



# Applying Behavioural Insights to Cross-government Data Sharing

---

## Final Report to ADR UK

Daniel Bogiatzis Gibbons, Philipp Friemann, Eva Kolker, Eleanor Collerton, Alex Sutherland

16 December 2020

Funded by:



# Table of contents

---

<b>Table of contents</b>	<b>1</b>
<b>1. Executive summary</b>	<b>3</b>
<b>2. Introduction</b>	<b>7</b>
Structure of this report	8
The ADR UK Model	8
<b>3. Barriers to data linkage, sharing and deposition</b>	<b>10</b>
3.1 Resource barriers relate to both availability and prioritisation by senior staff	10
Project Management	10
Stakeholder Management	11
Data Quality and Management	12
IT Infrastructure and Skills	13
Knowledge	13
3.2 Judgments: risks, rewards and uncertainty	14
Legal risks and uncertainty	15
Predicting public perception	17
Rewards may accrue to other public bodies or may not be strong enough	19
<b>4. Behavioural problems and solutions</b>	<b>20</b>
4.1 Resource Barriers: Intensity and Prioritisation	20
Planning failures	20
Friction costs	25
Allocation of attention	26
4.2 Judgments: risks, rewards and uncertainty	28
Loss aversion and status quo bias	28
Present bias	29
Availability heuristic	31
Desire for control	33
<b>5. Structural solutions</b>	<b>35</b>
Data preparation and engineering resource	35
Advising on data collection	36
Software library	38
Getting advice on legal gateways	38
Approvals process for the ONS SRS	39
<b>6. How to involve researchers, non-departmental bodies and councils</b>	<b>41</b>
Non-departmental bodies	41

Research organisations	41
Local authorities	44
<b>7. Ideas for future linkages</b>	<b>45</b>
A national system for monitoring rates of child maltreatment	45
Data sharing for violence prevention	46
Vulnerable consumers and consumer/household finance	47
Other ideas	48



# 1. Executive summary

---

Linking administrative data for research is key for innovation in policy development, for example by enabling the evaluation of policies or answering questions about segments of the UK population. ADR UK (Administrative Data Research UK) was created by the ESRC (Economic and Social Research Council) to facilitate the linkage of public sector data and to make it available to researchers in a safe and secure way. A key challenge in this effort is how to persuade government departments as well as other public bodies acting as data controllers to share and link data. ADR UK commissioned the Behavioural Insights Team (BIT) to identify key barriers to data sharing, as well as potential solutions. We conducted interviews with key stakeholders in government, desk research and a brief review of the behavioural science literature. Some of the barriers we identified are structural or technical. Others, however, relate to the cognitive biases and other psychological factors that drive human behaviour. We identified two overarching types of barriers:

- **Resources:** Participating in data linking projects is resource-intensive in terms of IT capacity, project management and staff time. While these challenges, as well as the solutions we identified, are in part structural, there are also behavioural factors that can hamper effective project management.
- **Judgements of risk, reward and uncertainty:** Organisations, their senior leadership teams and legal advisors need to make a series of judgements on whether and how to proceed with projects based on the risks and benefits they perceive. These teams have to navigate legal risks, public perceptions, and uncertainty around how the data will be used by external stakeholders (such as researchers). The key drivers, as well as solutions, in this domain are predominantly behavioural.

In the summary table below we give an overview of key barriers in each domain and recommended solutions. We highlight ideas that could be actioned immediately in *italics*. In the last section of the report, we describe potential data linkage projects that ADR UK could develop in the future. These include but are not limited to:

- **Child maltreatment:** Developing a coordinated approach for sharing and working with aggregated child maltreatment data in England to understand the landscape of risk and better inform policies to respond.
- **Violence prevention:** Creating a centralised repository for linking up crime and health data (such as A&E admissions) to make data sharing easier and enable better identification of violence hotspots.

- **Protecting vulnerable consumers:** Creating national datasets to understand which consumers (or groups of consumers) are considered vulnerable across multiple sectors (and the respective regulators).

**Table 1: Overview of main barrier domains and solutions**

Barrier	Solutions
<p><b>Resources:</b> Lack of resource to devote to project and stakeholder management</p>	<ul style="list-style-type: none"> <li>● <i>Developing a ‘data protection’ starter pack that includes project management templates (including for project ‘pre-mortems’), appropriate legal gateways for data linkage, open-source software solutions and terms of reference for specialist skills.</i></li> <li>● Appoint a business partner to help researchers &amp; departments navigate linkage and access processes, and link them up with other help services. This should be done by ADR UK, as DCMS’ focus is only on legal gateways not the holistic process of data linkage. This system is being internally used within the Department for International Trade (DIT) currently, with their “Data, Development and Innovation” team acting as guides through linkage and access processes for other teams within DIT.</li> </ul>
<p><b>Resources:</b> IT and human Infrastructure to enable data quality to make linkage worthwhile and conduct linkages</p>	<ul style="list-style-type: none"> <li>● Create data engineering fellowships for data collection, maintenance and linkage as in HMCTS. These could be funded by ADR UK or more broadly by UKRI, along with host HMG departments. In practice these could take the form of part of PhD funding (in a <a href="#">CASE model</a>), post-doctoral fellowships, or mid-career fellowships.<sup>1</sup></li> <li>● Provide advice on data collection.</li> <li>● Create a software library or discounted access arrangements for needed software.</li> <li>● Set up an advisory panel of experts, similar to the Trials Advice Panel. As with TAP, this would need to be linked with the Cabinet Office to signal the level of buy-in and importance to less engaged departments.</li> <li>● Create a cross-government data catalogue (including unlinked datasets) to allow for a better idea of what could potentially be linked.</li> </ul>
<p><b>Resources:</b> The research application process for linking data and accessing linked data</p>	<ul style="list-style-type: none"> <li>● <i>Change the ESRC application process for projects involving linked data to include how researchers will use and publicise (including depositing) the data.</i> Our understanding is that standard ESRC conditions already include depositing new data holdings created as a result of grant funding into the UK Data Service. However, some well-funded studies have not deposited data.</li> </ul>

<sup>1</sup> Noting that to allow greater flexibility these would need to be financially viable for non-HEI.

	<ul style="list-style-type: none"> <li>● <i>Increase transparency on what is available by:</i> <ul style="list-style-type: none"> <li>○ <i>Making the Secure Research Service (SRS) catalogue easier to understand and including information in the catalogue on whether and how different datasets can be linked within it.</i> This might include e.g. what the IDs are in each dataset, the work required to link datasets together and the match percentages between datasets. So if someone wants to link variables together (or these are suggested to them) then it makes clearer what data linking is needed.</li> <li>○ <i>Devising sector-level guides on how to use and link different datasets in the SRS and where outside data may be needed for specific cases (e.g. some HMRC data). One example of where this is happening already is DIT who are writing a guide for trade data.</i></li> <li>○ <i>Similarly, sector-level guides could be devised for partner data as well. For instance, there are also several pieces of work being funded in relation to having better linked data on children. A guide could be created bringing all of these together and made available on all partner websites, detailing the data resources available on understanding children's lives across the UK.</i></li> <li>○ Including a clear tracker of when current linkages are likely to be completed, and updating this on a regular basis (monthly or bi-monthly).</li> </ul> </li> <li>● Help research organisations both to conduct linkages and navigate application processes for linked data. This could be addressed by future open calls.</li> </ul>
<p><b>Judgements:</b> Benefits of linkage are not salient or 'top of mind'</p>	<ul style="list-style-type: none"> <li>● Create stronger incentives by creating a challenge prize for linkage, or advocating to HMT to incentivise linkage as part of spending review. The latter could include holding back percentages of project budgets until data is shared and deposited.</li> <li>● Create rapid prototypes using dummy or synthetic data of what could be done with linked data, and show real world policy changes. This could include ethnographic research with potential beneficiaries captured in video format to bring the benefits to life for government leaders. These should be deliberately not long-form academic papers, but short policy-relevant examples. These could be done by ADR UK funding academics and included in training for students and government researchers.</li> <li>● Hive off data linkage teams into special project teams, to shield them from shifting priorities (e.g. MoJ Data First).</li> <li>● Having central ambition/oversight of data linkage across government so that HMG Depts understand there is a</li> </ul>

	<p>purpose/strategy they are feeding into.</p> <ul style="list-style-type: none"> <li>• ADR UK could ask to attend cross-government meetings such as Heads of Profession for Statistics meetings and Director of Analysis meetings.</li> <li>• ADR UK seeks a senior policy ‘champion’ who can engage with other senior policy colleagues.</li> <li>• ADR UK could also work up a package of support for HMT to offer to departments to help them conduct data linkages.</li> <li>• <i>Involve data owners who may not directly benefit from a linkage (e.g. local councils) in decisions on what to link and what questions to answer using the data.</i></li> </ul>
<p><b>Judgements:</b> Legal risks and uncertainty</p>	<ul style="list-style-type: none"> <li>• <i>Developing voluntary standards for information governance, including shared DPIAs, consent forms and standard terms for data access that include deposit.</i> This could be incorporated into the ‘data protection starter pack’ mentioned above.</li> <li>• <i>Link up organisations that provide advice on legal gateways (including Public Sector Linkage team in DCMS) and on data protection (including ICO).</i></li> </ul>
<p><b>Judgements:</b> Public perception and trust</p>	<ul style="list-style-type: none"> <li>• Reframe narrative to highlight the costs of <i>not</i> doing linkage, especially in COVID-19, and test messages.</li> <li>• Making salient the extent to which data linkage is already happening - but framing this positively to minimise the risk that plain facts about data linkage feed into narratives about ‘surveillance’.</li> <li>• Conduct research to understand in greater detail the contexts and conditions under which people feel comfortable with their data being shared. This could inform the development of clearer principles around data sharing and reduce perceptions of risk in government.</li> </ul>
<p><b>Judgements:</b> Lack of trust between organisations and desire for control of process.</p>	<ul style="list-style-type: none"> <li>• Opportunities for project teams to co-locate or facilitate online collaboration through shared Slack channel. This would be short-term in nature to establish working practices and relationships but could then resolve to their own departments (i.e. going beyond transactional meetings and fostering a sense of joint effort).</li> <li>• Create a community of practice around data linkage.</li> <li>• Promote norms of operational transparency.</li> </ul>

## 2. Introduction

---

Data linkage and better use of administrative data are core to advancing research and policy in the UK. ADR UK (Administrative Data Research UK) is funding several initiatives and working hard to ensure that more administrative data is available and linked up to answer fundamental questions about our social and economic lives. A lack of linked data not only hampers the evaluation of existing policies, but compromises the targeting of policy and development of theories of change in intervention design, as well as limiting the ability to generate hypotheses or answer basic questions about the UK population.

A core challenge to this endeavour is convincing data controllers to share and link data - with significant amounts of population-level data held by UK government departments core to such efforts. It is notable that the Cabinet Office has contributed to a national geospatial data strategy,<sup>2</sup> but not a broader administrative data strategy beyond that set out in a 2017 policy paper and the DCMS National Data Strategy.<sup>3</sup>

This report sets out BIT's engagement work with Whitehall, non-ministerial departments and other organisations. It has a particular focus on the behavioural barriers to data linkage and how these might be overcome. We also detail technical barriers and proposed solutions to these, as well as ideas for future data linkages and possible use cases for ADR UK to pursue.

