Transcript: Finding Grey Literature

Finding grey literature sources

There are three key properties of grey literature to consider when searching. They are inconsistent indexing in search engines, items move and disappear and items are not sold.

1. Inconsistent Indexing

If we look at a normal book item record you see that it has a set of subject headings, often between one and three, a set of authors, possibly including the name of an organisation as an author. It will also have a publisher listed in an imprint statement and a specific year it was published. These elements make up what is called bibliographic control. There are standards setting agencies that think very carefully about the way all of these elements are going to be presented and how you would search for that element. For example, for organisations there will be clear guidelines about how to handle names that change and merge.

Many repositories of grey literature, especially individual organisation websites, can lack this level of bibliographic control and may be oblivious to the need to deal with the possibility of the demise of the organisation that has created the item. When looking exhaustively for grey literature you may need to search for various forms of an organisation’s name and to be broader in your application of keywords or subject headings. Search engines may index the whole text of documents with relevancy ranking of results but may not have functionality to for terms in titles or summaries of the document.

For journal articles there are tools like the subject headings of Medline which are applied to articles from over 8000 different journals on a consistent basis. When looking for grey literature, each organisation (equivalent to each journal) can create their own set of subjects or have no subject indexing at all. The World Health Organisation IRIS site, Analysis and Policy Observatory and Victorian Department of Health website may all contain different terms for the concept of workforce for example. One may have health personnel while another refers to labor force. You do encounter this when searching the white literature but it is more a problem with grey literature as there are so many potential groups creating indexing for their sites.

2. Permanency : Items Disappear

Many grey literature items are presented on websites as essential material for a topic of current importance for the organisation or its users. For example, there are data sets being presented about the Covid-19 issue on many websites. Once the Covid-19 crisis dissipates many of those websites and data sets will also dissipate or be archived in other locations, if archived at all, as their requirement is not for preservation.

Sites will often change their file naming structure. This leads to 404 Not found results, a common occurrence when looking for grey literature references. I once was even unable to locate the results of a Victorian Royal Commission held by the Public Records Office, a
group that should be excellent at maintaining long term links to resources. You sometimes need to be quite creative in performing searches for older reports on specific websites and may need to contact the site support to track down files. In the case of the PRO, the site support was able to identify where the files had been moved to within their organisation. A good strategy when you cannot locate a report is to use archiving websites like the National Library of Australia’s Australian Web Archive https://trove.nla.gov.au/website?q= or the Wayback Machine on archive.org https://archive.org/web/web.php which show how websites looked at a particular point in time.

When collecting grey literature take care to obtain copies of the items you wish to include in your research at the time you read or view them so that you can protect against their future disappearance.

3. Availability : Items Not Sold

If we look at books, cars, flights or hotel rooms the property of them being saleable requires you to have a catalogue or a search engine with standardised terminology for buyers and sellers to meet in the marketplace. Examples of standardisation include barcodes, document numbers, airport codes, hotel codes and car model. Although it may look simple standardising record entry in this way and maintaining it consistently can be an expensive undertaking. With grey literature the item is often not for sale, just freely available on a website and therefore the producing organisations may not see a commercial requirement to have standardised terminology or filters for retrieving the items produced on a similar topic. Not being able to filter consistently makes it much more difficult to know that you have examined the resources of a website completely. It also impacts on the number of access points available which affects findability. For non electronic items, this also means that the literature is often available in small quantities such as having just one copy of a thesis in one University. Records may also only be found in search engines that are restricted to users from particular organisations, such as consultancy reports.
Search Strategy

Given what we know about grey literature properties, what do you need to do when developing a search strategy for grey literature.

First, think carefully about where the grey literature may be for the topic. There is often little to be gained from just looking everywhere and it is a large task. It is usually better to conduct a targeted search. You may use Google or another internet search engine to explore ideas and terms and gather easy to find sources.

Identify places to search for grey literature from journal articles on the topic. See which organisations and report sources they cite. Also, it can be important to talk to topic experts to discover their knowledge of organisations and repositories where reports and statistics reside. You can certainly try using Google Scholar and general Google with a variety of keywords.

To identify indexing when you’re using a search engine, perform a basic search on a key word relating to your topic, then look at some results and what level of indexing they have and what specific terms have been used for your topic. This gives you a good idea of how complex a query you will be able to perform using the search engine.

With your strategy consider two things.

1. **A one line search strategy**: You will often be only able to search on one line on search engines covering grey literature, especially with organisation websites, so you need to distil what you may have searched on Medline or Web of Science in a 20 line search into just one line. You may want to target very specifically the key term or terms relating to the topic you are interested in or you may use a mixture of terms but again using very specific phrases or limited to specific geographic areas.

   Many grey literature search engines will be searching across the full text of the reports and other documents stored on the site. They will use relevancy ranking to some degree to show you the best items first. It will probably not be necessary to scroll through all of the items from the result lists presented for grey literature searches but you should document what you did in terms of looking through result sets as you may need to report this in your methodology.

2. **Use site contacts**: When using a grey literature search engine or a website search engine, if it is a particularly important website for your topic you may want to contact the people who manage the search engine and discuss your requirements with them.

   They will be aware of the best way to gather materials out of their search engine. Usually the people involved in building or maintaining a search engine for a particular organisation or area of interest will be quite interested in talking to researchers who are using their search engine to perform that research. They may have other tools other than just the standard keyword search engine to scan all of the resources the website has available. They may also be able to supply you with spreadsheets produced by a report generator that is available from the back end of the website.