

# Data for Development Senegal: Report of the External Review Panel



April 2015

This Report has been prepared on behalf of Orange S.A., by the Institute of Business Ethics (IBE). It gives a brief overview of the Data for Development (D4D) Challenge in Senegal set by Orange which ran from April 2014 – April 2015. It summarises the discussions of the D4D External Ethics Panel (DEEP), of which Philippa Foster Back CBE, Director of the IBE, was a member. It also raises some wider issues which relate to maintaining high ethical standards in the analysis and application of Big Data, the insights of which could be leveraged in the continuing context of data for climate action (D4CA).

## Introduction

Following the success of the D4D Challenge in the Ivory Coast in 2013, Sonatel and the Orange Group made anonymous data, extracted from the mobile telecommunications network in Senegal, available to international research laboratories.

The D4D Challenge was a Big Data and innovation competition open to the research world in order to aid development in Senegal. The first objective of the Challenge, under the patronage of the Ministry of Higher Education and Research, was to contribute to the development and welfare of the population. For this purpose, five priority subject matters for research were defined as: *health; agriculture; transport/urban planning; energy; and, national statistics*. The requirements for the research were set in collaboration with the responsible ministries or partner institutions.

The Challenge also sought to contribute to three wider objectives: *advancing research in the field of Big Data; involving local stakeholders and guaranteeing benefits in education and development of the ecosystem of local start-ups; and, advancing anonymisation techniques to allow sharing of data that is relevant for society while respecting privacy*. Prizes were awarded with respect to these objectives.

Whilst the highest standards of research ethics were upheld by all participants in the challenge, Orange was concerned to maintain an ethical stance in this initiative. Submissions were therefore also assessed with regard to business ethics and intended application. This required a need to balance the risks and opportunities related to the application or use of the data. The IBE defines business ethics as “*the application of ethical values to business behaviour*”. This requires exercising discretion in a responsible way when faced with a dilemma. A course of action may well be legally permissible, yet could still be considered questionable or unethical. The perspectives of various stakeholders and other interested parties, such as individuals, businesses and communities, need to be balanced.

A small number of the papers submitted as part of the challenge needed to be assessed with regard to their actual possible application for development. These were considered by the D4D External Ethics Panel (DEEP) and are detailed below.

## Background Information on Ethics and Big Data

Whilst there are many definitions that can be used to define Big Data, it is commonly understood to incorporate the collection, storage, and analysis of large, diverse and complex datasets generated from a variety of sources including sensors, internet transactions and other digital sources, such as mobile networks.

The use of Big Data is on an ascending trajectory, encouraged by increasingly powerful processing systems, improved computation and storage capabilities, available at cheaper costs. This reflects the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data.

The collection and use of Big Data can play a crucial role in the development of modern society from many points of view, including problem solving, improving well-being, and generating economic prosperity, as highlighted by the D4D Challenge.

However, it can also blur lines of what is considered to be acceptable and has consequently raised a number of ethical questions. An additional challenge is that Big Data can enable understanding and modelling large scale human behaviour with a temporal and spatial granularity never achieved before. As a result, there is a growing body of global research related to ethics and Big Data.

In the context of development, the application of Big Data to policymaking and development programmes is

becoming particularly relevant as the use of digital communication technologies, and especially of mobile phones, has seen an exponential rise over the last decade. This is especially true in low and middle income countries (LMICs), and across Africa which claims to be the ‘mobile continent’ as users have embraced mobile communications to overcome weak or non-existent landline infrastructure.<sup>1</sup>

This creates a wealth of Big Data as a by-product which has the potential to make a significant contribution in addressing a variety of issues, such as those prioritised in the D4D challenge.

Such datasets are important as they are seen by both policymakers and researchers as a potential solution to the lack of reliable statistical data in some countries. In addition, it can be considered as ‘organically’ produced data as it is an automatic collection of actual behaviour rather than being embedded in institutional practices and biases. Therefore, it is usually not subject to censorship or manipulation by intermediaries for political reasons.

