Response to reviewers

Associate Editor

Comments to the Author:

Very important paper about a unique and an important initiative. Pease take into account the comments of the reviewers where possibly. Since the focus is on policy, if you need to extend content, prioritise the policy side and less the technical side. Including some figures would make it easier to read. Where you can, please add something of the use of the insights by policymakers and health professionals, for example the eHealth network. Also try to rethink some of the real challenges (as one reviewer notices, many of them are known challenges) of such a large and important initiative. There must be much to learn from.

The authors would like to thank the Editor and the reviewers for the extensive and valuable comments provided. We have addressed them in the new version of the paper. The track of changes is reported here below. The manuscript was also updated with the latest developments of the initiative, including the number of MNOs and countries covered, the outreach to the eHealth Network, recent references to the tools developed (i.e. Mobility Visualisation Platform) and to the Letter of Intent with MNOs. Finally, Figure 1 and Figure 2 were added giving insights about the analysis carried out in the framework of the initiative as suggested.

Reviewer: 1

Comments to the Author

Overall this is a very important and interesting initiative. I am glad to see it explained here and to note its unprecedented scale. The EC-JRC has done great work in partnering with the industry and adapting quickly to address important issues, even with all the hurdles posed by the very nature of working with different competitors at the same time.

I would recommend adding a little more on the impact or potential impact of the results obtained. It may be early and difficult to measure the impact of the initiative, but you can certainly clarify the potential it has, and how it can (and probably should) be used by policymakers in the response to the pandemic.

A subsection (3.4) on outreach was added to list the ways access to the findings has been facilitated to policymakers in the EU Member States. The text also reflects new reports published on how mobility insights are feeding new tools and scenarios in view of the new wave of the pandemic.

Furthermore, I would like to see a further explanation or a deeper dive into the heterogeneity issues. These stem not only from a technical and technological standpoint but also due to a lack of harmonisation in regulation across Europe, with distinct interpretations of ePrivacy and GDPR across the continent; furthermore there is sometimes ambiguity on what are "sufficient levels" of anonymisation. This poses a problem for the homogenous formats that would most definitely be helpful and drive initiatives like these further, faster. Of course this is just one of the aspects that contribute to the heterogeneity, but I believe it to be an important one.

The heterogeneity aspect identified by the reviewer for further expansion is indeed of primary interest. Yet, this initiative does not entail the evaluation of specific reception at national level of the EU directives in terms of ePrivacy and GDPR, and, even more so, their interpretation by MNOs.

The article is very interesting and significant, and it shows what we should be working on to empower policymakers with real data-driven decisions.

The authors are thankful to the Reviewer for the comments provided.

More detailed comments:

Page 3:

I would recommend clarifying the dates for these initiatives (e.g. the HLEG is from 2018-19), in order to clearly showcase the timelines and how the COVID-19 initiative is well aligned with previous work.

Dates have been added to clarify the timing of previous efforts with respect to the COVID-19 initiative.

Page 4:

This paragraph: "As opposed to openly available mobility data derived e.g. from social media, MNO data provide a level of granularity, frequency of update, representativity (coverage of large fraction of the population) and a higher level of transparency that make this dataset valuable in terms of insights that can be extracted about human mobility." - could easily be challenged by other players in mobility data that derive it from apps or social media. I would recommend giving concise but concrete justifications further on in the paper of why mobile data has all these advantages (no dependency on smartphones, more representative particularly of older populations etc) and add some of of caveats as well (in some cases only active events are captured, potential for data loss, potential for regional or MNO specific bias, etc)

The comparison between mobile positioning data and mobile app location data in Section 1 has been expanded. The main advantages of the CDR data are: increased granularity, increased representativity, existence of connectivity data, and higher level of transparency.

Page 8, in what way do you believe the MFAs will lead to a better balance between the public health and the socioeconomic impact? Could you also provide a short explanation here on how these are created, as I believe it would be hard to understand what these areas are just from reading this paragraph. Including a figure would be valuable in order to show what these represent

The discussion on how the MFAs are constructed has been expanded and a figure added (Figure 1) to exemplify the idea of using the MFAs rather than the administrative borders to contain COVID-19 cases.

