Data Gathering and Data Sources

Matthias Hügel

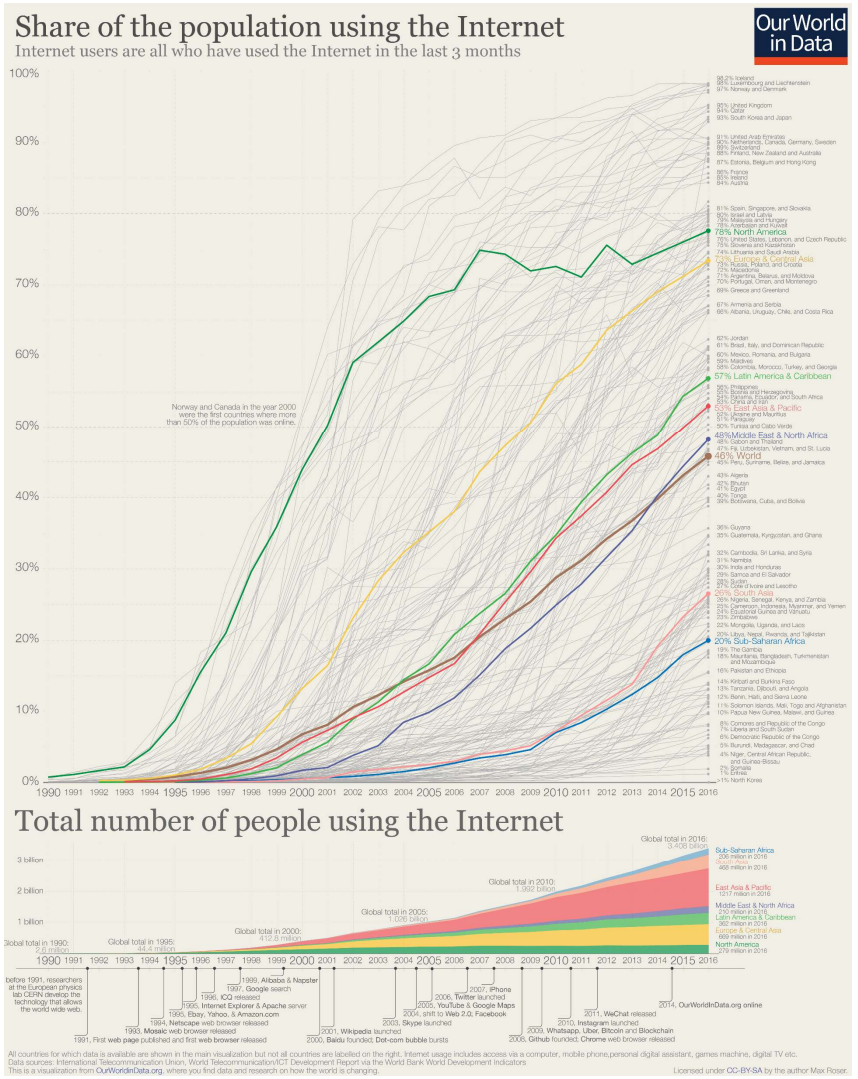
Chair of Microeconomics

Friedrich Schiller University Jena

“In God we trust, all others bring data.”

W. Edwards Deming (1900 – 1993)

American engineer and statistician



Data gathering

What does it mean?

- Process of collecting and measuring information on variables of interest in an established systematic fashion

For what purpose?

- To answer research questions, test hypotheses, evaluate outcomes

What is the overarching goal?

- To capture quality evidence that then translates to rich data analysis and allows the building of a convincing and credible answer to questions

Outline

Part I	**Types of Data**
Part II	**Methods of Data Gathering**
Part III	**Focus: Survey**
Part IV	**Group work until workshop 3**

Outline

Part I	**Types of Data**
Part II	**Methods of Data Gathering**
Part III	**Focus: Survey**
Part IV	**Group work until workshop 3**

1. Qualitative vs. quantitative data
2. Primary vs. secondary data

Comparison: Quantitative vs. qualitative data

	Qualitative data	Quantitative data
Definition	Non-numerical and descriptive data	Numerical and can be mathematically computed
Codification	Words and sentences	Numerical values in different scales
Aim	Describes certain attributes; to understand the 'why' and the 'how' behind certain behaviour	Tells us how many, how much or how often
Gathering	Through observations and interviews	By measuring and counting things
Analysis	Grouping data into meaningful themes or categories	Using statistical analysis

→ Depends on the aim, what kind of question should be answered with the data

Examples

How many companies in the Free State belong to the agricultural industry?

