HUD using a web-based platform known as Homelessness Data Exchange. Alternatively, some data can be directly input, such as the PIT counts.

Data presentation/sharing: Different data sources are combined to form a national picture of homelessness in the United States, which is submitted to the United States Congress as the Annual Homeless Assessment Report.

Information collected:

PIT data are counts of the homeless population in an area as well as information on specific subpopulations, including chronically homeless persons, veterans, and unaccompanied youth. A PIT count is composed of two parts: a sheltered PIT count which can usually be constructed from HMIS data and which is required every year, and an unsheltered PIT count, which is required at least every other year.

HIC data are an annual inventory of the beds, units, and programs designated to serve an area's homeless population. Counts are provided by household types served in the inventory (i.e., households with at least one adult and one child, households without children, and households with only children). HIC data are often pulled directly from the community’s HMIS.

Communities aggregate their HMIS data and submit it to HUD, including grant applications and annual performance reports for HUD-funded projects. HUD uses aggregated HMIS data from communities that have enough coverage and completeness to determine national estimates on the nation’s sheltered homeless population.

Strengths: Using multiple data sources leads to more valid estimates of homelessness in the United States and makes the task of estimating the scale of homelessness practical given the size of the country. Guidance is given to CoC on what methodologies are acceptable for production of PIT counts of unsheltered homeless people.

Weaknesses: National data linkage not possible given aggregate nature of data returned to HUD, however regional linkage at the CoC level has been undertaken in the past.

Reference material:
Main webpage where information obtained:
https://www.huduser.gov/portal/periodicals/em/summer12/highlight2.html


More information on HUD's coordinated entry requirements: https://www.hudexchange.info/resource/4427/coordinated-entry-policy-brief/
**Data system:** New York City Coalition on the Continuum of Care (NYC CCoC) Homeless Management Information System.

**Summary:** In New York City, the NYC Department of Social Services acting on behalf of the NYC Department of Homeless Services as the HMIS Lead Agency. In addition to meeting the requirements of the HEARTH Act, the goal of the NYC CCoC is to have 100% participation of all homeless service providers within New York City, regardless of funding stream. NYC CCoC uses the software product Foothold Technology Service (FTS) AWARDS as the basis for its HMIS. The HMIS system is referred to as a data warehouse; data are uploaded periodically rather than being live on a HMIS system for others to view. The systems for managing homelessness in the NYC CoC are in the process of changing to a coordinated entry assessment. There is therefore a move to the creation of real-time by-name lists of applicants for assistance, known as a Coordinated Assessment and Placement System.

**Purpose:** Meet requirements under the HEARTH Act; however later alterations to the approach of the NYC CCoC have moved toward coordinated assessment.

**Geographical coverage:** New York City.

**Population:** People who present at homelessness-related services that forms part of NYC CCoC funding. People experiencing domestic violence are exempt from having their data recorded on HMIS systems.

**Update frequency:** Monthly.

**Data inputting/collecting organisation(s):** Projects/services who are funded by NYC CCoC, however, encourage all homelessness-related services to contribute data regardless of funding. For-profit organisations are not eligible to apply for grants; however these organisations may be included in wider data collection. Types of organisation included in the system performance measures, and which are therefore funded:
- Emergency shelters
- Transitional housing
- Permanent Supportive Housing
- Street outreach
- Safe havens
- Housing (with and without support)
- Rapid re-housing

**Data collection method/storage:** Data collection takes the form of a ‘warehouse’ or pool of information rather than data sharing platform. Upload data in CSV format via secure file transfer system, or data are shared with the HIMS system if the local case management systems are compatible.

**Data presentation/sharing:** Regional reporting and data contributes to national report on homelessness. Data not visible to other organisations in area.

**Information collected:** Universal data elements common across all inputting organisations are:
- Name
- Social Security Number
- Date of Birth
- Race
- Ethnicity
- Gender
- Veteran Status
- Disabling Condition
- Project Start Date
- Project Exit Date
- Destination
- Relationship to Head of Household
- Client Location
- Housing Move-In Date (applicable to all permanent housing project types)
- Living Situation

**Strengths:** Data collections across the HMIS inputting organisations are standardised. Repeat service use is accounted for at the HMIS level.

**Weaknesses:** Anonymised data are submitted to government from each HMIS, therefore not able to explore repeat homelessness nationally. Warehouse design meant that data remain fairly siloed – until the coordinated approach was adopted.

**Reference material:**
Data system: Chicago Homeless Management Information System (All Chicago)

Summary: Since 2012, All Chicago has been elected by Chicago's CoC to be the lead agency and administer Chicago's HMIS. Unlike the NYC HMIS which acts as a central data depository and processing system, the Chicago HMIS also enables some sharing of information with other organisations dependent on appropriate data sharing agreements being in place. Organisations participating in the HMIS system must sign an agreement outlining their roles, which includes confidentiality and data access. As part of the move toward a coordinated entry system, the All Chicago system is being used to produce a by-name list, 'One List' – see separate entry for by-name list.

Purpose: All Chicago meets the requirements of the HEARTH Act to have a Homelessness Management Information System.

Geographical coverage: Chicago.

Population: People who present at homelessness-related services that forms part of Continuums of Care funding. People experiencing domestic violence are exempt from having their data recorded on HMIS systems.

Update frequency: Monthly if uploading data; set timeframes if inputting directly into system.

Data inputting/collection organisation(s): Projects/services who are funded by Chicago CoC, however encourage all homelessness-related services to contribute data regardless of funding. For-profit organisations are not eligible to apply for grants; however these organisations may be included in wider data collection. Types of organisation included in the system performance measures, and which are therefore funded:
- Emergency shelters
- Transitional housing
- Permanent Supportive Housing
- Street outreach
- Safe havens
- Housing (with and without support)
- Rapid re-housing

Data collection method/storage: Data are either entered directly into the HMIS via an interface or can be uploaded as a single file.

Data presentation/sharing: Regional reporting and analysis, and data contributes to national report on homelessness. Certain data are visible to others on the HMIS dependent on consents from person. A verbal explanation of the purpose of the HMIS is provided to clients in order to gain consent for inclusion in the HMIS. Each time a client reports to an agency they complete a consent form. For those under the age of 18, parental/guardian consent to share data is asked. Data can be accessed for research purposes.

Information collected: Universal data elements common across all inputting organisations are:
- Name
- Social Security Number
- Date of Birth
- Race
- Ethnicity
- Gender
- Veteran Status
- Disabling Condition
- Project Start Date
- Project Exit Date
- Destination
- Relationship to Head of Household
- Client Location
- Housing Move-In Date (applicable to all permanent housing project types)
- Living Situation

Strengths: Can share information across organisations.