Page 9 - in the Privacy, Commercial Sensitivity and Fundamental Rights, could you please clarify how the Reasonability Test was performed? This topic is extremely relevant for policy and so I believe it would benefit from a quick high level explanation of the method

In the framework of this initiative, as explained in the text, the JRC carried out a so-called "reasonability test" upon the reception of preliminary data samples from MNOs. The objective of the test is twofold: to

actively verify that the data specification in terms of origin destination aggregate data were respected and to assess whether or not the risk of re-identification of the individuals was reasonably low.

Page 10: the establishment of a multi-disciplinary group of experts is a great idea, but do you see this as a point of collaboration between industry and government bodies? It is unclear where the experts would be coming from but in my opinion the case would be stronger if it advocated for a private-public sector collaboration. Furthermore, and if you have space, could you address the issues of sustainability for this group? What you are proposing would be a short-term, long-term collaboration? Are there any potential caveats with funding this initiative?

A few of such experiments already exist. One of these is the Social Science One initiative, which is a partnership between academia and industry. The expert group works closely with experts within the industry to create a framework that allows access to the industry data preserving both the privacy of the users and the commercial value of the companies involved. Another example is the Big Data for Migration Alliance.

The discussion in the text has been expanded accordingly.

The Ethics Committee, who would be appointed? Once again I would say that the partnerships with the industry would be extremely relevant here. Furthermore, the Risk Assessment section could (and maybe should) be an entry point to the Ethical Assessment, as it will be quintessential to address potential unintended consequences and privacy risks.

Similar considerations would probably apply as for the previous question. The authors agree with the reviewer that this important topic deserves a dedicated and structured discussion on possible options.

The matters of a common standard across MNOs are extremely relevant, but do you see an issue with their implementation? Given the large costs of maintaining, curating, processing the datasets, how do you see the industry implementation motivation? There could be a problem here with misalignment of goals that would be important to address even if briefly. The publicly available insights would also pose a commercially sensitive risk for the MNOs, so I would recommend that the analysis of what can be made public should be left to the same multi disciplinary groups involving public and private sector, as I referred above.

The referee is correct. The whole initiative may be supervised under the umbrella of the group of experts, with the understanding that a "data for good" approach is not enough for producing official statistics (or complements to these) if the process going from data to products is not, for instance, transparent and uniform across MNOs.

Reviewer: 2

Comments to the Author

With the prevalence of COVID-19, the manuscript details how a Public Private Partnership on data sharing and exchange could effectively help combat against infectious virus in the epidemic or pandemic situations.

This paper highlights an on-going effort in a unique Business-to-Government (B2G) initiative between several Mobile Network Operators in Europe and the European Commission.

It is well explained in the paper that why MNO data is very important compared to other openly available data derived from social media in perspective of level of granularity, representativity and relevance to high human mobility. It is truly remarkable that data from 15 MNOs covering 22 EU member States and Norway have been transferred to the Commission on a daily basis which could be used to help quality of epidemiological modelling of virus spread and forecasting of the future pandemic. It would seem all but natural that this Initiative would have extensively used the AI and Big Data technology in analyzing and processing the data but no mentioning it. It is recommended to make an explicit reference to the use of the AI and Big Data technology.

The authors made use of various AI and Machine learning tools and methods to process the data e.g. specialised Python and R libraries. We added an explanation on the technologies that we used in Chapter 4: Data Security, Integrity. "Data received by the MNOs at high spatial and temporal resolution was not allowed to exit from the secure platform. All the data processing, analysis and storage took place remotely on the Unix secure platform using open-source technologies such as Python, R and PostgreSQL."

This B2G initiative clearly demonstrated the potential of data sharing and clear definitions and attributes of Mobility Data as well as insights of Mobility Data Products along with the challenges and recommendation should be fully considered for well preparedness of new virus waves.

One observation is that this initiative focuses on usage of anonymized data (due to privacy regulations) which has primary purpose of virus spread modelling. However, in some countries like Korea, usage of non-anonymized data are allowed under strict conditions (such as pandemic) that has proven very effective in terms of tracing and tracking of potentially virus exposed people, and alert and bring them to test for virus infections and treatments if necessary. The paper lays out important recommendations on what are critically needed to strengthen the preparedness for future pandemics. Acknowledging that there are different levels of tolerance and acceptance, it is of a recommendation of the Reviewer that their subsequent work would consider including the sharing of non-anonymized data for the purpose of public good.