However, there are challenges that need to be addressed with regard to such data. One such challenge is that the meaning of such data is not always simple or stable, and local knowledge is needed to understand how people are using the technologies in question. Another issue is that the statistical bias in proprietary data can be hard to understand and quantify. The D4D Challenge sought to mitigate these concerns by encouraging analysis by leading and reputable academic institutions and research laboratories. However, these concerns can never be fully mitigated and remain open issues which prompt some of the main ethical questions for individuals and for groups, surrounding the use of Big Data.

Furthermore, there are risks to individual or group privacy in the absence of a clear ethical framework or set of rules for handling and sharing digital data. At present, most legislation is aimed at individual privacy, however, group privacy is also a concern.

Other issues which require consideration are:

- Anonymisation techniques require more standardisation. Whilst complete global standardisation is unlikely, global experts should seek to characterise what is appropriate

and set a framework which ensure security as a minimum. This should also seek to accommodate a risk based approach which enables a trade off with the granularity of detail.

- There is less awareness in LMICs of the implications of making personal data public.
- Digital data protection has yet to be identified as a concern for a majority of LMIC governments. However, in Senegal this has been recognised, under the supervision of the “*Commission des Données Personnelles*” (CDP), from which the D4D Challenge received authorisation.

## The Orange Response

To ensure the anonymity of the three datasets released to researchers, Orange prepared them in two stages. First, was a local pseudonymisation in Dakar, second was a number of layers of anonymisation run in the Orange Labs in Paris.

The first data set was a series of matrices, counting the number of calls and texts between antenna per hour. The second was a low resolution mobility dataset which provided information on the mobility of a sample of 1% of the population at the level of the 127 prefectures of Senegal. The third was a finer mobility series of small samples, at the level of antennae, where the antenna and the time of calls had been modified in a random fashion.

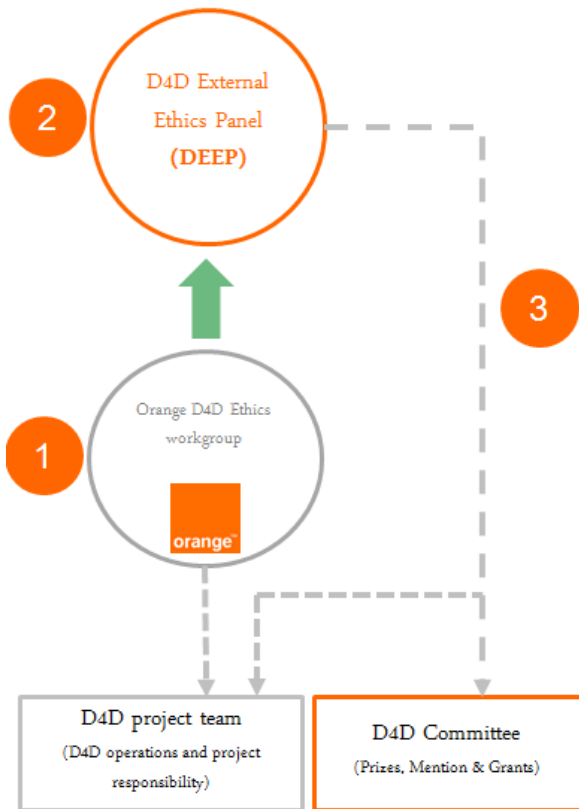
As part of the D4D Challenge, Orange set up a detailed governance structure to review all entries and ensure an ethical stance was maintained throughout. This is presented graphically in Figure 1 on page 3.

The D4D project team was supported by the Orange D4D Ethics workgroup, which consisted of Senior Managers, most of whom were not involved in the D4D project. Their responsibility was to propose the Sonatel-Orange preliminary position and initiate actions in cases of obvious need.

The D4D External Ethics Panel (DEEP) was also created. This was considered to be a key element of the governance structure. The DEEP was comprised of 14 external advisors with a balanced profile. A list of members of the can be seen in Table 1.

<sup>1</sup> See The Guardian 05/06/2014 – [Internet use on mobile phones in Africa predicted to increase 20-fold](#). Last accessed 24/03/2015.

**Figure 1** Governance Structure of the D4D Challenge



The DEEP was asked to provide advice to the D4D project team and the D4D committee, specifically with regard to papers which the internal workgroup were unsure about publishing.