Weaknesses: Anonymised data are submitted to government from each HMIS, therefore not able to explore repeat homelessness nationally. If clients choose to “lock” their files then this may alter their service experience.

Reference material:
Main webpage: https://allchicago.org/HMIS

Research/data use examples:
**Data system:** Online Navigation and Entry System (ONE System)

**Summary:** The ONE system is a new HMIS adopted for the San Francisco CoC and is run by the Department of Homelessness and Supportive Housing. San Francisco adopt a co-ordinated entry system for their homelessness services, which assesses the needs of the people and prioritizes them for a range of types of assistance, including immediate shelter and longer-term housing-focused programs. The ONE System itself is provided by the software solution/company Bitfocus, who also provide training in the system, data quality reports and technical assistance.

**Purpose:** Serve as the main repository for homeless client information, and function as the primary communication tool for service providers.

**Geographical coverage:** San Francisco.

**Population:** People who present at homelessness-related services that forms part of Continuums of Care funding. People experiencing domestic violence are exempt from having their data recorded on HMIS systems.

**Update frequency:** Continuous.

**Data inputting/collecting organisation(s):** Project/services within the San Francisco area. For-profit organisations are not eligible to apply for grants, however these organisations may be included in wider data collection. Types of organisation included in the system performance measures, and which are therefore funded:
- Emergency shelters
- Transitional housing
- Permanent Supportive Housing
- Street outreach
- Safe havens
- Housing (with and without support)
- Rapid re-housing

**Data collection method/storage:** Cloud based online integrated information management system.

**Data presentation/sharing:** Client data can be shared. In order to share information between organisations, a Release of Information form must be signed by the person using the service.

Regional reporting, analysis, and data contributes to national report on homelessness. Data are published in interactive infographic format on the Department of Homelessness and Supportive Housing website.

**Information collected:** Universal data elements common across all inputting organisations are:
- Name
- Social Security Number
- Date of Birth
- Race
- Ethnicity
- Gender
- Veteran Status
- Disabling Condition
- Project Start Date
- Project Exit Date
- Destination
- Relationship to Head of Household
- Client Location
- Housing Move-In Date (applicable to all permanent housing project types)
- Living Situation

Where more than one person – a household – is receiving services, a household entry can be created to which household members are added.

**Strengths:**
- Provides a co-ordinated entry system for San Francisco homelessness services.
- Enables sharing of information about people.
- Data inputters are asked to comment on the quality of key variables such as the persons Social Security Number and date of birth – which is useful for matching quality purposes.
- Standardised forms for inputting organisations helps maintain a level of data quality.

**Weaknesses:**
- Privacy notice adopted is lengthy (at 8 pages long) which calls into question whether consent is “informed”.

**Reference material:**
- Main webpage: [http://hsh.sfgov.org/overview/one-system/](http://hsh.sfgov.org/overview/one-system/)
- Documentation relating to the system are available at: [https://onesf.clarityhs.help/hc/en-us](https://onesf.clarityhs.help/hc/en-us)
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>Knoxville Homeless Management Information System (KnoxHMIS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary:</strong></td>
<td>As with other HMIS system in the United States, the Knoxville/Knox County covers all funded services under the Continuum of Care (CoC). The KnoxHMIS is a computerised data platform. The Knoxville/Knox County implement a coordinated entry system known as CHAMP--Coordinated Housing Assessment Match Plan--which enables people experiencing homelessness to be assessed and assisted by the most appropriate service within the CoC, the HMIS system helps facilitate the CHAMP approach. The University of Tennessee operates the KnoxHMIS system in a university-community partnership.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>Serve as the main repository for homeless client information, and function as the primary communication tool for service providers.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Knoxville, Tennessee</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>People who present at homelessness-related services that forms part of Continuums of Care funding. People experiencing domestic violence are exempt from having their data recorded on HMIS systems.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Continuous input into HMIS; quarterly updates of ‘data dashboards’</td>
</tr>
</tbody>
</table>
| **Data inputting/collection organisation(s):** | Project/services within the San Francisco area. For-profit organisations are not eligible to apply for grants, however these organisations may be included in wider data collection. Types of organisation included in the system performance measures, and which are therefore funded:  
  - Emergency shelters  
  - Transitional housing  
  - Permanent Supportive Housing  
  - Street outreach  
  - Safe havens  
  - Housing (with and without support)  
  - Rapid re-housing |
| **Data collection method/storage:** | Cloud based online integrated information management system. HMIS software provided by SharePoint, which a number of other HMIS system in the United States also use. |
| **Data presentation/sharing:** | Quarterly reports on homelessness within the CoC area are publicly made available on the Knoxville dashboard. The dashboard covers the following areas: emergency shelter bed capacity/usage; point-in-time counts on sheltered homelessness; returns to homelessness indicator; change in housing status (positive/negative/indeterminate); homelessness status and number of clients served; causes of homelessness. Identifying data can be shared without consent for a limited number of purposes, for example to prevent imminent threat of death or criminal investigation, whilst non-identifying data can be shared without permission for research. All other information/reasons to share data require consent, including use of data by other services for the purposes of service provision. ‘Data subjects’ are able to inspect their data by making a written access request. |
| **Information collected:** | Universal data elements common across all inputting organisations are:  
  - Name  
  - Social Security Number  
  - Date of Birth  
  - Race  
  - Ethnicity  
  - Gender  
  - Veteran Status  
  - Disabling Condition  
  - Project Start Date  
  - Project Exit Date  
  - Destination  
  - Relationship to Head of Household  
  - Client Location  
  - Housing Move-In Date (applicable to all permanent housing project types)  
  - Living Situation |
| **Strengths:**    | Data quality indicator on the public dashboard provides public accountability for data standards. Clients can be entered as alias’ into the system, if they refuse to provide this information for themselves/their dependents. University-community partnership model for the HMIS builds links between research and practice. |

**Reference material:**
Main webpage: [https://knoxhmis.sworpswebapp.sworps.utk.edu/](https://knoxhmis.sworpswebapp.sworps.utk.edu/)

**Research/data use examples:**
**Data system:** Clarity - Nevada Statewide Community and Homeless Management Information System (CMIS/HMIS)

**Summary:** The Clarity system is the HMIS system being adopted across the State of Nevada CoC. The Nevada CoC system have made strides to bring together services to form a coordinated entry system, with a corresponding change in how data are collected and shared across the system in order to achieve this aim. Clarity itself is provided by the software solution/company Bitfocus, who also provide training in the system, data quality reports and technical assistance. In order to share information between organisations, a Release of Information form must be signed by the person using the service. The coordinated entry system has been piloted as the new process for families experiencing homelessness, and San Francisco are beginning an Adult coordinated entry system.

