Mozilla's contribution to the European Commission's European Strategy for Data - Public Consultation

29 May 2020

Mozilla is the Corporation behind the Firefox web browser and the Pocket "read-it-later" application; products that are used by hundreds of millions of individuals around the world. Mozilla's parent is a not-for-profit foundation that focuses on fuelling the movement for a healthy internet. Finally, Mozilla is a global community of thousands of contributors and developers who work together to keep the internet open and accessible for all.

Since its founding in 1998, Mozilla has championed human-rights compliant innovation as well as choice, control, and privacy for people on the internet. In that context, this document sets out Mozilla's initial thoughts on the European strategy for data, and well as insights from our experience of stewarding data for ethical and socially good ends.

Our response addresses specific topics within section one and section two of the consultation. Specifically, we will provide additional insight on the following -

Section one:
- HARNESSING THE POTENTIAL OF OPEN DATA
- EMPOWERING USERS AND ADDRESSING COLLECTIVE HARMS
- DATA LITERACY

Section two:
- LEAN DATA PRACTICES
- DATA STEWARDSHIP AND FIREFOX
- THE POTENTIAL OF DATA INTERMEDIARIES

Section 1: General questions on the data strategy

Data can play a key role in a modern industrial policy, as well as serve as the basis for insights and innovations that advance the public interest. To achieve its objectives and ensure the trust of EU citizens, there are a number of key principles that the Commission should adhere to when implementing the strategy.

In the below we provide insights on how to manage a data operation, in a way that protects the rights of individuals, maintains trust, and allows for innovation.

HARNESSING THE POTENTIAL OF OPEN DATA

As part of our commitment to ethical data and working in the open, we publish the [Firefox Public Data Report](), a weekly public overview on the activity, behavior, and hardware configuration of Firefox users.

The purpose of the report is two-fold:
- **Empowerment:** We want to empower developers, journalists, and the overall public to better understand the state of the web and the direction of trends in web browsing.
- **Transparency:** At Mozilla, we like to say that we are 'Open by Design'. We believe in an open web, so data and insights from the public should be made public, so the public can benefit.

The Firefox Public Data report publishes non-sensitive telemetry data that we gather from individual Firefox installations, including data on the browser's performance, hardware, usage and customizations. All data undergoes an extensive review process to ensure that anything we collect is necessary and secure.

With this data, we aggregate metrics for a variety of use cases, from tracking crash rates to answering specific product questions (e.g. how many clients have add-ons? [33% at last count]().) In addition we measure the impact of experiments that we run to improve the browser. We make this data available to the public in a clear and intelligible manner because we recognise the power of open data in advancing research and innovation. For instance, to show what an internet outage looks like, we've recently released an [aggregate open dataset on Italy's mid-pandemic internet outage](). [We've also published novel data]()from our telemetry datasets to advance research around the efficacy of social distancing measures to combat the spread of the COVID-19 pandemic. The data captured changes to Firefox users' engagement over time, a potentially useful source of insight for researchers seeking to understand changes in individuals' daily habits as a means of understanding the impact of social distancing measures.

We believe the ethos underpinning the Firefox Public Data Report, and the operational principles that characterise it, are an important case study in how open data can advance the public interest while maintaining trust and privacy.

ADDRESSING COLLECTIVE HARMS

Mozilla believes that with regards to data, especially personal data, the individual must be empowered to make informed decisions about how their data is used, including when and by

---

whom it is collected, and for what purposes it is processed. It is also crucial that any value extracted from such data benefits the individual, and not only the entity collecting and processing the data.

As the GDPR protects largely *individual* rights, it will also be crucial to look at data protection and use on the societal level. In order to foster the development of data ecosystems where data can be leveraged to serve collective benefits, the legal and policy frameworks must also reflect an understanding of potential collective harms and how to mitigate them.

We would like to direct the Commission to Mozilla's response to the AI public consultation, in particular the narrative submission provided in the annex (which will be submitted shortly after this one). This document outlines in greater detail our view on mitigating the potential harms and negative outcomes of machine learning technologies, which relate to the need to assess the spectrum of risks and how they interact with one another.

