Anthropology with algorithms?
An exploration of online drug knowledge using digital methods

Lisa Jenny Krieg, Moritz Berning, and Anita Hardon

Abstract
Based on a study of more than twenty thousand reports on drug experiences from the online drug education portal Erowid, this article argues that the integration of ethnographic methods with computational methods and digital data analysis, including so-called big data, is not only possible but highly rewarding. The analysis of ‘natively’ digital data from sites like Facebook, message boards, and web archives can offer glimpses into worlds of practice and meaning, introduce anthropologists to user-based semantics, provide greater context, help to re-evaluate hypotheses, facilitate access to difficult fields, and point to new research questions. This case study generated important insights into the social and political entanglements of drug consumption, drug phenomenology, and harm reduction. We argue here that deep ethnographic knowledge, what we term ‘field groundedness’, is indispensable for thoroughly making sense of the resulting visualizations, and we advocate for seeing ethnography and digital data analysis in a symbiotic relationship.

Keywords
data ethnography, virtual communities, social media, big data, harm reduction
Introduction

In recent years, more anthropologists have been venturing into the digital domain, creating ‘netnographies’ (Kozinets 2010) or ‘virtual ethnographies’ (Boellstorff et al. 2012) of how people from Trinidad use Facebook (Miller 2011) or how protesters in the United States use Twitter (Juris 2012). In this article, we discuss how we utilized approaches from digital anthropology, the digital humanities, and new media studies in a case study of online drug knowledge, and show how a digitally literate analysis of online media can be integrated with ethnography. The case study presents our analysis of an online archive of more than twenty thousand reports of drug experiences posted on the website of Erowid Center, a drug education organization, and its Facebook page.

Anthropology and the internet

The ubiquity of the internet and digital media continues to transform human social life, and with it scientific research (see Kelty et al. 2008). Anthropology is no exception, with digital media generating new research topics (Bonilla and Rosa 2015; Escobar et al. 1994; Fader and Gottlieb 2015; Nuttall and Mbembe 2015; Wilf 2013) as well as methods for data collection and analysis (Burrell 2009; Fischer et al. 2013; Garcia et al. 2009; Murthy 2008; Wesch 2007). The potential for analyzing ever-growing volumes of new kinds of data leads to the need to update conceptual frameworks, as anthropologists (Escobar et al. 1994; Fischer et al. 2013) and digital humanities scholars (Liu 2012, 21ff.) have remarked.

Anthropologists have documented how digital media become integrated in local cultural contexts and diversify (Miller and Slater 2000; Wilson and Peterson 2002, 453). As numerous scholars have argued in recent years, the internet – far from being placeless – is becoming increasingly local (Coleman 2010, 496; Miller 2011; Rogers 2013, 13), with the emergence of ‘new’ cultures around technology, including hacker culture (Coleman 2010b), jazz improvisation through algorithms (Wilf 2013), and the ‘maker’ movement (Berrebi-Hoffmann, Bureau, and Lallement 2015; Lindtner 2014), thus creating new sites for anthropologists to study. For example, the open-source programming scene has been examined as a culture of gift exchange (Zeitlyn 2003). Anthropologists have also shed light on how entanglements of power, communication, and production change through digital media, in contexts ranging from social movements, protests, and democratization (Bonilla and Rosa 2015; Fader and Gottlieb 2015; Juris 2012; Postill 2014) to academic publishing, ownership, and copyright (Kelty et al. 2008), to issues of security, surveillance, and secrecy (Masco 2014; Nuttall and Mbembe 2015; Postill 2014; Yongming 2005) and the critical study of algorithms (Kockelman 2013; Seaver 2012; Wilf 2013).
Most anthropological research, however, does not capitalize on the data produced by digital media, but instead observes or participates in virtual spaces. In doing so, it resituates ethnographic methods in a digital context (Bonilla and Rosa 2015; Boyer 2010; Coleman 2010a; Wilf 2013), without assessing the added potential of digital media. This practice resonates with what new media scholar Rogers (2013, 19) calls digitizing or translating the offline, an ‘ontological distinction … between the natively digital and the digitized, that is, between the objects, contents, devices, and environments that are “born” in the new medium and those that have “migrated” to it’. This ontological distinction between ‘native’ and ‘migrated’ neglects the many dynamic interactions, transgressions, and historical entanglements between different types of online and offline data and media (Good 2013; Liu 2012, 16). While we do not want to adopt Rogers’s dichotomy, we do want to focus on the distinct properties of data that have recently emerged out of heterogeneous assemblages of media platforms, programs and algorithms, corporations, users, and media practices. These digital data and media practices have distinct affordances that can be put to use for anthropological research.

In their study of the Twitter hashtag ‘#Ferguson’ and the protests around the police killing of the unarmed African American teenager Michael Brown, Bonilla and Rosa (2015) treat Twitter hashtags as a field site. While their work leads to an insightful ethnography, it could have benefitted from taking advantage of the distinct affordances of Twitter, which allows for live monitoring of a hashtag. When downloaded, these tweets form a comprehensive sample of everything that was tweeted under a certain hashtag during a certain period. The dynamics of events could then be retraced and analyzed afterwards. Skype interviews, online surveys, participant observation of discussion boards, and informant recruiting through Facebook are other common examples of what Rogers would term the digitization or translation of ethnographic methods (see for example Garcia et al. 2009; Murthy 2008; Reich 2014).

In this article, we respond to calls by anthropologists to engage with new digital methods in their empirical work. Beaulieu (2004, 145), for example, asks anthropologists to examine digital ‘traces’ produced in digital environments, such as location data, hyperlinks, and uploaded pictures or videos. Such traces have emerged only recently and have distinct affordances. Pink and colleagues (2016, n.p.) criticize the ‘distinct lack of research’ of ‘bringing together, for instance ethnography and big data as complementary approaches’. Working with digital data and digital methods creates new opportunities for anthropologists to gain access to difficult fields, to explore informant-based semantics of large user-created data sets, and to discover new questions for further research. Using digital data and methods can also provide more context and lead to new hypotheses. Murthy (2008, 849) emphasized
this when he argues that the ‘combination of participant observation with digital research methods ... may provide a fuller, more comprehensive account’ of virtual cultures.