We have undertaken the following activities to inform this report (further detail in Annex 1):

- Arranged and conducted interviews with staff from three government departments, three non-ministerial government departments, one executive non-departmental public body, one local authority and two independent research organisations. We have also conducted less formal conversations off the record.
- Reviewed past ADR UK and external reports on data linkage, and linked these to the behavioural sciences literature.
- Reviewed documentation in relation to six case studies of successful data linkage.

---

2

<https://www.gov.uk/government/publications/unlocking-the-power-of-locationthe-uk-geospatial-strategy>

3

<https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy-better-use-of-data>; <https://www.gov.uk/guidance/national-data-strategy>. But for an example of departmental engagement with this see: [HM Courts & Tribunals Service \(2020\). Making the most of HMCTS data.](#)

**The overall message that came out of the work was a need to focus on the technical, organisational, behavioural and legal aspects of linkage rather than just producing analytical use cases.** Interviewees stressed that demonstrating the value of linkage was a necessary but not sufficient step to enable data linkage in government. There was also a general agreement that the current state of play is unacceptable. Accessing linked data should not take years or never occur at all, and there was support for enabling programmes like ADR UK. There is also a scarcity of resources available on examples of how other departments or research organisations have navigated these problems. The documentation for our case studies, for instance, overwhelmingly focussed on matching, data availability and sometimes legal issues, rather than organisational or software problems.

COVID-19 made directly approaching government departments more difficult, owing to the fact that many of BIT's contacts are directly involved with the response. Similarly, those at the top of government (Cabinet Office, PMIU) are almost exclusively focused on COVID-19, so trying to cut through was very difficult. That said, COVID-19 offers an opportunity to make the benefits of data-sharing much more salient to policy-makers as many are struggling with this themselves at the moment. Several interviewees noted that collaboration between departments on data issues including linkage has recently strongly increased. The UK research community has illustrated some of the barriers faced in a recent [open letter](#).

### Structure of this report

In the remainder of this report we set out our findings in five sections:

1. Detailing barriers to data linkage, sharing and deposit by HM government.
2. How to understand and solve barriers through behavioural science.
3. A sketch of how to solve the remaining technical and structural barriers.
4. Discussion of how to encourage the use and deposition of linked data by local councils, non-departmental public bodies and researchers.
5. Ideas for future linkages.

In Annex 1 we detail activities conducted to produce this report.

### The ADR UK Model

ADR UK exists to promote the availability of public sector data for policy-relevant research. It is made up of three national partnerships (ADR NI, ADR Scotland and ADR Wales) and the Office for National Statistics (ONS), coordinated by a UK-wide Strategic Hub.

For the purposes of this report, several elements of ADR UK's operating model are worth keeping in mind:

- ADR UK financially and operationally supports government departments and other public bodies to either share data to link themselves or be linked by the ONS (Office for National Statistics), for the purposes of creating de-identified datasets that can be used for research. Therefore, this report does not cover linkages for the purposes of providing public services to individuals, which usually requires personal data.
- The sharing of data between departments for data linkage is legally underpinned by Part 5, Chapter 5 of the *Digital Economy Act* which allows for sharing de-identified data for the purposes of research.
- The original data is often personal and sensitive, but must be de-identified before being made available to researchers.
- ADR UK has funded the expansion of the ONS' Secure Research Service (SRS), which is a system of facilities to enable safe access to data by accredited researchers (the service is summarised here: <https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme#introduction>).
- The other partnerships within ADR UK (ADR Wales, Scotland and Northern Ireland) also maintain secure research services. These are summarised here: <https://www.adruk.org/our-data/our-data-services/>

## 3. Barriers to data linkage, sharing and deposition

---

In this section, we summarise the key barriers found in our work (interviews, desk review) to data linkage, sharing and depositing data by the UK government. We categorise barriers into two main groups: 1) *resource barriers*, particularly relating to prioritisation and 2) judgement of *risk, reward and uncertainty* around data sharing.

### 3.1 Resource barriers relate to both availability and prioritisation by senior staff

The first set of barriers relates to the resources required from data controllers to participate in data sharing and linking projects. In summary, participating in data linking projects is resource-intensive in terms of IT capacity, project management and staff time.

Based on the discussions and review we have conducted, data sharing and linking projects face considerable obstacles in keeping 'top of mind' of government departments. These projects are often not part of statutory duties or high on the political agenda. This is compounded by the need for collaboration: if one stakeholder (or staff within one stakeholder) have issues with prioritisation, the entire project may stall. This combination of high resource intensity, low prioritisation and the need for collaboration results in these projects often not being considered. And when they do start they take much longer than anticipated and sometimes are not completed.

Five barriers exist under this heading:

- Resources and capacity to do effective project management,
- Resources for stakeholder management,
- Resources to enable sufficient data quality and management,
- IT infrastructure and skills,
- Having staff with sufficient knowledge of data linkage and the skills to execute or procure such projects.

#### Project Management

While there are many unique components to data linkage projects, they have a significant amount in common with any costly government endeavour. They require

careful planning, budgeting and management to complete but there can be significant overruns if this does not occur.<sup>4</sup>

At the *planning and budgeting stage*, budgets may be overly optimistic or not provide for any unforeseen roadblocks or delays. Plans may be overconfident in their timescales or views of departmental capacity to undertake particular analytical tasks such as matching. Further, as there is coordination required across departments there may not be sufficient plans in place to set up sustained communication about the linkage efforts.

At the *execution stage*, there can be a shortage of talented project managers and potential for the divided responsibilities between departments to hamper progress.

Linking projects often set up project steering groups to get all stakeholders together to make decisions. This makes project management more time consuming, as you need to ensure you have the right stakeholders around the table, to set meetings that work with everyone's diary, and to communicate with everyone about the project's progress.<sup>5</sup> We also know from other types of collaborative decision-making in different contexts (such as risk management) that the right level of representation is needed in order to progress projects.

Data linking projects rely on multiple departments moving through processes such as information governance, data cleaning and verification at the same pace. This interdependence means that if one stakeholder is held up, the whole project can be stalled.

For example, information governance processes can be very resource intensive.<sup>6</sup> Every dataset's data controller will have a separate information governance process, as there are no standard processes across organisations (up to each organisations' legal judgement). The more datasets that are involved in the linkage, the more complex this becomes to manage.<sup>7</sup>

## Stakeholder Management

Related to project management, interviewees commented that successful linkages involved significant work in managing stakeholders:

- For instance part of the reason Ofsted obtained regular access to linked data from DfE was not just that they had both a service level agreement but that their Chief Statistician has regular catch-ups with the relevant head of profession in DfE.

---

<sup>4</sup> See BIT's earlier report on applying BI to project management:

[www.gov.uk/government/publications/using-behavioural-insights-to-improve-project-management](http://www.gov.uk/government/publications/using-behavioural-insights-to-improve-project-management)

<sup>5</sup> [Office for Statistics Regulation \(2018\). Joining Up Data for Better Statistics.](#)

<sup>6</sup> Ibid.

<sup>7</sup> Ibid.

- It is also necessary to manage relationships within one organisation, especially with senior stakeholders. However, as their attention is scarce this is also very difficult to achieve. Outreach could focus on more than just demonstrating value and also consider incentive structures, for example attaching the creation and deposition of linked data to project funding from HMT.
- Relationships also need to be managed with internal information governance teams, with one interviewee remarking that their data linkage project had likely succeeded in part because they sat on the same floor as their Information Governance (IG) team and therefore had established a level of familiarity.

Barriers that existed to relationship management were:

- It is very time-consuming as it requires and relies on maintenance of informal social networks and establishing trust between and across organisations.
- Other demands on staff time (both in the project team and steering group) can slow projects as organisations prioritise statutory duties or political priorities over data linkage projects, which rarely rank highly on the organisation's agenda.<sup>8 9</sup>
- High levels of turnover (both in the project team and steering group) can result in the new person needing to be re-convinced of the value of the project.<sup>10</sup>

## Data Quality and Management

Ensuring that the underlying data needed for linking is of good quality is important for the quality of subsequent linking and analysis. If data is missing, inconsistent, or not uploaded (i.e. on paper somewhere), then it increases the difficulty of and resource requirements for data linking.<sup>11 12</sup>

Preparing data for linking is resource-intensive and often takes longer than expected, involving data cleaning, verification, and anonymisation. If this needs to occur across multiple data controllers it increases the complexity and resource intensity of the project.

---

<sup>8</sup> Administrative Data Research Network (2016). Mid-Term Review Report. Research Project Management.

<sup>9</sup> [Local Government Association & East Sussex County Council \(2018\). \*Matching data to support troubled families. Digital Transformation Programme Case Study.\*](#)

<sup>10</sup> Ibid.

<sup>11</sup> Gilbert, R., Lafferty, R., Hagger-Johnson, G., Harron, K., Zhang, L.-C., Smith, P., Dibben, C., & Goldstein, H. (2018). GUILD: GUIDance for Information about Linking Data sets. *Journal of Public Health, 40*(1), 191–198.

<sup>12</sup> Green, E., Ritchie, F., Mytton, J., Webber, D. J., Deave, T., Montgomery, A., Woolfrey, L., & Chowdhury, S. (2015). *Enabling data linkage to maximise the value of public health research data* (Public Health Research Data Forum). Wellcome Trust.

There are also substantial demands created by depositing linked data, e.g. anonymising the data and creating sufficient data documentation. Further, departments may need to strengthen their processes for accessing data and are at different stages of creating application processes. DfE, for instance, has a highly structured process.

## IT Infrastructure and Skills

Getting the **required software and hardware** is not straightforward, both internally within the data controller for storage, cleaning and verification, and for ensuing compatibility for linking and data processing:

- Any software needs to be able to run on existing IT systems and hardware. Local authorities and some government departments may have issues with legacy hardware and software.<sup>13</sup>
- Some software storage or coding solutions keep data in a way that means data is not compatible with other datasets for linking or sharing (often proprietary software).<sup>14</sup>
- Some software companies offer teaser rates to get through the door of procurement offices; when they subsequently increase prices, data projects stall because the software has become unaffordable (and teams have to re-procure software and start again).<sup>15</sup>

Data linkage requires specialist skills. For example the lack of unique identifiers across public sector datasets means that data linkage requires programming (to perform probabilistic matching).<sup>16</sup> This means departments either need to build in-house expertise or outsource data linkage work.

When specialist matching software is purchased, it may not even be used. For instance, East Sussex Council ended up manually matching individuals for the Troubled Families evaluation. In an interview, staff from the council said they had previously bought matching software but they found it was not sufficiently accurate for their purposes.

## Knowledge

There are knowledge prerequisites to setting up data linking projects:

- Knowing what datasets exist, especially in other government agencies or departments given the lack of a cross-government data catalogue;

---

<sup>13</sup> Based on previous BIT experience on data projects with local authorities.

<sup>14</sup> [Office for Statistics Regulation \(2018\). \*Joining Up Data for Better Statistics\*.](#)

<sup>15</sup> [Local Government Association & East Sussex County Council \(2018\). \*Matching data to support troubled families. Digital Transformation Programme Case Study\*.](#)

<sup>16</sup> Gilbert, R., Lafferty, R., Hagger-Johnson, G., Harron, K., Zhang, L.-C., Smith, P., Dibben, C., & Goldstein, H. (2018). GUILD: GUIDance for Information about Linking Data sets. *Journal of Public Health*, 40(1), 191–198.

- Knowing what datasets are suitable for linking for example which other datasets contain relevant IDs; and
- Knowing what questions to ask of linked datasets, or which could be asked, especially for policy teams rather than analysts.

