Drowning in data? A practical guide for beginning the data-to-theory move

Thomas Gegenhuber (Johannes Kepler University Linz) and
Leanne Hedberg (University of Alberta)

You just finished your fieldwork. The outcome of your data collection is tremendous. You conducted and transcribed over thirty interviews, collected over fifty documents, wrote rich fieldnotes on your observations and, last but not least, kept a thorough research journal capturing your initial impressions and thoughts. The next step?: diving into the data and beginning to analyze it. It's funny (not really) how just a short time ago you were pleased as pie at how much data you were able to collect. Now, you look at that very same mass of data but with a paralysing sense of overwhelm. Doubts arise. You feel like you are drowning in a vast sea of data. Questions arise: What now? Where do I possibly begin?

Rest assured, whether you are a PhD student or a more experienced researcher, facing the data behemoth is always overwhelming. Every time. Yet, the task is likely to be even more intimidating for early scholars. It is at this very point – approaching your dataset after collection – where we hope this blog post will be helpful. The insights provided in this post, aimed to help answer the questions of "What now?" and "Where do I possibly begin?", are based on a qualitative methods workshop given at the University of Alberta by [Prof. Trish Reay](https://business.ualberta.ca/about/contact-us/school-directory/trish-reay)[1] (who has, by the way, given us permission to share her wisdom).

True, there is a large body of literature on qualitative research design – how to collect, organize and analyse data. Yet, despite this vast amount of information, what's ironically yet glaringly missing, in our humble opinion, is *practical advice* on how to actually (and very specifically) begin analysing one's data towards the goal of (hopefully) turning an initial idea, and subsequent data, into a paper…that's published (but let's not get too far ahead of ourselves).

The goal of the data analysis is to climb up the ladder of abstraction – from minute details to broad themes - enabling you to convincingly convey, to an audience, a new (theoretical) understanding of a phenomenon. Related to foundation of this climb, it is said that a great theoretical story begins with an intriguing empirical puzzle. Consider the work of Reay and Hinings (2009) as an example. The intriguing empirical puzzle in this case was to understand the significant change that took place, over time, in the Alberta, Canada healthcare field. The theoretical story that emerged from this study was in answering the question: How did individuals, and ultimately organizations, deal with the imposition of a conflicting "business management" agenda into an agency traditionally steeped in a focused agenda of "medical profession"? Part of why this question is intriguing is because so many organizations around

---

[1] https://business.ualberta.ca/about/contact-us/school-directory/trish-reay

the world have been faced with such a situation. Theoretically, this paper added to the literature a better understanding of how actors deal with co-existing and competing institutional logics.

Assuming that you marched out to collect your data with a similarly intriguing question in mind, we now provide you with the specific (and practical) steps to begin attempting to answer your question through the analysis of your data. The following illustration is an overview of Trish's suggested steps on how to approaching this analysis:
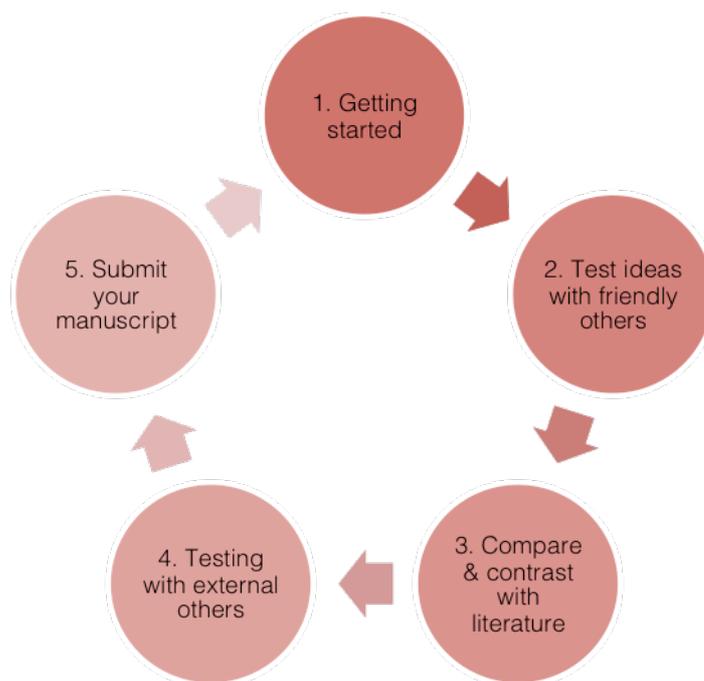
1. Get Started

Getting started is the single most important action you can take. In very practical terms, this means choosing four or five key interviews and documents. How do you know if the interviews or documents are "key"? Let the answers to the following questions be your guide: Which ones stand out to you the most? Which ones intrigue you the most? Above all, keep in mind that what is far more important, at this point, than picking the "right" interviews and documents to start with, is *simply starting*. This means that, in the absence of having a sense of which interviews or documents to start with, you can just randomly pick *any* four or five. Really.

