

This is the reference list for page 28 of the 2016 Insights for Impact report.



¹Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802–5805. doi:10.1073/pnas.1218772110

²Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. P. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, 26, 159–169. doi:10.1177/0956797614557867

³de Montjoye, Y. A., Radaelli, L., Singh, V. K., Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347, 536–539. doi:10.1126/science.1256297

⁴Jariyasunant, J., Abou-Zeid, M., Carrel, A., Ekambaram, V., Gaker, D., Sengupta, R., & Walker, J. L. (2015). Quantified Traveler: Travel feedback meets the cloud to change behavior. *Journal of Intelligent Transportation Systems*, 19, 109–124. doi:10.1080/15472450.2013.856714

⁵Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489, 295–298. doi:10.1038/nature11421

Insight headline	Big Data as a way of assessment
Theme	Big Data
Domain	Assessment
Proposed by	Lenka Knapová

Primary citations (max 2 – 1 original study; 1 review)

¹De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. *Proceedings of the 5th Annual ACM Web Science Conference*, 35, 47–56. doi:10.1145/2464464.2464480

²Jashinsky, J., Burton, S. H., Hanson, C. L., West, J., Giraud-Carrier, C., Barnes, M. D., & Argyle, T. (2014). Tracking suicide risk factors through Twitter in the US. *Crisis*, 35, 51–59. doi:10.1027/0227-5910/a000234

Most recent significant citation (2011-2015)

³Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. P. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, 26, 159–169. doi:10.1177/0956797614557867

Highest dissemination

³Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. P. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, 26, 159–169. doi:10.1177/0956797614557867

50-word summary of insight (non-technical)

Risk factors, personal characteristics, mental states, and trends for individuals and populations can be inferred from online behaviour (e.g., language expressed on social networks). This approach can help identify individuals and groups of interest (e.g., in need of intervention) and is more cost-efficient and faster than administering surveys.

Headline findings & critical numbers (simplify if overly technical)

Language and emotions expressed on Twitter can predict US county levels of atherosclerotic heart disease mortality more accurately than a model of 10 demographic, socioeconomic, and health risk factors (correlation with mortality rates $r=.42$ for Twitter, $r=.36$ for the 10 common factors)³.

Individual vulnerability and population levels of depression can be predicted based on behaviour recorded through Twitter (accuracy for individuals 70%; correlation with population data $r=.51$)¹.

Twitter posts indicative of suicide risk factors can predict suicide rates on the US state level ($r=.53$)².

Cautions & limitations

Most research is based on Twitter users as the data is publically available. However, Twitter users constitute only a portion of the total population (approx. 20%). Combination of more data sources (e.g., other social networks, purchase history, smartphone data) would be beneficial to provide even more accurate predictions. The issue of privacy needs to be addressed to obtain more data.

Insight headline	Predicting personal characteristics from digital footprints
Theme	Big Data
Domain	Personality (focus on ethics)
Proposed by	Lenka Knapová

Primary citations (max 2 – 1 original study; 1 review)

¹Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802–5805. doi:10.1073/pnas.1218772110

²de Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. (2013). Predicting personality using novel mobile phone-based metrics. *SBP'13 Proceedings of the 6th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, 48–55. doi:10.1007/978-3-642-37210-0_6

Most recent significant citation (2011-2015)

³Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108, 934–952. doi:10.1037/pspp0000020

Highest dissemination

¹Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802–5805. doi:10.1073/pnas.1218772110

50-word summary of insight (non-technical)

Personal information (e.g., sexual orientation, political views, and personality traits) can be predicted from even basic digital footprints. While this information might be used to offer more personalised products and services, it raises questions with regards to privacy and possible discrimination based on this information.

Headline findings & critical numbers (simplify if overly technical)

Sexual orientation can be predicted from Facebook likes (accuracy for males 88%, females 75%)¹.

Political views (democratic/republican) can be predicted from Facebook likes with 85% accuracy¹.

The personality trait openness to experience can be predicted from Facebook content ($r=.46$ with self-reports)³.

Levels of neuroticism can be predicted based on cell phone logs and gender with 63% accuracy².

Cautions & limitations

The accuracy of predictions should be kept in mind. As of now, even the predictions of variables with only two outcome values (e.g., male or female, democrat or republican) are still inaccurate in 10 to 30% of cases. Thus, crucial decisions on the level of an individual should not be based solely on a specific digital footprint (i.e., Facebook likes). Combination of more data sources such as other social networks, browsing data, purchase history, and smartphone data could provide more accurate predictions. Specific predictors of personal information might lose or change their predictive value over time (for example liking Pokémon used to indicate introversion, but this changed with the introduction of Pokémon Go that is popular across the whole introvert-extravert spectrum).