### 3.2 Judgments: risks, rewards and uncertainty

The second set of barriers that have surfaced from our work relate to the perception of risks and rewards involved in data sharing and linking.

For data sharing and linking projects, organisations, their senior leadership teams and legal advisors need to make a series of judgements on how to proceed based on the risks they perceive in the project and their own appetite for risk. These teams have to navigate legal risks, changing public perceptions, and uncertainty around how the data will be used by external stakeholders (such as researchers).

The default for public sector organisations appears to be *not* to share or link data, and thus if organisations do nothing, data remains unshared and unlinked. A recent meta-analysis<sup>17</sup> found that defaults are more likely to affect behaviour where they reflect the status quo. Defaults work because they are typically easier to follow and can implicitly convey the socially desirable behaviour. Moreover, it is (often) the legal default that data sharing should not be allowed, unless an enabling legal gateway exists.

More importantly, however, decision makers face an uncertain and skewed risk-reward balance. Potential benefits of data linkage may not materialise until several months or years down the line and may feel unclear or intangible. Moreover, the benefits may also accrue primarily to another body (for example, if crime data is needed for an evaluation of an educational programme). On the other hand, the consequences of data linkage gone awry can be severe. Such circumstances mean that organisations will be very cautious about the data linkage projects they engage in. The immediacy of risks combined with possibly unclear time-horizons for rewards (or no tangible benefit) tip 'balancing' scales towards risk being much more salient.

The combination of poorly understood but real risks, along with problematic defaults and a lack of correctly aligned incentives may prove strong deterrents to involvement with linkage efforts.

We have identified four major barriers under this heading:

---

<sup>17</sup> Jachimowicz, J., Duncan, S., Weber, E., & Johnson, E. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), 159-186.

1. Legal risks and uncertainty,
2. Predicting public perception,
3. Lack of control over data processors,
4. Rewards may not accrue to those that need to contribute or may not be realised on short enough timescales.

Few if any of these issues may be novel to ADR UK given its existing efforts, but do bear repeating, especially with a focus on decision-making.

### Legal risks and uncertainty

Departments and local authorities vary in levels of awareness of the legal difference between data protection law (which requires a legal basis to *process* data) and requirements around data sharing between government bodies (which require *legal gateways*). We will not retread ground covered elsewhere on the GDPR, but focus on the legal gateway problem as this is a more unique problem to data linkage across departments.

There is no blanket 'legal gateway' for data sharing and linkage that fits all situations. The *Digital Economy Act 2017* contains enabling powers for data sharing for research purposes but explicitly excludes health and social care data. There was also a perception that de-identification was required prior to sharing but that made sharing and linkage more complicated (see footnote on this point).<sup>18</sup> One of the interviewees remarked that departments and local authorities often default to legal gateways they are familiar with, even if that restricts the processing that they can conduct - which also suggests a need for greater clarity and guidance on that point.

There also may be confusion about the distinction between data sharing for *operational* reasons (usually with personalised data) and for *research* reasons (de-identified data if using the *Digital Economy Act* research gateway or any DEA-accredited data processor<sup>19</sup>).

Ultimately, each case is a judgement call and there are many uncertainties and

---

<sup>18</sup> It is worth noting the two-stage feature of the DEA. The first stage is processing to prepare data for disclosure. There is no suggestion in the DEA that data must be de-identified at this preparatory stage. The legislation expressly says the data will be personal data at this preparation stage, which is how linking/matching can be completed. The data can, and should, be identified data at this preliminary stage. Once linking/matching has been done, the second stage kicks in. The data are now made reasonably unlikely to identify an individual. Once that has been done, and the personal IDs or specific addresses have been removed, then the data are now sufficiently prepared for disclosure to a researcher.

<sup>19</sup>

<https://www.statisticsauthority.gov.uk/about-the-authority/better-useofdata-statistics-and-research/betterdataaccess-research/better-use-of-data/list-of-accredited-processors-under-the-research-strand-of-the-digital-economy-act>

risk-based judgements that need to be made<sup>20</sup> when sharing data with other government departments and subsequently preparing de-identified, linked data for deposit. The primary difficulties *perceived by interviewees* were:

- The legal case for sharing varies on a case-by-case basis.
- If sharing falls under the research powers of the *Digital Economy Act*, it needs to be in the 'public interest' which is open to interpretation.<sup>21</sup>
- Data needs to be reviewed to see if it has been truly de-identified prior to use by researchers (or deposit for use). There is, at the time of writing, scant legal precedent about what constitutes truly de-identified data.<sup>22</sup>
- Before data are made available to researchers, it often needs to be shared with other government departments with some personal identifiable information intact for matching individuals.<sup>23</sup>
- When depositing or linking data, information given to participants may need to be reviewed to see if consent was granted to use data for the purposes required by the project (could include both historical and current consent forms in some cases).<sup>24</sup> If not, then remedial action may be needed to inform participants of a change in legal basis for their data being processed.
- There needs to be a specific purpose for the sharing and linkage, which may impede exploratory projects if they are not crafted/presented in ways that make clear the purpose is exploratory.

Beyond these points, laws and regulations on data sharing have changed substantially in the last few years. Teams might assume that the legal context will continue to shift in future, which increases uncertainty surrounding data sharing and linking projects. Certainly the introduction of GDPR, even with a two-year implementation window, may have impacted on the ability of researchers to collect and process data.<sup>25</sup>

---

<sup>20</sup> [Office for Statistics Regulation \(2018\). \*Joining Up Data for Better Statistics\*.](#)

<sup>21</sup> The language of 'public interest' might arise from [GDPR](#). If the sharing makes use of the Digital Economy Act, the research must serve the *public good*. There is a [Code of Practice](#) which elaborates the meaning of Public Good in this context, and all decisions of the [Research Accreditation Panel](#) are open to examination to see examples of what has been found, independently, to be serving the public good. However, some departments may not be sufficiently aware of either the Code or the availability of the precedent-setting decisions made within its scope, and advice and assistance is still needed.

<sup>22</sup> 'Deposit' might relate to the data owning department depositing data with ONS (for example), or ONS depositing the data in the SRS for access by the researchers. The first can contain identifiers; the second needs to be de-identified.

<sup>23</sup> This bullet suggests that the sharing of linkable data has to be between the data owners first. That is an option, but again the point of the DEA is to make that unnecessary. The two or more data owners can send the data to ONS SRS who do the matching for them.

<sup>24</sup> [Local Government Association & East Sussex County Council \(2018\). \*Matching data to support troubled families. Digital Transformation Programme Case Study\*.](#)

<sup>25</sup> A salient example being the use of [biobank data](#), but see also [Health Ethics and Policy Lab \(2019\)](#).

While government departments and agencies do have access to legal advice, this still requires asking the right questions of lawyers and realising that linkages are legal for research in the first place. Analysts may be aware of these issues, but greater outreach and training could be provided for policy staff.

As an interviewee who worked at a non-departmental body said “the law is pretty enabling, but there is a natural caution not to do ‘the wrong thing’ with data”. This caution significantly slows down the ability to share data. For example, they had worked with an external organisation to develop a new risk model. There was then legal wrangling with that organisation on who could access data and data sharing. This required considerable tenacity on the part of the research team within the non-departmental body to resolve.

Relatedly, ethics and consent processes are not evenly developed across government. In the English NHS (noting that NHS data are routinely linked to support research in Wales), there are many steps in [ethics processes](#) which need to be navigated to enable linkage.

### Predicting public perception

A recent literature review by ADR UK shows that the public are generally supportive of data linkage as long as (1) the research is clearly in the public interest, (2) there are safeguards in place to protect privacy and prevent data theft, and (3) there is transparency about what the data is used for and trust in the relevant public bodies.<sup>26</sup> Trust is a particularly complex factor as it is impacted by the extent to which the other elements are in place. For example, trust can be increased through transparency - but only if the public also has faith that the data is treated responsibly, safely and in a way that produces important public benefits. News reports of data breaches, whether driven by intentional or unintentional acts, also serve to undermine trust. While the default approach now seems to be transparency about breaches and follow-ups with those affected, it's not clear whether these actions bolster or undermine wider public trust, or ameliorate the impact of breaches in the first place.<sup>27</sup>

As such, there is no blanket support for data sharing and public servants are often

---

<sup>26</sup> [ADR UK \(2020\). Trust, Security and Public Interest: Striking the Balance - A review of previous literature on public attitudes towards the sharing and linking of administrative data for research.](#)

<sup>27</sup> Note that this is something that can and should be evaluated using observational data or experimentally to understand the impact of different approaches to reassurance, messaging and so on (see Section 4). One observation is that whenever private companies report data breaches they make very clear that financial data (e.g. bank account details) are not part of the breach, which may reassure many of those affected. However, it may also be that segments of the general public do not understand the implications of seemingly innocuous data being made public, particularly if then combined with other public data sources.

concerned about the need for public support and how to gain it. It is challenging to effectively communicate benefits and safeguards to the public. The public may have a hard time distinguishing between different types of end uses, which can make garnering public support more difficult.<sup>28</sup> In addition, it is hard to communicate what different data safeguarding measures mean (such as the difference between depersonalised data where directly identifying information is included and de-identified data where there is no or very low reidentification risk) or how emerging Privacy Enhancing Technologies work in practice.<sup>29</sup> This adds a layer of complexity to convincing the public that data sharing and linking can be done in a way that preserves privacy.<sup>30</sup> Because of these complexities public bodies do not always communicate as transparently about the data linkage work as they might otherwise - and this in turn can negatively impact on the public's trust when stories about data theft or misuse appear in the media.

To overcome some of these challenges a recent report by the Centre for Data Ethics and Innovation (CDEI) proposes the development of a framework for trustworthy data sharing. This would include gaining greater clarity around how the public interest is defined and judged as well as the conditions under which it is appropriate to share data in the public interest, whether with or without user control.<sup>31</sup> Importantly, the CDEI also advocates more proactive engagement with the public and transparency about individual data linkage projects. Similarly, based on its aforementioned literature review, ADR UK has decided to move beyond projects aimed at understanding the general acceptability of data sharing towards more granular research with the public to understand how, why and when data sharing for particular use cases should be done. Behavioural science and experimental research can play an important role in developing effective data sharing policies and strategies for communicating these to the public. We describe initial ideas in Section 4.

A related challenge is that it is difficult for the data controller to know if the data processing entities will use the data appropriately.<sup>32</sup> In economics, this is referred to as the 'principal-agent problem'. The 'agent' (here the data processor) is acting on behalf of the 'principal' (here the data controller), but the agent has their own interests at heart, which may not always align with the principal.

There are at least three ways that this manifests in ADR UK's model:

---

<sup>28</sup> [Harper, H., Egan, M., Cornel, P., Ter Meer, J., & Wittels, A. \(2018, October 10\). Data sharing and the importance of choice architecture in healthcare: New results. \*Behavioural Insights Team Blog\*.](#)

<sup>29</sup> [The Royal Society \(2019\). Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis.](#)

<sup>30</sup> [Office for Statistics Regulation \(2018\). Joining Up Data for Better Statistics.](#)

<sup>31</sup> [Centre for Data Ethics and Innovation \(2020\). Addressing trust in public sector data use.](#)

<sup>32</sup> Green, E., Ritchie, F., Mytton, J., Webber, D. J., Deave, T., Montgomery, A., Woolfrey, L., & Chowdhury, S. (2015). *Enabling data linkage to maximise the value of public health research data* (Public Health Research Data Forum). Wellcome Trust.; *Joining Up Data for Better Statistics*. (2018).