Analysis of digital data can be facilitated by computational methods that increase ‘the capacity to collect and analyze data with an unprecedented breadth and depth and scale’ (Lazer et al. 2009, 722). For example, mobile phone data has been used for early recognition of health risks such as depression (Madan et al. 2012), while Twitter activity has been correlated with election outcomes (Gayo-Avello, Metaxas, and Mustafaraj 2011) or a movie’s box office revenues (Asur and Huberman 2010). We agree with computational social scientist Lazer and colleagues (2009, 721) and new media scholar Rogers (2013, 21) that research on the intersection of these fields should not be left to Google, Facebook, or the NSA, but should be pursued by the social and cultural sciences.

At the same time, the hype surrounding digital media and big data has been criticized as superficial, simplistic, and negligent of ethical concerns and issues of power (Crawford, Miltnner, and Gray 2014); as a mythology with exaggerated expectations (Boyd and Crawford 2012); and as serving the economic interests of the postindustrial state (Liu 2012, 10). Lazer and colleagues (2009, 722) and digital humanities scholar Liu (2012, 24) have also pointed to the gap between what large data sets have to offer and the current difficulties to conceptually accommodate their analysis.

Our research project has benefited from work done by the Digital Methods Initiative\(^1\) and the Utrecht Data School\(^2\). These research centers, among others, have developed tools and research protocols that can be used by social and cultural scientists to analyze digital data (Hochman and Manovich 2013; Niederer 2013; Rogers, Sanchez-Querubin, and Kil 2015; Schäfer 2011; Woltering et al. 2015). Such tools can be used for analyzing Instagram images according to their color composition (Hochman and Manovich 2013), for creating networks of Facebook users engaging with public Facebook posts (Rieder 2013b, 350), or for analyzing subcommunities of online forum participants (Seale, Ziebland, and Charteris-Black 2006). So-called application programming interfaces (APIs)\(^3\), easy-to-use digital tools like Netvizz (Rieder 2013b), and specifically programmed data ‘scrapers’ facilitate access to large data sets and help visualize them for analytical purposes.

\(^1\) [https://wiki.digitalmethods.net/Dmi/DmiAbout](https://wiki.digitalmethods.net/Dmi/DmiAbout)


\(^3\) Interfaces that allow access to applications such as Twitter, with the possibility of downloading specified data.
In this case study, we show how digital data offer new possibilities for anthropologists, different from data that is collected offline. We examine one popular drug education website, Erowid, whose virtual community would be difficult and time-consuming to study offline. We show how big data sets, containing information such as posted text narratives, posting times, Facebook page connections, and users’ comments and likes, can be downloaded, analyzed, and visualized, providing access to large volumes of semantically rich, informant-based data that can be integrated well with ethnographic knowledge. Rogers (2013, 19) coined the term ‘online groundedness’ to refer to ‘a research practice that learns from the methods of online devices, repurposes them, and seeks to ground claims about cultural change and societal conditions in web data’. As anthropologists, however, we do not strive to ground research fully online. Instead, we aim to achieve what we call ‘field groundedness’ — a deep ethnographic knowledge of a field that is not demarcated by the boundary between online and offline.

In our current digital world, people, objects, and metaphors not only move between continents due to globalization – as Marcus (1995) observed in his advocacy of multisited ethnography – but also between servers, apps, and digital media devices. To study connections and movements between on- and offline spaces, we use a definition offered by Burrell (2009, 189), who describes the field as ‘a network composed of fixed and moving points including spaces, people, and objects’ with ‘observable connections performed by participants’. Howard (2002, 561) emphasizes that this does not assume proximity or even spatiality in a physical sense, and calls on ethnographers to choose important nodes of the network ‘instead of choosing territorial field sites’.4

We selected Erowid for study, as our offline ethnographic inquiries found it to be a frequently used and trusted drug website. Several other projects in which the authors are involved provided the necessary field groundedness for this research. Due to limitations of space, we will not dwell much here on these ethnographic accounts, which can be accessed in other publications (van Schipstal et al. 2016; Berning and Hardon 2016). Moritz Berning was engaged in three parallel projects of ethnographic data collection before and during this digital analysis. From February to April 2015, he pursued a virtual ethnography of an online community focused on ‘designer drugs’ (synthetic drugs), and collected data through lurking, participant observation, online conversations with community members, and seven in-depth interviews via Skype and in person. He is currently involved in research on the dosing of licit and illicit drugs, entailing fieldwork in Amsterdam and Berlin clubs, festivals, and private

after-parties, and studying videos posted on YouTube. Our methods and findings were also discussed with ethnographers of the Chemical Youth research project at the University of Amsterdam, who work on topics such as psychedelics in recreational and medical settings, party drugs, peer harm-reduction strategies, and drug use in challenging work situations, such as sex work or working as a DJ at nightclubs.

Erowid as organization, community, and archive

Erowid started out in 1996 as a website, a database or ‘drug library’ (Bogenschutz 2000, 251) dedicated to publishing unbiased information about licit and illicit psychoactive substances ranging from green tea to LSD. Today, the site includes scientific sources, visual information, reports of user experiences, sources from popular culture, as well as information about meditation and altered states of consciousness.

Founded by a couple working under the pseudonyms Earth and Fire, and financed by donations, Erowid became a nonprofit educational organization in 2005. Erowid’s political agenda includes harm reduction and drug research, and the organization has been cited numerous times as an expert source in scientific journals (Ambrose et al. 2010; Corazza et al. 2012) and by the media (Witt 2015). Erowid intervenes in contemporary public health discourses as well as in ‘counterpublic’ discourses about the risks and benefits of certain drugs (Barratt, Allen, and Lenton 2014; Murguía, Tackett-Gibson, and Lessem 2007), many of which circulate information in online message boards, blogs, and alternative newspapers.

