



Tuesday, October 30th 2018

To whom it may concern,

This is a statement in support of the approach undertaken by the OpenDreamKit project in developing a toolkit for a mathematical Virtual Research Environment.

I was myself involved in designing the original application for OpenDreamKit, and then worked on the project for two years as a site leader (University of Zurich). Most of my work was concentrated on Work Package 6 (semantic interoperability between different components of the toolkit).

When writing the application, we emphasized the interest of an open and interoperable approach for mathematical research, but were also keenly aware of the interest of our toolkit for mathematical teaching. For instance, the ODK approach helps smooth the transition for students onto research, or it helps existing research networks co-develop significantly better courses.

I left the ODK collaboration 18 months ago, and am not representing the collaboration in any way through this letter.

Since leaving the collaboration, I have been building up a nonprofit based out of Geneva promoting personal data protection rights. This nonprofit, [PersonalData.IO](https://personaldata.io), has very quickly earned international recognition for uncovering the Cambridge Analytica scandal.

Beyond advocacy work, we are also now building tools to help emerging collectives make the personal data of its members more useful to them. We are for instance working now with the Independent Workers Union of Great Britain to ascertain the existence of biases in Uber's ride dispatch algorithm. This is the first glimpse of a large amount of work to be done reducing information asymmetry in the gig economy.

The ODK toolkit has - somewhat unexpectedly - proved itself very useful for such efforts, in ways that directly devolve from the core principles structuring the ODK project.

Because our use cases are directly tied to advocacy, which was outside of the initial scope of ODK, we thought it might be valuable to list some of the gains we are experiencing in that context due to the ODK approach:

- ODK's deployment model is very flexible. It accounts for large and structured research projects (the typical VRE use case), but also much smaller research collaborations, classroom settings, or even temporary workshops. In the personal data ecosystem, the deployment model is directly tied to the trust delegated to the various actors involved in data processing. We have found this flexibility in deployment choices very valuable in building alternative trust structures in the personal data ecosystem. For instance, we can directly offer advanced tools to any Uber driver, helping them regain a copy of their personal data and derive insights from it. Crucially, through a MyBinder deployment, this does not require the driver to entrust anyone with their data: certainly not us and not even their labor union! At a later stage, we can recentralize *some* of the data with the union, to help gain collective insights. The transition path there is completely smooth from a technical standpoint, and can be driven by the circumstances, *i.e.* levels of trust achieved by the various stakeholders.
- [PersonalData.IO](#) cannot do everything alone. Natural allies (to build visualisations, to forensically assess discrimination, etc) tend to use tools in the ODK ecosystem (Jupyter, for instance), which helps make the reclaimed data useful faster, while still minimizing the trust required from the individuals.
- In fact, the companies we are currently reclaiming data from themselves tend to use the same tools as us (such as Jupyter). Where they see limitations, we have observed that they build new software frameworks, while keeping interoperability with their current setup. They then release these extensions of their technical stack as open source software. This is the case for Uber, for instance, which has released [kepler.gl](#) and [sharedstreets.io](#). This helps us quickly and efficiently plot the data that Uber drivers actually care about most, in ways familiar to them (earnings, for instance, overlaid on a typical Uber map).
- The semantic interoperability layer of ODK has not yet proven useful, but we fully expect it to be at some point. Indeed, as we aggregate data from multiple sources, it would make sense to align concepts and reuse analyses and visualisations (think of how much translates from Uber to Lyft). This mode of use is similar to that envisioned in OpenDreamKit, although the mathematical semantics are not as rich. However, as we move on from the relatively rudimentary problem of aligning data, we expect to find eventually a lot of value in the current ODK semantic interoperability work. Indeed, emerging questions around algorithmic accountability will naturally lead us to formalize experimental



hypotheses around the automated decision-making processes under investigation.

While we are operating as a nonprofit, and still building a full methodology, the only real obstacle to us right now is in GDPR enforcement and the default enterprise culture of opacity surrounding personal data processing. As GDPR enforcement becomes smoother and awareness grows, that difficulty should be smoothed out, and we anticipate new commercial actors to align with the same methodology and likely use the same tools.

We hope this account of our use of the OpenDreamKit ecosystem will have been useful to you, and remain available should you have any questions.

Sincerely,


Paul-Olivier Dehay
paulolivier@personaldata.io
PersonalData.IO