1. When government departments need to share data with each other or with third-parties like research organisations, these organisations are ‘agents’. The concern around control here is likely about data protection, and the legal liability that data controllers retain for the actions of processors.
2. When data is eventually de-identified, departments still lose control over the messages created using that data. For example, it could be that the data reveals something politically unpopular or embarrassing.
3. Finally, while the ‘5 safes’ concept in the ADR UK operating model is designed to mitigate some of these risks, awareness of this in the interviews was patchy with the consequence that there may be perceived principal-agent problems.

### **Rewards may accrue to other public bodies or may not be strong enough**

The benefits of data linkage, especially those created for research purposes, may not accrue to the bodies that need to submit data. As an example mentioned in interviews, local councils often need to collect and contribute data for national evaluations and other data linkage projects, but do not always receive targeted advice or outputs at a local level. Similarly, if linked datasets are primarily used for academic research, departments may not view them as relevant to their policy priorities.

Even if benefits do accrue to those public bodies, they are likely to do so on slow timescales, since it takes a long time to link data and then to conduct research after the linkage is complete. This may be a significant problem in persuading civil servants and politicians to support linkage efforts as they will usually be working more reactively on shorter timescales. Overall, decision makers may simply perceive potential benefits as intangible and not strong enough.

## 4. Behavioural problems and solutions

---

In this section, we outline the behavioural science concepts which link these different barriers and identify potential solutions. We recognise however that not all issues are behavioural in nature, and hence also address structural solutions in Section 5.

### 4.1 Resource Barriers: Intensity and Prioritisation

The first category of behavioural barriers relate to the prioritisation of the resources needed to do the work and how behaviours of the civil servants involved can lead to a lack of sufficient resources to deliver the project successfully. The barriers in this section are:

- Barriers relating to planning: notably failure to forecast risk, planning overly optimistically, discounting the future and the planning fallacy.
- Small 'frictions' which cause barriers to starting and completing projects.
- Allocation of attention to linkage projects.

#### Planning failures

Below we detail behavioural issues that particularly relate to planning of data linkage projects.

##### Problem

There are two main concepts that specifically relate to failures of planning and project management - overconfidence and optimism bias. Some concepts that are discussed in the following subsections are likely to be relevant to planning failures as well, but we focus on these two concepts here.

##### *Overconfidence*

A number of studies have shown that individuals have a tendency to overestimate their abilities, plans and likelihood of success.<sup>33</sup> This overconfidence bias has been observed in many contexts including among those deemed experts in their fields and officials within government settings. One study involving officials from a range of climate change agencies found that the more experienced an individual was, the more overconfident they became in their own abilities. As overconfidence increased,

---

<sup>33</sup> E.g. Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134.

so did the likelihood of taking riskier decisions and choosing not to have risk-mitigation plans.<sup>34</sup>

### *Optimism bias*

People tend to be overly optimistic about the time needed to complete a task. The term planning fallacy is used when this optimism prevails despite an individual having knowledge and experience of previous delays in similar tasks. This individual is likely to continue to underestimate the time, costs and potential risks of the task they are planning for.

This phenomenon, first described by Kahneman and Tversky, has been used to explain a number of large scale public project delays or failures.<sup>35</sup> Optimism bias occurs because people take an 'inside' view of the task at hand and focus on its specific qualities, rather than taking a broader 'outside' view and incorporating experiences from other, similar projects when making plans.

Data linkage projects are similar to other large scale and/or complex projects in that they require accurate budgeting and planning, as well as appropriate resources and management to see them through to completion. Team members with experiences in their own department may put together project plans which are overly optimistic and underestimate potential risks (for example by not taking into account extra challenges from working collaboratively with other departments). In cases where projects struggle to gain support due to the resource needed there may also be a desire to have an optimistic budget in order to increase the project's appeal.

Within data linkage projects, there may also be multiple departments needing to work together, which often have different processes and expected timelines for certain tasks. Moreover, a planner overseeing a project may create a timeline which does not take into account past experiences of different departmental projects (or project ideas that failed to be realised) and associated roadblocks.

In addition to distorting the planning of time and resource, optimism bias also can also affect forecasting of risks to the project. Effective risk forecasting is a key component of successful project planning as it allows for better preparedness when roadblocks do occur. Individuals are biased to believe they are somewhat immune to negative events or things going wrong, which makes accurately forecasting risk to their projects difficult at times.

---

<sup>34</sup> Liu, X., Stoutenborough, J., & Vedlitz, A. (2017). Bureaucratic expertise, overconfidence, and policy choice. *Governance*, 30(4), 705–72.

<sup>35</sup> Kahneman, D., & Tversky, A. (1977). *Intuitive prediction: Biases and corrective procedures*. Decisions and Designs Inc Mclean Va.

Optimism bias in the original consideration of risk can also affect behaviour down the line when executing the project:

- First, a team creates an unrealistically optimistic forecast.
- Second, when executing the project the team may face potential ‘losses’ from an unexpected issue that has arisen, because the time taken or results delivered differ from their (unrealistic) expectations.
- Third, as a result teams might make decisions on the project which are not rational, such as continuing the project due to a strong aversion to losses or taking sensible but risky courses of action to avert further delays.

This is caused by the interaction between optimism bias and a concept which we will discuss later called loss aversion which refers to the human tendency to feel losses more strongly than gains of a similar size.<sup>36 37</sup>

### Solutions<sup>38</sup>

We discussed previously how the biases that people have around their own abilities, specifically in relation to overconfidence, can lead to issues arising when planning and managing a project. We will now outline processes that can be integrated into the project management cycle to help minimise the impact of these biases.

The main solutions that we propose going forward are to:

1. Train relevant staff to better forecast project risks through reference class forecasting,
2. Keep two estimates of risks and resourcing,
3. Building in opportunities to change course and revisit assumptions,
4. Map the sequencing of the project using swimlane diagrams and,
5. Conducting project pre-mortems.

We think that these solutions could be bundled together into a data linking project starter pack that includes project management templates corresponding to these five solutions as well as data protection templates, example consent forms if needed, open-source software solutions and terms of reference for specialist skills (to be discussed later).

In cases where ADR UK is not funding projects directly, it could:

---

<sup>36</sup> Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292.

<sup>37</sup> Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., ... & Gibson, S. P. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 1-12.

<sup>38</sup> These solutions largely come from a prior piece of work called ‘Behavioural Government’, which sought to apply insights from psychology and behavioural economics to government decision making itself rather than the behaviour of the general public. [Hallsworth, M., Egan, M., Rutter, J., & McCrae, J. \(2018\). Behavioural Government: Using behavioural science to improve how governments make decisions. The Behavioural Insights Team.](#)

1. Discuss with departments where they need support, and set out that ADR UK or related organisations such as DCMS are able to lend support on data protection, skills, project management and legal gateways.
2. When discussing support requirements, ADR UK could discuss the project plans, which might encourage greater robustness.
3. When facilitating linkages across departments, it could facilitate meetings where the different departments present and discuss their plans and forecast risks together.

### *Reference class forecasting*

Reference class forecasting can be used as a tool to ensure that a wider range of experience is taken into account when planning a project; it helps to move the team planning the project from an inside to an outside view. By using empirical evidence from previous projects, the overoptimism that normally creeps in over timescales and budget can be corrected early on.

The UK government's green book sets out the process for calculating these optimism bias estimates for infrastructure projects.<sup>39</sup> The process involves selecting the category which your project falls into and observing the relevant adjustment percentages suggested. There has been further work done by the Institute for Government on applying these principles to more social based programmes and projects.<sup>40</sup> Furthermore, these corrections can be applied in a simpler way by using a standardised optimism correction. A figure referenced previously is to add around 25-30% to all current projection of costs or timescales.

### *Keeping two estimates*

Another technique used to improve judgement is to have two estimates when predicting the outcome of a project in relation to its timescale and budget.<sup>41</sup> The project team would generate both a central estimate and also incorporate a more pessimistic estimate to their plans. This would read as a high-cost, low-impact estimate, similar to a 'worst case scenario' outcome. Having this secondary estimate should help teams reduce reliance on their central estimate and to correct any overoptimism present.

Having a more pessimistic secondary estimate would also allow an opportunity to revisit the project as a whole and decide whether to keep proceeding, helping to

---

<sup>39</sup> [HM Treasury. \(2018\). The green book: Central government guidance on appraisal and evaluation.](#)

<sup>40</sup> [Atkins, G., Davies, N., & Bishop, T. K. \(2017\). How to value infrastructure: Improving cost benefit analysis. Institute for Government.](#)

<sup>41</sup> Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, 20(2), 231–237.

mitigate against the 'sunk cost fallacy'.<sup>42 43</sup> Once the secondary estimate has been reached, it could automatically trigger a discussion over next steps and whether the project justifies continuation.

### *Breakpoints*

An issue that often arises in large scale projects is a fear to do a U-turn or call off a project when so much resource has already been invested into it. Confirmation bias plays a role here, whereby individuals will look for and interpret information to support the views they already hold.<sup>44 45</sup> A way to address the impact of confirmation bias is to add in formalised 'breakpoints' at the project planning stage. These points are intended to be opportunities to discuss the current state of a project and if it is running as planned with a view to take direct action if not. This technique has been used in surgery to help reduce instances of medical errors.<sup>46</sup> It is hoped that incorporating breakpoints would lead to less of a backlash to any U-turn, as the whole project process is viewed in a more open way throughout.

An example of this in practise is known as 'collaborative red teaming', which is a technique taken from the military.<sup>47</sup> The premise is that once a project plan has been developed, groups within the team can split off and become challengers of specific parts of the plan, which may need further work or debate. There can be multiple teams all tasked with a different aspect to challenge so as to avoid individuals being viewed as outsiders which could lead to their concerns being dismissed.

### *Map the sequencing of the project using swimlane diagrams*

A good way to get teams to think through these issues might be to have them map the structure of the data sharing project - for example, whether the project is sequential or a 'hub and spoke' model where small component tasks are conducted and then fed back to the centre. This can help to understand the resilience of the work to potential disruptions<sup>48</sup> and how much redundant standby resource needs to be built into the project to move work forward if delayed. One useful method of doing

---

<sup>42</sup> Roth, S., Robbert, T. & Straus, L. On The Sunk-cost Effect In Economic Decision-making: A Meta-analytic Review. *Bus Res* 8, 99–138 (2015).

<sup>43</sup> McAfee, R.P., Mialon, H.M. And Mialon, S.H. (2010) Do Sunk Costs Matter? *Economic Inquiry*, 48: 323-336.

<sup>44</sup> Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many cities. *Review of General Psychology*, 2(2), 175–220.

<sup>45</sup> Jerit, J., & Barabas, J. (2012). Partisan perceptual bias and the information environment. *Journal of Politics*, 74(3), 672–684.

<sup>46</sup> Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. H. S., Dellinger, E. P., ... & Merry, A. F. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *New England Journal of Medicine*, 360(5), 491-499.