With regards to protecting against harms, a crucial aspect to be considered is that individuals should not have to shoulder the burden of understanding and mitigating all the potential harms that could arise from the development of data ecosystems. Any harm resulting from these ecosystems is likely to be too pervasive to be tackled by single complaints, due mainly to the existing asymmetries of power between individuals, private companies, and governments. Such an approach would also put a disproportionate burden on users to protect themselves, as the past few years have demonstrated that the average user struggles to understand how to manage their data, due to the complexity of e.g. terms of services and technological solutions.

EMPOWERING USERS

While we agree that enhancing data literacy is a laudable objective, the above section illustrates why data literacy is not a silver bullet in mitigating the risks and harms that would emerge in the unbridled data economy. Data literacy - i.e. the ability to understand, assess, and ultimately choose between certain data-driven market offerings - is effective only if there is actually meaningful choice of privacy-respecting goods and services for consumers. If the only choice consumers have is among an assortment of privacy-invasive goods or services, then data literacy will amount to very low material impact (unless consumers simply withdraw from the market). Unfortunately many products available today - including e.g. many smart home devices - do not meet basic security standards, as evidenced by the Mozilla Foundation "[Privacy Not Included](#)" buyer's guide. While many individuals may want to protect their privacy, or make use of the data that can be collected about them, there aren't many readily available options in the market. In short, it's a chicken and egg situation: the market must not only respond to demand, but it can also model good practices and 'data literacy' to help foster consumer demand for these types of products and services.

Section 2: Data governance

We agree with the Commission's assertion that the use of data in the society and the economy raises a series of questions of legal, ethical, organisational, and technical nature. Many angles need to be looked at in order to fully reap the benefits of the use of data without harm. In the below, we provide a series of examples from our own product and policy work that we believe can serve as useful guidance for the Commission in developing a forward-looking and trustful model for data innovation.

LEAN DATA PRACTICES

Our product development ethos is guided by a belief that transparency and user-control is key to ensuring individuals trust our products and our ability to protect their interests. Earned trust can drive a virtuous cycle of adoption, while conversely, mistrust created by even just a few companies can drive a negative cycle that can damage a whole ecosystem.

As a company steeped in the open-source tradition, we make conscious efforts to help other stakeholders improve their approach to data use, by setting best practice standards and providing resources and tools for others to build on. One such resource is our Lean Data Practices (LDP) toolbox.

The LDP toolbox helps companies and projects of all sizes to earn trust by staying lean and being smart about collecting and using data. Lean Data Practices are not principles, nor are they a way to address legal compliance - rather, they are a framework to help companies think about the decisions they make about data. They do not prescribe a particular outcome and can help even the smallest companies to begin building user trust by fostering transparency and user control.

We have designed Lean Data Practices to be simple and direct:

1. stay lean by focusing on data you need,
2. build in security appropriate to the data you have, and
3. engage your users to help them understand how you use their data.

We have even created a toolkit to make it easy to implement them.

We use these practices as a starting point for our own decisions about data at Mozilla. We believe they could serve as a useful basis for companies, organisations, and public administrations in using data in a more ethical and trust-building manner. Moreover, we encourage the Commission to

build on the spirit of Lean Data Practices when building its data policy framework for the EU in the coming years.

More information on Mozilla's Lean Data Practices initiative can be found [here](#).

DATA STEWARDSHIP AND FIREFOX

In the same spirit as our external-facing Lean Data Practices initiative, we aim to give the individuals using our services meaningful control over their own data. To that end, we have developed a dedicated [data stewardship program](#), which reflects our commitment to data collection grounded in four key pillars:

- **Necessity:** We collect only as much data as is necessary when we can demonstrate a clear business case for that data collection
- **Privacy:** We give users meaningful choices and control over their own data
- **Transparency:** We make our decisions about data collection public and accessible
- **Accountability:** We assign accountability for the design, approval, and implementation of data collection

Mozilla's data stewardship program is core to how Mozilla approaches user data. Any employee who is interested can volunteer and help make sure that our collection efforts stay true to our goals and in line with our policies. It's not tied to a specific team and it gives employees an avenue to participate in furthering our mission of openness and transparency.