→ Quantitative data

What challenges are companies in the agricultural industry currently facing in the Free State?

→ Qualitative data

Comparison: Primary vs. secondary data

	Primary data	Secondary data
Definition	First hand data collected by researchers directly from main sources	Already collected data (through primary sources) and made readily available
Timestamp	Real-time data	Past data
Process	Very involved in data collection process	Quick and easy to collect
Availability	Crude form	Refined form
Sources	Questionnaires, experiments, interviews, …	Websites (databases), publications, …
Specific	Always specific to the researcher's needs	Depends solely on the kind of secondary data the researcher was able to gather

	Advantages	Disadvantages
Primary data	• Data collected specific to the problem under study • Up to date information • Unique data	• Costs of gathering • Time-consuming
Secondary data	• (Mostly) cheap • Quick access	• Might be outdated • Might be irrelevant

→ Depends on the research question that is intended to be answered
→ Often the combination of both is the optimal way

Examples of secondary data sources

International level

OECD Database:

https://data.oecd.org/

National level

Department Statistics of South Africa:

http://www.statssa.gov.za/

Cluster-related

German cluster platform:

https://www.clusterplattform.de/CLUSTER/Navigation/EN/Home/home.html

Outline

Part I	Types of Data
Part II	**Methods of Data Gathering**
Part III	Focus: Survey
Part IV	Group work until workshop 3

1. Questionnaire
2. Interview

Gathering primary data I

Questionnaire to obtain quantitative data

- Questions differ in their objective
  - Questions measuring seperate variables (e.g. preference for political party)
  - Questions aggregated to a scale or index (e.g. personality traits)
- Qustions differ in their end
  - Open-ended vs. closed-ended questions
- Questions differ in their scale
  - Dichotomous, nominal, ordinal, continuous

Gathering primary data II

Interview to obtain qualitative data

- Interviews differ regarding their structure
  - Structured interviews
  - Semi-structured interviews
  - Unstructured interviews
- Depth of detail
- Process
  - During interview: Recording
  - After interview: transcription, evaluation and categorizing the answers

Gathering primary data III

Other methods of data gathering

- Experiments
- Observations
- Case studies
- …

Examples of the Cluster-Literature

Primary data / Quantitative data / Questionaire data

- Bell, G. (2005): Cluster, networks, and firm innovativeness. Strategic Management Journal, vol. 26, pp. 287-295.

Primary data / Qualitative data / Interview data

- Giuliani, E. & Bell, M. (2005): The micro-determinants of meso-level learning and innovation: Evidence from a Chilean wine cluster. Research Policy, vol. 34, pp. 47-68.

Secondary data / Quantitative data / Focus: Biotechnology sector U.S.

- McCann, B. T. & Folta, T. B. (2011): Performance differentials within geographic clusters. Journal of Business Venturing, vol. 26, pp. 104-123.

Secondary data / Quantitative data / Focus: R&D active firms in Germany

- Grashof, N. (2021): Spill over or Spill out? – A multilevel analysis of the cluster and firm performance relationship. Industry and Innovation, vol. 28, pp. 1298-1331.

Outline

Part I	Types of Data
Part II	Methods of Data Gathering
Part III	**Focus: Survey**
Part IV	Group work until workshop 3

1. Basic steps in the survey process
2. Survey types
3. Designing questionnaire items

Survey = Questionnaire ?

- Questionnaire: any written set of questions
- Survey: both the set of questions and process of collecting, aggregating, and analyzing the responses

	Survey	Questionnaire
Individual questions	✓	✓
Delivery of questionnaire	✓	✗
Analysis of responses	✓	✗

Source: SurveyMonkey

Basic steps in the survey process

1. Define objectives
   - Determine what you want to know and why.
   - Think about who will look at the results.
2. Define the population and choose a sampling frame
   - Look for an existing sampling frame or create a sampling frame.
   - Consider probability and nonprobability sampling strategies.
3. Design a data collection strategy
   - Evaluate time and budget constraints.
   - Estimate available resources.
4. **Develop a questionnaire**
   - **Write the questions.**
   - Pretest the questionnaire.

Basic steps in the survey process

5. Collect data
   - Monitor responses.
   - Employ follow-ups as needed.
6. Manage the data
   - Create a codebook.
   - Export the data.
   - Clean the data.
   - Transform the data.
7. Analyze the data
8. Disseminate the results (for example, write a report or give an oral presentation)

Survey types

- Different types of surveys possible
  - Mail
  - Telephone
  - Face-2-Face interview
  - Online
- With different advantages and disadvantages; for a detailed distinction see Sue & Ritter (2007), chapter 1

Designing questionnaire items: Validity

- Put simply, valid questions measure what they are supposed to measure

Example

You want to know about respondents' magazine reading habits.