**Purpose:** In addition to meeting HUD reporting requirements the CMIS/HMIS explicitly states that data and reporting:
- Help those experiencing homelessness in Nevada achieve economic self-sufficiency.
- Help those at risk of homelessness maintain economic self-sufficiency.
- Preserve community funding and resources.

**Geographical coverage:** Nevada

**Population:** People who present at homelessness-related services that forms part of Continuums of Care funding.
People experiencing domestic violence are exempt from having their data recorded.
Other community service users.

**Update frequency:** Continuous.

**Data inputting/collecting organisation(s):** Projects/services within the Nevada State area. For-profit organisations are not eligible to apply for grants, however these organisations may be included in wider data collection. Types of organisation included in the system performance measures, and which are therefore funded:
- Emergency shelters
- Transitional housing
- Permanent Supportive Housing
- Street outreach
- Safe havens
- Housing (with and without support)
- Rapid re-housing
- Street outreach
- Safe havens
- Housing (with and without support)
- Rapid re-housing
- Fire and ambulance services can enter information about the transport of homeless people to local A&Es.
As the system covers community services, then a wide array of inputting organisations participate.

**Data collection method/storage:** Cloud based online information case management system.

In order to maintain data quality, data are entered onto the system within specific timeframes, dependent on the type of client/service.

System has a low threshold for missing data dependent on the variable type, ranging from 0 per cent accepted missing data to maximum of 5 per cent missing data.

**Data presentation/sharing:** Client data can be shared with other services on the system with the permission of the client.
Regional reporting and analysis, and data contributes to national report on homelessness.
Notifications are placed in sign in and intake locations in order to notify people of data usage. In addition to these notices, Release of Information form must be signed by the person using the service. Consent expires after 7 years.

**Information collected:** Universal data elements common across all inputting organisations are:
- Name
- Social Security Number
- Date of Birth
- Race
- Ethnicity
- Gender
- Veteran Status
- Disabling Condition
- Project Start Date
- Project Exit Date
- Destination
- Relationship to Head of Household
- Client Location
- Housing Move-In Date (applicable to all permanent housing project types)
- Living Situation
| Strengths: | Streamlined referrals and coordination of case management from the integration of data from across the CoC. Lead to new methods of service delivery, including the use of iPads by outreach workers, to a new method of identification that sped up services for regular users of the system. |
| Weaknesses: | No explicit reference or mechanism for research use of data. Potential burden on smaller organisations in order to maintain high levels of data quality. HMIS is obliged to work with organisation to rectify data issues, however if they persist then access to the HMIS can be revoked which can jeopardise the organisations funding. |

**Reference material:**

Main webpage: [http://nvcmis.bitfocus.com/](http://nvcmis.bitfocus.com/)

Nevada Department of Veterans Services (2014) Veteran Homelessness in the State of Nevada. Nevada: Nevada Department of Veterans Services
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>CARES of NY Regional Homeless Management Information System (CRHMIS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary</strong></td>
<td>Corporation for AIDS Research, Education and Services (CARES) of NY, Inc. is a not-for-profit organization that provides planning and administrative services to communities fighting homelessness. In 2003, as the lead agency for coordinating the Continuum of Care groups in 6 Capital Region counties within four Continuums of Care, CARES was asked by providers to assume responsibility for the HMIS development and implementation for the Capital Region of New York State, and the regional HMIS database was established in 2004. The AWARDS database provides the interface for data entry.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>Meets the HUD requirements to produce unduplicated counts. Promotional materials highlight the potential to understand the nature and scope of homelessness, as well as allowing the development, fostering, and maintenance of regional collaborations. HMIS data also serves to assist in the development of programs addressing the needs of people who are homeless.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>System covers 13 Continuums of Care within the New York state area</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>People who present at homelessness-related services that forms part of Continuums of Care funding. People experiencing domestic violence are exempt from having their data recorded.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Continuous.</td>
</tr>
<tr>
<td><strong>Data inputting/collecting organisation(s):</strong></td>
<td>Projects/services within CoC covered by the CRHMIS. For-profit organisations are not eligible to apply for grants, however these organisations may be included in wider data collection. Types of organisation included in the system performance measures, and which are therefore funded: Emergency shelters, Transitional housing, Permanent Supportive Housing, Street outreach, Safe havens, Housing (with and without support), Rapid re-housing</td>
</tr>
<tr>
<td><strong>Data collection method/storage:</strong></td>
<td>The AWARDS system used as the basis for data collection is a web-based platform. Data collection interface for agencies can be tailored, to an extent, including custom forms and reports. However, customisation may come with an additional cost.</td>
</tr>
<tr>
<td><strong>Data presentation/sharing:</strong></td>
<td>Agencies can either opt into or out of giving their clients the option to consent to data sharing. If the agency opts into data sharing, then clients are given 3 options: (1) restrictive but still shared, (2) less restrictive, and (3) no sharing at all. Basic demographics and program information are shared. Only those with rights to view data within a program can drill down into the program specific details. CRHMIS recognises those 18 years old or older as being data owners and therefore able to consent. For minors and incapacitated/disabled adults, the parent/legal guardian has the ability to consent on their behalf. CARES Inc. provide quarterly and yearly reports on the aggregate data collected within the HMIS to the corresponding CoCs. CARES have made a decision to prevent certain agencies from data sharing, including agencies which primarily deal with: mental health, HIV/AIDS, substance misuse, and youth programs that only serve minors. Continuum-wide, aggregate data will be provided to HUD annually as required through the HMIS Annual Performance Report, Annual Homeless Assessment Report and CoC Grant Application.</td>
</tr>
<tr>
<td><strong>Information collected:</strong></td>
<td>Universal data elements common across all inputting organisations are: Name, Social Security Number, Date of Birth, Race, Ethnicity, Gender, Veteran Status, Disabling Condition, Project Start Date, Project Exit Date, Destination, Relationship to Head of Household, Client Location, Housing Move-In Date (applicable to all permanent housing project types), Living Situation</td>
</tr>
<tr>
<td><strong>Strengths:</strong></td>
<td>Each year, Executive Directors determine which agency programs can/cannot share data. This gives agencies the opportunity to review and change data sharing practices. When inputting personal data, data inputters are asked additional data quality follow up questions. This could help when linking data. AWARDS has additional functions which can be used to send client level information securely, and set up client-meetings.</td>
</tr>
<tr>
<td><strong>Weaknesses:</strong></td>
<td>The patchy coverage of the CRHMS over the State means that data sharing is limited to direct case management uses, as it would not provide total coverage needed to explore movement of users and multiple service use across the State. Data sharing enables viewing of top level information, which can be used to determine whether client already exists on the AWARDS database and what program they last engaged with, however access to detailed information is limited to specific users. Scheduling functions within AWARDS can be cumbersome to use.</td>
</tr>
</tbody>
</table>