We believe the encouragement of these kinds of internal processes on the part of companies handling data should be a core component of the European Commission's data strategy. In the effort to build a deeper data ecosystem amongst European companies, the Commission cannot neglect its obligation to ensure that that diligence, trust, and accountability are integrated into the governance aspects of this new industrial approach.

THE POTENTIAL OF DATA INTERMEDIARIES

We welcome the Commission's exploration of novel means of data governance and management, under the broad umbrella of 'data intermediaries'. Mozilla is currently looking into different types of data governance models, such as trusts, as we believe this concept may hold promise. However, there are a range of challenges and complexities associated with the concept that will require careful navigation in order for trusts to meaningfully improve the state of data management and to achieve a truly ethical and trustworthy data ecosystem.

A couple of examples to provide further context around the kinds of approaches we are looking into are the following:

- **Data Trusts:** Similar to a land trust, a data trust is an independent intermediary between two parties: the people creating data ("data subjects") and the companies collecting that data ("data collectors"). The trust would have a fiduciary duty towards its members, and would negotiate data use with companies according to terms set by the trust. Different trusts might have different terms, and people would have the freedom to choose the trust that most aligns with their own expectations. Some data trusts already exist: for instance, UK Biobank, a charitable company with trustees, is managing genetic data from half a million people.

- **Data Cooperatives:** Another proposed approach is the data cooperative model. U.S. citizens first organized themselves into trade unions and credit unions as a counterweight to powerful companies and banks like Standard Oil and J.P. Morgan in the mid-19th century. The result was a complete rebalance of power between large and small players in the economy. Similar to a credit union, a data cooperative would have a fiduciary duty to manage and protect access to the personal data of its members. Because it could also run internal analytics about its members, the data co-op would also be in a strong position to negotiate better services for its members. Some data co-ops already exist: [The MIT Trust Data Consortium](#) has demonstrated a pilot version of this system.

While data intermediary models and related data governance approaches have the potential to provide the guardrails within which ethical and trustworthy data solutions can emerge in the EU, they have potential to aggravate, or create new and complex challenges of their own. Considerable work will need to be done to ensure these intermediaries do not duplicate the systemic problems that already exist today. For instance:

- **Security:** To the extent that data trusts act as custodians of data, this will require impeccable levels of security and data management; there is a major risk that unless this can be guaranteed data trusts could broaden the attack surface for abuse and misuse, amplifying the negative impact on the individuals and the collective. Major tech companies today, in part because their business models depend upon users giving them data, are able to dedicate significant resources and expertise to protecting that data from compromise. Yet even this is not enough in all instances - and thus this must go hand in hand with lean data practices to ensure that risks are not bigger than they need to be.

- **Legal:** There are a range of legal complexities to consider, from cross-jurisdictional challenges (e.g. within the EU, Europe, and globally), to consent models. On the former, and given the global nature of the economy and society, how would trusts legally interact, particularly in the case that some jurisdictions may rely on individual rights, while others on the collective? On the latter, it cannot be taken for granted that the challenges with data

literacy and the ability of users to truly understand how their data is being managed wouldn't manifest in a data trust model as well.

- **Trust:** If rights can be assigned to an entity, how can it be ensured they will behave in the best interest of the individual/collective? Put another way, if there are scruples with current data managers, how can it be ensured that the next generation's data managers will behave ethically and in the best interest of the individual?

While the intellectual concept of data trusts holds a lot of promise, and is a useful exercise to explore how data can be managed in more ethical and trustworthy ways, the devil is always in the details. We thus encourage the Commission to explore these concepts with caution. We remain at the disposal of the Commission to further discuss and share the outcomes of our exploratory research on this issue should it be of interest.

Conclusion

The EU's ambitions to realise the potential of data for industrial and social policy ends should not require that norms and standards of data protection must be compromised. On the contrary, developing a new data economy where trust and data protection are at the fore provides a key opportunity for the EU to provide global leadership. We thank the European Commission for this opportunity to contribute to this consultation, and we look forward to working with the EU institutions to develop a data ecosystem.

For any further information, please feel free to contact us at brussels@mozilla.com.