Researching ICT Companies:

A Field Guide for Civil Society Researchers

Nathalie Maréchal, University of Southern California
Sarah T. Roberts, University of California, Los Angeles

February 15
2018

Researching ICT Companies:
A Field Guide for Civil Society Researchers

Nathalie Maréchal
University of Southern California

Sarah T. Roberts
University of California, Los Angeles

1. Introduction

Information and communications technology (ICT) companies like Amazon, Apple, Facebook, Google, and Twitter are vitally important to billions of users around the world, not only in their day-to-day personal and professional lives, but also in their ability to shape social and political reality. Yet there is a pervasive lack of clarity around the policies and practices that govern user engagement on these platforms and sites, and the values that undergird them. This information is of great importance to policy researchers and civil society advocates, particularly in the wake of numerous recent events that have put the relative power and opacity of ICT companies in the spotlight. Access to information about them is often incredibly difficult to obtain, when it is available at all. These difficulties are faced by many categories of people interested in researching ICT companies, from academics to journalists, and from civil society advocates to policy researchers.

In this white paper, we outline some of the challenges we have identified as being particularly acute for policy researchers, as well as strategies for working through (and around) those issues. Advocating for civil society, human rights, and democratic values today often requires understanding the role played by ICT companies in deciding what kinds of speech are allowed (or not) on various platforms, in complying (or not) with government requests to restrict content or for user information, and in lobbying governments to enact (or not) various laws and regulations. Under the UN Guiding Principles on Business and Human Rights, companies are expected to respect human rights even as nation-states retain primary responsibility for protecting human rights. As is true of many UN norms, the Guiding Principles lack a formal enforcement mechanism, so other, often soft measures have been employed in order to enact results, or even to simply gain information. Although often with various end goals in mind, journalists, researchers and global civil society organizations share the common need to know more about these practices, policies and internal guiding principles that influence the behavior and and outcomes of platforms. For this reason, this disparate group with varying constituencies have developed shared techniques to obtain information about ICT companies' policies and practices, and, importantly, to influence them. This includes sustaining demands for engagement, "naming and shaming," shareholder advocacy, litigation, and more. These strategies all hinge on civil society groups knowing what companies are up to.

2. Barriers to Access

When parties approach firms with the desire to learn more about their practices and functions, they are typically met with either a well-trained public relations officer who intends to reveal very little, or a non-response. Reasons that firms often invoke for this lack of access that firms often invoke include trade secrecy and user privacy, although there are others, some of which are detailed below:

- Companies are primarily accountable to their shareholders, not the public, and may not fully understand their obligations under the UN Guiding Principles.

- Company representatives may not fully understand the relationship between their company's activities, human rights, and other social concerns.

- Researchers & company representatives often speak completely different languages. Sometimes this is literally true, but even if everyone involved speaks English, for example, civil society researchers and company staff tend to come from different educational, social and professional backgrounds and don't necessarily start from a place of common understanding.

- ICT companies tend not to have a culture of transparency, and resist sharing what their consider to be their proprietary data.

- Unlike with public entities that may be subject to Freedom of Information-type statutes (at least in some countries), companies typically have no legal obligation to disclose information that civil society is interested in.

- Many civil society organizations lack staff with formal research training, access to research databases or to data analysis software.

Understanding these factors, and potential avenues of redress for them, can lead to developing a strategy to opening dialogue where they may have previously been none. As social scientists with extensive experience researching ICT companies and human rights within both civil society and academic settings, we wrote this guide to help civil society researchers identify the most appropriate research methods for their research questions and offer strategies for overcoming many of the obstacles that they encounter along the way.

3. The Literature Review: Why Do It, and How?

The odds that you are the first person researching a given topic are slim to none. Any research project should start with an exhaustive search for existing work on the topic. Academics are trained to write a "literature review" that summarizes and critiques previous work on the topic. You don't have to produce something that is publication-ready, but the process of writing a literature review will help you internalize and remember what others have written about your topic.

It may be a reasonable first pass to simply google a variety of keywords and collect the resulting materials, but, for better or for worse, some of the richest and most important information is only available through other kinds of portals, behind paywalls and in other repositories that will not make their way to the top of a typical commercial search results. These can include journal articles, white papers, book chapters and other resources. If you are affiliated with an academic institution, you can and should make use of the vast electronic databases and other resources to which your campus undoubtedly has access, and you should enlist the help of a research librarian at that institution, who can give you tips on which databases might be best and how best to search.

