Lessons from ethnographic fieldwork in the Data Science Environment: Challenges and opportunities of studying up, over, across and through

Anissa Tanweer, PhD
Research Scientist, eScience Institute

QUAL Seminar
Nov 15, 2018
Research synopsis [15 minutes]

Reflecting on method [30 minutes]

- *Studying up*: navigating the blurred lines between research participants, colleagues, mentors and sponsors
- *Over*: managing mountains of qualitative data
- *Through*: picking a path through grounded theory development
- *Across*: finding an authentic voice when speaking to multiple disciplinary communities

Questions and comments [15 minutes]
Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren’t seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, “It was like arriving at a conference reception and realizing you don’t know anyone. So you just stand in the corner sipping your drink—and you probably leave early.”
Research Questions

1. How is data science scaling its reach across social sectors and problem spaces?
2. How are practitioners responding to the ethical crisis facing data science of the social?
How is data science scaling its reach across social sectors and problem spaces?

Anissa Tanweer, 2018
“Cyberinfrastructures”
Edwards et al., 2007; Ribes & Finholt, 2009

“Laboratories without walls”
Finholt, 2002
“We need to avoid boutique solutions so they can be scaled.”

- Panelist at national conference on data-intensive cross-sector collaborations
<table>
<thead>
<tr>
<th><strong>Infrastructure Characteristics</strong> (Star &amp; Ruhleder, 1996)</th>
<th><strong>Exostructure Characteristics</strong> (Tanweer, 2018)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embeddedness</td>
<td>Portability</td>
</tr>
<tr>
<td>Transparency</td>
<td>Transience</td>
</tr>
<tr>
<td>Learned as part of membership</td>
<td>Learned as part of development</td>
</tr>
<tr>
<td>Becomes visible upon breakdown</td>
<td>Breakdown is expected</td>
</tr>
<tr>
<td>Reach or scope</td>
<td>Same</td>
</tr>
<tr>
<td>Links with conventions of practice</td>
<td>Same</td>
</tr>
<tr>
<td>Embodiment of standards</td>
<td>Same</td>
</tr>
<tr>
<td>Built on an installed base</td>
<td>Same</td>
</tr>
<tr>
<td>Infrastructure Characteristics (Star &amp; Ruhleder, 1996)</td>
<td>Exostructure Characteristics (Tanweer, 2018)</td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Embeddedness</td>
<td>Portability</td>
</tr>
<tr>
<td>Transparency</td>
<td>Transience</td>
</tr>
<tr>
<td>Learned as part of membership</td>
<td>Learned as part of development</td>
</tr>
<tr>
<td>Becomes visible upon breakdown</td>
<td>Breakdown is expected</td>
</tr>
<tr>
<td>Reach or scope</td>
<td>Same</td>
</tr>
<tr>
<td>Links with conventions of practice</td>
<td>Same</td>
</tr>
<tr>
<td>Embodiment of standards</td>
<td>Same</td>
</tr>
<tr>
<td>Built on an installed base</td>
<td>Same</td>
</tr>
</tbody>
</table>
Infrastructure vs. Exostructure

Anissa Tanweer, 2018
“I’m hoping the whole thing gets built from scratch next year.”

- Mark, ORCA Project Lead
Mark’s vision for a “trusted data platform.”
How is data science of the social responding to ethical crisis?
Vernacular theorizing
&
Pedagogy of vernacular theory

- Thomas McLaughlin, 1996
Academic theories of sociomateriality

Vernacular theories of sociomateriality in data science of the social

Anissa Tanweer, 2018
The medical model of disability
The social model of disability

Michael Oliver, 1990
*The Politics of Disablement*
INTUITIVENESS
Ability for non-experts to readily understand and work with data
Primary stakeholder to consider:
OSM manual contributors

RELEVANCE
Importance of data for various geospatial informational needs
Primary stakeholder to consider:
End users of AccessMap

INTEROPERABILITY
Ability to integrate with other sources, applications, etc.
Primary stakeholder to consider:
Developers (including AMOS)

COMPUTATIONAL EFFICIENCY
Time, space, and resources required to support data
Primary stakeholder to consider:
OSM Organization

1. Curb ramps
By making curb ramps points, losing some precision compared to polygons, but making ramps easier to map.

2. Curb ramps
By mapping each ramp precisely, gaining important information but making it harder to automatically add data.