**Reference material:**
Main webpage: [https://caresny.org/about-cares-inc/](https://caresny.org/about-cares-inc/)
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>Ohio State Human Services Warehouse (OHSDW)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary:</strong></td>
<td>OHSDW aggregates information about the homeless population in Ohio from individual Continuum of Care (CoC) systems across the state to provide a more complete understanding of who experiences homelessness and how this population accesses services throughout the state. OHSDW's work provides the data required to measure existing need and craft practical solutions to meet that demand.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>OHSDW's mission is to strengthen collaborative efforts to develop a state-wide comprehensive strategy to alleviate the interrelated issues of poverty through the analysis of cross-system data related to homelessness and at-risk populations.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Ohio State, which contains 9 CoC</td>
</tr>
</tbody>
</table>
| **Population:**  | People captured under individual CoC systems, which includes people accessing:  
  - Emergency shelters  
  - Transitional housing  
  - Permanent Supportive Housing  
  - Street outreach  
  - Safe havens  
  - Housing (with and without support)  
  - Rapid re-housing |
| **Update frequency:** | Presumed annual. |
| **Data inputting/collecting organisation(s):** | Data are collected/input by CoC as part of their HMIS data collection systems, which are themselves a requirement of receipt of funding from central Government. Participation of CoC in the warehouse is voluntary. |
| **Data collection method/storage:** | Participating CoC export their data in CSV or XML format. Extracted data elements are similar to the data types sent to HUD. All personally identifying information will be encrypted to protect the identity of any individual. The data warehouse will be physically housed at the Ohio Housing Finance Agency (OHFA). The data warehouse is governed by a steering committee composed of eleven members including one representative from each of the participating CoCs, OHFA, and from the Coalition on Homelessness and Housing in Ohio (COHHIO). COHHIO will be an ex-officio member of the OHSDW Steering Committee and will not have voting privileges. These constituents will act on behalf and with the authority of their respective agencies/CoCs to make decisions for the use and functions of the OHSDW and all data and other information entrusted to the collaborative group. |
| **Data presentation/sharing:** | Access to de-identified raw data will be possible but limited. Local communities will only have access to aggregate reports specific to their region and their community upon approval of the data warehouse steering committee. It is intended that reports will be run, the contents of which will be expressly identified and agreed upon between the participating CoCs. Future developments will be to link homelessness data to other routinely collected data, i.e. medical service data from the Ohio Mental Health and Addiction Services. |
| **Information collected:** | Universal and project specific data elements will be submitted to the OHSDW, including:  
  - Race  
  - Ethnicity  
  - Gender  
  - Veteran Status  
  - Disabling Condition  
  - Project Start Date  
  - Project Exit Date  
  - Destination  
  - Relationship to Head of Household  
  - Client Location  
  - Housing Move-In Date (applicable to all permanent housing project types)  
  - Living Situation |
| **Strengths:**   | Combining CoC data enables a state-wide response to homelessness to be produced. Operation of warehouse is designed not to disrupt local level data collection processes. Ohio state CoC are invited to be a part of the steering committee for the OHSDW. Enables unduplicated counts for the state, rather than the individual CoC. Participation in the warehouse is free and the system designed to accept data that is readily available. |
| **Weaknesses:**  | Voluntary participation means that total coverage may not be possible. Data are not for use for case management purposes, i.e. guiding individual plans of care. |
Reference material:
Main webpage: http://ohiodatawarehouse.org
Initial report base on data:
Frequently asked questions document produced by OHSOW:
http://ohiodatawarehouse.org/documents/OHSOW_FAQs.pdf
Data system: Michigan's Statewide Homeless Assistance Data online Warehouse (SHADoW)

Summary: The SHADoW is a data warehouse bringing together human services (health) data and homelessness data, in order to provide a de-identified client level resource. 59 Continuum of Care (CoC) from across Michigan State have come together as part of the data warehouse, covering roughly 500 homelessness agencies. As one of the few states with a state-wide HMIS technology, Michigan was uniquely positioned to integrate homeless management data and public sector support services data using warehousing tools. The SHADoW project was able to build upon learning from other Michigan state data warehousing projects. The SHADoW project has a single steering group in order to provide governance and oversight for the project. Stakeholders are present at a policy level when making key decisions, whilst IT staff are also included.

Purpose: Enable cross jurisdiction counting and tracking of homeless individuals, and inform on public policy issues related to homelessness.

Geographical coverage: Michigan State.

Population: People interacting with homelessness-related services, covering both state and public sectors.

Update frequency: Unknown, however all data are deleted before re-creating updated data, therefore assume that this is not more regular than annually.

Data inputting/collecting organisation(s): Organisations working with homeless people, across a range of services, including:
- Emergency housing and shelter programs
- Community action and other emergency services
- Health care providers
- Mental health and drug treatment providers
- Youth programs
- Churches
- Food pantries

Data collection method/storage: Contributing agencies retain control over the data, and HMIS participating agencies can opt out of newer ‘builds’ of the data set. A subset of information from CoC HMIS systems are uploaded to the warehouse that then matches and de-identifies the data, retaining instead a random number to identify an individual. Data are stored as an ‘instance’, i.e. a point-in-time collection of data. SHADoW uses a ‘kill and fill’ method to build the linked data set, whereby no data are retained from previous instances of the data upload, i.e. all data are deleted and replace.

Data presentation/sharing: Data use is restricted for planning and research. From the outset, it has been stipulated that the data would not be used for case management or agency evaluation. Data are accessed via an interface that enables querying of the data – WEB intelligence, a BusinessObjects product. Queries of data returns counts, where the counts are greater than 5. Data sets can be accessed subject to signing a Data Use Agreement. Research access to the de-identified data set is strictly controlled by the SHADoW Board that includes representatives from the homeless service sector and HMIS, state departments and other professionals whose contribute expertise to evaluate data requests.

Information collected: In order to match data from different sources, the following personal data are requested:
- Client name
- Date of birth
- Social Security Number

As data come from diverse sources, data are transformed within the warehouse into a consistent form that includes:
- Person Characteristics
- Event Information:
  a. Age of Consumer
  b. Date of Event, i.e. entry into housing support
  c. Event Type
  d. Cost of Event (as appropriate)

Strengths: Built-in statistical disclosure controls as queries only return data where counts greater than 5. Designed specifically to enable research uses of the data and answer policy relevant questions. SHADoW project has strong leadership under the Statewide HMIS project, which had a history of leading large projects.

Weaknesses: Data are deleted and then re-downloaded, which may be resource intensive and does not allow for revisions in data to be monitored. Geographic information has been limited intentionally in order to protect privacy, however this may limit research uses.