Erowid is also a community of members who upload and share their drug experiences. This practice of collective sharing adds knowledge about substances often not studied in formal frameworks (see Berning and Hardon 2016; Móró 2014). Biomedical scientists initially saw the website as a threat to public health and as undermining biomedical expertise (Boyer et al. 2001). Today, researchers acknowledge the website’s positive role in providing vital information about the risks and benefits of otherwise under-researched substances (Van Hout and Hearne 2015; Wax 2002).

Erowid can be placed in a tradition of sharing knowledge about self-experimentation with drugs, one that has found expression in textual form (Huxley 2004; Rätsch 2005; Shulgin and Shulgin 1991) and in other art forms such as weaving and music (Labate and Cavnar 2014; Shanon 2002).

---

5 Verifiable via [http://greatnonprofits.org/](http://greatnonprofits.org/)
**Ethical concerns**

Analyses of digital media and large datasets pose challenges in terms of privacy and ethics, complicated even more by the technicalities involved (Murthy 2008, 841; Garcia et al. 2009, 77; Lazer et al. 2009, 722). Digital social media have been compared to Foucault’s (1977) panopticon, said to ‘encourage users to turn the gaze upon themselves or to actually invite others to do so’ (Lupton 2012, 236). They are considered as part of a neoliberal ‘surveillance society’ (Lupton 2012, 235) where the sharing or exposure of data enables citizens to surveil each other and to be surveilled by the state (Nuttall and Mbembe 2015, S323). Large exclusive data sets may pose a risk to those whose data is collected when the data is controlled by private companies (Van House and Churchill 2008, 306; Mager 2012, 770) or by governments (Nuttall and Mbembe 2015, S318). Crawford and colleagues (2014, 1666) critique the voluntary exposure of private data, which they assert is encouraged by a rhetoric of the good citizen who is expected to support social and scientific progress.

Ethical considerations for online research are entangled with the ambivalence of the internet as a private and/or public space, leading to the question of whether online data needs to be treated as confidential (Garcia 2009, 73ff). The question of open or limited access plays a role in such considerations (Garcia 2009, 74). While we agree with Wilson and colleagues (2002, 461) that ‘ethical principles – of showing respect for people under study, of protecting their dignity and best interests, of protecting anonymity or giving proper credit, and of obtaining informed consent – apply online as well as in face-to-face contexts’, we concede that receiving informed consent from every individual may be impossible given the size of very large data sets or because users are no longer active. Addressing ethical concerns in digital contexts remains a tightrope walk.

When working with large data sets – even those publicly available on the internet – individual privacy can be protected by refraining from singling out individuals by means of aggregation (Bond et al. 2013) or by modifying quotes so that they can no longer be traced with simple Google searches. The supposed anonymity of large data sets can be deceiving, as was seen in the search query release from the global mass media corporation AOL in 2006, when journalists succeeded in identifying an individual based on a set of anonymous search queries by 650,000 users (Rogers 2013, 30). If data analysis reaches down to the level of individuals, particular care is necessary. In such cases, Garcia and colleagues (2009, 75) suggest considering the norms of the online community in question. Erowid’s main focus is the distribution of knowledge generated by a community of users who voluntarily share their experiences. In accordance with these values, we shared our insights gained through analyzing Erowid’s trip reports by publishing an interactive website and promoting it on Twitter, which was subsequently retweeted by Erowid. Afterwards, we contacted the Erowid
founders via email, and exchanged thoughts and ideas. As we did not pursue our analysis to the individual level, the privacy of Erowid users is protected in the presentation of our findings.

Digital analyses
We analyzed content from the Erowid website with digital methods specific to the data created and shared through Erowid’s online platform. Our ethnographic interest was to explore Erowid as a node for drug knowledge and use. We decided to look at the three roles of Erowid: as an organization, a community, and an archive, and we gathered data from each of these sites. To understand Erowid’s ecology of relations as an organization, we examined its connections to other Facebook pages, using the digital media tool Netvizz and the network visualization program Gephi to generate a visual representation of the network (Rieder 2013a). To explore the Erowid community, we more closely examined engagement on Erowid’s Facebook page, that is, ways Facebook users respond to Erowid posts. To understand Erowid as an archive, we used computational language analysis to visualize the content of the Erowid website, more specifically the ‘trip reports’, or drug experience reports, posted on https://www.erowid.org.

Analyzing Erowid’s organizational ecology
Our analysis of Erowid’s organizational ecology focused on Erowid’s environment, by exploring Erowid’s relationship to other organizations, institutions, and communities through its public Facebook page Erowid Center. When an individual Facebook user or a representative of an organization on Facebook ‘likes’ a page, it establishes a connection between their page and the ‘liked’ page. As one of the largest social media platforms, Facebook is interesting for social scientists because it contains data based on behaviors such as ‘liking’ that take place within the platform (Rieder 2013b, 347).

The practice of liking is binary: a like-connection either exists or it does not. There is no ‘almost-liking’ a page. But the intentions behind liking (and not liking) are often not apparent. The omission of liking – purposeful or not – establishes boundaries between communities that can be objectified in network visualizations and subgroup identifications. We agree with new media scholar Rieder (2013b, 347) that Facebook engagements ‘revolve around elements that have cultural significance – liking a page of a political party is more than “clicking”’.

To analyze Erowid’s ecology, we downloaded the page-like network of Erowid Center’s Facebook page with the application Netvizz. In 2009, new media scholar Bernhard Rieder
developed Netvizz as an application integrated into the Facebook platform to extract raw data of pages and groups (Rieder 2013b). The application requires a Facebook login and is subjected to Facebook’s privacy regulations. It runs on the servers of the Digital Methods Initiative at the University of Amsterdam and is designed to make computational tools more accessible to researchers. We then visualized the downloaded data with the open-source network visualization software Gephi 0.8.2 Beta. A Facebook page-like network shows the like-relations of public pages to other pages as a directed-network graph. Pages transgress the online-offline divide: they can represent offline organizations, institutions, and public personalities, as well as online communities. (Individuals have Facebook ‘profiles’ not ‘pages.’) In the resulting network, each page corresponds to one node; one ‘like’ from page A to page B establishes a link, a so-called edge, between the nodes. The network shows an ecology surrounding the focal page, clustering into different topical areas, with many possibilities for analysis (Rieder 2013a). Similar methods to visualize and analyze such an ecology, which we will not examine here, might include the generation of a Twitter-follower network or an issue map (Rogers, Sanchez-Querubin, and Kil 2015).