The authors would like to thank the reviewer for the recommendation. A sentence has been added "The Ethics Committee should take into consideration the culturally shared privacy norms...".

* Note to the Authors: The Broadband Commission hosted by ITU and UNESCO (www.broadbandcommission.org) has published a report on the Epidemic Preparedness which also highlights an importance of MNO data sharing to combat the Epidemic/Pandemic –

- ITU-UNESCO Broadband Commission Working Group Epidemic Preparedness

(www.broadbandcommission.org/workinggroups/Pages/Epidemic-Preparedness.aspx)

- Full Report

(www.broadbandcommission.org/Documents/publications/EpidemicPreparednessReport2018.pdf)

- Executive Summary
(www.broadbandcommission.org/Documents/workinggroups/BBCOM%20WG%20Epidemic%20Preparedne ss%20Final%20Print%20(Executive%20Summary).pdf)

A reference to the report was added in the manuscript (introduction Section). The authors are thankful for the comments provided.

Reviewer: 3

Comments to the Author
The authors describe outcomes and experiences of using CDR-based data from 15 MNOs in 22 European countries. This is an unprecedented effort, and the paper is promising in that a lot of organizations and researchers could learn from the authors experiences. Nevertheless, I would recommend the authors to move beyond well known lessons and dig deeper into their insights and recommendations to make this paper worth publishing. The paper title suggests a focus on challenges and lessons learnt from B2G experiences in the context of MNOs, the European Commission and covid-19. Nevertheless, 2/3 of the paper are devoted to present a summary of results from other papers that the authors have already published (including one in this journal).

Despite devoting a good part of the paper to the summary of results, these felt insufficient and were hard to follow at times given that so many details are omitted. Just to name a few, the "common denominator" or the "Trusted Smart Statistics concept" were simply named and references provided. However, without clarifying these, it is hard to understand and evaluate the core technical contributions of the mobility data products. The summary should be self-contained and clearer. Having said that, I do not think that this is the intended main contribution of the paper, despite the fact that a considerable number of pages are devoted to it. Nevertheless, given that the authors wrote about it, and appear to want to include it, my recommendation would be for the summary to be understandable as a stand-alone.

On the other hand, challenges and recommendations, which according to the title are the main contribution of this paper were shallow and not novel most of the times. Some of the challenges mentioned: end-to-end encryption for data transfers, non-disclosure agreements, limited access, communication to address reputational damage or data retention horizons are well known and already in place for many B2X collaborations (with X being government, international organizations or research organizations).

Similarly the recommendations were also not very novel: working group of experts, ethics committee, communication committee. The last paragraph was possibly the most interesting in which the authors correctly point to the need of defining a shared data standard, a la GTFS, but for CDR data.

In my opinion, the paper is neither a good summary of the technical contributions, not a thorough discussion of the challenges and lessons learnt. However, I think that putting together data from 15 MNOs and 22 countries is an unprecedented effort. I am sure the authors might have experienced challenges that no other researchers have been exposed to before, and these challenges are the ones that the authors should surface in their paper.

First, I would recommend the authors to shorten the summary description and make it crisper. Some information is not needed, while other critical points as mentioned above are left to the reader to guess. In

addition, to make this paper novel, the authors should probably delve more into the challenges, recommendations and general lessons learnt from their experience.

We would like to thank the reviewer for the comments provided. Indeed the focus on results was one of the requirements of the special issue and we believe they are extremely useful to understand not only the challenges related to the process and framework of the initiative, but also related to the extraction of information and outreach. The results are a very light summary of three technical notes that expand and give details (by the way, the authors have not published yet in this journal), and only 1 page out of twelve, which we believe is a good balance in favour of challenges faced and recommendations.

Next, I describe a few suggestions that I think would be extremely interesting and novel contributions from a policy and data perspective:

1.   Data sharing protocols were briefly described. It would be important to further discuss how were the B2G agreements implemented, what were the negotiations necessary to put in place such a large scale collaboration? A qualitative analysis of expectations and pain points across participants would be extremely interesting.