Reference material:
Main webpage: https://www.hudexchange.info/resource/1696/hmis-data-warehousing-curricula/
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>Homeless Individuals and Families Information System (HIFIS).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary:</strong></td>
<td>The HIFIS software feeds into Canada's national monitoring system, the National Homelessness Information System (NHIS). Employment and Social Development Canada (ESDC) developed the HIFIS software. ESDC owns all rights, titles and interest in the HIFIS software and related documentation. Licence to use the HIFIS software is granted by ESDC in return for non-identifiable personal information related to the Service Provider and its clientele. The unique selling point of the HIFIS software is that it can be tailored to suit the inputting organisation's needs. The latest version of HIFIS is version 4 and has a greater emphasis on data sharing and service co-ordination. HIFIS was built was built based on a series of broad coalitions between government, university researchers, service providers, front-line workers and the homeless themselves.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>HIFIS is a downloadable software package that helps organizations that support people who are homeless or at risk of becoming homeless.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Canada.</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>People interacting with homelessness-related services, primarily homeless shelters.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Continuous into local software; De-personalised data are exported to the NHIS every 3 months.</td>
</tr>
<tr>
<td><strong>Data inputting/collecting organisation(s):</strong></td>
<td>Organisations working with homeless people, across a range of services.</td>
</tr>
<tr>
<td><strong>Data collection method/storage:</strong></td>
<td>HIFIS can act as an organisations local information management system. The software is customizable to suit local data collection needs. Automatically uploads de-personalised data that contributes to the NHIS.</td>
</tr>
<tr>
<td><strong>Data presentation/sharing:</strong></td>
<td>Cases can be searched. Newer version of software enables cross-community sharing of client data. Reports can be run at different levels, which can be used for operational purposes. Option available in latest version of the HIFIS software to generate reports in graphical format. HIFIS users are required to sign a Data Provision Agreement (DPA). The DPA details the responsibilities of all service providers that are entering data into the HIFIS software with respect to data collection, privacy and security. It also outlines the Government of Canada's responsibilities and obligations related to the use and protection of data.</td>
</tr>
</tbody>
</table>
| **Information collected:** | Contains roughly 1,000 different pre-set fields, however, most HIFIS licensed organisations use 20-30% of the fields. Indicative fields include: Demographics of clients, housing histories, health, education, record of support given, case management, and housing placement information. The following information is shared with the national homelessness data collection:  
- Shelter information.  
- Client information, including income, education, housing (date of birth is only personal information shared with national data collection).  
- Client shelter information, i.e. reason for using shelter, dates of use  
- Turn-away information, i.e. not giving assistance. |
| **Strengths:** | HIFIS software can be tailored to organisational needs which potentially increase buy in from the homeless sector. Participation is voluntary and the software is free, also includes a service helpdesk. Collection of data relating to turn-aways could help assess equality practices of organisations. |
| **Weaknesses:** | National level data as part of National Homelessness Information System is anonymised so no linkage to other data possible. |

**Reference material:**


**Research/data use examples:**


**Data system:** Calgary Homeless Management Information System (HMIS).

**Summary:** The Calgary area has adopted a homeless-serving system-of-care, being an inter-related network of agencies and people working together to ensure those at risk of or experiencing homelessness have timely access to the right housing and the right resources. As the coordinator of Calgary’s homeless-serving system-of-care, the Calgary Homeless Foundation (CHF) uses its resources, expertise and database (HMIS) to develop an integrated system-of-care that creates greater ease and access to services. The Calgary HMIS is an electronic data management system that allows agencies to track key indicators on homelessness. Many organizations serving homeless persons in Calgary enter client information into the HMIS, such as age, health status, employment status and housing status upon initial intake. While the client is receiving services, updated information is entered either every three months during support, and/or upon exiting – dependent on the type of service. The CHF is responsible for overseeing the HMIS system, providing training and technical support, running reports and analysis to meet requirements of funders and for performance management, and sharing data with external researchers.

**Purpose:** Provides insights into homelessness, enabling providers to efficiently and effectively focus resources where they will have the most impact.

**Geographical coverage:** Calgary.

**Population:** People interacting with homelessness-related services.

**Update frequency:** Continuous.

**Data inputting/collecting organisation(s):** Organisations working with homeless people in Calgary who are funded by Calgary Homeless Foundation. Program types that form the basic building block of the coordinated system:
- Affordable housing
- Prevention
- Emergency shelters
- Rapid rehousing
- Support services only
- Outreach
- Short-term supportive housing
- Housing and intensive support
- Permanent supportive housing

**Data collection method/storage:** Web-based information sharing platform.

**Data presentation/sharing:** Cases can be searched across the whole HMIS. Data anonymised and used for research purposes. Client data can be shared across the HMIS. HMIS has a report creation functions.

**Information collected:** People engaging with programmes are given the option to have their information entered the HMIS anonymously (without their name), under the advisement that this will mean that services cannot be coordinated. Otherwise, the client is asked to sign a Release of Information consent form enabling the sharing of their information. Two forms are needed, one for within organisation sharing and one for external sharing amongst the other HMIS users. Standard personal information collected includes:
- Name
- Basic demographics
- Citizenship/migrant status
- Ethnicity
- Housing needs
- Self-reported health

**Strengths:** Consent process is nuanced, which allows the person to choose an option of consent that suits their concerns.

**Weaknesses:** Clients can opt out of having their data shared which may lead to the most risk averse/disadvantaged people not receiving joined up assistance.

**Reference material:**
- More information on the Calgary area’s approach to homelessness services: [http://calgaryhomeless.com/about/our-role/](http://calgaryhomeless.com/about/our-role/)
### Data system:
Shelter Management Information System (SMIS)

### Summary:
The City of Toronto piloted the use of a centralised information management system in selected emergency shelters in 2009, which was then adopted by all City-funded shelters by early 2010. Each shelter has its own interface with the system, and are authorised to view/edit information that pertains to their programmes. However, in order to coordinate bed placements across the city, some limited personal information is visible to other SIMS users. This reduces double booking of beds for the same client, and also enables information sharing which could help tailor services to client needs.

### Purpose:
SMIS is an information management system designed for City of Toronto funded shelters and has been designed to data-to-day shelter operations.

### Geographical coverage:
Toronto.

### Population:
Homeless people attending a City of Toronto funded shelter.

### Update frequency:
Real-time, used for day-to-day shelter management.

### Data inputting/collecting organisation(s):
Shelters serving homeless families, singles and young people in the City of Toronto area.

### Data collection method/storage:
SMIS is a web-based information management system. Data are based around clients, i.e. client intakes are added to a basic client entry. If case relates to a family, then a family intake is completed where the ‘head’ of the family is entered first and additional members are added.

