



Global Partnership for Sustainable Development Data

OPAL Case Study

The purpose of this case study is to better understand GPSDD's contribution to and outcomes to-date of the collaborative Open Algorithms (OPAL) Initiative.

To capture the dynamism of the collaborative process, the case studies are designed to explore outcomes at the various stages of the process which can be broadly categorized as: 1. Catalyze connections between stakeholders; 2. Coordinate to understand the issue, develop, and prioritize solutions; 3. Impact decision-making by using, sharing, adapting, and re-using data. This activity is considered to be in stage 2. We hope to do a follow up case study to further explore and document outcomes from stage 3.

This case study was written in collaboration with DataPop Alliance and includes direct inputs from Orange, IPAR, FIP, and panelists from the 2018 OPAL World Data Forum session, as well as secondary information gathered through a number of reports.

Data4Development Challenge

One of the biggest challenges of the data revolution is addressing how to unlock the potential of private sector data for public good purposes in a safe, ethical, scalable, and sustainable manner. Using 'big data' sources collected by private companies (such as "Call Detail Records" by telecom operators) for research and policy purposes has been hampered by legitimate ethical and commercial reasons. "Data challenges" and bilateral agreements have showed how computational analysis of these data, alongside traditional survey data and official statistics, could inform better decisions to curb poverty, inequality, epidemics, crime, traffic, waste, and more, by unveiling socio-economic outcomes and processes at levels of granularities and complexities never seen before. But these arrangements are costly and risky to scale, at a time of growing concerns over privacy, fears of growing digital divides, and unease about "black box" algorithmic decision-making. To date, there are no systems and standards to mitigate the associated privacy risks and capacity gaps to leverage these data for the greater good, ethically, at scalable, in a sustainable manner.

Response

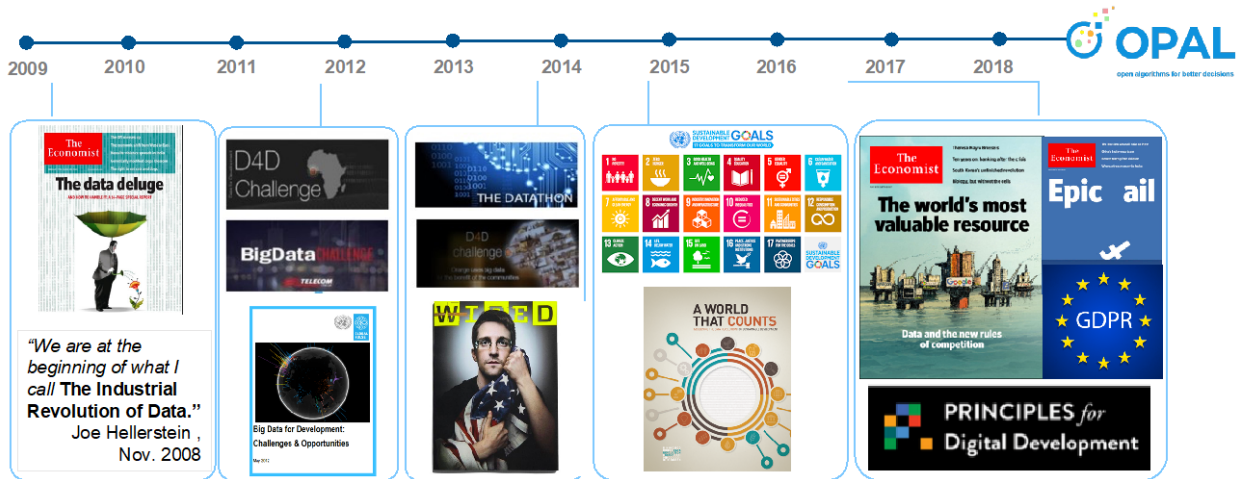
OPAL aims to fill that gap. OPAL (for "Open Algorithms") is a not-for-profit socio-technological innovation developed by a group of "data for development" pioneers around Data-Pop Alliance, Imperial College London, the MIT Media Lab, Orange, the World Economic Forum, in close partnership with Telefónica and the governments of Senegal and Colombia. At the launch of the Global Partnership for Sustainable Development Data (GPSDD) in 2015, conversations among participants around open data triggered the idea for OPAL, building on years of work by partners. It is designed to provide a far better picture of human reality to official statisticians, policymakers, businesses, and citizens, while fostering inclusion and inputs of all on the kinds and uses of analysis performed on personal data about them, in the context and in support of the Sustainable Development Goals and the Principles for Digital Development. As such, OPAL is seen by its founders "as a new paradigm for using private data for social good and a key milestone towards a vision where data is at the heart of fairer and more efficient 21st Century social contracts and systems."