Insight headline	Privacy issues in Big Data
Theme	Big Data
Domain	Privacy
Proposed by	Lauge Haastrup, Altan Orhon
Primary citations (max 2 – 1 original study; 1 review)	
¹ Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. <i>Science</i> , 347, 509–514. doi:10.1126/science.aaa1465	
² Hasan, O., Habegger, B., Brunie, L., Bennani, N., & Damiani, E. (2013). A discussion of privacy challenges in user profiling with big data techniques: The EEXCESS use case. <i>Proceedings of the 2013 IEEE International Congress on Big Data</i> , 25–30. doi:10.1109/BigData.Congress.2013.13	
Most recent significant citation (2011-2015)	
¹ Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. <i>Science</i> , 347, 509–514. doi:10.1126/science.aaa1465	
Highest dissemination	
¹ Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. <i>Science</i> , 347, 509–514. doi:10.1126/science.aaa1465	
50-word summary of insight (non-technical)	
Privacy has a measurable impact on behaviour and well-being. Attitudes toward sharing sensitive information are vulnerable to manipulation. As it is possible to infer information beyond what people may intend to share, the ubiquitous observation implied by Big Data is likely to have real-world psychological effects on a large scale.	
Headline findings & critical numbers (simplify if overly technical)	
People are willing to spend or forfeit some money for privacy, but they are more likely to do so if they already have privacy ¹ .	
When buying embarrassing objects online and having been made clear of different retailers' privacy policies, people were willing to pay a 5% premium to protect their privacy.	
People can be influenced to divulge information: Participants in an experiment were twice as likely to answer intrusive questions online depending on website design ¹ .	
The length and ambiguous wording of privacy policies lead to poor understanding of the protections they grant. More than half (54%) of the privacy policies in a sample were written in language incomprehensible to more than half of users surveyed (56.6%) ¹ .	
Increasing perceived control over sensitive information can increase willingness to disclose that Information ¹ .	
Individuals have different concepts of what information they consider private and, as such, what information they wouldn't want to share or have inferred about them ² .	
Cautions & limitations	
Privacy needs and attitudes toward disclosure vary between not only individuals but also cultures, meaning studies may not generalize to all settings. Several paradoxes concerning privacy and disclosure are known, for instance people who express high concern for privacy are likelier to divulge information. There are competing explanations for why and under what conditions paradoxes such as this appear. Additionally, the rate of technological advances makes it difficult to determine the amount of additional data that can be inferred from what is already available.	

Insight headline	Empowerment can improve impact of Big Data on individuals
Theme	Big Data
Domain	Health psychology
Proposed by	Eline Van Geert
Primary citations (max 2 – 1 original study; 1 review)	
¹ Hansen, M. M., Miron-Shatz, T., Lau, A. Y. S., & Paton, C. (2014). Big Data in science and healthcare: A review of recent literature and perspectives. Contribution of the IMIA Social Media Working Group. <i>Yearbook of Medical Informatics</i> , 9, 21–26. doi:10.15265/IY-2014-0004	
² Kaptein, M. C., de Ruyter, B. E. R., Markopoulos, P., & Aarts, E. H. L. (2012). Adaptive persuasive systems: A study of tailored persuasive text messages to reduce snacking. <i>ACM Transactions on Interactive Intelligent Systems</i> , 2, 1–25. doi:10.1145/2209310.2209313	
Most recent significant citation (2011-2015)	
³ Simons, C. J. P., Hartmann, J. A., Kramer, I., Menne-Lothmann, C., Höhn, P., van Bommel, A. L., Myin-Germeys, I., Delespaul, P., van Os, J., & Wichers, M. (2015). Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. <i>European Psychiatry</i> , 30, 900–906. doi:10.1016/j.eurpsy.2015.09.004	
Highest dissemination	
³ Simons, C. J. P., Hartmann, J. A., Kramer, I., Menne-Lothmann, C., Höhn, P., van Bommel, A. L., Myin-Germeys, I., Delespaul, P., van Os, J., & Wichers, M. (2015). Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. <i>European Psychiatry</i> , 30, 900–906. doi:10.1016/j.eurpsy.2015.09.004	
50-word summary of insight (non-technical)	
People can feel empowered by collecting and sharing their data and by getting understandable, personalised feedback and advice on which actions to take based on the data. This can lead to a bigger impact of the personal Big Data on individuals' behaviour and well-being.	
Headline findings & critical numbers (simplify if overly technical)	
Self-monitoring to complement standard antidepressant treatment may increase patients' feelings of empowerment. A substantial number of participants receiving self-monitoring with or without feedback showed an increase in empowerment (19% and 29% respectively) and only a small number showed a decrease in empowerment (4% and 0%). In the control group 17% increased, but 21% decreased in empowerment ³ .	
Achievement feedback from mobile devices can increase self-efficacy in individuals striving towards a goal ⁴ .	
Health information provided should be clear and comprehensible for individuals, even those low in health literacy, for them to benefit from it and to be able to make more informed health choices ¹ .	
Mobile devices can leverage contextual information to provide personalised behaviour change support at convenient moments. Tailored persuasive SMS messages were more effective in reducing snack consumption than random choice of persuasive messages ² .	
Cautions & limitations	
Many existing health-related mobile applications are commercial in nature and not evidence-based. It is not clear whether these applications have a genuine, positive influence on behaviour and well-being. Apps and devices are often used without professional guidance or oversight ¹ . Such involvement might be necessary ¹ for reasons of safety, efficacy, privacy, and improved impact.	
Policy Assessment 2 Index	