One of Trish's core recommendations is this: *stay as close to your data for as long as possible*. What, exactly, does this mean? It means that the most important phase of your data analysis is arguably in giving yourself plenty of time (weeks, months) to just read (and reread) through your data. Be with your data long enough so that it can begin to speak to you. Muse about your data broadly until you begin to notice one or two (or more) themes or patterns be-

ginning to emerge among the details of your data. What seems, initially, like an incoherent jumbled mass of words will, given enough time and attention, start to reveal itself albeit in a whisper at first.

Before you know it, you will have already begun the coding process, which in simple terms is whatever helps you to get from data to theory. And while "coding" may sound intimidating, boring and/or highly technical, it is - quite to the contrary - an inherently creative and fun process, especially once you relax and simply begin spending time with a small number of interviews or documents. In fact, this nature of data analysis – the opportunity to dive deeply into some aspect of a social world and to make sense of it - is exactly why researchers prefer qualitative research in the first place.

Because this phase of data analysis is essentially a creative process, we offer a quick suggestion: at this early stage, it is important to refrain from using a computer because of the possibility (probability) that it will hamper the creative process. This artist[2] insists that "using your hands" is key for the creative process. Using your hands can involve printing out your data and collecting various types of paper, along with scissors, glue, colored markers or pens, and a flipchart or pin board. Now, play with your data. Some suggestions to get you started are to make coloured notes in the margins with thoughts that jump out at you or to cut out quotes and paste or pin them up, beginning to identify aspects of your data that are either similar or dissimilar. This "play" time is actually the beginning of the categorization process using a sample of your data. Let this be a fun and even seemingly messy process. Allow the information that emerges on the canvas of your flipchart or pin board to give you an overview of your first glimpse on the data. It may be helpful to have someone work alongside you at this point, because it can provide much needed validation that someone else also sees what you see. Conversely, it can be invigorating to discuss the ways in which you see the same data differently. The thrust of this phase is just to begin generating thoughts and ideas about your data.

After this creative process, and from the viewpoint of your research questions, what similarities or dissimilarities have begun to emerge from within the first set of data that you selected? Given these, you can begin developing codes. Gioia et al. (2012) suggest working with 100+ codes is okay. Others, such as Creswell (2007) advise using not more than 30 – 50 codes. Trish sides with Creswell's approach because too many codes may be overwhelming and your codes should be broad enough so you can find them in the data repeatedly. Nevertheless, Trish points out that the debate on the number on codes, like many other topics in qualitative research, have their root in different philosophies. Hence, you should expose yourself to as many research philosophies as possible to find out what works best for you.

At this point, your research question is still broad and at a provisional stage (e.g. How do actors deal with institutional change over time?), which is perfect for providing you with a general direction of where to continue your search within your additional data. Over time, as

---

[2] http://austinkleon.com/steal/

you go through your data, allowing categories or codes to emerge, your research questions will become more focused. At this point, your theoretical knowledge on the topic is still broad but noticeably more developed than when you initially proposed your research.

2. Test ideas with friendly others

At this point in the process, you need to carve out a space where you can experiment with your data by testing your ideas, and receiving feedback, from *friendly others*. These friendly others can be your co-authors, your supervising team or fellow PhD students (potentially even from different disciplines) with whom you discuss your ideas and interpretations. The feedback you receive may be constructive in pointing out different interpretations of your data than what you had initially interpreted. The most important thing to be aware of at this stage is that the creative process of developing new ideas requires a *nurturing environment*. New ideas, in particular, are tender and require a culture of encouragement and support (Amabile, 1997). Harsh criticism at this point will slow down, if not stop, your greatly-valued momentum. At this and the following stages, you increasingly need to dig deeper into your data in order to leverage the richness of your *entire* research database.

3. Compare and contrast with the literature.

After testing your ideas with friendly others, you will have written down the resulting ideas in the form of a first (rough) draft of an in-depth case description or process story. Now is the time to become more critical. How does your story fit with extant theory and literature? What counterfactual explanations can you imagine? This is the point where you turn to the literature, get a better understanding of the current debates related to your research questions. Most importantly, this is where you identifying a gap within the literature. Although you definitely do not need to re-invent the wheel of the literature (e.g. contribute something that turns it on its head in an evocative yet plausible way), you do want to begin to understand what it is that you can add to the existing body of literature. Friendly others (e.g. peers and more experienced researchers) can also be of use at this point, helping to point you to relevant literature.

