Data justice

The data dilemma
Take a minute and think about all the data about you, which you have online. What kind of data is it? Do you feel it is an accurate representation of yourself, or does it represent only one part of you? Is this the part of you that you want to disclose online and share with the world? Or with the digital industry? Are you the owner of the data about yourself, and do you know how to claim its ownership?

In the context of the Love Data Week\(^1\), we reflect on these questions by referring to the concept of Data Justice, i.e., “datafication in relation to social justice”\(^2\), to think about how people are made visible or invisible, threatened or empowered, because of their digital lives\(^3\).

At the EPFL Library, for instance, the exhibition Data Detox shed some light on this topic (content freely downloadable\(^4\)). It clearly explains how and why it is essential to regain control over one’s personal data (digital identity check, cleanup of accounts, tracking cookies, etc.). Other approaches other than detox exist, and the industry is waking up. For instance, initiatives range from the Facebook VPN for kids\(^5\), or the deep-learning e-assistants like Oyot\(^6\) of the EPFL Innovation Park, or a paper on Equality of Opportunity in Supervised Learning\(^7\) in collaboration with Google. Another approach relies on fooling online surveillance: to shift the balance of power between the trackers and the tracked, browser plugins like TrackMeNot\(^8\) or AdNauseam\(^9\) employ different obfuscation strategies.

The decentralization problem
Many projects exist that harness big data or even AI for good\(^10,11\), like crowdAI\(^12\), or help creating civic campaigns, like the Fight for the Future\(^13\). But why do they even exist? Many risks concerning machine learning are recently being highlighted: while they drive the sales of these tech giants, the inherent biases of their algorithms (e.g., biased training datasets\(^14\), extremization of suggestions, news rankings, echo-chamber, etc.) are not yet correctly addressed\(^15,16\).

Even neutral, open and community-driven initiatives like Wikipedia struggle with data justice\(^17,18,19,20,21\), showing a polarization might seem contextual, but it is in fact a more general phenomenon. The well-studied rich-get-richer phenomenon persists and, paradoxically, decentralized structures amplify its scale\(^22\). The best example is the World Wide Web itself: deviating from its foundational scope of making academic information freely available to anyone\(^23\), large digital enterprises monopolize the data circulating on it, especially our personal data (e.g., Google, Facebook or Amazon in the West\(^24,25\), or Alibaba, Baidu or Tencent in the East\(^26\)).

FAIR data, fair data
In the context of sharing and publishing data, it is important to keep in mind, who will eventually have access to data. But, how to use of the published data as fair and neutral as possible? Context (scientific research, private information, etc.) and documentation (metadata, licensing, etc.) are key, but a critical reflection to share is necessary, for each dataset. That is why knowledgeable policymakers, as well as researchers and scholars, are increasingly pushing for the application of the so-called FAIR principles of datasets, which is to manage data in such a way to make them be Findable, Accessible, Interoperable, and Reusable\(^27\).

Open Data alone will not solve the issues ultimately related to the way we, humans, want to use data, but it contributes to make the processes more transparent and just at the social scale.
**GDPR in numbers**

Usually, national data protection authorities lead the investigations, while the other concerned authorities support them. If in disagreement, the European Data Protection Board will arbitrate. *Source: The European Data Protection Board.*

---

2. [https://datajusticeproject.net/](https://datajusticeproject.net/)
3. [https://doi.org/10.1177/2053951717736335](https://doi.org/10.1177/2053951717736335)
5. [https://techcrunch.com/2019/01/30/facebook-vpn-app-puts-spotlight-on-kids-consent/](https://techcrunch.com/2019/01/30/facebook-vpn-app-puts-spotlight-on-kids-consent/)
6. [https://www.privately.eu/#solutions](https://www.privately.eu/#solutions)
9. [https://adnouseam.io/](https://adnouseam.io/)
11. [https://aiforgood.itu.int/](https://aiforgood.itu.int/)
12. [https://www.crowdai.org/](https://www.crowdai.org/)
13. [https://www.fightforthefuture.org/](https://www.fightforthefuture.org/)
16. [https://nyupress.org/books/9781479837243/](https://nyupress.org/books/9781479837243/)
23. [https://www.w3.org/People/Berners-Lee/1991/08/art-6487.txt](https://www.w3.org/People/Berners-Lee/1991/08/art-6487.txt)