"It's A Grey Area": searching the grey literature on how local governments use real-world data:

[Mark Clowes](#) and [Anthea Sutton](#), ScHARR information specialists with two decades of combined experience of literature searching in the context of systematic reviews, reflect on the challenges of finding grey literature for the NIHR-funded [Unlocking Data](#) project (led by [Dr. Matt Franklin](#)).

Introduction

When searching for information, our usual practice as information specialists is to start with the published literature and a structured search strategy around a setting or population of interest (e.g. a patient requiring treatment; a problem waiting to be solved). Some projects, however, require a different approach. As part of the Unlocking Data project we have been conducting a mapping review of how local governments are accessing, linking, and using real-world data.

Local authorities do not have a well-established tradition of publishing peer-reviewed articles; so to find the type of case studies we wanted, we decided to start with the grey literature (i.e. information produced outside of the traditional commercial or academic distribution channels). Grey literature is notoriously difficult to find; it may exist in many different formats (e.g. organisational reports, newsletters, web pages), and searching for it can leave the professional information specialist with a nagging feeling of anxiety that they may not have found everything.

A particular problem is the "false positive" - lots of people are talking about data sharing, but that doesn't mean they've worked out how to implement it yet; we had to strike a balance between sensitivity (finding everything relevant) and specificity (minimising the "noise" from people talking about things they would like to do, rather than evaluating what they had done). We came to accept that we were unlikely to be comprehensive as even case studies which are relevant may not be fully reported - sharing arrangements between organizations may be announced, but never formally evaluated (sometimes, depending on the legislative frameworks involved, they may never be made public at all). Instead, we took a purposive approach aiming to find *up to* 100 possible case studies across a variety of domains (not just health) exploring how sharing different types of information (e.g. school attendance, rent arrears, and even library usage) could improve the commissioning of services to enhance the health and wellbeing of communities.

We were already aware of a handful of portals where potential case studies had been gathered already for a similar purpose to our own. These included the [HDR gateway](#); the Wellcome Trust's [Understanding Patient Data](#) site, and the [Local Government Association](#). We also made a list of our domains of interest and of local or national organizations' websites where we might expect to find such information. These sites were variable in quality and usability; some were structured in a way that allowed for browsing, while others relied on basic search functionality using one or two terms (rather than the complex Boolean strategies we use on databases like MEDLINE). We used specific search terms such as "data sharing", "linked data", "GDPR", "information governance", "routine data", "de-identified data" (in various forms and combinations). We also used Google advanced search to look

for terms occurring on a particular domain (e.g. gov.uk); although due to a lack of transparency about how deep within a site Google's indexing goes, we searched many of these sites using their own native interface as well. Searches of this nature take a very long time and it's easy to become lost in "rabbit holes" where web pages redirect you to other pages elsewhere. The list we had created a priori was crucial to our sampling strategy (and our time management) to make sure that we had at least attempted to find examples from all our domains of interest and not just the first we came across.

The case studies we found varied considerably (from a one paragraph summary with a contact e-mail address, to a 120 page PDF document) and there was little correlation between the quality of the reporting and the usefulness of the case study; though this is perhaps unsurprising, given that they were conducted by such diverse organisations and for many different purposes - some were reporting primarily for a local or internal audience, rather than curious researchers like us. In a depressing illustration of one of the other pitfalls of grey literature - its ephemeral nature - we also found a number of "404 File not found" messages where promising-sounding documents had been removed from websites. Webmasters often assume that everyone arrives at their site via the official home page and browses through the structure they have so painstakingly devised; the reality is that many users land mid-site via search engines, so may find out-of-date pages with broken links if these haven't been taken down.

Where case studies did not report the data we needed, we attempted to contact authors/project leads. Unfortunately, we retrieved several auto-reply messages saying that individuals had left their posts and when we tried to identify senior figures in the council who had been the project sponsors, these too had moved on (perhaps as a result of the local elections in May). To complement our searches we have sent out a call for further information via members of our steering group and their networks. We are now in the final stages of data extraction for our included case studies and will complete this by the end of September 2021.

Lessons learned when searching for grey literature

- Make a list of sources and/or topics you intend to cover before you start (and allow yourself enough time to work through them all)
- Use gateways, portals, and existing collections and reviews where these exist - don't reinvent the wheel
- Use a standard data extraction form to collect the key info you need from each included study. This will help you to deal with long documents pragmatically by searching or skimming through the text for what you need; you don't have time to read them in full.
- [This paper](#) by Claire Stansfield et al (2016) was useful in planning our search approach and we'd recommend it to anyone embarking on a similar process.

Funding

Read more on the topic of real-world data from Dr Mark Franklin here:
[Unlocking real-world data to promote and protect health and prevent ill-health in the Yorkshire and Humber region](#)