Finnish views on the data governance model - Towards decentralised data operating ecosystems with federated roles at EU level

Finland has been stressing the importance of the data economy, as data is the very basis of all digitalization and technological development. To promote digital transition and data-driven development Finland organised a High-Level Conference on Data Economy and produced Principles for a human-centric, thriving and balanced data economy[1] during its Presidency of the Council of the EU in November 2019. In these Principles Finland underlines the need to develop EU-wide actions to ensure that we can effectively access, share, act, trust, innovate and learn with data. This supports the approach that data is as findable, accessible, interoperable and reusable (FAIR) as possible.

In this input we aim to describe the next step that is required to build a securely functioning EU-wide data economy that facilitates both the development of artificial intelligence (AI) and the pooling of large quantities of data, but even more importantly, facilitates the evolution of a trustworthily functioning single data market and the free flow of data within the EU and in accordance with the requirements of protection of personal data.

With respect to the data economy it is important to bear in mind that data is a general and wide concept and the business needs of private sector entities differ from the needs of the public sector, as regards the principle of free flow of data. Also, the different sets of data that the public sector can possess and the level of re-usability of that data can differ based on the applicable access regimes within the Member States and whether the processed data includes personal data. Therefore the current fragmented nature of data and it's governance should be taken as a starting point when taking steps towards a joint understanding and a new governance model. However, the biggest change in mindset and most important feature in the new common governance is that although different applications may be needed also in the future, we however must aim towards developing a unified common approach on a common data governance model and supportive structures, which embraces all these different features and characteristics that are entailed in the data economy.

The European data governance

Achieving the goals of the European data policy depends on how well we are able to create the necessary models, rules of practise and infrastructures that enable seamless and secure data exchange. The goal should be to ensure the mobility and usability of data across national, sectoral and scientific boundaries. The key factor in achieving this goal is to set a vision for an efficient and functioning *data governance model* and then work relentlessly to make it a reality. This governance model should not be built from a technical perspective, instead it should enshrine the principle of

---

[1] https://dataprinciples2019.fi/

Id Versionumero

| Ministry of Transport and Communications | Office
Eteläesplanadi 16
FI-00100 Helsinki | Postal address
PO Box 31
FI-00023 Government
Finland | Telephone
+358 295 16001 | www.lvm.fi
firstname.lastname@lvm.fi
kirjaamo@lvm.fi |

human-centricity and thus always include first and foremost the end-user perspective. The model should be developed in an agile, sustainable and cost-effective way to support new forms of data driven innovations and businesses as well as technologies such as AI applications.

The governance model at EU level should support a decentralised data operating environment and strengthen the necessary federated coordination tasks and roles. The key requirement is to ensure, that data collection, storing and sharing should be fully decentralised, whereas there are some tasks that should or could be taken care of by a number of distinctive entities in a coordinated manner. In order to create the governance model we need to also define various roles related to the model. This also includes responsibilities and rights attached to the various roles.

The EU plays an absolutely pivotal role in developing a new governance model for data, as this necessitates strong cross-sectoral cooperation and collaboration between the Member States.To this end there is a clear demand to establish a forum for coordination and cooperation that should involve parties from the public and private sector. We need both policy-level decision making to steer the development towards functioning decentralised data exchange as well as harmonized mechanisms, terminology and tools that would increase interoprability. **The EU Commission should consider initiating an EU-level institutional structure to coordinate the interoperability specifications, identify specific roles in the data governance model and consolidate the development of the data economy.**[2] This cross sectoral work does not exclude the importance to continue proceeding also on sector-specific basis and the group can support in pinpointing these sector-specific needs. For example standardisation is easier to be reached sectorally.

Building trust in the data economy is a critical element of it's succes. There is need for respect for human rights, rule of law, security, reliability and the inclusion of all relevant stakeholders. Therefore trust of our citizens in services, technologies, devices and applications related to the processing of data must be secured. Therefore along the development of common governance framework, the EU should work towards common criteria on security against which the businesses should evaluate their practices and level of security. The mechanisms could be fostered by voluntary labelling and auditing by third parties. Data security and privacy by design need to be integral parts of business and service development practices.

Defining roles for data ecosystems

Sharing data is all about collaborating with others and creating new ecosystems around data. This requires supporting the development of data governance structures, as well as various actors such as companies willing to trust others, collaborate and share data. Trust in the marketplace is a cultural thing but also built through rules and regulations, means to enforce them and to see that the wronged party is reconciled. The EU-level data governance framework should aim to benefit from network effects, as attracting more companies to share data increases the attractiveness for current and new actors in the market. Rules and regulations are an important part of the toolset. Defining good and applicable roles and ways of managing those is a key success factor on how usable the final system will be.

---

[2] One functioning example of such group is the DTLF (Digital Transport and Logistics Forum) which is a group consisting of the Commission, Member States and stakehodler experts working with the logistics related data sharing issues.

There are other roles as well, like the end users and the service providers. However, in this paper we concentrate on the new roles with new responsibilities and rights that need to be defined, especially with regard to the coordination tasks. These tasks enable the decentralized, data-operating environment to function. It should be noted that the same entity might possess multiple roles.

Responsible data holders (not owners)

Instead of trying to define data owners we should clarify the role of data holders and rights to access data, since there might be multiple requests to claim for shared governance (e.g. companies controlling personal data). Data holders are responsible for managing the access rights and terms of use to data. Metadata and data should always be digital in machine-readable format, and if possible, available as real-time data. Data should be stored only once and collected from the original source by using where possible Application Programming Interfaces (APIs) accompanied with machine-readable access and licensing terms as well as information related to IP-rights.