⁴Achterkamp, R., Hermens, H., Vollenbroek-Hutten, M. (2015). The influence of success experience on self-efficacy when providing feedback through technology. *Computers in Human Behavior*, 52, 419–423. doi:10.1016/j.chb.2015.06.029

Insight headline	Using Big Data for social influence
Theme	Big Data
Domain	Social influence
Proposed by	Lauge Haastrup
Primary citations (max 2 – 1 original study; 1 review)	
¹ Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. <i>Nature</i> , 489, 295–298. doi:10.1038/nature11421	
² Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2013). Experimental evidence of massive-scale emotional contagion through social networks. <i>Proceedings of the National Academy of Sciences</i> , 111, 8788–8790. doi:10.1073/pnas.1320040111	
Most recent significant citation (2011-2015)	
² Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2013). Experimental evidence of massive-scale emotional contagion through social networks. <i>Proceedings of the National Academy of Sciences</i> , 111, 8788–8790. doi:10.1073/pnas.1320040111	
Highest dissemination	
¹ Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. <i>Nature</i> , 489, 295–298. doi:10.1038/nature11421	
50-word summary of insight (non-technical)	
Analysing social media networks with Big Data enables us to understand how and under which circumstances people influence each other online. This can be used to design interventions to alter behaviour across many domains, such as voting in elections or consumer behaviour.	
Headline findings & critical numbers (simplify if overly technical)	
Informing Facebook users of who among their close friends had voted in an ongoing election made users more likely to participate in the election themselves, generating 280.000 additional votes from the 6.3 million people whose voting information was available ¹ .	
It is possible to influence social media users to start using a product (movie/actor rating application) by showing them that their friends use the product. This method resulted in a 13% increase in product adoption. Further, it is possible to estimate how influential individual users are in making others adopt the application and how susceptible each is to being influenced by others. Thus, good intervention targets, i.e. users that are both susceptible and influential, and have susceptible peers, can be identified ³ .	
The emotions people express on social media have an influence on other users' emotions. Exposing users to fewer negative status posts from their peers influences their posting habits and they start posting more emotionally positive updates. Exposing users to fewer positive status posts from their peers has the opposite effect ² .	
Cautions & limitations	
No single social media platform covers all individuals or populations. Therefore, any intervention using one specific social medium will mainly reach only a specific population or a targeted group. Current effect sizes are small. The interventions do not influence any single individual very much, but can reach a large amount of people, making the intervention powerful nonetheless. Effects so far cover only a limited number of behaviours and lack assessment of long-term impact.	
Policy Assessment Index 3	

³Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, 337, 337–341. doi:10.1126/science.1215842

Insight headline	Transportation and Big Data
Theme	Big Data
Domain	Environmental, population, and conservation psychology
Proposed by	Altan Orhon

Primary citations (max 2 – 1 original study; 1 review)

¹Jariyasunant, J., Abou-Zeid, M., Carrel, A., Ekambaram, V., Gaker, D., Sengupta, R., & Walker, J. L. (2015). Quantified Traveler: Travel feedback meets the cloud to change behavior. *Journal of Intelligent Transportation Systems*, 19, 109–124. doi:10.1080/15472450.2013.856714

²Poslad, S., Ma, A., Wang, Z., & Mei, H. (2015). Using a smart city IoT to incentivise and target shifts in mobility behaviour—Is it a piece of pie? *Sensors*, 15, 13069–13096. doi:10.3390/s150613069

Most recent significant citation (2011-2015)

²Poslad, S., Ma, A., Wang, Z., & Mei, H. (2015). Using a smart city IoT to incentivise and target shifts in mobility behaviour—Is it a piece of pie? *Sensors*, 15, 13069–13096. doi:10.3390/s150613069

Highest dissemination

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, 6, 8166. doi:10.1038/ncomms9166

50-word summary of insight (non-technical)

Sensor networks deployed in “smart cities,” integrating data from social networks and smartphones, give administrators a rich, real-time view on mobility and communication in the urban landscape, which informs planning decisions, facilitates intelligent transportation routing, and drives persuasive sustainability programs that change the way people think about travelling.