In Gephi, the algorithm ForceAtlas2 was applied for spatially arranging the nodes. ForceAtlas2 is a layout based on two opposed forces: repulsion between the nodes and attraction through the ‘edges’ that link the nodes (Jacomy et al. 2014, 2). The movement emerging from this opposition leads to a spatial arrangement that separates communities, or clusters of nodes, thus enabling interpretation of the network (Jacomy 2014, 2). Frequently, nodes that have something in common also like each other more often than other nodes, which is why the ForceAtlas2 layout draws them together in clusters. These clusters of nodes have more connecting edges than the nodes lying further apart, and via these edges they are drawn together. Exploring the network and its clusters, which are created based on the collective practice of ‘liking’, can thus reveal insights about shared values and interests. Additionally, we used Gephi’s modularity algorithm (Blondel et al. 2008, 2) to color nodes belonging to different subcommunities.

The data we used to visualize the Facebook page-like network contains 1,641 nodes, corresponding to unique Facebook pages, and 9,368 edges connecting them. However, the meaning of the like-connection between pages remains unclear and can conceal diverse intentions, limiting what we can conclude from the data. The like network also masks the

6 [https://apps.facebook.com/netvizz/](https://apps.facebook.com/netvizz/)
7 [http://digitalmethods.net/](http://digitalmethods.net/)
8 [https://gephi.org/](https://gephi.org/)
chronology of relationships: it is always only a snapshot of the present in which traditional and stable relations appear identical to new, spontaneous, and unstable ones.

Figure 1

The Facebook page-like network graph (see figure 1) illustrates the ecology of Erowid’s organization with 361 nodes and 4,345 edges. The network is arranged in four peripheral, clearly separated clusters and two closely connected central clusters: the green peripheral cluster contains pages related to music, festivals, concert venues, and bands, which resonates with our team’s ethnographic work on festivals in the Netherlands and Germany. At many such festivals, organizations support ‘safe’ dancing by offering peer support in the case of problems caused by drug consumption. The connection to the field of public health is also visible through the big green node ‘Dancesafe’, the largest harm-reduction organization in North America. The light blue cluster on top contains pages related to art and spirituality and includes events and portals for spiritual art and poetry, a topic that the second author has encountered in the course of his research, but which none of our team has explored further. It reflects the historical entanglement between poetry and drugs, such as the twentieth-century Beat Generation (see, for example, Ginsberg and Schumacher 2015). Recognizing this as a field in its own right could extend research about psychedelics and creativity to the field of social and cultural studies (Kealen et al. 2015; Krippner 1985).

9  https://www.residentadvisor.net/features/2577
10  https://dancesafe.org/
A central pink cluster concentrates nodes on psychedelics, spirituality, and research including private research institutes and news portals on psychedelics and consciousness, an issue that one of our team’s ethnographers, Swasti Mishra, focuses her research on. The yellow cluster on the right, containing nodes on social and ecological activism, focuses on urban farms, educational sustainability initiatives, and intergenerational communities in the San Francisco area. The project’s ethnographers were familiar with the Bay Area as a center for the psychedelics community, but they were surprised by the range of social and ecological activism that was connected, a topic that could receive more research attention. The red cluster, the lower part of the central cluster, hosts our node of departure, Erowid Center, and revolves around psychedelics and drug legalization activism; a dark blue cluster features nodes on drug policy, nonprofit, and government organizations. Some of our team members were familiar with some of the nodes in these clusters, but had not focused on them specifically. In this case, the network revealed some interesting information, such as the connections between grassroots and institutionalized actors (for example, between ‘Dancesafe’ and several nodes in the blue cluster), and the fact that the majority of nodes contained a focus on legalizing marijuana.

This network can be analyzed by looking at each cluster individually and by exploring the connected nodes. It shows tightly interlinked clusters of social, political, and ecological movements, reflecting the values found in psychedelic drug culture (Lerner and Lyvers 2006). The visualization of these complex cultural and social entanglements contextualizes the different fields of psychedelic culture into a more holistic picture.

Analyzing Erowid’s community

Erowid’s community is hard to access, as the website itself has no social media functionality and does not provide much information about the authors of the trip reports. This is why we examined the Erowid Center Facebook page as an approximation of the Erowid online community (see also Wilson 2002). As the number of Facebook page likes was relatively small at the time of analysis (roughly three thousand followers), alternative analytical approaches based on Erowid’s Twitter presence, or on the page traffic of Erowid’s website, could have been used as well. For the coherence of this article, we limited our analysis to Facebook.

We used Netvizz to download all the posts on the Erowid Center page. We conducted three analyses on the page data, combining qualitative and quantitative methods. First, we plotted

---

11 For more information about her research, see https://vimeo.com/147143157.
page activity and user engagement on a timeline. Second, to get deeper insights into the sources of knowledge circulated and engaged with by the community of users, we analyzed the web domains in the links users posted, manually categorizing them into eleven different themes: academic, activism, books, drugs, lifestyle, news, science, social media, spirituality, sports and stores. The topical categorization exposed the most common websites from which drug knowledge is shared in posts on the Erowid Center page. Third, we accessed the larger Erowid community’s engagement through likes, shares, and comments, in order to understand which topics trigger the strongest interest.

The downloaded page data consisted of a tabular file with 491 posts, including comments and other forms of engagement, with anonymized user names, dating from 19 August 2013 to 18 July 2015, which we analyzed with Microsoft Excel. The most significant limitations of the data were the size and activity of the site. At the time of analysis (summer 2015), the page had a little more than three thousand followers and was thus relatively small. Analysis of the site’s activity showed that engagement on the page only started picking up in the spring of 2015. Most of the posts on the page were written by the site’s administrators, while users’ involvement was only visible through their likes, shares, and comments (summed up as ‘engagement count’).