If you're not affiliated with a university, you may have trouble accessing journal articles on your topic. Even if you are, there are other resources you should know about. Here are a few of them:

- University repositories: Seek out preprints/postprints at university digital repositories and/or scholars' websites: more and more, public research universities, federal funding bodies and other institutions have implemented requirements for scholars to share the results of their research with the public. Compliance with these rules can be achieved in a number of ways, often in the form of scholars uploading pre-publication versions of their work to institutional repositories and/or to their own websites. (Example: the University of California's open access eScholarship portal: https://escholarship.org/)

- Can't find what you're looking for at a university open access portal? Email the author of an interesting-looking article and ask them to send you a copy, explaining why you can't just access it directly. The author may have a preprint/postprint on hand (more and more researchers are negotiating this right with publishers) that they can direct you to or send you.

- Not affiliated with a university? Don't give up: public and university libraries may provide database access to members of the public- ask a research librarian for help!

- Google Scholar: This is a great starting point for seeking out scholarly research on a given topic, but it ought not to be the only site for researching, ah, research. In fact, there are plenty of critics who point out that the very best or most relevant information may not necessarily filter to the top of your results page. Google Search can definitely identify papers that are highly cited, but that is not necessarily an indication of quality (or even of agreement with a given paper), so the reader should beware. That having been said, it's an excellent starting point that can help lead you (often, via a given paper's citations) to other resources. Google Scholar can often also link directly to copies of papers that are available on the open web- although it does not determine whether or not those papers ought to be there.

- Google Scholar Alerts: While this feature is a subset of those available in the aforementioned Google Scholar, it is worth a mention of its own. You can configure a series of queries to alert you to new scholarly research related to the topic or issue you are tracking. This can be incredibly helpful and the results will come directly to your inbox. Of course, such alerts also exist for Google search; you may wish to create alerts using that tool, as well, in order to track your topic in the popular press and elsewhere on the web.

- Meta-repositories: SSRN (OWNED BY ELSEVIER); arXiv (for STEM); PubMed (NIH/NLM- National Institutes of Health National Library of Medicine)- The NIH Public Access Policy implements Division F, Section 217 of PL 111-8 (Omnibus Appropriations Act, 2009) - moderation/gatekeeping varies at each one, so your mileage will, accordingly.

- Many scholars make PDFs of their work on sites like Academia.edu and ResearchGate. However, these start-ups are starting to monetize user-generated content and user data, and some scholars are turning away from these sites as a result.

- Certainly, many simply ask friends/colleagues/interns with university credentials to download PDFs for you. Yet this path is fraught with peril: doing so will likely break contractual terms of service for the person providing you with access and you could be putting that person in danger of liability. There is even case law in the United States where firms were prosecuted for doing just this. With all the other options out there, we recommend them and cannot endorse this one.

4. Research methods, what they're good for, and their limitations

Once you have a good grasp on the preexisting research on your topic, you're ready to start collecting data. There are many different ways to research ICT companies. In this section we describe several such methods, explain what they're good for and what their limitations are, and provide examples of studies on ICT companies that use each method. (Note that this is in addition to citing existing work, including journalistic coverage and leaked documents that are publicly accessible.)

Companies' own public disclosures

One strategy for researching companies is to focus on the information that they do make publicly available. Corporate websites, press releases, legally required disclosures (such as Securities and Exchange Commission filings), and public speeches by company executives can provide a wealth of information. A growing number of projects worldwide systematically evaluate ICT companies' disclosures and commitments pertaining to freedom of expression, privacy, and corporate governance with respect to human rights in order to benchmark companies against one another. A growing number of advocacy groups issue regular scorecards that rank companies according to their performance on predetermined indicators, spurring a "race to the top." Such projects are not just about "naming and shaming," but can also call out those who perform particularly well.