Data holders, or controllers as defined in the GDPR working on their behalf, are able to control the authorizations for data sharing via APIs but the restrictions on data sharing should be based on well-defined reasoning at the corporate policy level and should enable as widely as possible third-party access to data and thus generate value creation for the ecosystem. The functioning of ecosystems is based on the principle of decentralized data management. For this reason, it is important that the governance model supports a decentralized model throughout the data life cycle including production/collecting, storing and sharing.

Metadata operators

One of the most important roles for coordination would be the data directory or repository managers ("metadata operator"), who maintain a directory or repository on data collected and made available for reuse by different actors. Although there already exists multiple examples of metadata management, there is still a grave need to bring coherence and scalability across all sectors in order to facilitate higher data usability, innovations, development of new services and the use of artificial intelligence. These services should ensure easy findability and provide metadata on data and its terms of use in a machine-readable format.

Attached to this issue is the need for the individual organisations and businesses to publish information of the data they collect, store and use – like a free-form data balance sheet[3]. This would be highly useful to other companies, researchers and public instances, as it would help other instances to recognize what kinds of data are collected. It would also be revealing to the citizens, who often are the source of the data. Undoubtedly, companies, especially SMEs, would need help in providing such information, and that could be another task for "metadata operators".

---

[3] Organisations shoud publish information on the kinds of data they collect, store and use – like a free-form data balance sheet. This would be highly useful to other companies, researchers and public instances. It can be informal or more structured one, such as data balance sheet by a Finnish bank OP. This would also help other instances to recognise what kinds of data are collected.

MyData operators

"The individual in control of their own data" is not only a philosophical basis, but needs to be very concretely a key design driver for the governance model.The difference between personal and industry data is not always clear. For example IoT data can be very sensitive as almost all data reflects individual human activities and can be used for tracking. Furthermore, we should not allow overly confident or simplistic discussion about anonymisation. Non-personal data can lead to the tracking of individuals, as it's very hard effectively to anonymise data.[4] When we design a data governance model on the European level, we have to acknowledge this and start with a quite broad understanding of individual data sovereignty.

Our recommendations aim to encourage the development of such data spaces in a coordinated way, taking fully into account the requirements of protection of personal data. This has shown to be especially important in the current Covid-19 crisis and its aftermath.

Balanced and fair relationships between people and organisations do not emerge automatically in personal data ecosystems. There needs to be some explicit methods for human-centric governance to guarantee that individuals' rights established in the GDPR, especially article 20 are followed. For this we need mechanisms to guard and follow the operators function and we need ***trusted personal data intermediaries*** - so called Mydata operators[5].

Identity managers

It is also worth to consider bringing closer European-level identity management practises for citizens, companies, products, services, and their digital twins (cyber-physical systems) in all sectors in order to enable holistic next-generation digital systems development. The electronic identification approach is an initiative to acquire trust and a secured status. As we have electronic identification for citizens, a similar coordinated approach should be taken into use to manage data linked to identifiers of companies and their products and services.[6] This underlines the need for effective service providing on identity/identifier management, authentication and authorization/consents. The concept of digital rights and permits should be introduced together with the electronic identifiaction.

Intermediary platforms creating interoperability

Interoperability refers to systems' ability to work with other systems and products now and in the future. We need legislative interoperability, organizational interoperability

---

[4] To the extent that personal data are processed, the respect for fundamental rights and the requirements of a high level of protection as set out in the EU's General Data Protection Regulation (GDPR) must be fully taken into account. In particular, the principles of data minimisation and proportionality should apply.
[5] https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf
[6] **Elements of fair and functioning data economy: identity, consent and logging** is an step to contribute to the development of a fair and efficient data economy. This agreement provides requirement specifications for three key elements of any internet service using personal data and in which the agency for that data is within the individual whose data is being utilised: identity, consent and logging. ftp://ftp.cencenelec.eu/CWA/CEN/IHAN/CWA17525.pdf

as well as semantic interoperability such as vocabularies and ontologies and we need technical interoperability.[7]

Interoperability is vital to allow companies and individuals to migrate their own data and services between service providers in practice.[8] They should also be able to combine data from different sources without facing implementation issues. Interoperability can be leveraged through standards, formats, protocols and regulation. There are already systematic approaches in some sectors towards interoperability, but we need a holistic approach to solve all the remaining interoperability issues that still exist. If the challenge of interoperability cannot be solved, the aspiration of establishing a genuine data marketplace remains trivial. However, we need to make sure not to mix agile interoperability with full harmonization of systems as that often tends to lock future developments and innovations.

New intermediary platforms that aim to create interoperability have started to emerge. They create concrete tools for example to make APIs and services more interoperable with each other. These platforms can also offer additional services and promote data quality by request of the data holder or they can integrate or pool data into larger quantities. However, the key role of intermediary platforms' role is only to intermediate. That is, not to collect data but to make it flow between the parties in the ecosystems, as data marketplaces' or 'data brokers' do. There are solid examples of soft infrastructure for intermediaries in data sharing. It is highly advisable to build data spaces based on these pieces of industry-driven soft infrastructure.[9]

Companies should be supported with packages with easy-to-follow instructions, contract and process templates as well as education of the whole process of data handling, security and business models.

---

[7] European Interoperability Framework (EIF) creates a good basis for the work.

[8] Sitra's IHAN® project aims create a method for data exchange and to set up European level rules and guidelines for ethical use of data. The Blueprint document contains detailed functional requierements for all IHAN ecosystem functional components in end user, service provider and data provider levels. https://media.sitra.fi/2018/12/22091907/ihan-blueprint-2-5.pdf

[9] Such as Gaia-x (https://www.bmwi.de/Redaktion/EN/Artikel/Digital-World/data-infrastructure.html), Platform of Trust (https://platformoftrust.net/en/) and Value Net for agricultural purposes (https://www.cinia.fi/en/frontpage.html).