Executive summary – PrivacyTech White Paper
A new governance for personal data in the XXIst century

This document is an executive summary of the PrivacyTech White Paper that was presented at the French parliament on April 10th, 2019. The White Paper is a collaboration of more than 50 organizations (big corporations, startups, politics, academics, NGOs, etc.) from 14 countries.

A new digital society (p. 6)

We believe that personal data is one of the most essential societal issues of the XXIst century. Personal data is not solely about technology as it encompasses also legal, design, social and economic issues. At the same time, in a technological world, **technological standards are key as they shape actual practices**, so there is a need to ensure that the values we want to prevail in our society are well embedded in the technologies we create and use every day. We want a responsible and ethical digital society that benefits people, businesses and society.

A fair and efficient digital society requires personal data to flow freely under the strict control of individuals themselves. It is perfectly **aligned with the EU Digital Single Market strategy**. As the importance of personal data in society continues to expand, **it becomes increasingly urgent to make sure individuals are in a position to control their personal data**, but also to gain personal knowledge from them and to claim their share of benefits. Healthcare, smart cities, environmental protection count among the many examples of tomorrow's challenges requiring an optimal circulation of data between organizations and, at the same time, a use that is well understood, accepted and most of all put under the control of individuals.

From GDPR and beyond (p. 14)

EU GDPR (General Data Protection Regulation) paved the way for this new era by defining a new legal framework for personal data, ahead of technology. **GDPR strengthens trust in the Digital ecosystem** that sometimes appeared as boundless. GDPR has introduced strong legal tools we can use and build upon:

- **Data portability** is today an enforceable right for every EU resident. It asks organizations to open-up their information system and allow individuals to get back their data in order to transfer such data to other organizations. The new portability right is an unactivated trigger for data circulation.
- There will not be trust in the data circulation ecosystem if it does not come with the right protection. GDPR also provides structural elements for protection like **privacy by design** principles, **the key notion of consent**, minimization, transparency, etc.

But GDPR is not enough**. We need to translate GDPR values and principles into applicable and widespread technological standards for it to truly benefit society, following a top down approach.** We also need to help organizations that try to implement GDPR principles to give feedback to law-makers and regulators. As the pace of technology is rapidly increasing, **we need to start a process of adaptive regulation, following a bottom up approach**. The best example is the right to portability itself; it is not enough today, and it could be improved to better facilitate data circulation.

Potential and benefits of data portability (p. 21)

Free flow of personal data under individual's control **can "rebalance" the relationship between people and organizations that hold their data**. It can shake up what seems to be a *de facto* monopoly for data holders and simultaneously bring more transparency and trust to the ecosystem. **People can acquire or develop their digital empowerment** by promoting their autonomy from a specific actor in the data ecosystem.

Free flow of personal data is also **a source of development and growth through its creation of new services and enhanced competition**. In 2016, Ctrl-Shift consultancy estimated the market size, for the UK only, would reach £16.5bn or 1.2% of its economy. If we extrapolate roughly to the whole EU, the figure reaches €125bn. This figure will of course have to be refined, but the impact on the global economy will be significant.

Transparency and the exercise of rights are **core principles of Data Ethics**. The transparency of algorithms and the possibility to control the criteria of their decisions is a fundamental democratic requirement. At the same time technology is strengthening tremendously. We need to empower citizens, educate them and put them in control so that this new power can be used in a durable and accepted way. **The free flow of personal data based on individuals' control will furthermore allow a more ethical development of Artificial Intelligence**.

Why is portability a complex issue (p. 29)

There are two kinds of portability: **competition portability** where a person gets data back from a service to send it to a competitor, in order for the consumer to switch operators (a new bank, insurance company, streaming service etc.) and **complementary portability** where a person gets data back from a service to send it to a complementary service, to simplify its experience or benefit from combined services. **Markets will be more fluid and pro competition** as they will avoid data "lock-in" strategies and **more open and innovative as new entrants like startups will finally access data** that was kept by large platforms and/or incumbents.