Global Partnership for Sustainable Development Data

The Genesis of OPAL:

A decade of experimentations, expectations, skepticism, controversies, and many questions



OPAL is built on two complementary tracks that reflect and foster these objectives:

- A Technology track, with a privacy-preserving open source platform and open algorithms, developed by Imperial College London, MIT Media Lab, Orange Labs, Telefónica, and communities of local developers. A defining feature of OPAL is its “question-and-answer” approach to data access and analysis, where computation of key indicators (such as population density, mobility and poverty estimates) is “pushed out.” Certified open algorithms run directly on pseudonymized data that remain on the servers of the partner companies, behind their firewalls, and only aggregated statistics are made available to selected users. With OPAL, no sensitive data ever leave the servers to be exposed to theft and misuse.
- A Governance track, with local oversight bodies known as *Council for the Orientation for the Development and Ethics* (CODE), participatory design and testing with local users, and capacity building activities, led by Data-Pop Alliance and the World Economic Forum, and local partners. This aims to ensure that use cases are relevant to local needs and constraints, respectful of local ethical norms and regulations, and that capacities, incentives, and connections are strengthened within local ecosystems.

Tariq Khokhar @tkb · 2h

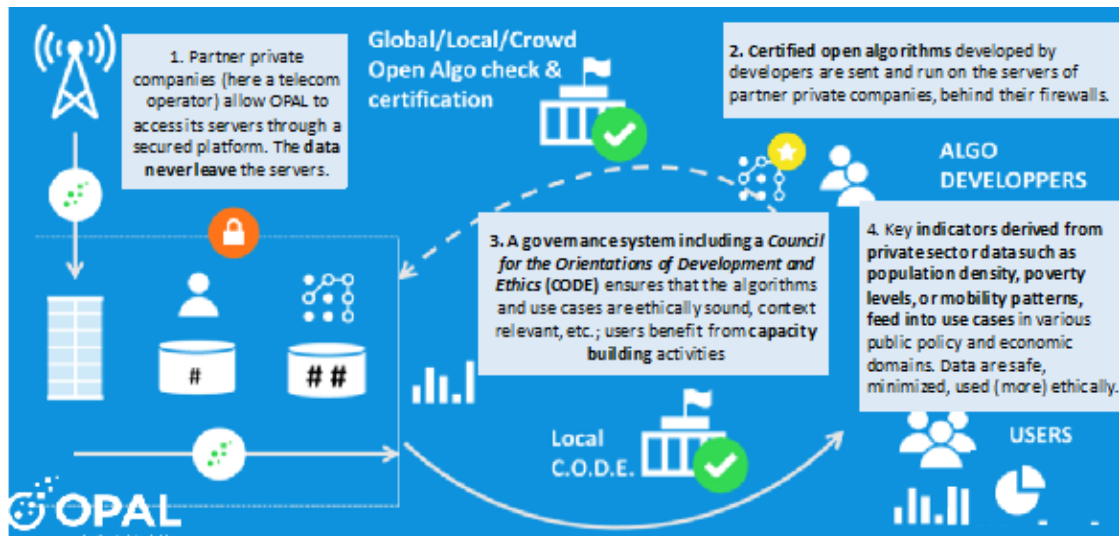
Shoutout to the OPAL project - "bring the algorithm to the data" - more at: opalproject.org #UNDataForum





Global Partnership for Sustainable Development Data

The future OPAL ecosystem



GPSDD, in partnership with the World Bank's Trust Fund for Statistical Capacity Building, provided OPAL with a \$250 thousand seed grant as part of the Collaborative Data Innovation for Sustainable Development Pilot Funding (now the Innovation Fund)¹. This grant was the first substantial increment of funding that took the OPAL idea to action and was used to conduct a user needs assessment in Senegal and Colombia, design the first white paper with MIT, hire a program coordinator, and improve ability to fundraise further. OPAL later successfully raised 1.5 million euros from the *Agence française de développement* (AFD), with additional support from the World Bank, and the Sustainable Solutions Development Network (SDSN) for the minimum viable product (MVP) phase.