<sup>47</sup> [Ministry of Defence Development. Concepts and Doctrine Centre. \(2013\). Red teaming guide \(2nd ed.\).](#)

<sup>48</sup> [Baran, P. \(1964\). On distributed communications networks. \*IEEE transactions on Communications Systems\*. 12\(1\), 1-9.](#)

this is a swimlane diagram which distinguishes job sharing and responsibilities for subprocesses in a larger project.<sup>49</sup>

### *Conduct project pre-mortems*

They could also engage in 'pre-mortems' (a technique originating from Gary Klein<sup>50</sup> and evaluated in diverse settings<sup>51</sup>) where teams do scenario planning as if the project has already failed, and try to find plausible reasons why. This can be useful to ensure plans are not overconfident and forecastable problems are avoided. It is worth noting that this differs from a typical risk analysis because the framing is not 'what could go wrong?' but rather 'assume the project has failed, why might that be?'. Klein, Sonkin and Johnson<sup>52</sup> identify five conditions for a successful 'pre-mortem':

1. Reframe the problem as allowing the group to identify weaknesses and hidden assumptions in existing plans. However, it needs to be clear that this is a scenario for planning purposes, not real expectations for the future.
2. A cognitively diverse group of people in the meeting representative of all the people who will be involved in the process.
3. An environment of psychological safety, where individuals feel free to talk openly about potential problems.
4. A sense of group equality within the meeting, where all contributions are valued equally. It is also key that all group members participate.
5. A sense of urgency.

## Friction costs

### Problem

The more friction costs (small inconveniences or steps) are involved in a task, the less likely it is that the task will be accomplished. Even small, seemingly minor frictions can have a disproportionate impact on our behaviour. For example in a trial we conducted with HMRC, taking one additional click out of a process had a large effect on form completion.<sup>53</sup>

There are multiple frictions in setting up data sharing and linking projects: separate information governance processes, staff turnover, data quality issues, and IT inter-dependencies. Making it easier to navigate these hassles in the process can help teams to complete data sharing and linking projects.

---

<sup>49</sup> [https://www.lucidchart.com/pages/tutorial/swimlane-diagram#section\\_2](https://www.lucidchart.com/pages/tutorial/swimlane-diagram#section_2)

<sup>50</sup> Klein, G. (2007). Performing a project premortem. *Harvard Business Review*, 85(9), 18–19.

<sup>51</sup> [Peabody, M. J. \(2017\). Improving planning: Quantitative evaluation of the premortem technique in field and laboratory settings \(Unpublished doctoral dissertation\). Michigan Technological University, Houghton, MI.](#)

<sup>52</sup> Klein, G., Sonkin, P.D. and Johnson, P. (2019). [Rendering a Powerful Tool Flaccid: The Misuse of Premortems on Wall Street](#). Working Paper.

<sup>53</sup> [UK Cabinet Office \(2012\). Applying behavioural insights to reduce fraud, error and debt.](#)

## Solutions

The key theme to solving friction costs is to 'make it easy'. Possible ways that ADR UK can do this within data linkage projects include:

1. As previously mentioned, develop a data linking project starter pack.
2. Create voluntary standards for information governance that meet legal requirements but also streamline reviews and assessments for data linking projects (such as a shared DPIA).
3. Create common consent forms for linkages that require primary data collection or template data sharing agreements when consent is required. As for examples from projects covered in the interviews:
  - a. Troubled Families: MHCLG related ICO advice on how to write consent forms for families, though this still creates burdens in terms of having to translate that advice.
  - b. Mental Health Liaison and Diversion: The programme encountered significant delays because there was no cross-government agreement on consent wording, even though several legal teams had been tasked with coming up with a form that could be used.
4. Appoint a data linking 'business partner' from ADR UK to work with teams who are engaged in data linking projects. Each government department and local authority should have a named contact in ADR UK that helps them navigate the process and signposts them to relevant materials and resources (such as those proposed above). This is a system employed by the Centre for Longitudinal Studies to help navigate their own processes and internally within the Department for International Trade to help navigate the ONS SRS processes and other data linkage issues.

A secondary theme is to 'make it transparent'. While not wanting to discourage government departments from planning data linkage projects, it is important that realistic amounts of time and budget are set aside for them, and that this is clear up-front. ADR UK could assist with this, for example, by clearly setting out what the frictions are likely to be, as well as the key phases of a data linkage project. This could be augmented, for example, by ADR UK offering ad hoc advice and support to pull together and review where projects have progressed to, and work with HMG departments to set out a road map for data linkage projects from planning to completion.

## **Allocation of attention**

### Problem

Humans have a limited amount of mental resources in terms of attention and

particularly when under pressure or stress our attention can narrow significantly.<sup>54</sup> <sup>55</sup> Public servants (like the rest of us) can therefore only cope with a limited “cognitive load” and where they direct their attention influences which issues and solutions are most likely to be prioritised and acted on. This allocation of attention is also often driven by public pressures and political concerns (and often rightly so, given that politicians have been elected to be responsive to their electorate). This means, however, that issues that are not high on a political agenda often get de-prioritised.<sup>56</sup>

Data sharing and linking projects are rarely at the top of a political agenda. Even in cases where they are, they are easily replaced by other priorities (such as statutory duties).

### Solutions

First, a solution could be to ensure that team leaders (e.g. analytical leads) that might be involved in data linkage have time to do forward thinking. For senior leaders, making it salient (i.e. ‘top of mind’) is important, because this decreases the pressure on them to have to actively remember it themselves. Explicitly making space in their schedules for conceiving and running data linkage projects is another aspect which relieves the cognitive load of scheduling and worrying about these tasks. This could be done by encouraging departments to set this as personal objectives, or as part of regular government cross-departmental analytical meetups.

It also helps to create structures and incentives that give analysts and leaders time to identify opportunities for high value data linking projects. For example, high performing data analysts could be given 0.5 - 1 day per week to identify such opportunities, just like staff at Google were granted during the first years of the company’s existence. Setting up a cross-departmental community of practice with external researchers may be particularly effective. In parallel, there would need to be other measures in place to ensure that the identified opportunities are then converted into real projects.

Second, ADR UK could advocate for government departments hiving off data linkage teams into special projects teams so that they can work outside of mainstream policy pressures.

Third, ADR UK should, especially for its funded projects, make timelines for data linking projects transparent and dynamic by adding a ‘live tracker’ on its website that

---

<sup>54</sup> Chernev, A., Böckenholt, U. and Goodman, J. (2015), Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25: 333-358.

<sup>55</sup> Zimmerman, F.J. & Shimoga, S.V. The effects of food advertising and cognitive load on food choices. *BMC Public Health* 14, 342 (2014).

<sup>56</sup> [Hallsworth, M., Egan, M., Rutter, J., & McCrae, J. \(2018\). Behavioural Government: Using behavioural science to improve how governments make decisions. The Behavioural Insights Team.](#)

lets stakeholders know where in the process a data linking project is and how long it has been at that point. The public nature of the tracker would also provide accountability for ADR UK and HMG. This would need to be done in a sensitive and consultative way, but a lack of this live tracker is a serious obstacle for research organisations wanting to make use of the linked data ADR UK is enabling access to.

Lastly, efforts to show the value of data linkage should be concentrated on organisations that very rarely engage in linkage even for their own purposes where there is likely a default against doing linkage.

## 4.2 Judgments: risks, rewards and uncertainty

Our perception of risk and rewards does not occur in a vacuum: it is dependent on the context we find ourselves in. Our environment, including what we think others are doing and how certain information is presented to us, can shape how risk averse our behaviour is.<sup>57</sup> A person's or organisation's perception of risk may be influenced by several contextual factors.

It was continually noted in interviews how averse departments and other government bodies are to any level of potential loss. A powerful example from our interviews is that in the Troubled Families (TF) evaluation, despite the ICO itself endorsing the effort and providing regulatory assurances, many local authorities were still worried about sharing data.

### Loss aversion and status quo bias

#### Problem

We dislike losses more than we like gains of an equivalent amount - for example, a salary increase of £500 feels less good than a reduction in salary of £500 feels bad. Research suggests that the pain of losing is twice as powerful as the happiness derived from gains.<sup>58</sup> For decision makers in government the potential losses from a data sharing project gone awry will loom large, too.

Due to loss aversion we are risk averse in the domain of gains in relation to a subjective reference point - put simply, if things are going 'well', we are not keen to take risks which could result in losses. Conversely, when things are not going well, we become more willing to take risks. The upshot of this is also known as status quo bias - people tend to prefer the status quo (to maintain the existing state of affairs).<sup>59</sup>

---

<sup>57</sup> Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-292.

<sup>58</sup> Ibid.

<sup>59</sup> Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, 5(1), 193-206.

Decision makers in government are also likely to stick with the status quo given the potential losses from a data sharing project gone awry.<sup>60</sup>

### Solution

A solution here might be to narratively reframe *not* linking data as a cost. Data linking projects are often framed around the harm to the public that can come from inadequate data protection (for example, re-identification from linked data). This could be tested experimentally by devising different messages to send to civil servants, and seeing which would be more effective in persuading them to go ahead with data linkage. In these experimental tests, it would be important to explore how different groups respond to distinct messages, as information governance staff, for instance, are likely to be persuaded by different messages than analysts or data scientists.

ADR UK should work to highlight the harm to the public (and public finances) that comes from *not* linking datasets: duplication of work, higher costs, lack of improvement or oversight of programmes, etc are all consequences of keeping data unlinked and unshared. Framing these examples of harm as stories which impact people - either front line staff or service users - can bring attention to the issue of data linking. This could be particularly powerful in the context of COVID-19.

Within current linkage projects, the Troubled Families team noted that innovation is viewed as inherently risky, so they found it important to keep key stakeholders updated, outline the benefits and provide reassurance on the risks.

## **Present bias**

### Problem

A related problem is posed by the fact that people place a disproportionate value on costs and benefits that occur in the present compared to those that occur in the future.<sup>61</sup> This means that, if a task will benefit us in the long term but has costs in the present, we are less likely to do it (this is why, for example, it is difficult for us to get up early in the morning to exercise).

Data linking projects have clear short term costs in terms of staff time, IT, and data management. Yet the benefits of these projects are far in the future (for example, when a policy question using the data has been answered or a research project using the data is completed). In addition, long term benefits may not be clear (e.g. the results of the research) and they may not directly relate to the data controller.

---

<sup>60</sup> One also needs to consider that in the case of data-sharing, government departments may use legislation specific to their policy area (and over many years), meaning that legal teams develop significant expertise with that legislation, hence the 'status quo' is to continue using familiar legislation.

<sup>61</sup> See, for example, Green, L., Fry, A.F., Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science*, 5(1).

Bringing project benefits to the present, delaying costs so that they are incurred in the future or making long-term benefits salient, can help organisations navigate present bias.

### Solutions

First, to solve this we can consider trying to bring the benefits forward to the present, for example HMT could make part of project funding contingent on data linkage (where this is relevant).

Second, the non-financial benefits of data linkage need to become more salient. This could be done by:

- Presenting organisational leaders with a prototype (using dummy data) to demonstrate what insights linked datasets could show and what questions it could help answer. This could be done by ADR UK funding researchers to develop dummy or synthetic datasets to make these points, and to help train students and researchers to analyse the data. In certain cases researchers could also conduct brief ethnographic research with potential beneficiaries of the linked data. Recording research findings in accessible formats such as video could help bring to life the potential benefits to organisational leaders. Research shows that people feel greater empathy and are willing to expend more resources to help specific, identified individuals compared to a larger impersonal group of people (known as the 'identifiable victim effect').<sup>62</sup>
- Holding a visioning workshop with stakeholders to get them to envision future benefits of linked data. This could include participants writing a letter to themselves describing what they perceive to be the key benefits of the project and sending these letters to participants 6 months or 1 year into the project.
- Having testimonials from successful teams at different stages of the project, and remote coaching and support.
- Considering who delivers the message around the benefits of data linkage, in particular trying to pick a trusted person for that team. It may be useful for that person to be in a strategy role in that department or organisation, to overcome the sense that it's nobody's priority in that organisation. It may be that the strategy for engagement needs to explicitly shift to senior policy colleagues to build buy-in from them to then help raise the profile of the work.
- Creating a community of interest of different departments and teams involved in data linkage projects.