- *What It's Good for:*

  - Empirically-based: it is difficult for companies to argue against reports that cite their own publications
  - Lends itself to advocacy campaigns and partnerships

- *Challenges & Limitations:*

  - Measures what companies say they do, not necessarily what they actually do
  - Most useful if done year-on-year

- *Examples:*

  - Who Has Your Back? (Electronic Frontier Foundation)
  - Corporate Accountability Index (Ranking Digital Rights)
  - Donde Estan Mis Datos (R3D)
  - Dependent Yet Disenfranchised: The Policy Void that Threatens the Rights of Mobile Users in Arab States (SMEX)
  - Lumen Database (Berkman Klein Center)

Technical testing of software and hardware

A complementary strategy uses technical testing and analysis to verify how software and hardware products actually meet their commitments and/or user expectations. Open Effect's investigation into the security of popular fitness trackers revealed that most products in this category leak user data in ways that are highly susceptible to exploitation. Similarly, the consortium behind the Digital Standard is developing methodologies for systematically assessing consumer products' adherence to predetermined best practices.

- *What It's Good for:*

  - Reveals deceptive practices
  - Incentivizes companies to improve their products
  - Paves the way for regulatory action to protect consumers

- *Challenges & Limitations:*

  - Requires collaboration between researchers with advanced technical skills and with expertise in socio-technical systems
  - Companies may argue that such testing violates their End User License Agreements, Terms of Service, etc. We recommend seeking advice.

- *Examples:*

  - Every Step You Fake: A comparative technical analysis of of fitness tracker privacy and security (Open Effects)
  - The Citizen Lab's research on South Korea's "Smart Sheriff" app

- *What It's Good for:*

  - Surfaces data that would otherwise be completely inaccessible
  - Creates engagement and awareness among internet users
  - Generates compelling, real-life examples that illustrate broader concepts and principles

- *Challenges & Limitations:*

  - Can't make any claims about representativeness of the data: "the plural of anecdote is not data."
  - Relies on unpaid labor by volunteers of variable motivation and qualification

- *Examples:*

  - Onlinecensorship.org
  - Free Basics in Real Life, by Global Voices Advox

Crowd-sourced user reports

Studying social media companies can be especially challenging because user experiences are highly personalized in ways that are opaque, unaccountable and all but impossible to replicate. It is becoming increasingly clear that this has troubling implications for privacy, freedom of expression, and democracy itself. Onlinecensorship.org is a crowd-sourced project that solicits reports from users who have experienced content censorship on social networking sites operated by Facebook, Google and Twitter. The multi-disciplinary team then analyzes the submissions and periodically issues reports that qualitatively analyze social media censorship. Similarly, the public interest journalists at ProPublica have launched a Political Ads Collector tool, a Chrome browser plug-in that allows Facebook users to track political ads even if they are not part of the ads' target demographic.

Interviews and other ethnographic methods

Typically qualitative. Use techniques and methods from a variety of academic disciplines, most notably anthropology (but other fields employ heavily, as well). Includes both "approved" methods (formal interviews with high-level employees, content analysis of public speeches) and "unorthodox" interviews of staff who may not have official approval to talk to researchers.

- *What It's Good for:*

  - Gaining rich "deep-dive" insight into firms' inner workings
  - Learning information that is not available in any other context
  - Triangulating and confirming information gained from some of the previously described approaches

- *Challenges & Limitations:*

  - For academic researchers, this kind of work is often considered "human subject research" and, as such, is subject to approval, oversight and governance by Internal Review Boards (IRB), ethics panels and others (see the next section)
  - It can be very difficult to access people and contexts in order to obtain these kind of data, where interviews, observation or other ethnographic methods are to be employed. "Going through the front door," however, is not the only option, but researchers must be cognizant of the personal risk that participants and informants may be taking when it comes to working with them. At best, employees are likely under non-disclosure agreements (NDAs) and violating those can have severe consequences for them

- *Examples:*

  - Roberts, S. T. (2016). Commercial Content Moderation: Digital Laborers' Dirty Work. In S. U. Noble & B. Tynes (Eds.), The Intersectional Internet: Race, Sex, Class and Culture Online. Peter Lang Publishing. Available as an open access preprint: https://ir.lib.uwo.ca/commpub/12/
  - Jørgensen, R. F. (2017). What Platforms Mean When They Talk About Human Rights. Policy & Internet. http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/boundary%2520OII.pdf

THIS LIST IS NON-EXHAUSTIVE. Seek out research studies and get inspired!

5. Research Ethics

For academics with university affiliations or appointments, there are often internal ethical standards that must be met before research involving human subjects is conducted. In the United States, these ethical determinations often involve the submission of a proposed study, and a description of its expected impact, to a body called the Institutional Review Board (or IRB). Researchers from outside academia may not need to go before such a body, but it's still a good idea to have some ethics and politics around the kind of research you intend to do, and upon whom. Recommendations from the Association of Internet Researchers (AoIR) include:

- When making ethical decisions, researchers must balance the rights of subjects (as authors, as research participants, as people) with the social benefits of research and researchers' rights to conduct research. In different contexts the rights of subjects may outweigh the benefits of research.