Leveraging OPAL's alignment with GPSDD's mission and network, the OPAL concept was presented in a number of meetings with GPSDD country partners including Senegal, Kenya, Colombia, and Philippines. GPSDD enabled connections through GPSDD convened national workshops and engagements with high-level political figures. While it is impossible to precisely distinguish the effect of these connections, GPSDD contributed to the creation of a community of trust among OPAL partners given that most of them are also GPSDD members in working towards a shared vision of data for development.

Outcomes

OPAL started in mid-2017 in Colombia and Senegal with pilots leading to Minimum Viable Products in 2 phases (MVP1 and MVP2). These pilots are implemented in partnership with Senegal's *Agence Nationale de la Démographie et la Statistique* (ANSD) in Senegal, and Colombia's *Departamento Administrativo Nacional de Estadística* (DANE) and *Departamento Nacional de Planeación* (DNP), and 2 major telecom operators, Sonatel and Telefónica Colombia.

¹ <http://www.data4sdgs.org/news/announcing-funding-10-development-data-innovation-projects>. The



Global Partnership for Sustainable Development Data

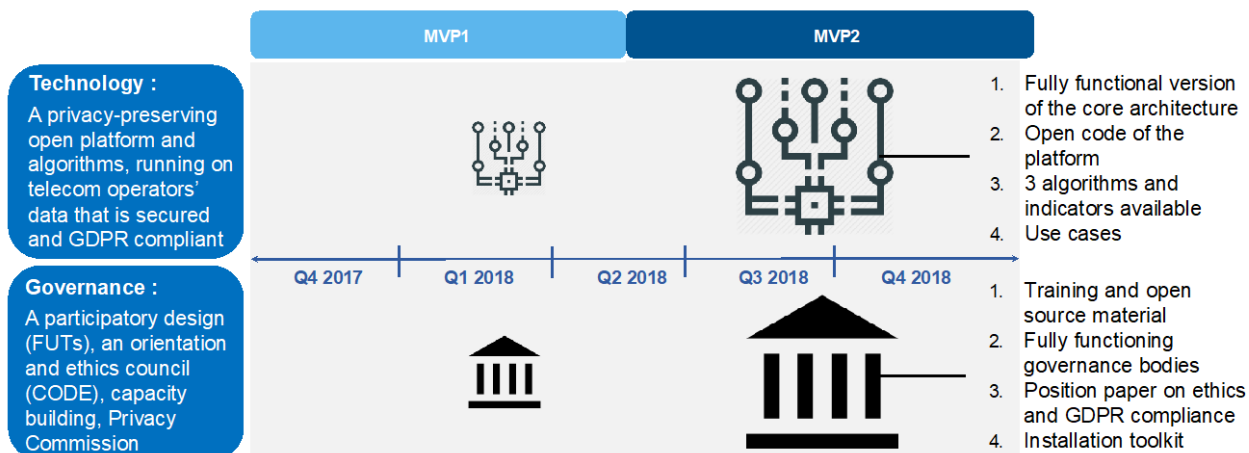
COLOMBIE SÉNÉGAL 23 MAI 2018

PROJET OPAL : LE BIG DATA AU SERVICE DU DÉVELOPPEMENT



The pilots have achieved important milestones along its Technology and Governance tracks in 2018. In terms of Governance, these have included: in-depth user needs assessments, capacity building activities; signatures of all key contracts, set up of the local CODEs, greenlight from the Data Protection Authorities, and contracting and interviewing of initial Friendly User Testers (FUTs) of the MVP1 platform in Senegal, including ANSD, AFD, the NGO IPAR, and the University Politecnico di Milano. On the Technology side, the main steps have been the development of demo versions with 3 algorithms by the telecom partners in both Colombia and Senegal, the installation of servers in both countries, availability of 2 years of CDRs in Senegal and 3 years in Colombia, and end-to-end functioning of the platform on real data from Senegal, undergoing security tests. By end of the pilot phase, key deliverables will allow OPAL to transition to its Beta phase.

