OPAL's vision to leverage data for societal development*

Alex 'Sandy' Pentland † and Emmanuel Letouzé $^{\sharp}$

February 2017

The Open Algorithms (OPAL) project is a socio-technological innovation to leverage private sector data for public good purposes by "sending the code to the data" in a privacy preserving, predictable, participatory, scalable and sustainable manner. It has two main objectives: providing a far better picture of human reality to official statisticians, policymakers, planners, businesses, and citizens, while enabling greater inclusion and inputs of all members of societies on the kinds and uses of analyses performed on data about themselves. OPAL, developed by a consortium composed of Data-Pop Alliance, Imperial College London, the MIT Media Lab, Orange and the World Economic Forum, builds on years of work of this group and others, and is a key milestone towards realizing a vision where data is at the heart of societal development around the globe, in support of the UN Sustainable Developments Goals and democracy.

OPAL came out of the recognition that accessing 'big data' sources held by private companies for research and policy purposes has been a conundrum. To date, one of the most valuable of them, known as Call Detail Records (CDRs), collected by telecom operators for billing purposes, have been accessed and analyzed externally either through data challenges such as the Orange's D4D Challenges or through bilateral arrangements under Non-Disclosure Agreements. These types of engagements have offered ample evidence of the promise and demand. A large body of academic literature has showed how computational analysis of CDRs (but also banking data), typically alongside traditional survey data and other official statistics, can help capture socio-economic outcomes and processes at high levels of geographic and temporal granularities and degrees of complexities—including disease spread, poverty, literacy, crime, as well as optimize public service delivery and transportation systems, notably.

At the same time, there are risks associated with using such personal, connected data. Concerns for privacy and security have grown as the notion of data anonymization was increasingly tested, as evidenced by research that has come out of MIT, and the nature and extent of the surveillance activities of the US National Security Agency publically revealed by Edward Snowden, fueling fears of an Orwellian future. The prospects of growing imbalances between groups that have access to the data and capacities and those who do not, and the resulting concentration of power, are also worrisome. A related criticism is that algorithms increasingly used to make policy decisions are akin to 'black boxes' concealing rules and procedures that cannot be subjected to public scrutiny and redress.

In light of these obstacles and requirements, OPAL's vision is the development of a new type of techno-institutional system built on trust that leverage private sector data to foster transparency, agility, accountability and inclusion while respecting privacy and security. To that end, OPAL will reflect and foster a paradigmatic shift to turn Big Data on its head and save it from itself. The first step is to send the algorithms to the data, not the other way around, so that data is not exposed to theft and misuse. The second step is to co-design how big data algorithms are used, so that they served local needs and respect local standards, instead of imposing external perspectives and expertise.

^{*} Additional information can be found on the OPAL project website www.opalproject.org and questions sent to contact@opalproject.org

⁺ Professor, MIT; Academic Director, Data-Pop Alliance

^{*} Director, Data-Pop Alliance; Visiting Scholar, MIT Media Lab

OPAL will start with pilots in Senegal and Colombia with 1.5m Euros in funding from the *Agence française de développement* (AFD), in partnership with their National Statistical Offices—respectively ANSD and DANE—and leading local telecom operators—Orange-Sonatel and Telefónica Colombia. In its initial phase of OPAL's deployment, requests for predetermined indicators—e.g. population densities—will be sent via the platform through predeveloped algorithms running on the companies' data servers, behind their firewalls, in a privacy-preserving manner, and results will be made available via an interface. Local engagement and empowerment will be central to the development of OPAL: needs, feedback and priorities will be collected and identified through local workshops and discussions, and their results feed into the design of future algorithms. These algorithms will be open, therefore subject to public scrutiny and redress. A local advisory committee on ethics and development—*CODE* in French and Spanish—will provide guidance and oversight to the project, to ensure it abides by key ethical principles. In addition, training will be delivered around the project to foster its use and diffusion as well as capacities and connections more broadly.

Early results and thematic use cases in Senegal and Colombia are expected to show the value of such an approach and the need for replication and expansion. For example, Senegal has faced catastrophic floods in recent years that have created irreversible damage, killed dozens of people and washed out large swaths of crops. These not only have immediate effects, but also lead to suboptimal risk-minimizing investment decisions. Colombia faces challenges of its own, including regular landslides and well as high levels of violence and inequality as it emerges from five decades of civil conflict. There as well, OPAL will help provide better indicators and insights and instill dialogue between groups.

OPAL could help save and improve lives. In the short and medium terms, OPAL will help produce standard CDR-based indicators at high levels of temporal and geographical granularity that will in turn allow estimating key statistics at local levels. These are expected to improve policies over time. OPAL will also allow local institutions, individuals and groups to weigh in on the kinds of statistics they deem important to monitor. In the medium to long term, our vision is for OPAL to help monitor some of the 17 Sustainable Development Goals and their 169 targets and contribute to better policies. Finer grained analysis may be monetized under a *freemium* model to be defined over time.

OPAL has been officially endorsed by a wide array of Senegalese and Colombian stakeholders from the public and private sectors as well as academia and civil society, including the Minister of Finance of Senegal, the Head of the National Planning Department in Colombia, both Heads of the National Statistics Agencies, the CEOs of the local telecom operators, as well as many research and non-governmental organizations. The same model will be applied in other countries where OPAL will be deployed. OPAL should become and provide a set of standards for how to locally and globally leverage private sector data to foster development and democracy. Over the period 2017-20, funding needs for 3-5 additional pilots including all related research, development, training and engagement activities are estimated in the range of \$5-8m.