The mixed-methods analysis of the posts on Erowid’s Facebook page provided insights into the dynamics of user activity and the circulation of knowledge through page administrators and followers (figure 2). The domain analysis (figure 3) shows a diverse picture of knowledge sources: more than half of the posts (53 percent) linked to news sites such as huffingtonpost.com, theguardian.com and nytimes.com. Twenty percent of the links led to drug-specific sites, such as erowid.org, ecstasydata.org, and shulginresearch.org. Lifestyle sites such as vice.com and nylon.com made up 8 percent of the links, and science and technology sites such as scientificamerican.com and wired.com accounted for 7 percent. A small number of links led to social media sites, social activism sites, academic publishing pages, online stores, and sites on spirituality, books, and sports. This collection makes for quite a diverse set of sources from which knowledge is gathered and then distributed via the Erowid Center page. Interestingly, the linked pages are not limited to drug-specific domains and include sites that do not necessarily subscribe to Erowid’s agenda. This analysis confirmed the emerging knowledge of Moritz Berning and other ethnographers on our team: many of the people we encounter in this research can be characterized as experienced, conscious, and educated drug users, similar to those described as ‘psychonauts’ (Doyle 2011), ‘e-psychonauts’ (Davey et al. 2012), or ‘drug savvy users’ (Van Hout and Hearne 2015, 31). The variety of sources in users’ posts reflects the interests of this type of user.
To better understand the Erowid community, we explored what kinds of knowledge are most often redistributed by users. Here the picture looks a bit different. First, we focused on the ten posts with the highest engagement count (figure 4). For these posts, the number of likes and shares was generally high, while the number of comments was low. The posts revolved around anniversaries associated with drug history, supporting Erowid, and studies about the benefits of LSD. In order to identify more controversial topics, we looked at the ten posts with the most comments. These controversial posts discussed topics such as drug policy and harms and benefits of drugs (figure 5). We concluded that the knowledge redistributed by users, through shares and likes, largely followed their agendas and beliefs, while more controversial engagement happened through commenting.
Figure 4

Figure 5

Analyzing Erowid’s archive

Our analysis of Erowid’s archive is based on the trip reports. These reports in Erowid’s ‘experience vault’ constitute a text-based, formalized collection of knowledge, a database that we think of as a digital archive (Manoff 2004, 10). It has also been called a ‘drug library’ with a high level of scientific accuracy (Bogenschutz 2000, 252). Trip reports undergo a process of submission, selection, and editing before they are published on the website (Witt 2015). The archive is filled through decentralized user engagement (see Liu 2012, 25), via voluntary submissions, but managed centrally by Erowid. It is an ongoing project, an open archive to which more trip reports are constantly added. As an accessible source of information, it is meant to educate the public about drugs, resonating with Joyce’s (1999) view on the central role of public archives for the liberalization of society.

Erowid’s trip report archive offers numerous possibilities for quantitative and qualitative analysis, which we certainly did not exhaust. We focused on experimenting with computational methods, such as natural language processing, to access the qualitative content. We began with a simple count of how many reports exist per substance, and...
combined this with a sentiment analysis of all reports on each substance, using the ‘Pattern’ module of the programming language Python (Smedt and Daelemans 2012). This type of analysis assigns values to evaluative words with the aim of extracting positive or negative sentiments from a text. As a method it has significant limitations, such as the inability to recognize irony, abbreviations, or slang. Its results thus have to be considered with care (Kennedy 2012, 437).

We proceeded to analyze the content by extracting the most frequent word pairs, or bigrams, from the reports for each substance, using the Python module NLTK. As an alternative method for accessing the content, we created keyword co-occurrence networks (Šišović, Martinčić-Ipsić, and Meštrović 2014) with the Python modules NLTK and networkx, as well as with Gephi. These were built by extracting the ten most frequent keywords from each report as network nodes, and establishing edges between those keywords/nodes that appeared together in the same report. The result was an undirected, weighted network graph that ordered words according to (partially) meaningful semantic clusters of practices, phenomena, drug use contexts, and effects. Creating the bigrams and the keyword co-occurrence network involves a highly subjective process of ‘tweaking’, or manual fine-tuning. Both bigrams and keyword networks involve ‘stopwords’, words that need to be ignored for the analysis, which typically include the most frequent words of a language (such as ‘I’, ‘have’, ‘over’, ‘the’, ‘is’). Lists of stopwords were tweaked until the results were satisfying. Tweaking the co-occurrence network also included filtering out less connected nodes and deleting nodes that distorted the results.

Finally, we conducted an analysis of the report metadata, which included the name(s) of each drug reported on in a particular report. We created a co-consumption network to explore the phenomenon of ‘polydrug use’, which has been described as a practice of simultaneous drug use that ‘maximises effects, balances or controls negative effects and substitutes sought after effects’ (EMCDDA 2002, 39). For each report in which the ‘substance’ field contained several substances, these substances were created as nodes and edges (links) were established between them. The goal of this analysis was to create a network representation in which substances that were frequently used together would cluster closely together.