No valid question: *How many magazine subscriptions do you have?*
Purchasing a magazine subscription is not a valid indicator of magazine *reading*. Many people purchase magazines and never read them, while others read magazines without ever paying for them.

Valid question: *How many hours did you spend reading magazines last week?*

Threats to validity: Social desirability

- Respondents to give the “right” answer rather than the real or valid answer to a survey question to conform to social norms
- Do you vote? Do you go to museums? Do you do volunteer work? Do you give money to charities? → high risk of biased answers
- Ways to reduce social desirability bias
  - Highlight anonymity and confidentiality

  “Remember all of your responses are anonymous and will be kept confidential.”
  - Employ face-saving strategies (permission to behave socially unaceptable)

  *“Everyone gets angry now and then. How many times last week did you find yourself getting angry**?”***
  - State that the behavior you are asking about is not unusual

  “A recent study found that 80% of college students have cheated on an exam. Have you ever cheated on an exam?”

Threats to validity: Inaccurate Estimates

- Respondents rather estimate information about past behavior or events than be able to give a precise value
- Key to valid measurement of factual information is to ask respondents focused questions covering a limited range of time and situations
- Ways to improve accuracy:
  - Ask about specific behavior within a limited, recent time period

  Poor: "How many miles have you driven since you received your driver's license?"
  Better: "How many miles did you drive last week?"

  - Ask respondents to think about a specific event rather than a category of events

  Poor: "In an average week, how much do you spend on groceries?"
  Better: "How much did you spend on groceries last week?"

  - Ask respondents about their own behavior, not the behavior of others

Threats to validity: Nonattitudes

- Not every issue that is important to you will be important to the participants in your survey
- Responses despite no opinion
  - Do not want to admit to being uninformed (→ social desirability bias)
  - Feel the need to "help" researcher by completing all of the questionnaire items
- Ways to prevent uninformed answers:
  - Make it socially acceptable for respondents to say they are unfamiliar with the topic

  "Some people are interested in politics and some are not . . . would you say you are interested in national politics?"

  - Use filter questions
  - Provide an explicit "no opinion" choice as a response option

Important additional readings

- Sue & Ritter (2007), Chapter 3 „Sampling“
- Sue & Ritter (2007), Chapter 5 „Designing and Developing the Survey Instrument“
- Sue & Ritter (2007), Chapter 6 „Conducting the Survey“
- Sue & Ritter (2007), Chapter 7 „Processing and Analyzing the Survey Data“

- What data do you think is important in managing a cluster, and in particular, what data is important in managing an Agriculture cluster in the Free State?

- How might these be collected (primary/secondary data)?

Outline

Part I	Types of Data
Part II	Methods of Data Gathering
Part III	Focus: Survey
Part IV	**Group work until workshop 3**

Development of a questionnaire

- Focus on primary data (we will work with secondary data during Workshop 3)
- Main objective: Development of a questionnaire
- Groupwork: Division into 4 groups
- Target group of questionnaire: Actors within a potential agriculture-cluster within the Free State
- Presentation of each developed questionnaire at Workshop 3

What should the questionnaire contain?

Section	Content of section	Exemplary item of section	Implemented by:
A	Socio-demographic information and business features	Age, Gender, Company position, … Industry, Size of company, …	All groups
B	Performance measures	Profit, Growth rate, Innovations, …	All groups
C	Innovation process	Relevance of external knowledge, Relevance of scientific knowledge, …	All groups
D	Network		
D1	Value-chain	Supplier, Customer, …	Group 1
D2	Technological knowledge	Information sources regarding new technologies, ….	Group 2
D3	Market knowledge	Information sources regarding market situation, …	Group 3
D4	Cooperation	Ressource sharing, learning, ….	Group 4

What should be done by when?

Time	Task
Today	Divide into the 4 groups by yourselves
Until Workshop 3	Develop your questionnaire
Until Workshop 3	Upload your questionnaire to the cloud
During Workshop 3	Present your questionnaire

References

- Bryman, A. (2005): Research Methods and Organization Studies. London & New York: Routledge.
- Kabir, S. M. (2016): Methods of Data Collection, Chapter 9, pp. 201-275.
- Sue, V. M. & Ritter, L. A. (2007): Conducting online surveys. Los Angeles: Sage.
- Yilmaz, K. (2013): Comparison of Quantitative and Qualitative Research Traditions: epistemological, theoretical, and methodological differences. European Journal of Education, Vol. 48, No. 2, pp. 311-325.