- Ethical issues may arise and need to be addressed during all steps of the research process, from planning, research conduct, publication, and dissemination.

- Ethical decision-making is a deliberative process, and researchers should consult as many people and resources as possible in this process, including fellow researchers, people participating in or familiar with contexts/sites being studied, research review boards, ethics guidelines, published scholarship (within one's discipline but also in other disciplines), and, where applicable, legal precedent.

- The greater the vulnerability of the community / author / participant, the greater the obligation of the researcher to protect the community / author / participant.

- Because 'harm' is defined contextually, ethical principles are more likely to be understood inductively rather than applied universally.

- Because all digital information at some point involves individual persons, consideration of principles related to research on human subjects may be necessary even if it is not immediately apparent how and where persons are involved in the research data.

Journalism is of course bound by its own ethical practices and procedures, which may be different in application from those at play in a university setting. Part of the ethical considerations for journalists, in particular, is to fight for inclusion of their academic and other sources and to demand that editors keep those attributions in.

6. How to get the right training

The research methods described above can be challenging to execute -- there is a reason that academic researchers spend so much time in graduate school! But that's no reason for civil society to give up on doing research. Indeed, NGOs are often closer to the social problems in question than academics, and can produce high-impact reports much faster than scholars working within the peer-review process. Moreover, academics don't always have the skillset, institutional support or desire to build advocacy around their research. It is vital that civil society produce intellectually rigorous research, but how? We have identified four broad strategies for organizations to acquire or develop suitable research skills:

1. Hire for the research skills you want
The most straightforward way for an organization to ensure that its staff has the requisite skillset for a given research project is to proactively hire for it. If you want to conduct a network analysis of Twitter users (for example), recruit staff with the relevant statistical and network mapping expertise. Organizations that can't afford such a hire can consider hosting a summer graduate fellow or partnering with an established academic (see point 4 for more on partnerships).

2. Develop research projects that leverage your staff's existing skillset
In contrast, this strategy leverages the skills of existing staff to create value. Maybe your organization's program associate did ethnographic research as part of her Master's thesis—you might consider having her do an ethnography of how journalists in your region engage with social media, for example.

3. Train staff in new skills
A third option is to enhance your current staff's skillset, whether by providing tuition assistance for graduate school, building the development of new research skills into their job descriptions, or paying for specific professional development experiences like conferences, summer schools, or webinars.

4. Partner with others whose skillsets complement your existing ones
Collaboration across professional boundaries for amplified pressure points/uptake of work/(other outcomes): this is greater than the sum of its parts; almost always necessary or at least makes the potential for achieving goals greater (or the results better disseminated).

About the authors

Nathalie Maréchal is a PhD candidate in Communication at the University of Southern California's Annenberg School for Communication and Journalism. Her research examines how information and communication technologies impact human rights, social movements, and geopolitics. Between 2014 and 2017, Nathalie held a series of fellowships at Ranking Digital Rights, a non-profit research initiative housed at New America's Open Technology Institute that works with an international network of partners to set global standards for how companies in the information and communications technology (ICT) sector should respect freedom of expression and privacy. Website: nathaliemarechal.net

Sarah T. Roberts is an assistant professor in the Department of Information Studies at the University of Southern California, Los Angeles. Prior to joining UCLA in 2016, she was an assistant professor in the Faculty of Information and Media Studies at Western University in London, Ontario. Professor Roberts is internationally recognized as a leading scholar on the emerging topic of commercial content moderation or CCM, a term she coined to define the field study around the large-scale, industrial and for-pay practice of social media user-generated content adjudication. Website: illusionofvolition.com

About the Internet Policy Observatory

The Internet Policy Observatory (IPO) is a project at the Annenberg School for Communication at the University of Pennsylvania. The overarching goal of the program is to deepen the reservoir of researchers and advocates in regions where Internet freedom is threatened or curtailed and to support the production of innovative, high-quality, and impactful internet policy research. The IPO facilitates collaboration between research and advocacy communities, builds research mentorships between emerging and established scholars, and engages in trainings to build capacity for more impactful digital rights research and advocacy.