The review panel was given an agreed framework against which to consider papers. The full framework is available as an Appendix. There were five potential outcomes for papers submitted. These were: (1) *Promotion* – for papers that were exceptional in the way they treat the ethical application of the research. Such papers were eligible for the ‘Ethic Mention’. (2) *Normal publication* – for papers which required no restrictions, even if there were risky aspects. (3) *Ask to consider adjustments* – for papers that require small amendments before publication, but without these they will not be published. (4) *Do not publish* – for papers which may be too sensitive or risky for general publication, but could merit presentation to appropriate authorities. (5) *Content and methodology verification* – for papers in which the review panel had further questions surrounding the methodology or conclusions. This was a prerequisite for all other considerations.

**Table 1** Members of the DEEP

| Name                    | Institution & Location  | Profile      |
|-------------------------|---|--------------|
| Lucy Bernholz           | Stanford University (US)  | Academic     |
| Philippa Foster Back    | Institute of Business Ethics (EU)                                   | Business     |
| William Hoffman         | World Economic Forum (Global)                                       | Institution  |
| Johannes Jutting        | Paris 21/OECD (Global)  | Nat'l Stat   |
| Robert Kirkpatrick      | United Nations (Global)   | Institution  |
| Emmanuel Lulin          | L'Oréal (EU)  | Business     |
| Ulrich Mans             | Leiden University (EU)  | Academic     |
| Mark Nelson             | Stanford University (US)  | Academic     |
| Yaye Fatou Camara Niang | Commission des Données Personnelles (AF – Senegal)                  | Institution  |
| Nuria Oliver            | Telefonica (EU & LAT AM)  | Business     |
| Juliana Rotich          | Ushahidi, BRCK (AF)   | NGO Business |
| Olivier Sagna           | Ministry of Higher Education and Research of Senegal (AF – Senegal) | Institution  |
| Jean-Philippe Vanot     | ParisTech Institute of Technology (EU)                              | Business     |
| Pat Walshe              | GSMA (Global)   | Business     |

The governance process included some screening as to the credibility of the researchers and the legitimacy of the use made of the data provided. Prior to submission, each entrant received terms and conditions to which they had to agree before access to the data was given. In total 260 research teams registered, and 59 submissions were received on time for the challenge and reviewed by the internal panel against the criteria above. Of these, six were referred to the DEEP.

### Referrals to the DEEP

Six papers, in the areas of *health*, *national statistics* and *privacy*, were referred to the panel. Questions arose related to matters of application of the data, as opposed to matters of research ethics. In these cases, due to the

applied nature of the research, the DEEP was keen to highlight to the research teams that robust anonymisation was not sufficient to ensure high ethical standards. This highlighted a need for a broader agenda in which data scientists and researchers continue to be educated about ethical issues in the application of their research. It was suggested that this agenda should be picked up at the institutional level, and pushed further.

Of the six papers, one was deemed to be appropriate for publication. A further two were recommended for publication after clarification from the research teams. Two papers required more substantial amendments as the applications were not considered to be appropriate, but the processes were considered to be interesting. It is likely that these papers will be published as separate pieces of scientific research. One paper was deemed to be inappropriate for publication. In all of these six cases, the discussion around the issues raised was an open one, and feedback was given to each of the research teams either by email or phone conversation. As this was an open innovation challenge, established with the intention of publishing almost all of the submissions, extra care was taken than could have been expected in a more controlled environment.

These papers raised a number of concerns for the DEEP. In addition to privacy concerns, the DEEP also discussed political concerns related to a regionalisation project, social unrest in Ebola analysis, and the perception/risk of surveillance, closely related to the privacy concerns. The DEEP used a risk-benefit based approach to assess the merits of various submissions. These were related to various constituents of society including individuals, groups, businesses and public institutions/politics. The analysis of potential benefits and risks to these groups of society as proposed by Bill Hoffman (World Economic Forum) and Amparo Ballivian from the World Bank is also available as an Appendix.