Making the benefits salient and relevant to partners is particularly important with bodies that tend to be data exporters more than they are data users, especially local authorities where the burden from creating new datasets can swamp their ability to

---

<sup>62</sup> Lee, S., & Feeley, T. H. (2016). The identifiable victim effect: A meta-analytic review. *Social Influence*, 11(3), 199-215.

do their own analysis. In this context, two staff from East Sussex Council we interviewed remarked that relevant examples of how research can be used to drive policy and practice were particularly important. Senior leadership at councils are unlikely to support efforts that just help further academic literature, and are very focussed on what will directly impact the citizens of the local community.

One way to overcome this would be to have ‘data exporters’ such as councils involved in developing research questions that might be answered by academics using data linking, so that they have more of a stake in making it work.<sup>63</sup> It may also be useful to involve them more closely in deciding what linkages are conducted, and providing products that they can use to conduct policy-relevant analysis of local data (given the generally lower analytical capacity of local councils compared to central departments or academic institutions).

Third, it is necessary to ensure that all project participants benefit from a linkage using a ‘carrot and stick’ approach. On the stick side, ADR UK could advocate to HMT to incentivise collaboration by holding back a percentage of relevant project budgets until data is deposited as part of spending review settlements, or making release of funds for area A conditional on deposit of data in area B. This would require the development of a light-touch process to assess whether data linkage is possible and/or cost-efficient. On the carrot side, ADR UK could also work up a package of support for HMT to offer to departments to help them conduct data linkages.

## Availability heuristic

### Problem

When judging the risk of a particular event, we need to judge not only its impact (what are the consequences if the event occurs) but also its likelihood (how likely it is that it will happen). In estimating how likely something is, we have a tendency to rely on the shortcut of how easily it comes to mind. This is known as the ‘availability heuristic’. For example, after the 11 September terrorist attacks, many people opted to drive instead of fly to distant locations: they assumed the risk of a plane crashing was high given how easily they could think of an example. Studies have since found that it is likely many more people died in road fatalities than would have died if they’d flown due to this misperception of risk.<sup>64</sup>

For government leaders and legal teams, the high profile of recent privacy legislation

---

<sup>63</sup> Similar to Areas of Research Interest

<https://www.gov.uk/government/collections/areas-of-research-interest>.

<sup>64</sup> Deonandan, R., & Backwell, A. (2011). Driving deaths and injuries post-9/11. *International Journal of General Medicine*, 4, 803–807.

(such as GDPR) or data sharing that garnered negative headlines<sup>65</sup> may be what comes to mind when they think about data linking project risks. This may result in them overestimating the chances of these negative outcomes happening in the future. Similarly, the availability of these negative incidents in their memories may also lead them to *underestimate* the level of public support for data sharing.

### Solution

A simple and more short-term solution to this problem is likely to be correcting for misperceptions about the risk, for example by:

- Testing what messages best convey the benefits and risks of data linkage to the public, which could be done through an online experimentation platform.
- Sharing examples of data linkage that have gone well and avoided common problems, and how they managed to do that.
- Use peer messengers as champions for data linking. Many organisations present case studies of what successful data linking projects have achieved. We think incorporating testimonials from staff who worked on these projects could make the case studies more salient. If possible, it could be helpful to provide legal team testimonials for successful linking projects and to share these testimonials (and potentially contact information) with legal teams considering similar issues. It may also be useful for these messengers to be accompanied by high-level sponsors from each department.

A longer-term solution would be to help civil servants understand the conditions under which the public considers data sharing (un)acceptable. This is not to dictate or restrict what linkages are undertaken, but it can clarify where more work is needed to engage the public. As mentioned before, ADR UK wishes to conduct more detailed research with the public to understand when, how and why data can be linked in different policy domains. Such research can lead to the establishment of clearer policy principles around data sharing, engender greater public trust and reduce the perceived risk of data linkage in government. Qualitative research and deliberative forums such as citizen juries are crucial to help understand people's preferences about data sharing in different contexts. However, we believe experimental research informed by behavioural science also has an important role to play in revealing people's preferences and understanding of the issues involved.

---

<sup>65</sup> <https://iapp.org/news/a/ico-uk-government-reported-495-data-breaches-during-fiscal-year/>

One example of this is where BIT helped develop an evidence-based choice architecture for the National Data Opt-out Programme (NDOP), which aims to allow people to opt-out of their data being used for research and/or planning in the NHS. We designed a series of online experiments and found that most people did not understand the difference between planning and research. Based on these findings it was decided to offer people the choice to opt-out of both at the same time, rather than separately.<sup>66</sup> This is an example of how empirical research can inform government policy. Online experiments with representative samples can be an effective way of assessing people's understanding of content (e.g. policy proposals), preferences and decision-making. Such experiments could be used to understand whether people feel comfortable with their data being shared in particular scenarios and how information needs to be presented to them to address potential concerns.

## Desire for control

### Problem

Human beings have a deep-seated desire for certainty and control, which they don't like giving up.<sup>67</sup> To give up control in the face of uncertainty requires trust: building trust between partners in a business alliance, for instance, has been found to make partners more comfortable giving up control.<sup>68</sup> In interviews people familiar with data linking projects noted the importance of trust between organisations and end-users of the data (such as researchers), possibly for similar reasons.<sup>69 70</sup>

### Solutions

A solution may be to help build up trust between organisations. A simple way of creating trust between relevant government staff is to provide opportunities for people to get to know each other, for example through communities of practice (see above) or meetup events.

Another solution is to create default norms around operational transparency. Operational transparency entails deliberately disclosing how a process works to increase empathy and understanding between customers and a business, or between citizens and government. A famous example involved travel-booking website Kayak.com; Kayak showed users its 'work' as it searched, fanning flight options across the screen as it ordered them by price. Experimental evidence

---

<sup>66</sup> [Harper, H., Egan, M., Cornel, P., Ter Meer, J., & Wittels, A. \(2018, October 10\). Data sharing and the importance of choice architecture in healthcare: New results. \*Behavioural Insights Team Blog\*.](#)

<sup>67</sup> Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in cognitive sciences*, 14(10), 457-463.

<sup>68</sup> Das, T. K., & Teng, B.-S. (1998). Between trust and control: Developing confidence in partner cooperation in alliances. *Academy of Management Review*, 23(3), 491-512.

<sup>69</sup> Green, E., Ritchie, F., Mytton, J., Webber, D. J., Deave, T., Montgomery, A., Woolfrey, L., & Chowdhury, S. (2015). *Enabling data linkage to maximise the value of public health research data* (Public Health Research Data Forum). Wellcome Trust.

<sup>70</sup> Administrative Data Research Network (2016). Mid-Term Review Report. Research Project Management.

showed that customers are more satisfied with websites that are ‘operationally transparent’ in this way.<sup>71</sup>

---

<sup>71</sup> Buell, R. W., & Norton, M. I. (2011). The labor illusion: How operational transparency increases perceived value. *Management Science*, 57(9), 1564-1579.

## 5. Structural solutions

---

In this section, we address solutions that are not behavioural in nature, but rather are ways that ADR UK could contribute to fixing technical or structural problems. We have chosen not to concentrate on problems like having a common ID spine.

The five structural elements analysed here are:

- The need for more data preparation and engineering resource,
- The need for good data collection in the first instance,
- The need for software to do data linkage and analysis,
- The need for high-quality legal advice,
- The approvals process for using the ONS SRS.

### Data preparation and engineering resource

#### Problem

Several interviewees expressed that they wanted national efforts to promote data linkage to go beyond the idea that departments and agencies are unwilling to share data but to recognise that in many cases data is not in a state to be shared or linked.

In general, most datasets (outside some high-profile examples like the National Pupil Database) are created to answer specific questions, which may be for operational or policy questions. Therefore, analysts will put them together in an *ad hoc* way and there are rarely enough people who are dedicated to maintaining or cataloguing them. Even analysts in a department might not know about the quality of particular datasets or what is in the data.

For example, in the Longitudinal Educational Outcomes (LEO) linkage there was no overall data documentation, which meant that the Institute for Fiscal Studies (IFS) analysts had to spend a significant amount of time attempting to interpret the meaning of data fields.

Another interviewee brought up the example that MoJ analysts were quite unfamiliar with courts data and unaware in advance of issues like fields that may not be filled in or might only be filled in by some courts. As no documentation had been kept as to why these fields were missing, it was also unclear if this was a systematic or random error. As ADR UK will be aware firsthand, the Data First project illustrates the scale of the task at hand in criminal justice data.

Writing high-quality data documentation and understanding how the data was collected requires both a high level of skill, sufficient time and preferably, limited staff turnover.

### Solution

One suggestion on how to fix this would be to ringfence resources to ensure data quality, for example data engineering fellowships like those created by HMCTS. These could be funded by ADR UK or more broadly by UKRI, along with host HMG departments. In practice these could take the form of part of PhD funding (in a CASE model), post-doctoral fellowships, or mid-career fellowships. This was also suggested by several interviewees as a potential activity for ADR UK to coordinate.

In general, several interviewees expressed that they wanted ADR UK to help with the technical and legal aspects of linkage rather than produce further analytical use cases.

### Advising on data collection

#### Problem

While ADR UK is focused on data linkage, to ensure linked datasets are of high quality it may make sense to increase efforts on improving the quality of data collection as well. To make ADR UK's efforts effective, this needs to occur in important target areas alongside demonstration projects in policy domains with existing high-quality linked data. This is for three reasons:

- It takes a considerable amount of time to improve the quality of data collection, and waiting to engage departments on doing this will slow down the creation of linked datasets in very important policy areas. As a powerful example, one interviewee brought up that a lack of data linking mental health disorders and crime has affected the ability to form basic hypotheses at the policy making stage. For example, without the right data it is not possible to know if people with certain conditions use healthcare services significantly more than the general population. A lack of linked data also affects the targeting of policy and development of theories of change in intervention design.
- In BIT's experience of convincing departments and other public bodies to use evidence, we have found that examples from other policy domains are not always seen as persuasive or relevant.
- Interviewees' responses reflected the demand for help with these issues now, and that it is important to not just focus on demonstrating the value of data linkage.

Issues brought up included:

1. Newly collected datasets need to be set up to link well with other data.

2. Data sharing purposes should include linkage and deposit, but often do not. It is necessary to specify a purpose for sharing, linking, storing and processing data under both data sharing and data processing legislation. If this purpose is set too narrowly or precludes data deposit, this can prove impossible to reverse. This is a particularly severe problem if data is collected from individuals on the basis of informed consent.
3. Datasets (including linked datasets) need to attempt to anticipate future uses of the data. For example, in the Troubled Families programme, users were asked to look at the incidence of knife crime, as their data sharing agreements and privacy notices allowed for analysis to inform policy direction. However, the dataset did not have a specific enough crime code, so they needed to return to ONS to get an additional level of code added. As this was a high profile evaluation, MHCLG had the political 'clout' to do so but it did delay the analysis and would have likely blocked the analysis in the case of a less high-profile project.
4. Data linkages need to reflect the need for answering specific policy questions, which aren't necessarily at the stage where a specific research question has been formulated.
5. Where individual-level data cannot be deposited, this may severely limit the usefulness of the dataset. For example, with Troubled Families, due to the design of the consent forms the original dataset has to be destroyed, but a derived dataset could be kept. The deposited dataset is at a reduced level of specificity to limit the risk of identification, but this means that other data linkages will not be possible with this dataset.