4. Test ideas with external others

Now that you have initial ideas from a part of your data and you also have ideas on where you might be able to make a contribution to the literature, it is time to test your ideas via conferences and workshops. This is a great opportunity to see how your potential audience responds to your work and can also be a great source of inspiration. In general, feedback is delivered with friendliness and clearly supportive intentions. Sometimes, however, the criticism may feel harsh. But keep in mind that any feedback from external others, even if it's not delivered in the most diplomatic and encouraging manner, may include useful nuggets or challenge you to think in novel ways about your paper. Presentations at these types of events are not only part of the collective creative process of developing your paper, they are also a opportunity to position yourself within the community. What this means is that, as you present, you will begin to make a name for yourself related to your empirical setting and/or theoretical focus.

Speaking of theoretical focus, at this point it is this and your research question that take the main stage. Information that seems interesting to you, yet is irrelevant to the main story, needs to be pushed backstage and not visible to the audience. Even in a study that relies on a small number of interviews, you cannot tell everything – it would confuse and overwhelm the audience. Streamlining your argument in accordance with the data is crucial for crafting a convincing argument. Once again, friendly others can be of great help here. You can present to several willing peers to get feedback on how to best tell the story of your research-in-progress.

5. Submit your manuscript

Finally, you have analysed your data and received really helpful feedback, which you've incorporated into an actual paper. Now you can submit your work to a suitable journal. Before you do submit the paper, use multiple search engines to find your own paper. Make sure that you remove any full papers and references that may have been posted on the web following conference and workshop presentations. This will help to uphold the double-blind peer reviewed process. Also, it is important to state clearly, in your letter to the editors, if you used the dataset that had been used, by you or by others, for a separate paper that was submitted for publication. If this is the case, you will need to convince the editors that your submitted paper is different from the other(s), whether this is because you used a different part of the data set or because your research questions and/or theoretical framing is clearly different. Be aware, however, that if you are using a similar data set for a different paper, you won't be able to reuse your methods section. Rather, you will need to rewrite it.

Present your data and your approach in a rigorous way. This may involve using, as role models, a handful of highly cited papers that use the same methodology as you did and that have been published at top-journals. This will give you an idea of much attention needs to be placed on clearly articulating and illustrating your analysis process so that your audience is convinced of the claims you have extracted from your data. Clearly, it not possible that several different researchers who are given the same data set would arrive at *exactly* the same results or conclusions. However, if we, as readers, were to follow the steps of data analysis that you describe in enough detail, in your methods section, it needs to be plausible that we would arrive at a *similar* conclusion. For your findings section, what's most important is that you convey a compelling *story*. Finally, in your discussion section you must be very explicit about how you contribute to theory. Even if you not have actual propositions in the text, Trish suggests developing propositions within your discussion section in order to distil the essence of your work. The use of propositions can serve as a structure for writing the discussion section and can guide you in explaining your key arguments to your audience.

Ideally, you will receive the joyous news of a "revise and resubmit". But it's also likely, regardless of the stage of your academic career or even given a history of R&R success, that you may be rejected. Although this initially feels like failure or defeat, it is important to take a brief break from this particular project. This means that you should take a while to process the rejection. There are many factors that determine acceptance or rejection at a particular jour-

nal, and some of these factors are completely outside of your control. After taking time away from your project, look at the reviews again. Very normal response after reading reviews for the first time include some variations of the following: I am such an utter failure; I am not cut out for this career; Those reviewers are jerks! After taking a break from your project, you hopefully start to see them as helpful. Consider that these individuals took their precious time to carefully read your paper and to express, the ways in which it can be better. They essentially put effort into helping you along in your career, even if that help was an honest "this idea has no future". Now you can go and integrate this very helpful feedback into a rewrite of your paper and submit it elsewhere or you can place the paper ever so gently in your desk drawer and re-orient your attention towards another project that just may have more potential. It's been greatly comforting for us to hear stories from several highly published authors about their "failures" (having had their papers rejected). These stories tell us that it is part of the process and it really is a process, one that we will be in throughout our careers as academicians.

References

Amabile, T. M. 1997. Motivating Creativity in Organizations. *California Management Review*, 40(1): 39–59.

Creswell, J. W. 2007. *Qualitative Inquiry & Research Design*. Thousand Oaks, CA: SAGE Publications.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. 2012. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, 16(1): 15–31.

Reay, T., & Hinings, C. R. 2009. Managing the Rivalry of Competing Institutional Logics. *Organization Studies*, 30(6): 629–652.