Adopting such an approach provided a variety of perspectives for the DEEP to consider and helped with understanding on the overall position to take on some papers. This more balanced approach was useful for one paper in particular which used the data set to propose the constitutions of administrative regions, according to social ties and geographic considerations. This was considered to be a sensitive political topic by the DEEP, which posed a risk of creating social unrest. The question considered was not a question about the

ethics of the research, but rather the appropriateness of advertising the results of an application of social and semantic views that have potentially significant political implications. Therefore, despite the potential benefits associated with improved regional understandings in Senegal, the paper was deemed not to be appropriate to be published as part of the D4D Challenge. Instead, it was agreed that it should be shown to the appropriate authorities as a way of introducing a more evidence based approach to such a complex topic.

Another issue that arose was in addressing privacy concerns. The issue of group privacy was considered by the DEEP to be the most material concern arising from the challenge. From the reviews it became evident that there is an increasing need for greater clarification of the data ethics and the processes for assessing and balancing risk/benefit impacts at the community and group level (in addition to impact assessments at the individual level). National Security and Commercial applications were the two most talked about domains of Big Data where it was necessary to identify and target unique users. However, the D4D challenge was more concerned with groups, as is the field of development in general. This specifically played out in the balance between the rights of the individual and the benefits of groups as a consequence of the collection and use of Big Data.

The greater visibility Big Data provides to some groups of people can have a hugely positive impact. Examples include the possibility of tracking the spread of a disease faster, or bringing relief to a disaster area more effectively. However, there can also be downsides which require consideration, especially when operating in countries with limited regulation and potentially weak government. Such datasets could be easily acquired by companies to plan ethically questionable marketing strategies, or armed groups could take advantage of this information to target specific sets of people with their military operations.

This problem is not only relevant when using data collected through mobile phones, but also affects other sources, such as social media, where users are not representative of the population at large. Therefore, groups with more visibility might accidentally be favoured, or indeed discriminated against, at the expense of those less visible. This may result in failing to detect potentially relevant issues.

The second aspect is related to the privacy of individuals and can be seen as directly related to some of the above considerations.

## Summary and Conclusions

As the D4D project outlines, Big Data has the potential to influence greatly sustainable development. This will also continue to increase exponentially as new data sources become available. It is imperative that such research is conducted to the highest ethical standards, in relation to research ethics, business ethics and application. These issues should be considered before the use of such datasets.

A particularly sensitive challenge is related to the lack of a regulatory framework or widely accepted benchmark on the privacy of data for individuals and for groups and its subsequent use or sharing. Another is the consideration for the publication of scientific results, the application of which might be considered more sensitive in certain cultures. This is especially true in some LMIC countries, but it is not limited to them.

In order to provide a sound ethical framework to the project, Orange appointed the DEEP as part of its governance structure. The panel served as a self-regulatory body that ensured the highest ethical standards on this topic were met, both at the national and international level.

Orange's project represents an important step forward in the production of an appropriate framework for the use of data to inspire research standards, assist government and business and facilitate development. The D4D experience highlights the necessity of such a common understanding, demonstrating how difficult the interaction between data scientists and the actual needs of development programmes can be. As highlighted above, even high quality research production, based on the available datasets, aimed at improving living conditions in LMIC countries, can produce ethical issues with regard to the use of such findings and their practical implications on the populations involved.

A further issue in need of consideration which arises from this project is the need of a closer collaboration with public policymakers who could bring their local expertise to the table in order to coordinate the efforts and underpin the common goal of development. This

would also be useful to provide a safe regulatory and operational environment for private companies to engage in partnerships in these areas helping them also to limit their exposure to reputational risk.

A collaboration between Orange and the Commission des Données Personnelles of Senegal is under discussion to progress on this issue.

### **Contact Information**

For further information about the D4D Challenge and the findings of the DEEP, please contact Nicolas de Cordes at Orange: [nicolas.decordes@orange.com](mailto:nicolas.decordes@orange.com).

For further questions related to the IBE, please contact Philippa Foster Back: [research@ibe.org.uk](mailto:research@ibe.org.uk).