### Solutions

ADR UK could leverage the expertise of the ESRC, its own staff and academics to help with the five raised issues:

1. *Ensuring that newly collected datasets are set up to link well with other datasets:* If a dataset will need to be linked in future for research or operational reasons, it needs to contain relevant IDs or, where possible, other information such as name, DOB, or address.
2. *Ensuring that data sharing is, preferably, not limited for use on a particular project (where legal):* One bolder option to explore would be to include data deposit terms by default in DSAs between government departments, and to test out variants of these terms. ADR UK data infrastructure partners work like this with data owners already, however it would be useful to advocate for this to be a broader default.
3. *Ensuring that future uses of newly collected datasets are contemplated:* Consider having 'pre-mortems' for linkage projects, for instance in the form of round-table events to brainstorm ideas for policy questions the work could

answer. This is often done ad hoc or post-hoc and it could be more prospective. These would also help to keep track of the benefits of what ADR UK is delivering.

4. *Data linkages need to reflect the need for answering specific policy questions:* One way of solving this issue is to have the roundtables in (3) occur even before sign-off to help build the case. This is especially important where data owners are engaging on a goodwill rather than funding basis.
5. *Individual data cannot always be deposited for legal reasons:* One alternative that could be investigated is the use of synthetic datasets, where after a linkage has occurred a copy of the dataset is made that has the same statistical properties but none of the original personal data. This would not work if more datasets need to be merged at a later date, but would still be very useful for research and policy purposes, where departments would not consent for the original linked data to be deposited.

### Software library

There are three ways that ADR UK could help:

1. For data preparation, ADR UK could either provide a list of vetted data engineering APIs or encourage their centralised procurement.
2. For software needed to do data linkage, like transfer, storage and matching software, ADR UK could provide a list of vetted solutions.
3. For analysis of existing linked datasets, it could work with its infrastructure partners to ensure programmes such as Stata, Tableau, R, SPSS, SAS and Stan are all available.

ADR UK could also help by producing or seeking advice on contract negotiations for specialised software to avoid problems such as government departments being ensnared by teaser rates.

### Getting advice on legal gateways

While the ICO, UKSA (regarding research linkages) and DCMS (regarding operational linkages) all provide advice on the legal issues around data linkage, this advice could be better publicised and joined up. In particular:

- DCMS has responsibility for the section of the *Digital Economy Act* that concerns the linkage of personalised data for providing services to specific individuals. DCMS also helps government departments, non-departmental public bodies and local councils find other gateways where appropriate - though they may not be aware of this service.
- UKSA has responsibility for the section of the *Digital Economy Act* that concerns the linkage of de-identified data for research purposes.
- ICO has responsibility primarily for data protection and the GDPR.

It might be useful to create simplified guidance in the form of a decision tree or easy to use workbook that helps departments to think through these issues and to know who to contact.

It may also be worth embedding timely reminders to government departments when they are considering undertaking new linkage efforts to think about alternative legal gateways and not be stuck in existing ones - perhaps by putting these in forms justifying the use of the gateway.

## Approvals process for the ONS SRS

### Problem

The specific issues articulated in interviews were:

1. Two interviewees expressed that their engagement with linked data is limited by the perceived need to have specific research questions in mind prior to accessing the data. This was viewed as defeating the point of exploratory research, which is partly about refining research questions.
2. The process of linking outside data into the SRS and of removing data from the SRS were viewed as cumbersome and frustrating, particularly for rapid-fire analyses in the context of, for example, crises like COVID-19.
3. Perhaps more worryingly, it was also expressed that if more data linkages are deposited in the SRS, this is likely to overwhelm the ONS' approvals staff, given the administrative burden of the current system. This is particularly true of checks needed to ensure that no disclosure of individual data occurs.

### Solutions

In order:

1. There are a few ways of getting around this:
  - a. Clarify what a reasonable research purpose means in guidance on accessing linked data, and by working with government departments to ensure that it allows for exploratory research.
  - b. Create exemplar research questions that are exploratory in nature but that could be adapted by government departments wanting to do more exploratory research.
  - c. Create synthetic datasets as explained above which have the same statistical properties as the underlying data.
2. As approvals also need to be sought from government departments one solution could be to have a standing group of decision-makers who can rapidly approve use cases for the linked data. 'Fast track' approvals for government departments could also be considered. While we realise that the approvals

process on the UKSA side is very rapid already, in cases where data owners need to give approval, this can significantly slow down the delivery of projects.

3. ADR UK could give more assistance to departments setting up their own access to the SRS, especially smaller departments and public bodies who may find the process difficult. In addition, the burden placed on ONS' approvals staff could be reduced by finding ways to automate the checking of outputs for disclosiveness. We recommend dedicating some developer time to understand if it is possible to build an approach that allows the majority of checks to be automated.

## 6. How to involve researchers, non-departmental bodies and councils

---

In this section, we cover issues arising in interviews and case studies that involved organisations other than central government departments, in turn: non-departmental bodies, research organisations and local authorities.

### Non-departmental bodies

Unlike government departments, often these bodies do not own data. For example, the Competition and Markets Authority (CMA) typically only holds data obtained from private companies for the purpose of doing market studies, and this data cannot legally be kept past the end date of those studies. Similarly, Ofwat purchases water consumption data from a private company, receives complaints data from the Consumer Council for Water and receives data from the Environment Agency on water drinking quality.

Some of the data linkage they facilitate is between private companies. For example, Ofwat and Ofgem have been encouraging the sharing of vulnerable consumer lists between energy and water companies.

The exception here is that bodies like Ofsted, the CQC (Care Quality Commission), and FSA (Food Standards Authority) hold data at the regulated body level (e.g. school level) on inspection outcomes.

They represent strong potential customers of data linkage, and there is particular interest in understanding the needs of vulnerable consumers. As such, engagement might focus on using their needs to encourage departments that do own the data to become involved in data linkages or creating new datasets.

### Research organisations

Using linked data is clearly important to a significant share of both government-driven and purely academic quantitative social science. Further, researchers often lead data linkage projects - especially for impact evaluations - but the IFS also did much of the linkage to create LEO.

However, research organisations and universities face significant barriers to both using linked data and doing linkages themselves.

#### *Using Linked Data*

The open letter outlines four main barriers to using linked data:

1. The cost of using health data is prohibitive.
2. Lengthy approvals processes, for example a linked dataset with 3 data owners might require 3 sets of approvals. ADR UK could propose that departments move to a common approvals process (as DfE and MoJ already have for linkage), and in the meantime help researchers to navigate existing ones. It is worth noting that the SAIL databank and ADR Scotland (who are both part of the ADR UK partnership) already work on this model.
3. Lags in when datasets are finalised mean they are out of date.
4. There are barriers to actually accessing the data, as we have previously discussed in the subsection on the [SRS](#).

These points were discussed in the interviews. Additionally, the issue of misaligned grant processes was brought up. For example, a foundation or ESRC might give out a grant for a project that relies on a data linkage unlikely to be done within the timeframe required, or requiring significant effort on linkage to be undertaken in preparation of a bid which may not be successful. Interviewees also felt that it was unclear which linkages were finished. One action that could be taken to remedy this would be to increase the regularly updated information (i.e. progress updates monthly or bi-monthly in an easy dashboard) that is publicly available and/or provided to major grant funding bodies and research organisations as to the progress of live data linkages (as suggested above).

It may also make sense to further publicise ADR UK's own processes for applying for grant funding to use linked data and, as stated elsewhere, help applicants through the process through dedicated managers (if this is not already underway). It would be worthwhile auditing these processes and understanding whether/how they could be streamlined.

### *Linking Data*

Researchers are often involved in data linkage efforts, either doing the linkage for departments for a direct commission or as part of commissioned impact evaluations.

The following barriers exist to conducting these linkages and to depositing data:

- In ESRC applications, Principal Investigators (PIs) are asked about the policy impact of that specific work, but rarely if ever are prompted to think about how the data they collect could be used elsewhere. That could be fixed by setting a default in the application process.
- Research organisations often rely on default terms in Data Sharing Agreements which often say that data can only be kept for 1 year post-publication and do not allow for depositing data (noting that 'post publication' could mean a report, or a final project output such as a journal paper). It is worth noting that ADR UK data infrastructure partners don't have

this clause, but the issue primarily applies to new primary data collection, especially for evaluations.

- Terms imposed by ethics committees during primary data collection can make this even more difficult, e.g. by asking for consent forms where they are otherwise not needed. This creates problems regarding the GDPR (and pushing back raises risks of projects being cancelled). Liaising with ethics bodies to encourage them to think about being more open to ongoing usage of de-identified data after project completion might be an important step to remedy this.
- Grants made by non-UKRI funding bodies (e.g. many charitable foundations) do not make provision for depositing data (in particular for anonymisation and tidying), and neither do most government contracts for evaluations or research projects.
- Approvals processes are extremely time-consuming, with approximately 75% of one major project's budget being spent on approvals to access data, which is a deterrent to linking the data in the first place. This can be costly both financially and psychologically, with one interviewee remarking that they would not embark on future projects with linked data due to the bureaucracy of approvals processes. This was also viewed as deterring interdisciplinary research. While researchers might have good contacts and familiarity with approvals in their own area, this might not be true in other domains. This speaks clearly to the need for cross-disciplinary approaches to funding and research collaboration, which ADR UK is already proposing and promoting, but which may need further encouragement to increase uptake by universities.

One piece of engagement that might therefore be useful would be to work with departments to include depositing de-identified data as part of their default terms for commissioned research and data access, as well as working to link up grant funds to those project teams where strategically appropriate. Research teams are unlikely to apply for this money themselves in some circumstances, so it may be prudent to proactively seek them out and proactively seek out data linkage teams as with the MoJ-DfE data linkage project.<sup>72</sup>

Another model is to mirror the HMG Trials Advice Panel (TAP)<sup>73</sup> but for data sharing. The TAP is a group of professional researchers who provide support and advice to government on experiments and evaluation. A parallel body might be the Data

---

<sup>72</sup>

<https://www.adruk.org/news-publications/news-blogs/expert-team-appointed-to-test-potential-of-moj-dfe-linked-dataset-234/>

<sup>73</sup>

<https://www.gov.uk/government/publications/cross-government-trial-advice-panel-role-and-membership>

Advice Panel, consisting of those with experience in and outside of government offering support on, in this case, data sharing and data linkage.

### Local authorities

Local authorities (and institutions within them such as schools) are a valuable source of data that constitutes many administrative datasets like the NPD or Troubled Families dataset. They may also be good testbeds for linkages to check feasibility, as interviewees expressed concerns that data linkages in government have often been very large-scale and therefore highly risky.

On the other hand, councils can struggle to make linkage projects work, especially where they are not unitary authorities and therefore do not directly control service provision. They may also be disengaged from data linkage as they are more often 'data donors' having to expend a significant amount of money and staff time to submit data to national departments, with little direct return.

One interviewee wondered whether for demonstration linkages it may make more sense to also work with local authorities and associated local bodies (NHS trusts, police forces etc). An example cited was that Blackpool has a Better Start programme and good data on a manageable scale, but ambitious linkage efforts have not got off the ground there due to lack of technical capacity.

