My experiences with reproducible research

NORMAN BACKHAUS
NORMAN.BACKHAUS@GEO.UZH.CH
Defining reproducibility

• Technically not really possible since every data collection is unique to a certain extent
• Data collection that is geared towards comparing the results with previous research
• Data collection is only reproducible if enough parameters of the previous research are known

• In social scientific research reproducibility does not seek to reproduce the exact same result but to compare different situations with the same methodological approach
Failure and success with reproducibility

• Example of long term monitoring in the Swiss National Park
  - Early pioneers did remarkable research and went to great lengths to make it reproducible
  - But still there are many uncertainties as to the exact locations and methods they used
  - Or, places and methods are known, but not deemed adequate for contemporary research
  - Such things often only show after a long time

• Social scientific research: example of value generation in the Swiss National Park region
  - 1998 by Irene Küpfer, 2012 repetition (see below)
  - Both failure and success
Efforts to make our research reproducible

• In general with social scientific research
  - Making methodological approach as clear as possible
  - Not only provide questionnaire or interview guidelines but also critically assess what went well and what not

• With qualitative research
  - State the researchers’ positionality because it can influence the way responses are given
  - State the special context in which the data collection is embedded
  - Potentially remark what kind of reproduction would be interesting in the future
Which aspects are not reproducible?

• Both with quantitative and qualitative research
  - People change and cannot be questioned the same way
  - With panel surveys people (probably) remember the earlier questions
  - Positionality can be stated but almost not reproduced (even if it is the same person that conducts interviews)

• With qualitative research
  - In most cases interview transcripts should not be handed over to other persons (data protection)
  - Hence, even with the same data reproducibility is difficult if not possible
Reproducing others’ research

- We reproduced Küpfer’s 1998 study on value generation in the Swiss National Park and also Knaus’ 2011 study on the same topic in the Biosphere reserve Entlebuch
- Both authors gave us good information about their methods in their writings (and orally)
- Nevertheless an exact replication was not possible
  - Küpfer did not provide the exact locations where she collected data
  - For resource reasons we could not adapt the same kind of data collection
    - We had less people but more time
    - We asked for people’s expenditures from the past day rather than from the last week
    - We had different notions about what is a NP visitor and what not
  - Knaus’ questionnaire could be used well, however the context was different
Measures to make research reproducible

- Transparency regarding context, positionality and questions asked
  - Not always possible in journals
  - Clean up data sheets from quantitative surveys (excel, SPSS)
  - Find a place to store these data sheets (still uncommon to attach them to papers)

- Bear in mind the limits imposed through data protection
  - Anonymity and people’s right to their own data
  - Be careful about which kind of data can be stored

- State what would be interesting to be reproduced and what not