As of today, people in the EU do not ask for their right to portability which appears very abstract and unpractical to them. **There needs to be a large education phase with GDPR rights**. On the companies' side, portability is still under evaluated or misunderstood. **Most companies enforce an *a minima* strategy regarding GDPR** in general and they do not see the sheer opportunity portability represents. When they do see it, they estimate that portability ROI is not clear yet and that the risk of disintermediation is too high, so they do not propose portability to their clients, who in return do not understand it. It is a vicious circle.

Portability and data circulation must find a sustainable business model. There are options to explore like a B2B user-centric where portability is free for the end user (as required by GDPR) but services can exchange value with one another in a transparent manner and always under the control of the individual. Other business model issues are debated like the opportunity for the end user to monetize personal data by him/herself.

Legal aspects of data circulation and portability are not cleared by GDPR entirely. Many companies are still not comfortable with the idea of data circulation. They fear liability issues if they open-up their data to other organizations. Data circulation has also become a media issue since Facebook's Cambridge Analytica scandal. Many corporates fear negative press coverage that would result from a misuse of data circulation. **We need to define a liability model for data circulation** that would be based on GDPR and would reassure all the stakeholders of the ecosystem. For it to be accessible to any company, whatever its resources, it has to be defined as a standard monitored through a proper governance.

There are many technical issues related to portability such as authentication methods, data transfers and formats or consent management for data reutilization, etc. All those issues can be addressed through standards and tools based on those standards. **Several types of architecture are also possible for portability**. Most companies today propose portability through CSV spreadsheets, which is not secure and does not create the conditions for data circulation. **Such a portability that does not rely on consent for reutilization is inefficient and dangerous for people and companies**. A better option is the use of APIs (Application programming interfaces) that allow direct transfer, following consumer consent, from a service to another. But APIs are complex to develop on the companies' side and complex to control on the end users' side. A simpler version of APIs is the aggregator model such as with Google or Facebook that develop their own ecosystems. The problem is that ecosystems are not often compatible together, privacy is not always at the heart of those ecosystems and it is complex to control from a user perspective. **The alternative we propose is to build APIs centered around the user thanks to new tools, PIMS or Personal Information Management Systems. PIMS allow individuals to manage their data**. They come with different value propositions (consent management, portability, data storage, etc.) but are all aimed at giving more control to the end user and at limiting the costs for companies. There exist hundreds of PIMS all over the world, most of them are startups and many are based in Europe.

Practical use cases (p. 62)

Data circulation is cross-sectorial and cross-borders by definition. **Potential is tremendous for many aspects of our digital lives and of our lives in general**. It will help people to optimize their transports, especially at the city level where portability is becoming a key element of the smart city strategy. It is key in education and job market, facilitating administrative tasks and adaptive learning processes. It is already a major issue in healthcare for administration, a better patient follow-up and data exchanges within research programs. It will help banks or retailers to become life coaches and insurance companies to benefit from a broader set of data. It could change the way administrations work, making every citizen lives simpler. **There is an infinity of other use cases in all industries (Retail, Telecom, Real Estate, Entertainment, Human Resources, etc.)**.

The necessity of trust (p. 105)

Trust is the prerequisite for a stable personal data flow. Part B of the PrivacyTech White Paper addresses these issues through a series of expert articles proposing best practices to ensure that individuals have confidence in the exchange of their data.

This trust is mainly based on control and awareness.

Three main propositions are emerging:

- Individuals' control over **access** to their data;

- The control of individuals over the **use** of their data;

- Individuals' **awareness** of the use of their data.

Individual control over access to their data: Re-balancing powers (p. 106)

In this section, via three articles from three different private structures, we explore **new architectures for applications that use personal data.**

The objective of these new technical architectures is to make it **extremely complicated or even impossible for the application to access personal data without the authorization of the data subject**. It is a question of changing the current paradigm where the one who owns the server that processes the data has full power.

