PIECES ON OUR CRAFT

Data Collection Online: Techniques and Traps When Searching for (Treaty-Related) Data on the WWW

LEIF HOFFMANN
University of Oregon

Much of the recent literature on the Internet as a research tool has focused on negative effects on student papers and on the creation of quality criteria for Internet resources. Less attention has been paid to how instructors can assist their students in looking for and retrieving data for class assignments requiring Internet searches. As a by-product of my work on the International Environmental Agreements Database project (http://iea.uoregon.edu), I present here a list of techniques on how to find valuable raw data on websites and how to avoid common traps. The underlying premise is that, although the Internet is home to "a great deal of useless information," a targeted search can be a fairly efficient way of determining data availability and retrieving data for undergraduates and scholars alike. The guidelines are organized into five categories: preliminary steps, potential sources and indicators, organization-related issues, language, and technical concerns.

Preliminary Steps

- *Specify your research topic.*
- *Specify what kind of data you are seeking.*
- *(Re)read recent articles or books related to your research topic to pinpoint experts.* Start by reading their published work, and perhaps email them to tap their specialist knowledge.
- *Use Dissertation Abstracts databases, which may contain work referencing data sources.*
- *Use librarians.* Librarians are trained to locate information. They can direct you toward databanks and institutions that may hold relevant data. University libraries often provide access to subscription databases that provide search tools for data (or the data itself) not available to the general public.
- *Identify institutions.* Appropriate institutions will vary according to your research interests. Some institutions have a primary mission of collecting data, for example, U.S. Census Bureau or Eurostat. But data are also often gathered by government departments, non-governmental organizations, or individual researchers.

- *Check out the institutions' websites.* Even if data is not directly available, the website may indicate what types of data they collect. This may also lead to a "snowballing" search, in which one institution indicates or links to other institutions with the data you need. Always scan an organization's website carefully before contacting it for data. This will reduce frustrations on both sides. You will be better able to specify what you are looking for from them, and they will be better inclined toward a well-informed request.

Potential Sources and Indicators

- *Scan for key words* like data, statistics, compliance, *etc*. They may lead straight to data or indicate where relevant data have been collected.
- *Read reports accessible online*. Although annual reports often serve publicity purposes and may not disseminate data themselves, they can be helpful in clarifying that such data exist and where it can be obtained, for example, by carefully reading the source information for tables. Meeting and technical reports can be similarly useful. Don't read too much, however, into categories or titles; these don't necessarily bear out in the data contents of a report, and sometimes a blandly-labeled meeting report is a better starting point and source for data than a technical report. There is unfortunately no shortcut past reading the reports.
- *Be aware that you might end up with something other than what you think you have found.* My experience in researching (more recent) treaty agreements has been that the first databases established by treaty-related secretariats were in general not indicator databases, but rather generic databases cataloguing legislative acts, experts, and research institutions. They may still turn out to be useful—and may point to potential contacts for further steps—but, like with reports, the contents of databases are not always what you hope from the label.
- *Institutional websites with online forms for member states or individuals often have good data.* Although the data might not be directly available online, the form itself indicates that data exist. Treaty provisions that require data collection also are good indicators that such information has been collected.
- *If data have been collected for a narrow timeframe, parallel historical data usually exist.* Note the source and inquire directly for corresponding data.
- *Institutional documentation centers and libraries may—but don't always—have useful data.* Some websites list publications related to the institution's mandate but are written by other organizations and authors. Moreover, documents listed in an institution's online library search engine may only be accessible locally, requiring a personal visit or a direct inquiry to ascertain what the documents actually contain.

Organization-Related Issues

- *Data usually is spread out across different reports and electronic formats, requiring considerable time to assemble.* Before wasting time entering data manually, inquire whether you can obtain the discovered data directly from the institution in electronic format.
- *Identified data may not be centrally stocked.* Often data are collected and remain with agencies and departments in different signatory countries. Be sure to check the national websites of the government agencies or institutions that

might have provided the data to an international organization or its secretariat. Online links, however, might not be provided.

- *Responsibilities for data collection and assessment vary significantly by country.* Analogous data may be gathered in one country by a ministry, in another by a university, and in another by multiple agencies. Avoid the trap of equating a known institutional setting with the institutional setting in an unfamiliar country.
- *Be aware that data might exist but not directly with the identified institution.* Several sub- or sister organizations may make up and carry out responsibilities for an agreement. They may collect data for different issue areas, or one organization may collect data while others carry out other tasks. Moreover, institutions may neither collect data themselves nor rely on members (states, firms, or others) to do so, but instead outsource data collection to umbrella organizations or specialized private companies.
- *Always explore institutional organization charts if available.* They may provide information on whether branches exist which collect or store data.
- *Consider when the institution has been established.* Data might not be (easily) available before the starting date of the organization. Yet in some circumstances data for prior years might exist in the agencies or ministries of member states.
- *Always be ready to rethink the indicators and data you are looking for.*

Language

- *Websites frequently exist in several languages and not all versions contain the same information.* In non-English-speaking countries, English versions are frequently abridged. In these cases it becomes crucial to explore the foreign-language mirror sites, personal language skills permitting. Even if English is the website's main language, reports may be written in another language.
- *Ask yourself what search word(s) you want to employ when using a search engine integrated into an institution's website and establish a list of obvious and less obvious key words.* Am I looking for data or *datos, données, Daten*, etc.? Even in English relevant information can be stored under different names and categories. Does my understanding match that of the people who created the engine and the texts?
- *Record what words produce no hits.* This way when either sharing with other researchers or when contacting somebody at an institution you can pass on what didn't work, and note that data is organized in non-obvious ways.

Technical Concerns

- *Check out a website's site map.* It lets you ascertain that nothing major has been overlooked.
- *Keep server problems in mind.* Infrastructures vary and the server can temporarily be down. Access to the entire or parts of a website might sometimes be gained via other avenues, for instance indirectly via a search engine such as Google or Yahoo. Other times you must simply try again the next day.
- *Be mindful that website addresses and contents change.*
- *Save the information.* To make sure that you don't save a generic website address, open a link in a new window for the correct URL.
- *Create a document's directory list.* For example, use the columns in an Excel file to store information like the type of document, the website address, the name of the stored file, a short description, an evaluation of potential usefulness and a column for potential follow-ups.