Digital Humanities in the Nordic Countries, Helsinki, Finland, 7–9 March 2018.

Breaking Bad (Terms of Service)? The DH-scholar as Villain

Professor Pelle Snickars
Department of Culture and Media Studies / Humlab
Umeå university
pelle.snickars@umu.se

For a number of years I have been heading a major research project on Spotify (funded by the Swedish Research Council). Entitled, "Streaming Heritage: Following Files in Digital Music Distribution", the project has involved system developers Fredrik Palm, Roger Mähler, Andreas Marklund and Johan von Boer (at Humlab, Umeå University), as well as researchers Maria Eriksson, Anna Johansson (at Umeå University), Rasmus Fleischer (at Stockholm and Umeå University), and Patrick Vonderau (at Stockholm University). A guiding question for the project has been how people's practices and approaches toward cultural forms such as songs, books, or films—practices including the production, expression, and exchange of those cultural forms—are currently transformed under the shift from commodity ownership to digitized and commodified experiences (with Spotify as our prime example).

The project has taken a software studies and digital humanities approach towards streaming media. It has repeatedly engaged in reverse engineering Spotify's algorithms, aggregation procedures, and valuation strategies—and has hence repeatedly been non-compliant with Spotify's Terms of Service (ToS). Importantly, however, the project has been explicit and open about its methods and critical approach from the very start (in 2014). The project's research forms part of a tradition, well established within the social sciences and the humanities, which studies media actors critically. Studying industries critically means that our research is not *applied* research, that is, not conducted on any company's behalf against payment—but in the argumentative form and based on the standards of problem-oriented humanist inquiry. Different forms of public activism have, for example, been part of the project's 'interventionist' research design. In blog posts, newspaper articles, public service interviews—in both television and radio—the project's interventionist strategy has been outlined as based on digital methods, including the use of bots and the dissemination of self-produced sounds via Spotify.

The project in many ways followed calls to use the same tools that organize online information, due to Spotify's reluctance to share data. Our encounters with Spotify began with a brief conversation with (then) Head of Marketing, Sophia Bendz, in Stockholm in the fall of 2012. Asked if Spotify would not qualify as a regular media company—given its declared business interest of providing content to audiences, while selling those audiences to advertisers—Bendz rushed to praise Spotify's achievements as a tech company. A follow-up meeting with a company executive in Spotify's headquarters (before having received the grant funding), led to mutual expressions of respect, but little more. At this point, Spotify had already been made aware of our research interest to work with the company's data. Over the years, we individually met with engineers and marketers, data geeks and academics related to the company. But conversations often fell apart as soon as Spotify's 'tech identity' was questioned.

This inspired a research approach that would go beyond interviews, direct observation, and other standard methods of media studies research. Not that we researchers avoided talking to Spotify employees—we met a few of them over the years—but as with any other media company striving for a monopoly position, listening to company spin and "industrial self-theorizing" felt not enough. Complementing such 'frontend' inquiries with experimental 'backend' studies of digital media infrastructure, metadata generation, and aggregation practices, the project aimed to initiate public debate about the often subtly changing standards, values, and politics of cultural dissemination online.

Spotify once offered a technical solution for music distribution, yet the aggressive discursive framing of Spotify's operation as being primarily technological has tended to obscure its long-term entrepreneurial, financial, and culture-changing strategies. Since 2015, for instance, Spotify has implemented a plan—and the technology—to generate data based on its music streaming that allow to study human behavior at scale. The company acts, in other words, not only as a music provider, but also as a private data broker.

In order to study such data flows, our project's key methodological suggestion has been to 'follow files' rather than those making, using or collecting them. Through the notion of 'following files' our project developed a habitual way of working with digital methods at Humlab (Umeå University), to understand both Spotify's general streaming architecture, as well as particularities of the service interface, its music recommendation engine, or radio functionality. We deployed different types of digital methods—from computational tracking of all data transmissions that occur every time one presses 'play', to working with bots as virtual listeners that logs their 'behaviours', or text mining approaches of scraped Spotify job advertisements. In addition, the way in which we have studied music aggregation from the inside— by launching our own record label—has been yet another way of looking closer at some of the entries (or 'holes') in the 'Spotify black box', not the least by getting access to different monitoring services at aggregators, or concretely learning what kind of metadata categories that are available (and obligatory) when music is being bundled into packages of differentiated data. A first thing to note, for example, when going *under the hood* is that the Spotify infrastructure hardly appears as an uniform platform. Rather it is downright traversed by data flows, file transfers and information retrieval in all kinds of directions—be they metadata traffic identifying music, aggregation of audio content, playout of streaming audio formats (in different quality ratings), programmatic advertising (modelled on finance's stock exchanges) or interactions with other services (notably social media platforms). Spearheading the new data economy of the 21st century, Spotify in many ways resembles a sprawling network of interaction that includes musicians and listeners alongside other actors and interests that have little to do with cultural commodities or media markets in a traditional sense.

During the summer of 2017 I received an email from a Spotify legal counsel who was "concerned about information it received regarding methods used by the responsible group of researchers in this project. This information suggests that the research group systematically violated Spotify's Terms of Use by attempting to artificially increase plays, among others, and to manipulate Spotify's services with the help of scripts or other automated processes." I responded politely—telling him what we had done, and that we would gladly share our research results with Spotify. I got no answer. A few weeks later, however, I received a letter from the senior legal advisor at my university. Spotify had apparently contacted the Swedish Research Council with the claim that our research was questionable in a way that would demand "resolute action", and the possible termination of financial and institutional support. Our research group were asked to formally describe and rebut the claims—we wrote:

"Let us begin by firmly stating that neither the aim, methods nor results of our project were in any way designed or used to cause harm to Spotify or any of its users—or to benefit commercially from non-authorized access to the service's proprietary data. Our results do not reveal any detailed information about Spotify's proprietary algorithms or software, or disclose information that might be harmful if it ends up in the hands of Spotify's competitors. Our scientific research and its scholarly findings are thus not a competitive threat to Spotify as a company. We have never violated the integrity of any Spotify user, or collected any personal data related to Spotify users, or illegally shared copyrighted content via Spotify. We have respected Spotify's—and any other company's—wishes to protect the integrity of their service and brand. […] Spotify mixes three very different standards in making its claim: *methodical*, *ethical*, and *legal* standards. Spotify is concerned about a possible violation of its Terms of Service (ToS). The framework of *user agreements* (or ToS) is all Spotify has to make this claim. While Spotify's insistence on our adherence to its user agreements is well-founded, it is also off the point as we have ended all activities that could be understood as being in violation of these agreements"

DH-research is embedded in 'the digital'—and so are its methods, from scraping web content to the use of bots as research informants. Within scholarly communities centered on the study of the web or social media there is a rising awareness of the ways in which digital methods might be non-compliant with commercial Terms of Service (ToS)—a discussion which has not yet filtered out and been taken serious within the digital humanities. However, DH-researchers will in years to come increasingly have to ask themselves if their scholarly methods need to abide by ToS—or not. As Amy Bruckman has stated, it might have profound scholarly consequences: "Some researchers choose not to do a particular piece of work because they believe they can't violate ToS, and then another researcher goes and does that same study and gets it published with no objections from reviewers."

In September 2017, the Swedish Research Council as well as Umeå University decided to close the case; Spotify's emailed request were thus turned down. My paper will recount our legal dealings with Spotify—including a discussion of the digital methods used in our project—but also more generally reflect around the ethical implications of collecting data in novel ways. ToS are contracts—not the law. Still there is a dire need for ethical justifications and scholarly discussions why the importance of academic research justifies breaking ToS.