## Appendix: Proposed typology of actions to apply to the D4D papers

| Action   | Description   | Brochure | Prize | Poster | D4D Doc | Private* |
|--|---|----------|-------|--------|---------|----------|
| Promotion (Ethic Mention)                                  | A paper that would be exceptional in the way they treat Ethic aspect of the research will be eligible for the Ethic Mention<br>Orange Sonatel will choose to promote some work externally in a dedicated Brochure., of which the Ethic Mention laureate | X        |       |        |         |          |
| 2 Normal publication                                       | No restriction applied, even if risky aspects. Can be invited to present in poster session, in plenary session and is also fully eligible for prizes and Grants if the D4D Committee choose it  | X        | X     | X      | X       |          |
| 3 Ask to consider adjustments                              | Requires some wording or visual or content adjustment before decision.<br>Without further adjustment, Orange-Sonatel ask to Not publish it  |          |       |        |         |          |
| 4 Do Not Publish   | Orange Sonatel ask to not publish the document.<br>Some papers might be sensitive/too risky for a general publication, but could merit a presentation to the appropriate authorities. This will be evaluated on a case by case basis                    |          |       |        |         | X        |
| Also possibly ask for content and methodology verification | Question the methodology or conclusions, which is a prerequisite for all other consideration. A « bad » paper won't be considered nor published in the D4D document   |          |       |        |         |          |

## Appendix: Potential Benefits and Risks to Society, Bill Hoffman & Amparo Ballivian

| Group                    | Benefits  | Risks   |
|--------------------------|---|---|
| <b>Individual</b>        | <ul style="list-style-type: none"> <li>• Improve health and wellness outcomes</li> <li>• Freedom of movement</li> <li>• Increased earning power and access to employment opportunities</li> <li>• Identify integrity assurance</li> <li>• Fair and non-discriminatory uses of advanced analytics</li> <li>• Establishments of private spaces which are safe and protected</li> </ul>  | <ul style="list-style-type: none"> <li>• Reputational damage</li> <li>• Death or bodily harm arising from violence</li> <li>• Chilling effect on freedom of speech, association, etc</li> <li>• Detriment arising from monitoring or exposure of identity, characteristics, activity, associations or opinions</li> <li>• Personal, family, workplace or social fear, embarrassment or anxiety</li> <li>• Unacceptable intrusion into private life</li> <li>• Loss of liberty or freedom of movement</li> <li>• Damage to earning power</li> <li>• Other significant damage to economic and social interests</li> </ul> |
| <b>Community</b>         | <ul style="list-style-type: none"> <li>• Strengthened ability to uphold political, civil, economic and human rights</li> <li>• Strengthening of social trust and accountability</li> <li>• Ability to understand, act and adapt to crisis situations and public safety concerns with more precision and flexibility (e.g. disaster response, infectious disease)</li> <li>• Optimized food, energy and water resource allocation to address environmental issues from climate change</li> </ul> | <ul style="list-style-type: none"> <li>• Reputational damage</li> <li>• Damage to democratic institutions (e.g. excessive state or police power)</li> <li>• Loss of social trust (who knows what about whom?)</li> <li>• Discrimination</li> <li>• Economic interests</li> <li>• Loss of privacy/additional surveillance</li> <li>• Disrespect of culture</li> </ul>  |
| <b>Governments</b>       | <ul style="list-style-type: none"> <li>• National statistics: increased timeliness, accuracy, coverage</li> <li>• New partnerships and better engagement with private sector and NGOs</li> <li>• Reduction in economic and societal impact from crises situations</li> <li>• Increased growth in GDP per capita</li> <li>• Better use of resources</li> <li>• Promotion of transparency</li> </ul>  | <ul style="list-style-type: none"> <li>• Trade restriction/sanctions</li> <li>• Threats to national security</li> <li>• Threats to sovereignty</li> <li>• Political change and/or backlash</li> <li>• Loss of economic opportunities</li> </ul>   |
| <b>Private Companies</b> | <ul style="list-style-type: none"> <li>• R&amp;D</li> <li>• General market &amp; consumer research</li> <li>• Reputational improvement</li> <li>• New partnership and better engagement with government and NGOs</li> <li>• Reduction in business impact from crises situations</li> <li>• Business development</li> <li>• Internal culture development for Big Data usage</li> </ul>   | <ul style="list-style-type: none"> <li>• R&amp;D</li> <li>• General market &amp; consumer research</li> <li>• Reputational improvement</li> <li>• New partnership and better engagement with government and NGOs</li> <li>• Reduction in business impact from crises situations</li> <li>• Business development</li> <li>• Internal culture development for Big Data usage</li> </ul>   |