BA-Eindwerkstuk Seminar

Week 2: Data and Sources – Generation, Collection, and FAIR Principles

Dr. Steven Denney

Korean Studies
Leiden University

February 13, 2026

Review from Week 1

Exercise #1: Share Your Question

From last week's homework:

1. What is your research question (or top candidate)?
2. What is the motivation / problem / gap?

Discussion

Let's hear from a few of you. We'll revisit these questions at the end of today's class—they may shift once you think about data.

What Makes a Good Question – Revisited

Recall the five characteristics:

1. **Specific** — Clearly defined scope and boundaries
2. **Open-ended** — Cannot be answered with yes/no
3. **Researchable** — Can be investigated with available evidence
4. **Relevant** — Contributes to scholarly understanding
5. **Personal** — Connects to your genuine interests

Today we focus especially on #3: **researchable**. A brilliant question without the right data is unanswerable.

Questions to Avoid

Problematic Question Types

- **Leading** — Assumes the answer before research begins
- **Loaded** — Contains hidden value judgments
- **Yes/No** — Too narrow; closes off inquiry
- **Overly causal** — Often too ambitious for a BA thesis
- **Unfocused** — Too broad to research in 10,000 words

Diagnosing Bad Questions

Problem	Bad Example	Better Version
Leading	"Why has South Korea's education system failed its students?"	"How do South Korean university students evaluate the pressures of the national college entrance system?"
Loaded	"Why is K-pop so much better than Western pop music?"	"In what ways do K-pop production strategies differ from those used in the U.S. music industry?"
Yes/No	"Did the Sewol Ferry disaster change Korean politics?"	"How did civic responses to the Sewol Ferry disaster reshape activist strategies in South Korea?"
Causal	"How did Confucianism cause Korea's economic development?"	"What role do Confucian values play in South Korean corporate governance practices?"
Unfocused	"What is the Korean Wave?"	"How do Korean cosmetics brands use social media influencers to build consumer trust in Southeast Asia?"

From Question to Data

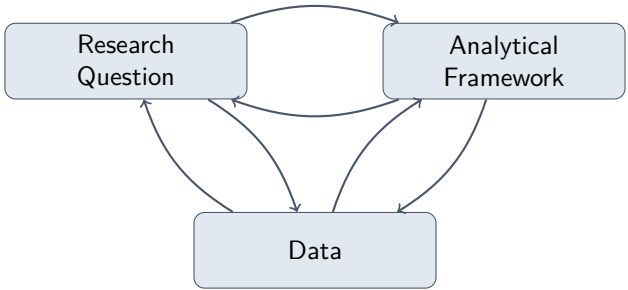
Today's Agenda

1. Review of Week 1 and research questions
2. The question–data relationship
3. Data collection and sources
4. Using Korean-language sources
5. FAIR principles for thinking about data
6. Workshop: Mapping your data

Today's Focus

What is your data, where does it come from, and how will you get it? We set aside analytics for now and concentrate on the *foundations* of your empirics.

The Iterative Research Process



You will move *iteratively* between question, framework, and data. This is normal. Having a somewhat **data-driven question** is not a bad thing—it's often how real research works.

Good Questions Need the Right Data

The Core Challenge

A good question without the right kind of data cannot be answered. A question must be answerable *in practice*, not just in theory.

Ask yourself:

- What evidence would I need to answer this question?
- Does that evidence actually exist?
- Can I realistically access it within the timeframe of a BA thesis?
- If not—how should I adjust my question?

Data Collection & Sources

Primary vs. Secondary Sources

Primary sources

- Original, first-hand materials from the time of an event or topic
- Created by participants or direct observers
- *Examples:* government documents, newspaper articles, interviews, diaries, speeches, legal texts

Secondary sources

- Analyses or interpretations of primary sources
- Created after the fact by other scholars
- *Examples:* academic journal articles, textbooks, book reviews, historiographies

Note

Your thesis needs **both**. Secondary sources frame your literature review; primary sources form your empirical evidence.

Qualitative vs. Quantitative Data

Qualitative data

- Descriptive, non-numerical
- Captures experiences, meanings, context
- Answers “how” or “why” questions
- *Examples:* interview transcripts, ethnographic notes, media texts, archival documents

Quantitative data

- Numerical, measurable
- Can be counted or statistically analyzed
- Answers “what,” “how many,” “how much”
- *Examples:* survey results, statistical datasets, frequency counts, polling data

Most BA theses in Korean Studies rely primarily on **qualitative data**, but some mix both approaches.

Types of Data – More Common in BAKS

Newspapers & Digital Media (Written)

- *Best for:* Tracking political discourse, media framing, public opinion over time
- *Examples:* Chosun Ilbo vs. Hankyoreh coverage of North Korea, Naver News comments on political events
- *Challenges:* Bias in sources, selecting a representative sample, data retrieval

Digital Media (YouTube, Blogs, Social Media)

- *Best for:* Analyzing culture, online activism, public engagement
- *Examples:* YouTube videos on K-pop marketing or North Korean propaganda, Naver blogs on historical memory
- *Challenges:* Platform restrictions, data ethics, fast-changing content

Types of Data – Less Common in BAKS

Interviews

- *Best for:* Studying personal experiences, expert opinions, community perspectives
- *Examples:* Interviews with activists on democratization, film directors on Korean cinema, Zainichi Koreans on identity
- *Challenges:* Language proficiency, access to participants, ethical considerations

Archival Research

- *Best for:* Historical studies, government policy analysis, institutional research
- *Examples:* National Archives records on Cold War policies, colonial-era legal documents
- *Challenges:* Misperception of what an archive is today (e.g., Naver, Wilson Center Digital Archive); language (e.g., Hanja)

Using Korean-Language Sources

Authenticity and depth: Korean-language sources provide local/Korean-based perspectives. They are of *particular importance* and should play a larger role in Korean Studies scholarship.