3. Sidewalks
By separating sidewalks from roads, easier to map but harder to name for directions.

4. Sidewalks
By mapping sidewalks as lines, provided a visual cue, but would increase size of sidewalk data.

5. Municipal Data
By vetting municipal data, slowing import process but avoiding bloating the map with reverted errors.
Navigating the blurred lines between research participants, colleagues, mentors and sponsors
"Mutual vulnerability"
I arrived to find that one of the Laboratory's leading weapons scientists had come to my talk wearing nothing but a loincloth and carrying a cane to which he had nailed an animal skull. He shook this at me and grunted whenever my presentation displeased him — which seemed to be quite often.

- Gusterson, 1997, p. 117
As he so aptly dramatized for me, the objectifying, exoticizing language of anthropology is as objectionable at home as abroad, and one is less likely to get away with it

- Gusterson, 1997, p. 117
OVER

MANAGING MOUNTAINS OF QUALITATIVE DATA

Anissa Tanweer, 2018
1. **Be kind to your future self** - Writing up notes doesn’t always happen as thoroughly as we ideally would like or intend. But add some layer of interpretation to “raw” notes right away, no matter how cursory. Even just a couple bullet about the most important observations of the day will help immensely during analysis.

2. **Slice and dice** - Cut and paste your notes along multiple cross sections. I.e. don’t just review chronologically, but gather all observations about particular settings, topics, persons, etc. and review those thematically.

3. **Start with your observations** - If you analyze interview transcripts first, you will end up with an interview study.

4. **More is less** - Memoing is in a way adding to the heap of material to get through, but the more you do of it, the easier it will be to cut through noise and find a signal in your data.

Anissa Tanweer, 2018
Picking a path through grounded theory development
<table>
<thead>
<tr>
<th>Phase</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germination</td>
<td>Grounded theory for ethnographic data</td>
</tr>
<tr>
<td>Distillation</td>
<td>Cartographic situational analysis</td>
</tr>
<tr>
<td>Refinement</td>
<td>Retroductive analysis</td>
</tr>
</tbody>
</table>
Germination: Grounded theory for ethnographic data
Charmaz & Mitchell, 2007

1. Simultaneous data-collection and analysis
2. Pursuit of emergent themes through early data analysis
3. Discovery of basic social processes within the data
4. Inductive construction of abstract categories that explain and synthesize these processes;
5. Integration of categories into a theoretical framework that specifies causes, conditions and consequences of the process(es).
Distillation: Cartographic situational analysis
Clarke, 2003

1. Situational maps
   Illustrate connections between relevant social actors and objects
1. Social worlds/arenas maps
   Identify the relations and negotiations between collective actors
1. Positional maps
   Draw out salient differences and relations between discursive perspectives
Refinement: Retroductive analysis
Ragin, 1994; Sæther, 1998

Constructing “images” or “idealized cases”
“Decision stories” (Eisenhardt & Bourgeois III, 1988; Maitlis & Lawrence, 2003)
ACROSS

Finding an authentic voice when speaking to multiple disciplinary communities

Anissa Tanweer, 2018
<table>
<thead>
<tr>
<th>Things I wrote about</th>
<th>Things I could have written about</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical data studies</td>
<td>Transportation studies</td>
</tr>
<tr>
<td>Theories of sociomateriality</td>
<td>Philosophy of ethics</td>
</tr>
<tr>
<td>Information infrastructures</td>
<td>Collaborative governance</td>
</tr>
<tr>
<td>Open data</td>
<td>Etc.</td>
</tr>
<tr>
<td>Practice theory</td>
<td></td>
</tr>
</tbody>
</table>
INTERDISCIPLINARY SCHOLARSHIP = CHRONIC IMPOSTER SYNDROME
Thank you. Questions?

With special gratitude to:
Gina Neff, Cecilia Aragon and Kirsten Foot
Brittany Fiore-Gartland
eScience Institute
Mark Hallenbeck and the ORCA team
Anat Caspi, Nick Bolten and the AMOS team
Gordan & Betty Moore Foundation
Alfred P. Sloan Foundation

Anissa Tanweer, 2018