### Data presentation/sharing:
Data are real-time which enables occupancy of shelters to be assessed, thereby enabling client referrals and allocation of vacant beds.
- Reports on bed occupancies can be generated.
- Basic personal information are shared across the SIMS network.
- Collection of information notice explicitly indicates that the data are shared across SMIS; used to enable administering of social assistance funds; shared with Toronto Public Health; and for research/analysis on de-identified data.
- Summary data are publicly available on the City of Toronto website for daily and monthly shelter occupancy.

### Information collected:
Shelters using the SMIS use the same data entry templates to ensure that data collection is consistent. A limited amount of client information is available to every Toronto SMIS user. If a client was previously registered in the system, then every intake worker can automatically see the following information:
- Client name (first, last and alias)
- Date of birth
- Gender
- Client number (generated by the system)
- Brief description of any health and safety flags
- Client status—inactive or active (i.e. currently assigned a bed within a shelter)

Includes a case management feature, limited to the inputting shelter, which can be used in order to create case notes and organise follow up with clients.

### Strengths:
Provides shelters with information (bed) management systems that may increase the efficiency of their services, compared to a reliance on paper based records.

- Basic personal information are shared across the system, thereby enabling shelters to ensure appropriate provision of services to individuals.
- System automatically checks whether the person is eligible for social assistance funding, the Personal Needs Allowance (PNA).

### Reference material:


### Research/data use examples:

**Data system:** Veterans Health Administration Corporate Data Warehouse (CDW)

**Summary:** The CDW is an integrated healthcare information system supporting the Department of Veterans Affairs Veterans Healthcare Administration (VHA). The CDW was developed in 2006, and was designed to include advanced data management, statistical, graphical, and business analysis functions. The system is structured such that updates of data from other databases are combined within a ‘data warehouse’. There are additional functionalities that enable the warehouse to be queried and to generate outputs. The CDW provides longitudinal data relating to over 22.3 million veterans, as well as supporting a large number of active users of the system across the United States.

**Purpose:** To bring together patient data from across the different Veterans Health Administration settings in order to drive improvements in veteran health care provision.

**Geographical coverage:** United States; covers over 1,000 medical settings

**Population:** Veterans enrolled in the VA healthcare system.

**Update frequency:** Daily.

**Data inputting/collecting organisation(s):** Healthcare providers across the Veterans Healthcare Administration.

**Data collection method/storage:** Data are uploaded from a number of corporate electronic health record systems at VHA settings, including: the Computerized Patient Record System, VHA Regional Data Warehouses, and VHA National Patient Care Database. Data are patient level, but include other entity types, such as prescriptions for medications. The aim of the CDW is to enable patient research, therefore entities are linkable to the patient/case. Data within the CDW have replaced other data collections; being administrative data based, then there has had to be some compromise/trade-off between the timeliness of the data in the CDW and the quality of the data.

**Data presentation/sharing:** Data warehouse can be queried in order to answer general reporting questions but can also return individual level data. Warehouse supports a range of other functions, including: predictive analytics and live flagging of ‘at risk’ individuals; graphical tools to aid in population health management; data mining to explore new relationships within the data; evaluation by comparing outcomes between groups, for example comparing outcomes between two drug groups. Functions have been developed which draw down data from the CDW in order to fulfil a variety of research and practitioner needs. Data have also been extracted from the CDW in order to form stand-alone data resources, such as the Veterans Affairs Patient Database 2014–2017.

VA data can only be accessed by VA-affiliated clinicians and investigators.

**Information collected:** Data warehouse combines over 60 domains of clinical information. These domains were agreed upon by a governing body composed of clinicians and operational leads, in order to provide data necessary for a broad range of purposes. Data items covered by the VHA include:

- patient demographics
- in-patient stays
- out-patient clinic visits
- pharmacy orders and fills
- laboratory results
- vital signs
- patient assessments
- aspects of cost

**Strengths:** Querying capability can help identify patients who need follow up. Combining data from multiple sites enable healthcare usage histories to be constructed, i.e. where services were being sought in different areas. Having data in a warehouse enables novel functions to be bolted on and draw down data from the system, improving ways of accessing and representing data for clinical decision making (in addition to its use for service design).

**Weaknesses:** Alert fatigue can lead to staff missing red flags if they are bombarded with excess amounts of data.

**Reference material:**
- Fihn, S. et al. (2019) Insights From Advanced Analytics At The Veterans Health Administration. Health Affairs. 33(7):1203-1211
Research/data use examples:
**Data system:** Virginia Longitudinal Data System (VLDS)

**Summary:** The Virginia Longitudinal Data System (VLDS) is an example of a number of other longitudinal data systems in the United States, funded under the Statewide Longitudinal Data Systems Grant Program. The aim of the program has been to help states make informed decisions to improve student learning and outcomes by designing, building, and maintaining longitudinal data systems. The grant programme was established under the Education Technical Assistance Act of 2002, and began providing grants in 2005. Over time, the grant, and the resultant data systems, have evolved to meet changing data needs; for example, integrating early childhood, then post-secondary education, and workforce data.

Funded in 2009 under the Statewide Longitudinal grant scheme, the VLDS is an example of a federated data model. This model of creating a data collection/reporting system differs from a centralised collection—i.e. where all data are brought together into a single database/set—in that the data remain in their host agency’s systems. Data are then extracted when a request is made.

**Purpose:** Primarily, to drive policy change to improve the education and employment outcomes through the provision of longitudinal data resources.

**Geographical coverage:** Virginia State.

**Population:** People who have engaged in some form of education.

**Update frequency:** As-and-when requested; data are not stored centrally.

**Data inputting/collecting organisation(s):**
- State agencies participating in the federation cover:
  - Education (schools, college, and Higher Education)
  - Employment
  - Social Services
  - Healthcare

**Data collection method/storage:** Federated data model meaning that data remain in their host institutions system and are automatically extracted by a ‘shaker’. Requests for data must be cleared by the host organisations.

This model of data provision was necessary as State laws prohibit the centralised collection of data, for example in a date warehouse type model.

Data are merged from the different participating agency data sources.

A data-broker manages requests for and provision of data.

**Data presentation/sharing:** Requests for data are made via a portal in order to manage requests.

Data results are de-identified, and can be downloaded by the requester. If the data have not been downloaded after a period of time, they are deleted from the system.

Researcher’s requesting data must have an agency sponsor who then creates a profile within the data portal. Requesting researcher completes a research proposal outlining the purpose and data needs.

As the data for each request are drawn directly from the live systems then they may change over time as the underlying system data is changed.

There is an option to retain unmatched records from agency databases, which can be used to create comparison groups and exploration of linkage bias.

Data are provided as separate data extracts from each system that the researcher merges to meet their own analytical needs.