Headline findings & critical numbers (simplify if overly technical)

A location-aware travel feedback application tested on commuters in San Francisco reduced average weekly driving distances by one third, which was found to correspond with an attitudinal shift driven by peer influences and increased awareness. Commuters described as frequent drivers were observed to walk 5 kilometres more in the third week of the program than in the first week¹. The SUNSET Project, tested on commuters in three western European cities, demonstrated that individuals could be induced to alter their travel schedules (i.e. to avoid rush hour) and to change their mode of transport (e.g. car to bicycle) by a location-aware travel incentives application that integrated user data and data from fixed sensor networks in the city. In Enschede (Netherlands) and Gothenburg (Sweden), approximately 15% of participating drivers shifted their departure times to reduce congestion, for which each received 100 points with no exchange value².

Cautions & limitations

The programs described above were short-lived, small-scale proofs of concept that require further validation^{1,2}. The feasibility and expense of integrating real-time sensor data from infrastructure (e.g. traffic lights) may vary considerably depending on installed technology and regulatory considerations. Integrating these data sources is important because data provided by volunteering smartphone users may not provide sufficient evidence for interventions targeting the transport network.

Policy Assessment Index

Insight headline	Using social media to identify emerging substance use patterns
Theme	Big Data
Domain	Substance abuse
Proposed by	Altan Orhon

Primary citations (max 2 – 1 original study; 1 review)

¹Cameron, D., Smith, G. A., Daniulaityte, R., Sheth, A. P., Dave, D., Chen, L., Anand, G., Carlson, R., Watkins, K. Z., & Falck, R. (2013). PREDOSE: A semantic web platform for drug abuse epidemiology using social media. *Journal of Biomedical Informatics*, 46, 985–997. doi:10.1016/j.jbi.2013.07.007

Most recent significant citation (2011-2015)

²Deluca, P., Davey, Z., Corazza, O., Di Furia, L., Farre, M., Flesland, L. H., Mannonen, M., Majava, A., Peltoniemi, T., Pasinetti, M., Pezzolesi, C., Scherbaum, N., Siemann, H., Skutle, A., Torrens, M., van der Kreeft, P., Iversen, E., & Schifano, F. (2012). Identifying emerging trends in recreational drug use; outcomes from the Psychonaut Web Mapping Project. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 39, 221–226. doi:10.1016/j.pnpbp.2012.07.011

Highest dissemination

³Baumann, M. H., & Volkow, N. D. (2015). Abuse of new psychoactive substances: Threats and solutions. *Neuropsychopharmacology*, 41, 663–665. doi:10.1038/npp.2015.260

50-word summary of insight (non-technical)

Social media analytics are being used to monitor trends in drug use and the emergence of new psychoactive substances in near-real time. This supports clinicians, administrators, and law enforcement in gauging attitudes, assessing readiness, and responding actively to potential threats much earlier than is possible with traditional monitoring methods.

Headline findings & critical numbers (simplify if overly technical)

Automated monitoring of drug users' public interactions on social media, broad web search data, and other sources are being used to detect trends in drug use^{1,2}.

In two years, the EU-funded Psychonaut Web Mapping project provided early warning of 414 new substances, including the now-banned mephedrone, which were discussed in online niches years before they became bona fide threats to public health².

New psychoactive substances designed to skirt drug laws are introduced to market continually with reports of new variants in Europe rising from fewer than 20 annually 2005-2008 to 101 in 2014.

These substances are sold openly online. As most of them are unstudied, knowledge of which ones are gaining popularity is important to risk assessment and response preparation^{2,3}.

Cautions & limitations

As this technology is based on data generated by internet users, it lacks the ability to pick up on trends not being reported online, such as those that might be present in communities with low internet access. People in these communities are among those most vulnerable to the health impacts of drug use. Self-reported drug use descriptions online must be approached with caution as they cannot be verified, especially if the substances in question are misidentified (e.g., fentanyl sold as heroin).