Four factors were seen as important to local authority involvement (and this chimes with BIT's experience on other projects):

- The usage of research and data to evidence funding bids and understand vulnerabilities in local communities.
- Provide examples of where research could be used to inform practice and policy, for example by enabling comparisons to other local authorities or informing the selection of interventions.
- There needs to be a degree of reciprocity between the national department or researcher benefiting from locally provided data and the council. For instance, on the Troubled Families programme, each local authority was provided with disaggregated local impact reports and cost-benefit analyses.

Engagement that could therefore be useful with local authorities is to identify LAs with interesting possible linked datasets to inspire future national linkage, and to publicise the use of linked data to address local policy concerns.

## 7. Ideas for future linkages

---

In this section, we detail ideas for future linkages that ADR UK could help to facilitate or fund. There are three more developed ideas relating to child maltreatment, violence prevention and vulnerable consumers. We then detail speculative ideas sourced from our interviews.

### A national system for monitoring rates of child maltreatment

Currently there is no centralised monitoring system for child maltreatment data in England. Police forces, children's services departments, and public health officials have to negotiate to share data and there is not a single approach to doing this. Yet each of those agencies - and other organisations besides - hold administrative data that on its own may be more valuable if combined with other data sources. Administrative data holds great promise in both helping to understand risk for maltreatment and evaluate interventions to improve outcomes through different systems.

Work in the United States has illustrated not only what can be achieved but also what is being lost by not taking a more data-savvy approach to child welfare.<sup>74</sup> Child maltreatment, in particular violence against children and exploitation, remain intractable while data are shrouded and/or siloed.

A team from the University of Oxford, led by Dr. Michelle Degli Esposti and Dr. David Humphreys, is working towards a national monitoring strategy for child maltreatment, enhancing and supplementing their previous work with aggregated, publicly available national data<sup>75</sup> by using a localised bottom-up approach, starting with Oxford/Thames Valley. We have introduced the team to ADR UK with a view to exploring funding options - existing research from the US initiative illustrates many of the possibilities this affords. The situation regarding child maltreatment in the UK and globally has been thrown into sharp focus by the COVID-19 pandemic; creating a sense of urgency for the need for better information to support practitioners and policymakers in effectively planning and responding to need as the UK likely cycles in and out of lockdown.

---

<sup>74</sup> <https://www.datanetwork.org/research/>. For a recent innovation on child protection monitoring data see <https://digital.nhs.uk/services/child-protection-information-sharing-project>.

<sup>75</sup> Degli Esposti, M., Taylor, J., Humphreys, D. K., & Bowes, L. (2018). iCoverT: A rich data source on the incidence of child maltreatment over time in England and Wales. *PLoS one*, 13(8), e0201223.

## Data sharing for violence prevention

There have been several initiatives pushing the use of shared data to help prevent violence. Perhaps the most successful version of this is the Cardiff Model, pioneered by Professor Jonathan Shepherd.<sup>76</sup> In the mid-1990s, Professor Shepherd initiated a public health and data-led approach to record de-identified information about people with serious injuries presenting at Emergency Departments (EDs). This led to the discovery that nearly two-thirds of violent incidents requiring treatment for serious injuries presenting at the ED were not reported to the police.<sup>77</sup> The Cardiff team then began an initiative to combine ED data with police and community safety partnership data in order to better understand the burden of violence on communities and act to prevent it. This work has led to national recording and data sharing standards for violence prevention in 2012, based on the original Cardiff approach,<sup>78 79</sup> and improved accuracy of prevalence estimates for violence.<sup>80</sup> It has been estimated that the practice of data-sharing in Cardiff alone generated nearly £7m in savings to health and justice systems, compared to similar cities.<sup>81</sup> More recently, ambulance data has also been shown to be a unique source of information on the volume, nature and location of violence,<sup>82 83</sup> and is increasingly viewed as a means to target interventions.<sup>84</sup>

While the above summary of the use of ED and ambulance data sounds like a resounding success, some of the barriers to successful data sharing identified in this report were present and took immense effort to overcome. For example, in one study, negotiations to share ambulance data took 12 months (of a 24 month project) with delays largely centring around cultural and legal differences between police and

---

<sup>76</sup> <https://impact.ref.ac.uk/casestudies/CaseStudy.aspx?Id=3669>

<sup>77</sup> Sutherland, I., Sivarajasingam, V., & Shepherd, J. P. (2002). Recording of community violence by medical and police services. *Injury Prevention*, 8(3), 246-247.

<sup>78</sup>

<https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1594-information-sharing-to-tackle-violence-minimum-dataset>.

<sup>79</sup> [Giacomantonio, C., Sutherland, A., Boyle, A., Shepherd, J., Kruithof, K. and Davies, M. \(2014\). Injury surveillance: using A&E data for crime reduction Technical report.](#)

<sup>80</sup> Gray, B. J., Barton, E. R., Davies, A. R., Long, S. J., Roderick, J., & Bellis, M. A. (2017). A shared data approach more accurately represents the rates and patterns of violence with injury assaults. *J Epidemiol Community Health*, 71(12), 1218-1224.

<sup>81</sup> Florence, C., Shepherd, J., Brennan, I., & Simon, T. (2011). Effectiveness of anonymised information sharing and use in health service, police, and local government partnership for preventing violence related injury: experimental study and time series analysis. *Bmj*, 342, d3313.

<sup>82</sup> [Sutherland, A., Strang, L., Stepanek, M., Giacomantonio, C. and Boyle, A. \(2017\) Using Ambulance Data for Violence Prevention: Technical Report. Santa Monica, CA: RAND Corporation.](#)

<sup>83</sup> Taylor, A., Boyle, A., Sutherland, A. & Giacomantonio, C. (2016). Using ambulance data to reduce community violence: critical literature review. *Eur J Emerg Med* 23(4):248-252.

<sup>84</sup> Ariel B, Weinborn C, Boyle A. Can routinely collected ambulance data about assaults contribute to reduction in community violence?. *Emerg Med J*. 2015;32(4):308-313.

healthcare.<sup>85</sup> The arrestee liaison and diversion programme referenced elsewhere in this report encountered significant delays and problems with data sharing for similar reasons. And the Cardiff model data share took years of repeated effort.

ADR UK funding and the ONS SRS infrastructure could bring significant efficiencies to data-sharing at the overlap of health and justice in England, as with similar projects in Scotland and Wales. In particular, providing a secure third-party location for data to be deposited and analysed would help to overcome some of the issues with sharing between police and healthcare services in particular.

### **Vulnerable consumers and consumer/household finance**

In discussions with the interviewed departmental bodies, and in previous discussions with the UK Regulator Network, a common theme that has come up is the lack of a national dataset to understand which types or groups of consumers are vulnerable across multiple sectors (and therefore regulators). This is important as without understanding and evidencing harms and how they might work across different regulatory domains, most regulations are aimed at the average consumer.

The CMA in particular is concerned that it cannot measure the impact of interventions or regulations on subgroups of consumers in typical datasets it receives. For example, if companies provide it with data from A/B testing, they typically don't measure any characteristics relating to vulnerability. Therefore, the CMA might rely on misleading proxies like where a consumer lives which can lead to incorrect inferences.

The CMA has previously tried a data linkage project in this area, by attempting to bring together customer records from regulated firms (obtained from the regulators which sit under the CMA) with longitudinal data held by Essex University (Understanding Society). This effort did not succeed due to legal issues between the regulators and academic partners. There also were issues relating to the need to share data with regulated firms to enable matching.

Interviewees also mentioned that this had affected their COVID-19 response, as they could not observe how consumer behaviour had changed in relation to COVID-19.

In a related suggestion, Nest Insight have an ambition to build a linked de-identified dataset of household finances, starting with Nest's data, but with a goal of linking to retail banking, HMRC, consumption/savings, transactions and credit reference data (and perhaps others) for the purposes of high quality public research. The plan is to first conduct observational studies with these data to understand how different aspects of household finance interact with one another and with policy changes – for example, does consumer debt increase when policy interventions lead to higher

---

<sup>85</sup> [Sutherland, A., Strang, L., Stepanek, M., Giacomantonio, C. and Boyle, A. \(2017\) Using Ambulance Data for Violence Prevention: Technical Report. Santa Monica, CA: RAND Corporation.](#)

saving (such as with pensions auto-enrolment). Such a resource would potentially be invaluable in supporting the evaluation of secondary and spill-over effects of policy interventions into consumer financial behaviour. Longer-term, such a dataset could also be used to improve the evaluation of behavioural trials. Effectively Nest has a large body of routine administrative data, with which Nest Insight already conducts research and analysis and which can also possibly function as the basis for further data linkage. However they view any initiative along these lines as something that would need to be done in partnership with multiple data providers and with a trusted third party to hold and manage access to the data.

Currently Nest Insight are working on three specific projects which could contribute to this broader ambition and help to make the use case to other potential partners and funders:

1. With other pension providers and the Pensions Policy Institute - effectively a research dataset to try and pull together administrative data across the Automatic Enrolment market to enable policy research into issues such as the impact of labour-market mobility on the growth in small dormant pension accounts.
2. With the FCA to merge Nest data and credit referencing data to explore interactions between pension saving and consumer borrowing (and consumption).
3. Collaborating with the University of Essex to include a consent question in the latest wave of Understanding Society (US), enabling Nest administrative records to be merged with US responses for research purposes.

## Other ideas

Below we present ideas for linkages which are more speculative:

- Developing a permanent version of the dataset created by the RAND team in the Mental Health Liaison and Diversion work.<sup>86</sup> More generally, we would advocate that any datasets created via commissioned research should contain contract/data clauses that default the data to deposit in ONS SRS. Both the L&D and Juvenile Cohort Study<sup>87</sup> research projects cost millions of pounds but the datasets are not being used.
- There are also other examples of datasets 'looking for a secure home', with one salient example being the Metropolitan Police Public Attitudes Survey (MetPAS). This was being held at the University of Essex and was on the [UK data service](#), but is now embargoed and is being held at MOPAC because of concerns about data protection due to the nature of the data (individual survey

---

<sup>86</sup> [Disley, E., Taylor, C., Kruithof, K., Winpenny, E., Liddle, M., Sutherland, A., ... & Francis, V. \(2016\). \*Evaluation of the offender liaison and diversion trial schemes\*. London: RAND.](#)

<sup>87</sup> [Wilson, E. \(2013\). \*Youth Justice Interventions – findings from the Juvenile Cohort Study \(JCS\)\*.](#)

responses). [MetPAS is used for operational and research purposes](#) and has been linked to other administrative data for analysis, but at present is not being as widely used because of the need for a secure location to deposit the data outside of MOPAC.

- Work with local authorities to conduct local data linkages as demonstration projects.
- Ofsted were interested in establishing data linkage to enable a 20 year longitudinal study with quality of education as an input to final outcomes. In particular, they were interested in whether Ofsted ratings predict later life outcomes.
- Converting existing HMG datasets from completed projects into data deposits. One example is the Juvenile Cohort Study, which collected routine administrative data on 10,000 young people in the criminal justice system and tracked their progress. The project completed just prior to a change in legislation that simplified sentencing for young people, meaning the data was seen as less relevant. However, the £1.2m spent on collecting the data would otherwise be wasted.
- Establishing de-identified transactional datasets from, for instance, Voca Link<sup>88</sup> and combining that with other datasets to better understand consumption patterns, fraud and the impact of changes to HMG workplace benefits. We know there is interest in central government in a similar idea.

---

<sup>88</sup> <https://www.vocalink.com/>.