



A New Take On Privacy By Design

— A project launched in Senegal and Colombia aims to prove it's possible to securely and ethically extract private sector data for the public good.

● **During the 2014 Ebola outbreak** in West Africa qualified researchers were prevented from accessing relevant mobile phone data that could have helped curb the epidemic despite the fact that mobile phone operators, mobile industry association GSMA, and United Nations agencies pushed for its release. At issue were privacy concerns. Data sources collected by private companies, such as call detail records by telecom operators, alongside traditional survey data and official statistics, combined with artificial intelligence, have the potential to unveil socio-economic information at levels of granularity and complexity never seen before. Such data could be used to inform decision-making on epidemics, poverty, inequality, crime, traffic, waste and more. But, without proper safeguards the same data could recoup sensitive personal information and tie it back to individuals. That explains the concerns and why no process is in place to allow such data to be used. Enter OPAL, short for Open Algorithms, a not-for-profit socio-technological innovation developed by a consortium that includes the MIT Media Lab, mobile operator Orange, the World Economic Forum, Imperial College London and the Data-Pop Alliance, a group backed by MIT Media Lab and the Harvard Humanitarian Initiative, in close partnership with Telefónica and the governments of Senegal and Colombia. Building on years of work conducted by this group and others, it aims to crack what it sees as one of the single biggest conundrums of the age of intelligent connectivity: How to ethically unlock the potential of private sector data

for the public good. “OPAL is trying to get at the causes of the world’s greatest ills and see what role data could play in solving them,” says Emmanuel Letouze, director and co-founder of the Data Pop Alliance. “We as a society have to find systems and standards that people will trust so that data can be used for good.” With the help of its team of engineers and social scientists, OPAL says it has found a way to ethically extract relevant information for the public good from an array of private companies, including mobile phone operators, banks, retailers, energy companies and logistic providers. The hope is that OPAL’s platform can serve as a model and eventually be adopted across the globe, says Nicolas de Cordes, Orange’s vice president, Marketing Anticipation, who has played a key role in the project from the start. OPAL is based on an open source platform developed by the MIT Data Trust Consortium and Imperial College London. The data stays within the premises of the private company and third parties only access it through Open Algorithms that provide a safe question-and-answer system. The questions are validated in advance by a board of advisors comprised of experts and local members of the community. Real-world deployment of OPAL started in mid-2017 in Colombia and Senegal with pilots leading to minimum viable products (mvps) in two phases. These pilots are being run in partnership with Senegal’s national statistics office (ANSDD), Colombia’s national statistics office (DANE) and national planning department (DNP), as well as two other major telecom operators, Senegal’s

Dakar, Senegal and Bogota, Colombia, are capital cities of the two countries currently piloting OPAL, a non-for-profit project that has found a novel way to extract private sector data for the public good. The project includes the participation of mobile operators Orange, the MIT Media Lab and the World Economic Forum, among others, in close cooperation with the governments of Colombia and Senegal.

Orange-Sonatel and Telefónica Colombia. Core funding for the MVP phase of €1.5 million was provided by the French development agency (AFD), with additional support from the World Bank, the Global Partnership for Sustainable Development Data and the Sustainable Solutions Development Network. There are plans to launch pilots in two more countries and one more industry by 2022. Getting to this point has been anything but easy. “Allowing you to use data while truly preserving privacy is a hard problem,” explains OPAL team member Yves-Alexandre de Montjoye, an assistant professor at Imperial College London and a special advisor to EC Commissioner Margrethe Vestager. One of the reasons call records have not been used for public good until now is that recent studies show that pseudonymization and standard de-identification techniques are not sufficient to prevent users from being re-identified in mobile phone data. Four data points — approximate places and times where an individual was present — are enough to uniquely re-identify a unique digital trace 95% of the time in a mobile phone dataset of 1.5 million people. This unique trace might sometimes be a key to then reidentify a single person. Opal turns the problem on its head, he says.

A Big Task

Setting up the contractual and institutional architecture of the project has also been a big task. “It has involved MOUs with telcos, an agreement with statistical offices, signing agreements with friendly user testers and launching an initial version of the platform in Senegal, plus setting up governance in the form of local councils to oversee ethical development, discuss legitimate use cases, and advise on use cases that are too sensitive or risky. “Now we are at the stage where we want to move into the beta phase and apply the data to things like poverty monitoring and education,” says de Montjoye. “That is going to be our focus in 2019 and 2020, as well as opening one additional country and expanding into the use of data from the electricity and/or banking sectors.” The French mobile operator joined OPAL because it believes anonymized telco data – along with data from enterprises in other sectors - can be used to help achieve the U.N.’s sustainable development goals, save lives during periods of crisis, and improve education and city services, says de Cordes.

Orange already has a commercial business that packages and sells telecom data. The business, called Flux Vision, analyzes population flows in real-time using data from Orange’s mobile network. It converts data from the mobile network into statistical indicators to analyze how often different geographical areas are visited and how people move around. In addition to location indicators such as density, provided by the mobile network, the service offers anonymous socio-demographic data such as age, gender and socio-professional category, with the aim of giving local authorities and businesses greater insight into the profiles of their customers and users. Some of this type of information – and a broader set of data from other



players - is badly needed by governments and NGOs from a not-for-profit source. When it comes to population density “the census is done every 10 years and in between you don’t have any sense of where people are – at any given point in time during the day,” says Letouze. “There is often major flooding in Dakar, and in Colombia there are frequent landslides but when you need to send rescue teams you don’t know where to send them.” Distress messages posted on social media are what Letouze calls “a false positive.” If people are tweeting it means they have cellular service and are probably better off than the ones no one knows about, he says. “That is why knowing the population distribution 10 minutes before or while a crisis is happening is really powerful and can help you put rescue teams where people need it most.” Real-time population density information can also help determine where new hospitals or schools should be located. All these applications will lead to better decision making, provided you also manage the bias in your data and the fairness and transparency of your algorithms, say Letouze and de Cordes. The key is calibrating the models so that researchers and governments have the information they need without being too granular and breaching both privacy and trust. “It is amazing what open algorithms can do,” says Letouze. “We are really at the forefront of showing how data can be safely used for good.” ●

J.L.S.