We downloaded an archive of the trip reports from github, a platform frequently used by the open-source community to share code and cooperate on projects, in a dataset in the form of

---

12 http://www.clips.ua.ac.be/pages/pattern-en

13 http://www.nltk.org/
The dataset was uploaded by another interested user to github in October 2014 and contains 20,534 reports, dating between 2002 and 2014. The structure and the categories of a report in the json file look like this:

```json
{  'author': {    'gender': 'male',    'name': 'xxx',    'weight': 93},  'title': "A title",  'dose': [{    'amount': {      'grams': 0.22,      'unit': 'mg',      'quantity': '220'},    'substance': 'DXM',    'administration': 'oral',    'time': '0'}, {    'substance': 'oral',    'administration': '20 mg',    'time': '0'}, {    'amount': '(powder / crystals)',    'substance': '20 mg',    'time': 'AMT'}],  'report': [free text],  'date': {    'experience': '2002-01-01T00:00:00.000Z',    'submission': '2007-01-23T07:00:00.000Z'},  'id': 1,  'erowid': {    'id': '0000',    'views': '0'}}
```

The dataset has a number of limitations. First, it contains numerous errors created by users who did not fill in the form correctly or accurately, which for example makes it difficult to state how many different substances are mentioned in the reports. Second, the dataset only includes accepted, edited, and published reports, and only those experiences reported in the first place. The rejected and unpublished reports are not part of the dataset, which introduces bias. The dataset thus does not represent actual drug consumption practice but the experiences of users who decided to write a report, and then only those reports that passed Erowid’s selection and editing process. Third, we relied on the github user who scraped and uploaded the data set. This involves a certain amount of trust, which is however not unusual in the open source community (Zeitlyn 2003).

Our exploratory analysis of the Erowid archive brought to light several layers of meaning. The substance count (figure 6) shows cannabis to be the most frequently mentioned substance in the reports, followed at a distance by psilocin mushrooms, Salvia divinorum (a hallucinogenic plant), MDMA (‘ecstasy’), alcohol, and LSD, among others. The sentiment analysis plotted on the same graph shows values that interestingly do not correlate with the frequency of reporting. The high value of MDMA resonated with the many positive reports about MDMA that Berning heard over the course of his ethnographic research, and reflected other ongoing research about MDMA and its effects (see, for example, Kamboi et al. 2015). For further insight, the reports of the substances with the lowest and the highest values should be qualitatively explored. Another issue that should be followed up on ethnographically is the rather low value of caffeine and the effect it might have in combination with other drugs.
The bigram lists (figure 7) dive deep into the content of the experience reports. The findings are associative and touch on practices of preparation, dosage, and consumption; on setting and environment; and on positive and negative effects. It should be taken into account that this method works with probability: it shows what is frequently reported and neglects outliers and extreme experiences, but still offers a glimpse into the worlds of meaning and practice surrounding specific substances. Again, this data resonated with our ethnographic research. For example, the ayahuasca bigrams aligned with Berning’s ethnographic experience of ayahuasca rituals. The terms ‘Santo Daime’ and ‘Don Jorge’ feature prominently, relating to the Brazilian syncretic churches formed around ayahuasca as a traditional medicine (Shanon 2002). ‘Mimosa hostilis’ and ‘banisteriopsis caapi’ describe parts of the plant used in the most common recipes for the ayahuasca brew that always contains two substances, one acting as a MAOI inhibitor and one containing the DMT responsible for the visions (Ott 1996). ‘Brown liquid’ and ‘wood stove’ have to do with the materiality of the preparation process, while ‘divine presence’, ‘ego death’, ‘solar plexus’, and ‘sexual energy’ describe the effects of the drug, resonating with Shanon’s (2002) phenomenological study of ayahuasca. Bøhling (2015) suggests we understand drug use as an intersection, an assemblage of many different factors, all of which contribute to the actual drug event. This was confirmed by the bigrams, which covered a variability of the prominent aspects of the ayahuasca experience and emphasized the medical-spiritual approach of many users (Leonti and Casu 2013). It also suggests that the relevance of the traditional contexts of ayahuasca is still high, even though more recently the substance is also consumed detached from a ritual framework (see for example Leonti and Casu 2013). A broad, associative overview, as provided by the bigrams, helps to understand drug use more fully as not only caused by the
human drug-using subject, but as embedded within a network of humans, non-human elements, the environment, and music.

<table>
<thead>
<tr>
<th>Ayahuasca (123 reports)</th>
<th>MDMA (1639 reports)</th>
<th>Cocaine (467 reports)</th>
</tr>
</thead>
<tbody>
<tr>
<td>santo daime (28)</td>
<td>dance floor (139)</td>
<td>heart rate (29)</td>
</tr>
<tr>
<td>mimosa hostilis (18)</td>
<td>pure mdma (91)</td>
<td>white powder (27)</td>
</tr>
<tr>
<td>don Jorge (17)</td>
<td>glow sticks (43)</td>
<td>heart attack (26)</td>
</tr>
<tr>
<td>martial arts (15)</td>
<td>heart rate (43)</td>
<td>crystal meth (24)</td>
</tr>
<tr>
<td>drinking ayahuasca (13)</td>
<td>jaw clenching (42)</td>
<td>addictive personality (18)</td>
</tr>
<tr>
<td>choose pleasure (12)</td>
<td>rave scene (41)</td>
<td>cocaine hc (17)</td>
</tr>
<tr>
<td>divine presence (12)</td>
<td>orange juice (37)</td>
<td>fat line (15)</td>
</tr>
<tr>
<td>syrian rue (12)</td>
<td>term memory (31)</td>
<td>grapefruit juice (15)</td>
</tr>
<tr>
<td>banisteriopsis caapi (11)</td>
<td>panic attacks (29)</td>
<td>san diego (15)</td>
</tr>
<tr>
<td>ego death (11)</td>
<td>blood pressure (27)</td>
<td>coke drip (14)</td>
</tr>
<tr>
<td>solar plexus (11)</td>
<td>positive energy (27)</td>
<td>cracker jacks (14)</td>
</tr>
<tr>
<td>wood stove (11)</td>
<td>short term (27)</td>
<td>damn computer (14)</td>
</tr>
<tr>
<td>black hole (10)</td>
<td>absolutely amazing (26)</td>
<td>regular basis (14)</td>
</tr>
<tr>
<td>dark brown (10)</td>
<td>mdma powder (26)</td>
<td>winamp visualizers (14)</td>
</tr>
<tr>
<td>mother earth (10)</td>
<td>nitrous oxide (26)</td>
<td>coke user (12)</td>
</tr>
</tbody>
</table>

