

Primary vs. Secondary Research Data

Research can feel confusing, but it becomes much easier once you understand the two main types of data you'll work with: primary research data and secondary research data. Think of them as **collecting self-generated information** versus **using what's already out there**.

Primary Research Data

Primary research data = you collect the data yourself.

This is *original, first-hand* information created directly by the researcher to answer a specific question or test a hypothesis.

In simple terms: You are the one going out and generating fresh data. Nothing exists until *you* collect it.

Examples:

- Asking people questions in **interviews** or **surveys**
- Running your own **focus groups**
- **Observing** behaviour in a natural setting
- Doing your own **experiments**

Definitions:

- Primary data is **unique, first-hand information** collected from the source by the researcher using methods like interviews, focus groups, surveys, and observations. [\[open.edu\]](https://open.edu)

When might you use primary research:

- When you can't find existing information that answers your question
 - When you want **up-to-date, tailored, or specific** insights
 - When your project requires new evidence, insight or original findings
-

Secondary Research Data

Secondary research data = using information someone else has already collected.

This means analysing or reviewing *pre-existing* data or published sources.

In simple terms: You're gathering information that's already been produced by someone else - like reading, searching, and selecting useful evidence.

Examples:

- Academic journal articles and other scholarly writing (e.g. books)
- Government statistics (like census data)
- Organisational records (e.g. medical records)
- Market research reports

- Surveys or interview data conducted by someone else
- Social media content (e.g., user comments or posts)
- Digital/web content (e.g., blog, online forums, Youtube video)
- Literary works (e.g., academic texts, auto/biographies, books, diaries, letters, novels)
- Broadcast material and periodicals (e.g., printed or online news, magazines, zines)
- Archival documents: unpublished/published material in public/private/digital archives or libraries (e.g., historical manuscripts, photographs, film footage, correspondence, pre-existing oral histories, adverts)

Definitions:

- Secondary research uses data **already in existence**.
- It can be raw data such as databases, censuses, pre-existing oral histories and social or economic data **which is not possible for an individual themselves to collect**; [\[open.edu\]](https://open.edu)
- The Open University describes secondary data as information previously gathered, and either **being used for a purpose other than that for which it was originally collected** or by **someone other than the researcher who collected the original data**. [\[open.edu\]](https://open.edu)

When should you use secondary research data:

- At the **start** of any project—it can be more time-efficient, cost-effective, instantaneous and helps you understand what's already known
- To provide background, define your topic, and identify gaps
- When you need large datasets or broad context that you could not collect yourself

Important considerations related to the ethical use of secondary data:

- Access to *some* secondary data may require specific permissions or registration (e.g., registering for a free reader pass at the British Library).
- Some unpublished archival documents fall under strict copyright rules (e.g., the 2039 rule) and therefore may require clearance.
- It is the researcher's responsibility to check privacy notices/policies (e.g., a social media platform's privacy policy and user agreement) and/or clear copyright before analysing, reproducing or distributing materials.
- Public/private domain distinction: Do not assume that what appears to be public domain data is always free to be used for research, as researchers should always consider the ethical challenges associated with determining whether data gathered from digital spaces (e.g. social media or forums) is publicly viewable or private ([British Psychological Society, 2021](https://www.bps.org.uk/ethics)). Any implications for anonymity, confidentiality, consent to access and data security should also be considered.

- As with primary research, secondary research can involve a researcher potentially encountering sensitive, risky or traumatising material. Before engaging in such a project, a researcher should consider their own risk factors, consult best practice guidelines (e.g. [Association of Internet Researchers' Risky Research Guide, 2025](#)), and model healthy practices to mitigate the emotional/psychological challenges associated with any data analysis.

Primary vs. Secondary Research: Quick Comparison

Feature	Primary Research	Secondary Research
Who collects the data?	You do	Someone else
What kind of data?	New, original	Existing
Control	High - you design the method	Lower - you rely on what's available
Best for	Specific questions needing new insights	Background knowledge, identifying gaps

(Comparison descriptions of data types.) [\[studocu.com\]](#), [\[nu.edu\]](#), [\[open.edu\]](#)

Final Tips

Think of **secondary research** as *reading what's already written*, and **primary research** as *actively generating the evidence yourself*. Both matter and have distinct ethical considerations, and using them together can make your work stronger, clearer, and more original.

Primary and secondary **sources** are different to primary and secondary data. Sources can be categorised as primary, secondary or tertiary depending on whether or not they are 'original' materials, however the distinction often depends on the context and subject area. The Library Skills for Learning page has more information on [Primary and Secondary sources](#).