Different choices are discussed:

1 - **Qwant case (p. 107):** the application never stores personal data on company-specific servers

Qwant presents the Masq architecture that **stores all encrypted data on the browser**. It is a serverless architecture that gives the application publisher no access to personal data.
All code runs only on the client's browser. Any new request for connection to Masq by an application must be confirmed by the user, which effectively ensures the protection of the user's data.
The Masq team will be limited by the technologies and programming interfaces available on the different browsers.

2 - **Amborella proposal (p. 111):** the application stores encrypted data and trusted third parties share decryption keys

In this article Amborella details its architecture proposal to ensure control and transparency of the data flows of connected objects: **the principle consists in separating the storage processes on the one hand and encryption on the other hand.**

The basic principle is to separate the storage and encryption processes. In other words, personal data must be stored encrypted, and the decryption means (key management) must be managed by an entity totally independent of the one that stores the data. Thus, accessing the stored data, whether as operator or simply as host, will require revocable authorization at any time, limited to a specific set of data, which will be administered by a trusted third party.

This solution also makes it possible to track each data exchange and use.

3 - **Whaller case (p. 116) :** Public and private social spheres

This article illustrates the architecture choices made by a social network (Whaller) to allow everyone to control the access and visibility of their data on the network.

The architecture of this platform of social and collaborative networks has been built with respect for three fundamental needs of organizations and individuals: privacy, relationship and belonging.

Whaller allows you to **have different profiles to attach to different networks** according to your social choices. Each profile is hermetically sealed to the others via the community with which it is associated.

This way, each person can control the visibility of their data according to the space in which it is deployed.

These various proposals make it possible to **open up a debate on a rebalancing of the powers of access to personal data** and thus restore trust by giving the individual the power to choose who has access to his or her personal data.

The individual's control over the use of his or her data: the technical standards of consent (p. 116)

This section focuses on how to **limit the use of personal data to specific and authorized uses by the individual**, once the information system has access to the personal data.

The aim is therefore to ensure **automatic and technical compliance with "consent" within the meaning of the GDPR** : by establishing **standards and technical rules** to ensure that the authorization is respected.

Experts, lawyers, engineers, researchers address this subject in various articles:

1 - **Feedback on advertising consent and current and new standards (p.117)**

We all have experience in **authorizing advertising cookies**, this article looks back at the technical implementation of these authorizations, the first figures and proposals for other standards.

2 - **Consent within a sensitive data exchange ecosystem :**

This subject is addressed in a series of **three articles to address the key issue of respecting consent: how to ensure that the entire data exchange chain respects consent**. The example of digital health with **sensitive data exchanges amongst connected objects** is taken to illustrate the issues and implementation (p. 121).

Standards to be developed are proposed for **exchanging**, **verifying and updating consent** (p 127).

The solution of using a **trusted third party to manage consents** is discussed (p. 124).

In order to ensure compliance and proof of authorizations, an article details the **relevance of using blockchain technology** (p. 133).

These articles provide **technical and legal solutions to ensure that the person's authorizations are respected**. Evidence and assurance that authorizations are being respected can build trust in the exchange and sharing of data.

Individual awareness of the use of their data: the challenges of transparency (p. 140)

This section focuses on a fundamental aspect of control over personal data and therefore trust: **transparency**. It aims to **provide tools, reflections and methodologies to ensure that people have a real understanding** of how their data are used.

First of all, the articles address the **crucial and central place of transparency in the GDPR**, both as a principle (p 141) and as a condition of Privacy by Design (p 151). The **empirical method** is proposed to assess understanding, an essential criterion for the conformity of transparency (p 147).

Then **precise techniques and tools are detailed** to ensure effective transparency and put design at the service of understanding:

- Representation **icons** (p 157),

- **Legal Design Patterns**: the principles to be applied to inform about personal data and current limits (p 164).

Different **methodologies** are discussed to design the interfaces and the concept of **"Data Ethics"** is highlighted to guide transparency choices (p 185).