Figure 7

The keyword co-occurrence network (figure 8) is even more associative than the bigrams and should be understood as a research tool rather than a finding. Similar to the bigrams, it offers a glimpse of phenomena frequently associated with a substance and can be tweaked to fit one’s research interests. It can complement ethnography by putting known practices into a context, or be used to discover unknown phenomena as a starting point for further research. The keyword co-occurrence network of the substances GHB and GBL, for example, offers insights into dosing practices: the keyword node ‘cap’ on the upper left hints at the imprecise and risky way of dosing the substance using a bottle cap instead of with a needleless syringe. We are familiar with this dosing practice from ethnographic research, but the size of the node in the network makes it look much more common than we previously assumed. This caused us to begin a new research project specifically on dosing practices.
The co-consumption network of the Erowid trip reports (figure 9) allows new perspectives on the phenomenon of polydrug use, which has been observed in many cultural contexts (Harden and Idrus 2014; Labate and Cavnar 2014). From previous ethnographies and ongoing research, as well as other literature (for example, EMCDDA 2002; Scholey et al. 2004), we are familiar with people combining several substances to shape the drug experience according to their wishes: for a softer or a more intense trip, for an easier come-down, and for dealing with hangovers. The visualization in figure 9 allows us to identify clusters of co-consumed substances, based on the frequency of their combination. These clusters roughly correspond to chemical groups but also correlate with drug use practices, (sub)cultures, places and settings, and styles of music.
We identified four rather distinct clusters: 1) The pink cluster on top features loosely connected pharmaceuticals, brand-name painkillers such as tramadol (an opioid) or diazepam (benzodiazepine), antidepressants, and substances used for off-label party consumption. 2) The antidepressants merge into the green cluster on the left, which contains substances related to ‘cognitive enhancement’ (Frati et al. 2015). Herbal substances such as kava, kratom, and ginseng, used as mild anti-anxiety medications, are represented in the upper part of the green cluster. The lower green cluster features synthetic ‘smart drugs’ such as piracetam, adrafinil, and aniracetam, which are used for cognitive enhancement (Frati et al. 2015). It appears that an organic vs. synthetic discourse is present, similar to one about food that is ethnographically studied by members of our team (Mazzacano, D’Amato, and Falzon 2014). 3) The small yellow cluster at the bottom contains plant substances such as Syrian rue, a category called ‘ecodelics’ (Doyle 2011, 26), including ingredients for ayahuasca, its analogues (‘anahuasca’), and variants (‘pharmahuasca’, ‘vaporhuasca’). The substances in this yellow cluster are closely tied together because they are mostly used in an ecodelic or consciousness exploration context. 4) The dark blue cluster features drugs used in recreational party settings, including the most commonly consumed licit and illicit drugs:

cannabis, alcohol, tobacco, mushrooms, LSD, MDMA, and salvia. In the lower part of that cluster are substances that can be termed ‘designer drugs’ (EMCDDA 2015). Among the insights generated by this visualization is that cannabis is the leading co-use substance; the tightly connected party cluster surrounding it hints at the high frequency of these drugs’ co-use, which our ethnographers can confirm. The fact that drugs with similar effects cluster closely together was more surprising. It reflects a practice where similar substances are consumed together, rather than those with complementary effects (Ives and Ghelani 2006). The co-use network provides an entry point for further questions about and investigations into drug co-consumption.

Discussion

In our research journey, we followed, as suggested by Beaulieu (2004, 145), the digital ‘traces’ left by internet users: the pages they liked on Facebook, the comments they left on a Facebook page, and the texts they shared in Erowid’s experience vault. Inspired by Rogers’s (2013, 19) notion of ‘natively digital’ data, we chose to focus on data created within and through recent heterogeneous digital platforms, while not following his strict dichotomy of native vs. digitized. We believe the methods explored in this article can be used fruitfully by ethnographers of drug cultures to provide a ‘fuller, more comprehensive account’ of what happens online, as Murthy (2008, 849) hopes. The digital tools we employed allowed us to explore some of the broader dimensions of Erowid’s ecology, community, and archive. But instead of giving answers, they mainly raised questions that need to be pursued through more in-depth digital and offline ethnography. In this section we discuss the advantages and limitations of these methods.

The Facebook page-like network we used as a proxy for Erowid’s organizational ecology gave us an overview of the ideologies and social and ecological movements to which Erowid is connected. It both confirmed and broadened our ethnographic knowledge. The Facebook page-like network is a tool that can quickly contextualize the field and make relationships visible in a way that can support ethnographic fieldwork if used during the early stages of a project. Used after fieldwork, it can add validity to generalizations and be used to re-evaluate theories and assumptions. The limitations inherent in the medium of Facebook discussed above – the exclusive focus on public pages as actors, the multiple possible meanings of a ‘like’, and the lack of chronology – remain, and need to be taken into account.

Our analyses of the data downloaded from the Erowid Center Facebook page revealed which topics and controversies the Erowid Facebook community consumes and engages with. In our opinion, the value of the potential outcome outweighs the technical barriers that must be surmounted to employ these methods. Facebook’s privacy regulations limit access to data that are not clearly public and thus also the research questions that can be posed. The particularities of the Facebook medium pose further limitations; communities use the platform in different ways and Facebook does not play an important role for them all.

The analyses conducted on the texts in Erowid’s experience vault give glimpses into the language and experience of drug users who report these. As such, they are a research tool for anthropologists that fits the associative nature of fieldwork and the ‘poetic dimensions of ethnography’ (Clifford 1986, 26). The bigrams and keyword co-occurrence networks provide access to a vast body of informant-based semantics that can be used not only to contextualize known phenomena from the field but also to discover formerly unknown practices and experiences. Analyzing large data sets such as the trip reports permits the mapping of new substances’ effects, an advantage that has been used by Daniulaityte and colleagues (2015) in their semantic analyses of tweets about the use of novel forms of cannabis. Finding significant keywords hinting at risky drug practices, like the ones relating to dosage practices for GHB/GBL, is another possible use of natural language processing and keyword co-occurrence networks in the area of harm reduction and drug education. The analyses of the trip reports raised numerous questions, which is why we suggest using these data analysis tools within an iterative research process.