The Letter of Intent signed by MNOs and the European Commission was added to the introduction, together with the way the initiative started: "Initiated by means of an exchange of letters, the terms of cooperation between MNOs and the European Commission are outlined by a Letter of Intent (European Commission and GSMA partners on Data4Covid: https://www.gsma.com/gsmaeurope/resources/d4c/)".

2.   Can "trusted intermediaries" only be european commission groups? What are the opportunities for others, and what does it mean in terms of data transparency and equity? Do we have to assume that workers at the european commission are always neutral? In other words, shouldn't other types of intermediaries such as advocacy groups be able to see this data and extract policy findings (e.g., human rights organizations would benefit from using location data)?

As mentioned when responding to a similar comment by another reviewer, a subsection (3.4) on outreach was added to list the ways access to the findings was facilitated to policymakers in the EU Member States.

3.   This journal focuses on the intersection of policy and data. It would be advisable to thoroughly discuss the ways in which the mobility data products and insights revealed by this research project were used for policy. How were the EU or national policies during covid-19 guided or informed by the outcomes of these analyses? Time and again, the research community and research units at MNOs have published papers with findings for potential social impact, and time and again these have been mostly ignored. What is it that makes this collaboration successful from a policy perspective and why? It would be extremely important to understand what is the secret for success in the CDR data to policy pipeline, if the authors have been successful at that.

The authors find this question interesting. Outreach activities are still ongoing, reports are currently being published, however it remains difficult to quantify how the findings have been used.

4.   Another important component aligned to policy and data is that of comparing findings to other sources. If the insights identified by this project were in fact used for policy, what were the insights that

could only be identified by CDR data and not identified by other sources of information? For example, location intelligence companies also collect location data, or aggregators that bring together mobile app location data. How are these complementary (or not) to CDR data, and why? ██████████

The comparison between CDR data and mobile app location data in Section 1 has been expanded. The main advantages of the CDR data are: increased granularity, increased representativity, existence of connectivity data, and higher level of transparency.

5.   It would also be important to clarify the sustainability of the collaboration between MNOs and the European Commission, or with other "trusted intermediaries". These collaborations are often times ad-hoc, and only due to extreme events such as a pandemic. I would like to learn more about the authors' perspective on how these types of collaborations can be sustained over time and what would be best practices for that.

A reference to the agreement is now part of the introduction, giving additional details about the current status of the initiative. As mentioned in the conclusions section, the initiative is expected to somewhat simplify similar collaborations in the future. Finally, a reference to the EU Governance Data Act recently published was added in the text.

Other comments.
These are minor comments that I think would improve the paper if clarified.

1.   In the Introduction (page 4), the authors claim that getting "raw" mobility data increases transparency because insights are not extracted by MNOs but rather by external researchers. Nevertheless, it would be important to clarify the numerous pre-processing decisions that MNOs make prior to distributing the "raw" data to interested parties. I would in fact argue that the data provided by MNOs is not raw at all. It would be important to clarify the pre-processing steps followed by MNOs for full transparency. For example, some MNOs apply algorithms to identify gender or age, or home locations, and these might introduce noise into the data, or even worse, remove certain users.

The data received by the European Commission from the MNOs is indeed not "raw". This point has been clarified in the text.

2.   In Section 2, it would be important to clarify what do the authors mean by "harmonization". How was the "common denominator" defined? Also, please clarify the "Trusted Smart Statistics concept" mentioned as a better solution in the paper.

Clarifications of the terms "harmonisation", "common denominator" and "Trusted Smart Statistics" have been added to the text.

3.   The privacy challenge (page 9) left me wondering about how the "Reasonability test" has been carried out. How did the authors assess re-identification risks if they did not have the original data? Also in this section, it would be important to clarify what are the "legitimate business interests of the operators" that were protected, mostly to clear concerns and for full transparency

See answer to previous comments.

4.   There is no "Data availability statement" provided. Data was indirectly used for this paper, I just wanted to raise this point for the Editor, not sure if it would be required in this case.

The authors did not derive insights from the data in the framework of this manuscript, which is meant to describe the initiative, impact and lessons learnt. The reader is directed towards research on the analysis of the data throughout the paper.

The authors would like to thank all the reviewers once again for the valuable comments.