

THIS WEEK

EDITORIALS

WORLD VIEW Native Hawaiian culture is not opposed to modern science **p.7**

FAR FROM BLAND Intensively farmed vanilla endangers plant biodiversity **p.9**



FIRE AND ICE Carbon from Arctic wildfire equals annual emissions of Belgium **p.10**

Digital-data studies need consent

Anonymized data sets are growing and it is becoming easier to identify individuals. Research-consent procedures must be updated to protect people from being targeted.

People today shed data wherever they go. Data flow from their financial transactions, social-media platforms, wearable health monitors, smartphone apps and phone calls.

By tapping massive digital data sets collected by phone providers, technology companies and government agencies, researchers hope to reveal patterns in the data and ultimately to improve lives. Such studies range from an analysis of call records in Nepal that showed where people moved to following an earthquake, so that aid could be delivered; to estimates of pollution exposure based on location data from the Google Maps smartphone app. But relatively little attention has been given to the ethics of how this research is conducted and, in particular, how those who supply their data should consent to taking part.

In general, proposals for research involving people are vetted by guidelines rooted in the 1947 Nuremberg code and the subsequent 1964 Declaration of Helsinki. These are ethical principles forged after unconscionable Nazi experimentation during the Second World War. They demand that researchers obtain voluntary consent from people who understand the subject matter of the study well enough to make an informed decision about whether to take part. But informed consent is often not required for studies that access anonymized and pooled data.

One reason is that, in theory, such data are no longer connected to a person. But in fact, risks remain. Many studies have shown that individuals can be identified within anonymized and aggregated data sets. Last week, researchers from Imperial College London and the Catholic University of Louvain in Louvain-la-Neuve, Belgium, demonstrated in a paper published in *Nature Communications* (L. Rocher *et al. Nature Commun.* **10**, 3069; 2019) how it is possible to re-identify people, even when anonymized and aggregated data sets are incomplete.

One implication is that vulnerable individuals and groups — including undocumented immigrants, political dissidents or members of ethnic and religious communities — are at risk of being identified, and therefore targeted, through digital-data studies. A News feature in *Nature* in May described examples of potential unintended consequences of tracking locations of populations through anonymized, aggregated phone-call records (see *Nature* **569**, 614–617; 2019).

ASSESSING THE RISKS

Concerns about potential misuse also apply to anonymized and aggregated data derived from smartphone apps, social networks, wearable devices or satellite images. Right now, the decision on whether the benefits of digital-data studies outweigh the risks largely falls to the researchers who collect and analyse the data — and not to the people who are unwittingly taking part.

The Nuremberg and Helsinki principles for informed consent evolved to correct this imbalance. Yet consent is complicated in the age of big data. Unlike in most biomedical studies, researchers who use digital data sets rarely gather the primary data themselves. Rather, telecommunications companies, tech firms and national agencies collect the information and decide whether to allow research on it.

If people being monitored were given an option to share their data for study, the consent would need to be relatively open-ended. This is, in part, because studies of big data search for unexpected patterns. Moreover, they can lead to results, or to potential applications that cannot be predicted. For example, researchers studied anonymized phone records from millions of callers in Turkey to see whether the location and movements of Syrian refugees in the country could

“If consent is offered at all, it’s often no more than a box to tick in the terms and conditions.”

reveal aspects of their lives that might one day inform helpful measures. The researchers could not have asked participants to share their data for a defined purpose because the researchers themselves did not know where their studies would lead.

In the United States, studies using anonymized, aggregated data are allowed under the ‘broad consent’ clause of the Common Rule, the federal policy governing research on people. But broad consent does not equal informed consent, because participants don’t know how exactly and why their data will be used, nor will they be aware of potential harms. In the European Union, researchers using anonymized, aggregated data are exempt from complying with the General Data Protection Regulation.

If consent is offered at all, it’s often no more than a box to tick in the terms and conditions that few people read as they rush to activate their phone service or app. And big-data studies often disregard a crucial principle in other research involving people — that participants should be allowed to withdraw from a study at any time. That’s because it is technically very difficult to extract and remove a person’s data from a de-identified, pooled data set.

When properly carried out, informed consent — the gold standard in medical research — includes a conversation between clinical researchers and study participants. It is hard to imagine how such conversations could be replicated among millions of people signing on to an app, but that’s no reason to give up.

In the growing field of data governance, computer scientists, bioethicists and legal and human-rights scholars are concentrating on how to return agency to the people from whom the data derives. Ideas range from tagging the data as they are being collected, so that individuals can see how this information is being used, to creating institutional review boards capable of assessing the safety of big digital-data studies.

Conversations around digital consent are happening, but must be given more urgency. They need to be led by organizations that are independent of governments and industry, such as national data regulators, so that powerful interests do not dominate. That said, they should include companies that collect the data, as well as ethicists, human-rights organizations, national science academies and researchers who carry out studies using digital data.

The Nuremberg code was written to protect innocent people from the risks of harm. Those risks have not gone away, which is why there needs to be an updated set of guidelines fit for the digital age. ■