Avoiding bias: Relying only on English sources can skew research. Using Korean sources helps balance this and fill in gaps.

Academic expectation: Students and scholars of Korean Studies are expected to engage with Korean-language materials.

Practical Tip

Start identifying Korean-language sources *now*. Don't wait until the writing stage—they take more time to locate, read, and integrate.

Break

10 minutes

Think About Your Sources

Option A: Build a corpus

- Can you gather, collect, scrape, or locate your sources in a structured way?
- If so—do it! Build a corpus for computational text / DH analysis

Option B: Close reading

- You don't have to go the "big data" route
- You can read 12 sources closely and do excellent research
- What matters is that your material is **manageable** and **organized**

Workshop: What Is Your Data?

Mapping Your Data (10 minutes)

Exercise: For your research question, answer the following:

1. **What is your data?**
   - What type of evidence do you need? (texts, interviews, statistics, archival documents, media content...)
2. **Where does it come from?**
   - What is the source? (a specific newspaper, database, archive, organization...)
3. **How will you get it?**
   - Is it freely accessible online? Do you need institutional access? Do you need to generate it yourself (e.g., interviews)?
4. **How will you organize it?**
   - How will you store, label, and keep track of your materials?

If you can't answer these questions, your research question may need adjusting.

Discussion: Does Your Question Still Work?

After mapping your data, consider:

- Does the data you need actually exist and can you access it?
- Is the scope realistic for a 10,000-word thesis?
- Do you need to **narrow** your question to match available data?
- Or do you need to **broaden** it because the data opens new possibilities?

Remember

Adjusting your question in response to data availability is not a weakness—it is good research practice. The question and the data evolve together.

FAIR Principles

Why Think Structurally About Data?

Once you know *what* your data is, you need a framework for managing it well.

The **FAIR principles** provide a way to think structurally about your data—not just for others, but for *yourself*.

FAIR = Findable, Accessible, Interoperable, Reusable

Originally developed for research data management in the sciences, but the principles apply equally to humanities research.

Findable & Accessible

Findable: Your data should be easy to find—by others *and* by yourself.

- Use clear identifiers and rich descriptions
- Assign meaningful file names, not `data_final_v3_REAL.docx`
- Provide metadata: keywords, author, date, source

Accessible: Once found, data should be obtainable under well-defined conditions.

- Store data in a reliable location (cloud backup, repository)
- Note access conditions: Is it public? Restricted? Behind a paywall?
- Even if data can't be fully open, information on *how* to access it should be clear

Interoperable & Reusable

Interoperable: Data should be in understandable formats that allow it to be combined with other data.

- Use common, open file formats (e.g., CSV, plain text, UTF-8 encoding for Korean)
- Use standardized data schemas where possible
- Think about whether someone else could combine your data with theirs

Reusable: Data should be well-documented and licensed for reuse.

- Provide context: methods, definitions, collection dates
- Proper documentation ensures that *you in a few months* can still understand your own data
- Choose open licenses or sharing agreements where appropriate

FAIR in Practice: Your Thesis

What does FAIR look like for a BA thesis?

Principle	What You Can Do
Findable	Organize files with clear names and folders
Accessible	Back up your data; note where it came from
Interoperable	Use standard formats (UTF-8, CSV, PDF)
Reusable	Document your collection methods in your thesis

Further Reading

`https://en.wikipedia.org/wiki/FAIR_data`
`https://www.library.universiteitleiden.nl/researchers/data-management/fair-data`

Where It All Fits

Aligning Question, Data, and Methods

The lesson of Mullaney and Rea Chapter 3: a clear question and a matching method are the twin pillars of success.

What you should show your reader (especially for Assignment #1):

1. Here is my **question**
2. Here's **why it is important** (connecting it to the problem and gap in knowledge—and the literature to which it belongs)
3. Here's what **data** I will collect to answer my question
4. Here's **how** I will collect and analyze that data (*methods / analysis*)

Where Data Fits in the Thesis

1. **Introduction** (question, motivation, data & methods)
2. **Literature Review** (situating the question; exploring methods, theories, etc.)
3. **Methodology / Analytical Framework**
   - What data you collected, how you collected it, how you stored it (FAIR comes in here), and how—in as precise language as possible—you analyzed it
   - This should state clearly **how** you will answer your question
4. **Findings**
5. **Discussion and Conclusion**

See the Thesis Protocol on Brightspace for more information.

Exercise & Next Steps

Exercise #2: Draft Your Data Collection Plan

Write a short plan (bring to Week 3) that addresses:

1. **Your research question** (revised if needed after today)
2. **Type of data** you intend to collect
3. **Source(s)** — Where does the data come from?
4. **Access** — How will you obtain it?
5. **Organization** — How will you store and manage it?
6. **FAIR alignment** — How does your plan align with the FAIR principles (to the extent possible at this stage)?

This is exploratory. The goal is to start thinking concretely about your empirical foundation.

Looking Ahead

Week 3: Conducting a Literature Review

- Purpose and structure of a literature review
- How to connect your lit review to your question, gap, and framework
- Strategies for organizing and synthesizing literature

Key Dates:

- Assignment #1 due: March 13, 2026
- Start thinking about your supervisor and schedule a meeting

Course Website

`https://scdenney.github.io/baks_thesis-seminar/`