Ch11 Sources

**Key steps in finding data**

Think about it first

* What would the data contain? What variables might there be? What would they be called? What you call “spending” might be “expenditure” or “total amount”.
* Bear in mind that whenever you fill in a form online – such as registering your vehicle, or applying for a new passport, you are adding to somebody’s database. One day you may want to interrogate that database en masse – so make a note of who owns it, and what the input terms are – the language used on a web form is likely to become the names of the variables in the database itself.
* Who would gather and keep the kind of data you want? A government ministry? A national statistical office? A watchdog? An international organisation? The answer to this question will help you as you start to look for it.
* How can you get hold of the data? A lot of data is held as files on web portals for you to download, but an increasing number of datasets are held in databases which you need to interrogate via a webpage in order to generate a file you can download and work with.
* In the UK the [Electoral Commission](http://search.electoralcommission.org.uk/?currentPage=0&rows=10&sort=AcceptedDate&order=desc&tab=1&et=pp&et=ppm&et=tp&et=perpar&et=rd&isIrishSourceYes=true&isIrishSourceNo=true&prePoll=false&postPoll=true&register=gb&register=ni&register=none&optCols=Register&optCols=CampaigningName&optCols=AccountingUnitsAsCentralParty&optCols=IsSponsorship&optCols=IsIrishSource&optCols=RegulatedDoneeType&optCols=CompanyRegistrationNumber&optCols=Postcode&optCols=NatureOfDonation&optCols=PurposeOfVisit&optCols=DonationAction&optCols=ReportedDate&optCols=IsReportedPrePoll&optCols=ReportingPeriodName&optCols=IsBequest&optCols=IsAggregation) and the [Land Registry](https://landregistry.data.gov.uk/app/ppd/) are examples of sites where you “ask” for the files. Likewise, in the USA there are site like [USASpending.gov](https://www.usaspending.gov/search) which use the same interface system. In looking for them, this kind of site is not going to be findable using the “filetype:xls” google operator – you should try using a search string like “site:gov.uk intext:download” (along with a suitable search term, of course).
* For a complete guide to Google Operators – go to [googleguide.com](http://www.googleguide.com/advanced_operators_reference.html)

If you can’t find the data anywhere, maybe you should try building your own database – you could create a google or Microsoft form, and invite your audience to supply details. When The Guardian wanted to track deaths at the hands of the police in the USA, as well as using open datasets, they asked their readers to send them details of killings which hadn’t made it into the papers. [The Counted](https://www.theguardian.com/us-news/series/counted-us-police-killings) was the resulting project.