Finally, we discuss **transparency through sound in a world without interfaces**, that of connected objects (p 198) and the **limits of design alone to meet the challenges of understanding:** not everyone is equal when it comes to learning technologies (p 202).

We conclude this part on trust by pointing out the **conditions of trust: the assurance that access to and use of data are controlled, that authorizations are respected, must be enforced by open and proven technical rules, norms and standards.** For transparency and understanding, empirical methods are needed that validate and disseminate good practices.

Major standardization initiatives (p. 225)

Many initiatives have appeared over the past decade. GDPR has triggered many initiatives all over the world. Among the most noticeable are **the DTP (Data Transfer Project) from Google, Facebook, Microsoft and Twitter**. DTP started in 2018 and aims at offering a technological standard for portability, it is moving fast. The other major project is **Solid, created by Tim Berners Lee,** the inventor of the Web. Solid aims at offering a new architecture for the Web, decoupling data from applications and allowing individuals to choose where their data can be used and for which purpose.

The European Self Data Movements (p. 233)

There has been an existing movement in Europe, for years now, the Self Data. It is represented by different organizations all over Europe like **MyData, a NGO from Finland that spread over 20 locations over the world**, **La FING, an association in France** that joined the MyData network, **Stiftung Datenschutz, a Federal government-funded foundation from Germany** that works with the German government and industrials, and **Ctrl-Shift, a UK based consultancy firm** that works with several governments and advises companies on the topic. Recently, the European organizations started to converge and federate.

Why do we need Standards? (p. 248)

Allowing a free flow of personal data under the individuals' control implies **a complex and important work on interfaces between a huge number of heterogeneous systems** (companies and administrations Information Systems) and end-users. Standards are one of the best means that can be used to achieve it. **Standards will substantially reduce costs** for every company/organization, **be easier and more secure** for individuals and will bring trust in the ecosystem while **spreading ethical values and principles**.

We already identified key issues that require new standards: data transfer, authentication, security protocols, consent management, data formats / ontologies, users' rights, privacy by design, business models, transparency, data circulation liability framework, non-personal reference data, personal information management systems (PIMS), etc.

Limits of existing standards initiatives (p. 247)

The energy and the will to build new standards is already there. Many domains and expertise are involved: legal, technical, design, business, policy. Many sectors are involved: Mobility, Healthcare, Administration, Commerce, Finance and insurance, Entertainment, Energy, Telecom, Job market, Education, etc. **The main problem is that all those initiatives are partial as they mostly work by country, by sector, by expertise or through closed consortia**. Despite our energy and good will, **we are recreating silos**, which will be highly detrimental to the main goal. Personal data circulation and protection is a cross-sector issue and data have no boundaries.

Building a new governance body for personal data (p. 251)

The governance body we recommend creating would be an independent and international standard supporting organization where member organizations would define technological standards, terminologies and guidelines to allow free flow of data under the individuals' control. The approach **would combine a transversal approach with expert workgroups** on all the identified topics (technical, design, legal, business, etc.), **a sectoral approach with sector hubs** (mobility, finance, health, administration, retail, etc.) as well as a cross-sectoral group. A Technical Board would help coordinate the hubs and work groups with other standards organizations, legislators, regulators, academics, users, etc.

As personal data is a global societal issue, **we believe that the responsibility for defining technological standards does not belong to one actor of the ecosystem anymore but has to emanate from global coordination** of most of them. And that such standards should be based on self-assessment practices, reinforced by targeted audit mechanisms. **Therefore, the Governance Body would promote a mix of a top-down approach as it translates regulations into technological standards and a bottom up approach as it gives market feedback to legislators and regulators (adaptive regulation process)**. It must also ensure a thorough representativeness of its deliverables (end users, academics, incumbents, startups, etc.).

Timing is of the essence as the market won't wait and we need to build on GDPR today. It is key we, as diverse but complementary European and International entities, translate today some of GDPR guiding principles into values and tangible benefits for citizens and the economy. The aim of the governance body is to rapidly become global while spreading European values expressed in GDPR over individual rights through its standards.