Further limitations of these tools lie in layers of bias and in the technical skills needed to use them. Bias is introduced in the analysis of trip reports at three points: first, through the limited sample of drug users who choose to report their experiences; second, through Erowid’s selection and editing process prior to publishing; and third, during the process of data cleaning, tweaking, and myriad small analytical decisions taken by the researcher. In addition, the methods used to analyze the experience reports demand high levels of technical skill. They are thus only suitable for researchers who are computationally inclined, who work on teams with others who are, or who have the means to outsource such tasks.

The methods presented here are best used in a symbiotic and iterative way, while triangulating data from different sources. We believe that the findings derived from the use of digital tools can only be properly interpreted with previous knowledge presented in literature and generated through fieldwork. The digital tools discussed here do not produce stand-alone results that provide clear-cut answers. Rather, they are tools that help researchers explore large data sets and discover new questions. For anthropologists, they are only useful in a tight, symbiotic relationship with ethnographic fieldwork.
Conclusion

In a world of connectivity and the ever-increasing presence of digital data and devices; in a society in which Google, Facebook, and the NSA monopolize the analysis of connections between data, social behavior, and cultural values; and in an academic environment where the cultural study of digital data is dominated by the digital humanities and media studies, anthropologists remain under-represented in research both of and with digital data. This case study of Erowid Center, a drug education organization and website, is an attempt to combine ethnography with digital and computational tools, and thus serves as an exploration of data-based ethnography.

With drugs increasingly purchased and discussed online (‘iDrugs’) (Janíková et al. 2016, 102); with more and more knowledge distributed on the internet; with local and national drug regulations causing stark differences in legality and prosecution; and with online fora, websites, and social media platforms connecting people around the globe, Erowid is indeed proof of the inseparability of online and offline worlds in a ‘glocalized’ drug culture (Robertson 1994). These entanglements produce new media and new forms of data alongside new analytical tools to access them. These entwinements, and the availability of digital tools and data, we argue, offer possibilities for anthropologists to enrich ethnographic data with the analysis of semantic worlds at larger scales.

The findings of such analyses can offer glimpses into worlds of meaning and practice, and they can provide anthropologists with means to contextualize practices, behaviors, and connections. These data and their visualizations can help discover unknown phenomena and make relationships visible in a way that can guide ethnographic fieldwork. They also demand caution in protecting individual privacy and careful decision making about issues of grounding and the ethnographic field. A field defined on the basis of movement and connection, resonating with Howard (2002) and Burrell (2009), we claim, can accommodate global connectivity and phenomena that take place in online and offline spaces, and that are mediated by websites, servers, and smartphones. Digital data depend on ethnographic knowledge for interpretation, and they can provide insights for developing new research questions and to re-evaluate hypotheses and assumptions. In contrast to Rogers’s (2013) notion of ‘online groundedness’, which proposes to completely abandon the offline baseline for grounding digital data, we argue in favor of ‘field groundedness’, the grounding of digital data in deep ethnographic knowledge of the field.

In the context of anthropological drug research, we find the analysis of digital data to hold promise for the field of harm reduction. Future research in this direction could have ethnographers, data analysts, and medical practitioners working together to map the effects
of psychoactive substances in order to deal more efficiently with overdosing emergencies and to monitor risky drug use behavior. For anthropology at large, we see many opportunities in working with digital media. These media capture data in the moment and entangle individual needs, social practices, cultural norms, political agendas, and commercial interests. Working with digital tools demands a certain degree of creativity and methodological openness from the researcher. Such tools also enable anthropologists to participate in the re-appropriation and sharing of digital content and to contribute their analyses to researched communities.¹⁷ In no way do we think that digital data and digital methods are capable of, or are at all likely to, replace ethnographic fieldwork. There is no shortcut or substitute for gaining the deep, ethnographic understanding of society through human interaction. But digital data are available and digital media are used by people every day. There is no reason why anthropologists should not follow suit.

About the authors

Lisa Jenny Krieg studied cultural anthropology in Heidelberg and Jerusalem and completed her PhD on German social memory in autumn 2016 in the Department of Cultural Anthropology at Utrecht University. She then worked in the Department of Anthropology at the University of Amsterdam on the circulation of online knowledge about psychoactive drugs with digital methods. Currently, she works as a postdoc in the Department of Geography at Bonn University on the intersection of digital technologies, human–animal relations, and global species circulation.

Moritz Berning holds a BA in sociology and film studies from the Gutenberg-University in Mainz, Germany, and a M.Sc. in medical anthropology and sociology from the University of Amsterdam. He wrote his master’s thesis on collaborative harm reduction in virtual communities who experiment with new designer drugs. He works in the Chemical Youth project as a research assistant on collaborative experimentations and virtual drug spaces. His research interests are virtual ethnographies, new digital methods, drug cultures, and the history of psychoactives.

Anita Hardon earned her PhD in medical anthropology from the University of Amsterdam in 1990. She is a professor of anthropology at the Amsterdam School for Social Science

¹⁷ We published selected parts of our analyses on http://chemicalyouth.org/visualising-erowid/ and tweeted the link to Erowid Center, which then retweeted it. Our website also has a comment function that enables discussion, which sparked an article by Inverse Magazine (https://www.inverse.com/article/12536-erowid-analysis-finds-people-like-cannabis-no-matter-what-other-drug-they-take) that used the site as a research tool to draw their own conclusions.
Research and scientific director of the Amsterdam Institute for Social Science Research (AISSR). She has researched new reproductive technologies, immunization, AIDS medicines, and self-care in the Philippines, Malaysia, Uganda, Pakistan, and Indonesia. Her research has generated important ethnographic insights on the appropriation of these technologies in diverse sociocultural settings, their efficacy in everyday life, the role of social movements in their design, and the dynamics of care and policy making in their provision. In 2012, she was awarded a European Research Commission Advanced Grant for her current multisited Chemical Youth project.

References
Anthropology with algorithms?


Wesch, Michael. 2007. ‘Web 2.0... the Machine Is Us/ing Us’. https://www.youtube.com/watch?v=NLlGopyXT_g.