OPAL pilot phase key components and deliverables



Colombia Friendly User Tester

Fundación Ideas Para La Paz (FIP) was invited to become a FUT through previous engagement with DataPop Alliance. In addition to being involved in the user needs assessment process, FIP is a part of the governance mechanism (CODE) in Colombia as a civil society representative.

Data

FIP has identified three use cases that will generate new information to better understand the conflict in Colombia: 1. Displacement and/or mobility restrictions estimation; 2. Displacement of micro drug trafficking post-intervention in the Bronx; and 3. Extortion from prison population estimation. Traditional sources of information on conflict-related measures in Colombia are of poor quality and frequency with many data gaps. For reasons cited as internal national security, the military and national police only share data up to 2015. Therefore, there is a need for non-traditional sources of information to understand the conflict. Big data in the form of CDR, can uncover patterns that were otherwise not possible to detect. In particular, CDR allows displacement estimation through additional data on point of departure and arrival such as route, length of travel, and whether displacement happens in groups or



Global Partnership for Sustainable Development Data

scattered individuals. Similarly, CDR data can be used to uncover patterns that enable the identification of risk levels in displacing micro drug trafficking and which areas are more likely to become distribution centers as well as to identify and track which prisons are linked with individuals running extortion schemes through cell phone calls². For FIP, OPAL has enabled access to non-traditional private data that fills a critical gap in understanding key factors related to conflict dynamics in Colombia. Findings from these use cases will be used to share knowledge among the general public as well as inform recommendations for policy makers and design targeted interventions.

Skills and Knowledge

In addition to access to data from Telefonica, OPAL has provided FIP, a civil society organization with crowdsourcing power to engage and collaborate with Telefonica. It was noted that the collaboration enabled faster movement on the negotiation process. In addition, the partnership with OPAL has added value through sharing methodologies and processes in developing and applying the algorithms, particularly through strengthening skills in Python, SQL, and R programming.

Senegal Friendly User Tester

In 2017, through a 1-day workshop where the concept of the OPAL platform was explained, OPAL invited IPAR to become a FUT. The contracting procedures with IPAR and Sonatel took time to materialize largely due to the fact that it was a new type of relationship, with a private sector actor. However, it was noted that the process was inclusive and transparent and ultimately led to successfully signing the contract with Sonatel providing access to CDR data.

Data and Knowledge

For IPAR, the partnership with OPAL was the entry into big data. While IPAR has been working with other largely traditional data sources, this partnership allows access to new type of data and access to the private sector. In particular, the CDR data will be used to do a mapping of regional and weekly markets to gain insights on livelihoods, mobility behavior, and associated transaction costs. This project is helping break down cultural barriers to open data and setting the stage for more cooperation and collaboration between state and non-state actors.

Future Implications

In 2019, OPAL will enter a two-year Beta phase, expanding from a “proof of concept” to a “proof of market,” as well as to at least one more country. A minimum of \$2M in additional funding are sought to turn the Minimum Viable Products developed in Colombia and Senegal in the pilot phase into robust Beta versions with the enhanced technological features and governance functions. Key beta technological features will include performance optimization, a research environment, a bank of certified and documented algorithms, and service integration APIs. Key governance functions will include CODEs running in both countries on a quarterly basis, more advanced capacity building resources and activities, including online, support to key users such as both NSOs, and the setup of a long term viable business model based on a freemium and revenue sharing approach. These will help

² “Cases for the OPAL platform for monitoring the characteristics of the armed conflict and problems associated with citizen security in Colombia” – FIP knowledge management area document



Global Partnership for Sustainable Development Data

produce a body of use cases serving as proof of market. The deployment on OPAL in another country from a list of half a dozen options is also a priority.

By 2020, building on the lessons of the first 2 years of deployment, OPAL will have created public knowledge, know-how and a community of developers and users for leveraging the transformative power of private sector data in an ethical and scalable manner. In doing so, it will have paved the way for the expansion of OPAL to other industries such as banking, insurance, hospitals, and other geographies including up to 10 countries by 2025.

OPAL development and expansion plans 2019-22