**Information collected:** Data from partner databases is combined upon request, meaning that the total possible number of variables across these systems is large (over 2,000). Types of data covered include:
- student assessments/outcomes
- screening tests
- student funding sources
- wages and employment data
- juvenile detention and criminal justice
- fostering
- health surveys (covering health conditions)

There are limitations as to what kinds of cases are included in each data source. For example, education data relate to state funded schools, and only assessments that are nationally reported one, i.e. not local school assessments. Wage data only relate to those employers operating in Virginia who are required to report on Unemployment Tax.

**Strengths:** Having numerous departments as partners in the federation enables cross-cutting issues to be explored, such as health and education.

Data being retained within the host agency’s systems, and de-identified upon leaving that system, enabled the VLDS to overcome data protection laws in the State.

Data are drawn from live systems, therefore the returned data requests are up-to-date.
<table>
<thead>
<tr>
<th>Weaknesses:</th>
<th>Resource is required to clear data requests at host agency; dependent on the number of requests this could become burdensome. Data are encrypted for each different request, therefore cannot build up a longitudinal resource form annual requested extracts, also means that if additional variables are requested, a completely new extract has to be compiled. Data from external sources cannot be linked as data are de-identified.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference material:</td>
<td>Main webpage: <a href="https://vlds.virginia.gov/">https://vlds.virginia.gov/</a></td>
</tr>
<tr>
<td>Data system:</td>
<td>Stella P</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Summary:</td>
<td>Stella P provides dynamic visualisations of HMIS data for communities funded in the United States. Rather than being a different form of data collection, the system builds upon data already being submitted to HUD by communities as part of the Longitudinal Systems Analysis. There is no requirement for communities to use the system, however it has the potential to create measures of local (community) performance, and thereby drive work towards the goal of ending homelessness. As of 2019, Stella was still in production, however attest version of the application was being developed.</td>
</tr>
<tr>
<td>Purpose:</td>
<td>The purpose of Stella is to enable communities to assess how their homelessness systems are performing and model a system that enables a community to more fully address homelessness.</td>
</tr>
<tr>
<td>Geographical coverage:</td>
<td>Accessible to all States (in United States)</td>
</tr>
<tr>
<td>Population:</td>
<td>People interacting with homelessness-related services, covering both state and public sectors, i.e. those covered by the HMIS data</td>
</tr>
<tr>
<td>Update frequency:</td>
<td>As-and-when a community requires</td>
</tr>
</tbody>
</table>
| Data inputting/collecting organisation(s): | State agencies participating in the federation cover:  
  - Education (schools, college, and Higher Education)  
  - Employment  
  - Social Services  
  - Healthcare |
| Data collection method/storage: | Data are originally collected and stored by communities as part of their Continuum of Care HMIS system.  
The Stella system itself is a web-accessible service and is built into the same system used to upload data to HUD (HUD Exchange).  
Files from the Longitudinal Systems Analysis are combined to generate reports. |
| Data presentation/sharing: | Tool is for data visualisation; therefore a series of diagrams are generated automatically from the data. Outputs are not publicly accessible—though a community could decide to publish the outcome measures. |
| Information collected:  | Performance measures which are generated by Stella include:  
  - Number of day homeless  
  - Exits from the homeless system to permanent destinations  
  - Returns to the homeless system after exits to permanent destinations  
These three measures are created using household level data rather than individual level. Measures can be viewed by several different options, including:  
  - Time trend  
  - As a system map, showing movement through different services—which can itself be filtered by household type and services.  
  - For specific populations, i.e. first-time homeless, those fleeing domestic violence  
There is the ability to explore demographics of homeless people, which is the only data at the individual level. |
| Strengths:              | Automatically undertakes visualisation that a community may not be able to undertake given their own resources—demonstrates value-added  
Visualising data helps in goal setting and engaging with data.  
Provides whole system metrics of performance, rather than static points in time. |

Reference material:  
Main webpage: [https://www.hudexchange.info/homelessness-assistance/stella/](https://www.hudexchange.info/homelessness-assistance/stella/)
### Data system:
The BIWM formed part of series of projects aimed at designing a methodology for using homelessness services data in Poland. At the time of development, Poland focused on the use of headcount surveys of homelessness as a way of enumerating the homeless population. Further, as indicated by Wyngnanska (2015), there was a generally negative attitude to the sharing and centralisation of homelessness data collections.

The new methodology needed to generate new indicators of homelessness and enable researchers/policymakers/funders to understand the basic characteristics of the homeless population. As part of these projects, NGOs were engaged with in order to produce a standard template for data entry. A methodology of anonymising data was produced, as well as systems for collecting and processing data. Individual level anonymised data were then provided to a research team, who then combined these data sources to aggregate the data at the individual level, i.e. take multiple service interactions and generate person level indicators of service use. Like the operation of HMIS’ in America, the end product of these projects was a unique count of people who were homeless in the pilot area.

### Purpose:
The goals of the projects were to demonstrate that data sharing between NGOs was possible without generating privacy issues; to generate an indicator of “flows” of people who were homeless, being a count of unique individual using services; to demonstrate that such indicators could be created through software and appropriate methodology; and finally to create a standard for data collection.

### Geographical coverage:
Initially single municipality within Poland, however methodology could enable data sharing at multiple scales.

### Population:
Individuals who have engaged with a homelessness service, e.g. a shelter or health clinic, and those who were recorded as using a service because of homelessness, e.g. local welfare centre or hospital.

### Update frequency:
One off as part of a research exercise.

### Data inputting/collecting organisation(s):
Services working with homeless people.

### Data collection method/storage:
Data were collected using a free software database that was designed specifically for the project. Software acted as a local database for each participating service, and output required person level data for that service.

Developing the BIWM standard meant merging European data recommendations (i.e. the ETHOS definition of homelessness) with local level data collection practices.

### Data presentation/sharing:
The aim of the data standard was to develop an indicator of multiple service use; therefore, sharing of data was fundamental to its success.

In order to comply with data protection legislation, personal identifiers were required to be anonymised prior to leaving each service.

Additional computer coding was provided to services that anonymised personal identifiers through the creation of a unique person ID. This unique reference number was created in the same way by all services, and enabled linkage between data from different services.

### Information collected:
The BIWM standard comprised of a set of variables to be collected by all services adopting it, which included:
- Identifiers (name, date of birth, sex)
- Address details
- Citizenship
- Marital details
- Education
- Declared length of homelessness
- Household structure
- Relationships
- Type of service being accessed
- Living situation before and after service (following ETHOS definition)
- Reason for most recent episode of homelessness
- Support needs
- Source/type of income before and after service

Historical data were requested as part of the pilot project, covering a period of 3 years prior to the project start date.
### Strengths:
Anonymised (‘pseudonymised’) data were transferred between organisations, which satisfied data protection laws, and may ease service provider concerns about sharing personal information. Requesting data on people attending a service because they were homeless challenged assumptions that the homeless were heavily using the local welfare service. Lead to the production of new metrics of homelessness services use.

