Cases

> **Note**
>
> This document is still a work in progress.

Overview

In case-study research, cases represent discrete instances of a phenomenon that inform the researcher about it. The cases are not the subjects of inquiry, and instead represent unique sets of circumstances that frame or contextualize the phenomenon of interest. They must be selected with care, and in ways that serve the overall project's objective. This document outlines the case selection strategy for this research.

Case selection is crucial for delimiting the project's overall scope, and will significantly impact the extent to which the research questions may be adequately addressed.

Case Study Research

Cases usually share common reference to the overall research themes, but exhibit variations that enable a researcher to capture different outlooks or perspectives on matters of common concern. Drawing from multiple cases thus enables comprehensive coverage of a broad topic that no single case may cover on its own.

In other words, cases are contexts that ascribe particular local flavours to the activities I trace, and which I must consider to account fully for the range of motivations, circumstances and affordances that back decisions to perform activities and to implement them in specific ways.

Moreover, the power of case study research derives from identifying consistencies that relate cases to each other, while simultaneously highlighting how their unique and distinguishing facets contribute to their representativeness of the underlying phenomon. Case study research therefore plays on the tensions that challenge relationships among cases and the phenomenon that they are being called upon to represent.

It should be noted that case study research limits my ability to derive generalized findings across the whole field of epidemiology. As such, my intent is instead to articulate some significant aspects of data harmonization as they are represented at the cases accessible through this study. In other words, I aim to make certain under-appreciated social and collaborative commitments that underlie data-sharing initiatives more visible and to draw greater attention to certain sensibilities, attitudes, and apprehensions that are relevant to contemporary discourse on the nature of epidemiological data and ongoing development of information infrastructures designed to support data integration and re-use.

Comparative Factors

Comparison can be greatly facilitated if the bases for comparison are well thought out in advance. Here I will identify factors that will enable me to draw comparisons among cases so that I may more effectively meet the project's objectives

To reiterate, this project investigates the social and collaborative apparatus that scaffold data harmonozation initiatives in epidemiology. Through analysis of data obtained through interviews with various relevant stakeholders attached to data-sharing initiatives, the project will ascertain the actions taken and challenges experienced to mediate the varied motivations, needs and values of those involved. In effect, the project aims to articulate the collaborative commitments that govern the constitution and maintenance of epidemiological information commons, and to relate these to technological, administrative and epistemic factors.

A few strands of thought come to mind:

- How close-knit is the web of projects that are sharing data?
  - Are they already very familiar with each other?
  - How were they recruited?
  - Do any participants seem “forced”, or are they all a natural fit?
- How did data-sharing initiatives emerge?
  - Opportunistic? Problem-solving? Culmination of a long-standing desire? Accidental by-product?
  - How did these origins influence the social relations that came to shape the initiative?
- How are the commons governed?
  - What does the leadership look like, in terms of participants and their background, and the reasons for why they are in positions of power?
  - Is there a tier structure of community participation? What boundaries are being drawn, and why are they being drawn in the ways that they are?
  - What bounds the project as a whole?

Since a significant aspect of this work is to compare different approaches to data-sharing, and the approaches have not yet been systematically articulated, we will have to loosely define the parameters through which we initially characterize each potential case. While one could argue that the identities of cases should derive from the parameters of what we are looking to compare, this is not realistic, for a few reasons. First, we are limiting cases to partners of the Maelstrom project, as a way of reducing “overhead” with regards to discussion of technical and administrative discussion. Moreover, this is foundational work in that it aims to articulate the different approaches even as it compares them; the parameters simply have not yet been defined.

However, even through they have not been formally defined, the community does have a sense of different approaches to epidemiological data-sharing, and I will consult with key stakeholders to arrive at a consensus about which cases are worth approaching, and the rationales behind these selections.

The consultation process is meant to ensure that case selection adheres to community will and reasoning, while also ensuring that cases are logistically feasible. I will therefore ask for input from leading members of epidemioligical data-sharing initiatives who are familiar with the goals of the this project, and who are involved with the Maelstrom Project which establishes logistical boundaries around the scope of the project, to weigh in. The consultation will be guided by a few concerns:

1. Age, duration, longevity: this will reflect a capacity to address metrics or values surrounding "success" of data-sharing initiatives
2. Community membership: this will reflect the size and composition of the community, degree of separation of its members, and the mechanisms through which connections are managed
3. Support structures: the ways in which the initiatives are supported, i.e. funding models, reliance on specific platforms or initiatives that influence how they operate
4. Disciplinary trends: reflections on whether an initiative is emblematic of specific paradigms or value regimes that permeate the discipline
5. Historical or contextual factors: whether initiatives are meaningful in relation to broader trends, i.e. brexit's impact on trans-European funding

Fixed Cases

Maelstrom will serve as a "fixed point" that limits the scope of the cases' breadth, while also ensuring that participants (and myself) have a common frame of reference. This will greatly simplify the interviews by reducing the "overhead" of having to learn or be told about administrative or technical systems, which may distract from the core interest.

CITF will also serve as a fixed case. This is for logistical issues, since the grant meant to support the CITF Databank. But it is also so that the data can be compared with research on user experiences pertaining to CITF specifically, which will be performed concurrently with my own work.

Limitations and Sources of Bias

- Identifying prospective interview participants, sampling
- Willingness to participate
- Prioritizing local connections and in-person interviews?

Logistical Constraints

- 12 month time limit
- 12-15 interviews, 60-90 minutes in duration

Ragin (1992)

Ragin, Charles C. 1992. "Casing and the Process of Social Research." In *What Is a Case? Exploring the Foundations of Social Inquiry*, edited by Charles C. Ragin and Howard S. Becker, 217–26. Cambridge University Press New York.