### Weaknesses:
The request for historic data lead to time consuming and costly data inputting task, as historic data were mainly paper based. BIWM project found that the reliability of ETHOS typology of homelessness as a tool for measuring homelessness required clarification and further research.

### Reference material:
Data system: North Carolina School Works (NCSW)

Summary: In 2012, the North Carolina General Assembly enacted legislation to create the North Carolina Longitudinal Data System; bringing together workforce data with education data from all levels, in order to facilitate their link age and generate insights that can help improve education and workforce participation. In 2012, North Carolina received a $3.6 million under the Statewide Longitudinal grant scheme to facilitate the creation of a longitudinal data system. The NCSW is the resultant system and adopts a federated data approach with a data broker responsible for managing data queries and merging of data from partners involved in NCSW.

Purpose: Provide insight into education and workforce participation.

Geographical coverage: North Carolina state.

Population: People engaged in education and work.

Update frequency: Data queries compiled from agency data, rather than School data; therefore timeliness of data will vary.

Data inputting/collecting organisation(s):
- North Carolina Department of Public Instruction
- University of North Carolina General Administration
- North Carolina Community College System
- North Carolina Independent Colleges and Universities
- Labour and Economic Analysis Division of the North Carolina Department of Commerce

Data gathering method/storage: Data remain in the partner agencies' systems, and the data broker automatically transmits the requested data.

Students and staff are assigned a unique identifier (UID) by a separate system: NC Department of Public Instruction Unique Statewide Identifier (UID) System. Students/staff can be searched to find pre-existing UIDS; if no record exists, then a new UID can be generated.

Data presentation/sharing: The data broker is responsible for requesting data and can merge data from different partner organisations, if the request requires multiple data sources. Before data are released by NCSW, all partners supplying data for that request must approve it. The NCSW lacks a formal mechanism for requesting research access to the data. However, for non-partner data requests, individual de-identified data can be supplied, where the unique identifiers are encrypted differently for each data request (avoiding the ability to link data across different requests), or aggregate data are provided. Data for staff and students are linked by their UID.

Information collected: Education and workforce data, covering preschool, grade 20 or higher education, and workforce participation.

Strengths: Enables longitudinal evaluation of education programmes. Data quality is evaluated by the Data Quality Campaign; a non-profit advocacy organisation that assesses each state's longitudinal data system to determine their effectiveness.

Weaknesses: Lacks a formal process for external access to the data. Data exchanges are anonymised through the UID, and system is not designed to follow individuals.

Reference material:
Main webpage: http://slds.rhaskell.org/state-profiles/north-carolina#research-identifier
<table>
<thead>
<tr>
<th><strong>Data system:</strong></th>
<th>e-Estonia X-tee</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary:</strong></td>
<td>Estonia has developed an integrated approach to public sector - citizen interaction using technologies, or e-Services. These services span voting, tax, health, governance and residence, amongst others, and are underpinned by digital identities that enable users to interact with these different services seamlessly, whilst reducing duplication of data. Underpinning this digital development, is X-tee, initiated in 2001, and provides a means for data to be linked, shared, searched, and written across departments. X-tee is referred to as a ‘data exchange layer’ and uses a federated approach to data extraction and transmission, meaning that data remain in their original host systems, rather than being entered into a central database. In 2018, X-tee has expanded to include a data exchange system or ‘trust federation’ with Finland, known as X-Road. X-tee implemented a different way of governing data, whilst using already existent technologies.</td>
</tr>
<tr>
<td><strong>Purpose:</strong></td>
<td>Enable secure exchange of information across state agencies.</td>
</tr>
<tr>
<td><strong>Geographical coverage:</strong></td>
<td>Estonia, extended to Finland in 2018</td>
</tr>
<tr>
<td><strong>Population:</strong></td>
<td>Users of public services.</td>
</tr>
<tr>
<td><strong>Update frequency:</strong></td>
<td>Continuously; live data exchange</td>
</tr>
<tr>
<td><strong>Data inputting/collecting organisation(s):</strong></td>
<td>Public services.</td>
</tr>
<tr>
<td><strong>Data gathering method/storage:</strong></td>
<td>The system operates on a federated approach, where data remains in the producing organisations systems. In order for the federated approach to function, agencies are required to install a security service for the X-tee system.</td>
</tr>
<tr>
<td><strong>Data presentation/sharing:</strong></td>
<td>Data held by other agencies can be searched/queried, for example driver’s licences are no longer necessary as police can search the driver data base in order to check vehicle registration. In order to share data through X-tee, the separate data owning agencies must have agreements in place to enable sharing/viewing each other’s data. Text data can be exchanged across the federation, and whole files. All data are encrypted during exchange. Others can use data in state systems within the mandate of performing the business operations.</td>
</tr>
<tr>
<td><strong>Information collected:</strong></td>
<td>Information collected can vary on the agencies involved in the federation.</td>
</tr>
<tr>
<td><strong>Strengths:</strong></td>
<td>Data is stored where it is created. X-tee has been scalable as new services have developed and been added to the e-Estonia portfolio. The system uses technologies to undertake automated processes, freeing up services to work with the people; working hours are saved though shared data.</td>
</tr>
<tr>
<td><strong>Weaknesses:</strong></td>
<td>Requires agreements in place to enable data sharing, which can be a time consuming aspect of data share.</td>
</tr>
<tr>
<td><strong>Reference material:</strong></td>
<td>Main webpage: <a href="https://www.ria.ee/en/state-information-system/x-tee.html">https://www.ria.ee/en/state-information-system/x-tee.html</a></td>
</tr>
</tbody>
</table>
Appendix B:

Individual level data from commissioned services

Single incidence of database (i.e. organisations work area)

Single incidence of database (i.e. organisations work area)

Single incidence of database (i.e. organisations work area)

Housing options client database

Welsh Government data team

Payload data & person ID

Data validation and upload/download software (e.g. AFON)

Payload data, person ID & hashed identifiers

Governance structure: use of data and inclusion

Payload data & person ID
Transform individual level data extraction (Option 1)
Federated model of data access (Option 2)
Housing services database (Can be expanded to include other services over time)

Warehouse staging area: data are cleaned and consolidated

Data warehouse (Regional or national)

Payload data & identifiers

Statutory & HSG data

Payload data & identifiers

Identifiers and person ID

Housing Options client database

Payload data & identifiers
Central data warehouse (Option 3)
Welsh Government data team

Governance structure: data quality monitoring, use of data, and inclusion

Identifiers & person ID

Payload data & person IDs

Data management solution (Regional or national)

Single incidence of database (i.e. organisations work area)

Core data which is common across all incidences and can be accessed

Single incidence of database (i.e. organisations work area)

Single incidence of database (i.e. organisations work area)

Etc.
Integrated data management system (Option 4)