PUBLICATIONS

Rulebook for a fair data economy

The rulebook offers instructions and templates to facilitate data network building.

READ PUBLICATION

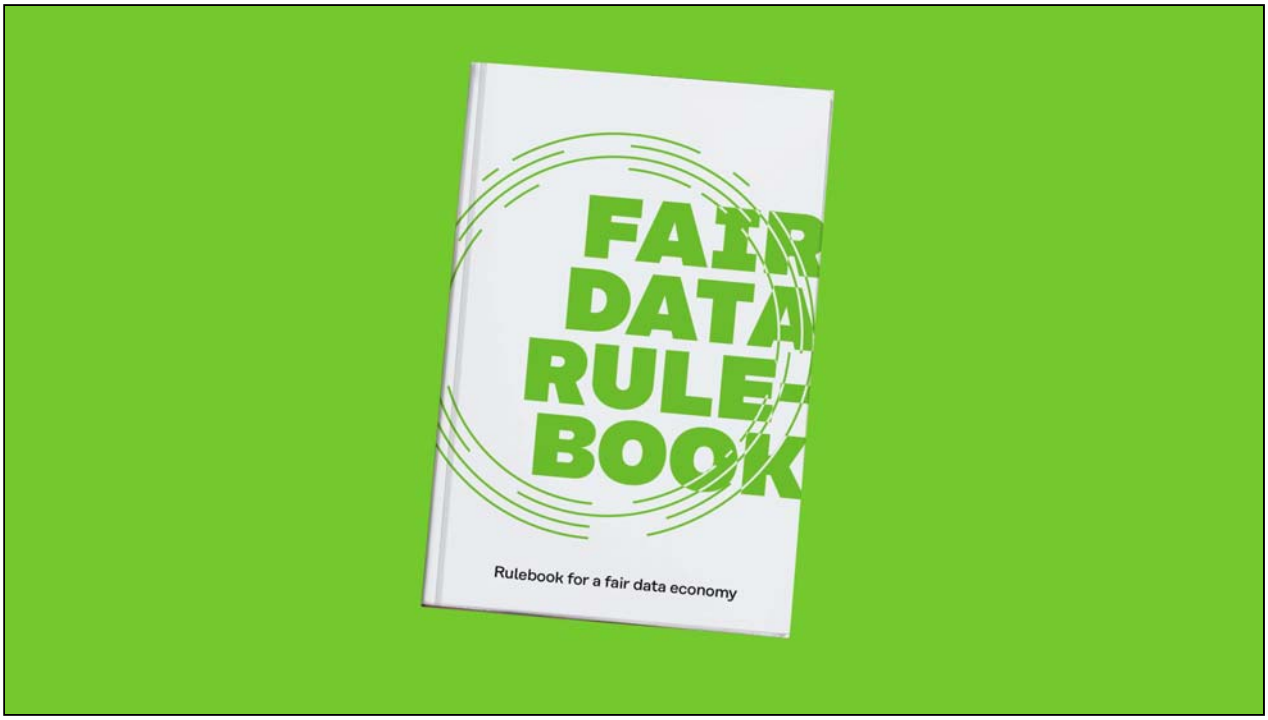


Image: Topias Dean / Sitra

WRITERS

Olli Pitkänen

Founder, CLO, 1001 Lakes Oy

Juhani Luoma-Kyyny

Sitra

PUBLISHED
August 31, 2022

CONTENTS

Close

Preface and templates

The rulebook for a fair data economy is a guide for creators of fair data economy networks. Agreement templates and other tools make it easier to build and join new data networks which highlight transparency in data sharing.

The rulebook contains:

- model agreement templates for legal, business, technical and administrative rules
- a range of control questions
- code of conduct templates

The rulebook consists of two parts: general content in the **part 1 below** and **editable templates in the part 2** (templates in a Word file and PDF).

Earlier versions:

The first Rulebook for a fair data economy was published on 30 June 2020 (version 1.1), and updated on 11 January 2021 (1.2), on 13 August 2021 (1.3) and published also in Finnish, updated on 20 April 2022 (1.4).

The version 1.2 of the Rulebook was published in Portuguese by the SPMS – Shared Services of the Ministry of Health in Portugal on 10 March 2021 and updated 30 March 2021:

- Manual para uma Economia Equitativa de Dados (1.2, PDF)

- Manual para uma Economia Equitativa de Dados (1.2, Word)

The Rulebook Working group:

The Rulebook Template was created by the Rulebook Working group under Sitra's Fair Data Economy theme.

The Rulebook Template has been actively contributed to by Olli Pitkänen (editor, 1001 Lakes Oy), Sami Jokela (1001 Lakes Oy), Marko Turpeinen (1001 Lakes Oy), Viivi Lähteenoja (1001 Lakes), Jyrki Suokas (Sitra), Juhani Luoma-Kyyny (editor, Sitra), Saara Malkamäki (Sitra), Anna Wäyrynen (Sitra and Adesso Nordics Oy), Jorma Yli-Jaakkola (Borenius Attorneys Ltd and Lexia Attorneys Ltd), Otto Lindholm (Dottir Attorneys Ltd), Jani Koskinen (University of Turku), Jussi Mäkinen (Technology Industries of Finland), Kai Kuohuva (TietoEVRY Oyj, Fortum Oyj), Jutta Suksi (VTT), Jari Juhanko (Aalto University), Kari Hiekkanen (Aalto University), Antti Kettunen (TietoEVRY Oyj), Petri Laine (Hybrida), Kari Uusitalo (Business Finland), Pekka Mäkelä (University of Helsinki), Meri Valtiala (The Human Colossus Foundation), Anna-Mari Rusanen (Ministry of Finance), and Sari Isokorpi (Medifilm Oy).

The Data Security Operating Model has been created on the initiative of Digipooli (Technology Industries of Finland), the organization of the Service Security, and with the support of The National Emergency Supply Agency. The Data Security Operating Model has been developed by 1001 Lakes Oy's experts Olli Pitkänen, Sami Jokela and Marko Turpeinen and Digipooli's pool secretary Antti Nyqvist. Digipooli members have actively participated in the work by presenting their views in workshops and commenting on the model.

Recommended citation: Sitra (2022), Rulebook for a Fair Data Economy, version 2.0

Part 1: why and how to use a rulebook

1. Introduction to Part 1

1.1 Why and when you should use a rulebook for data sharing

The purpose of this Rulebook Template as a whole is to provide an easily accessible and usable manual on how to establish a data network and to set out general terms and conditions for data sharing agreements. This Rulebook Template will help organisations to form new data networks, implement rulebooks for those data networks, and promote the fair data economy in general. With the aid of a

rulebook, parties can establish a data network based on mutual trust that shares a common mission, vision, and values.

A rulebook also helps data providers and data users to assess any requirements imposed by applicable legislation and contracts appropriately in addition to guiding them in adopting practices that promote the use of data and management of risks. However, despite the Rulebook Template, it is important to note that the parties still need to check for themselves that all the relevant legislation, especially on the national and subnational levels as well as specific legislation regulating the data in question, is considered.

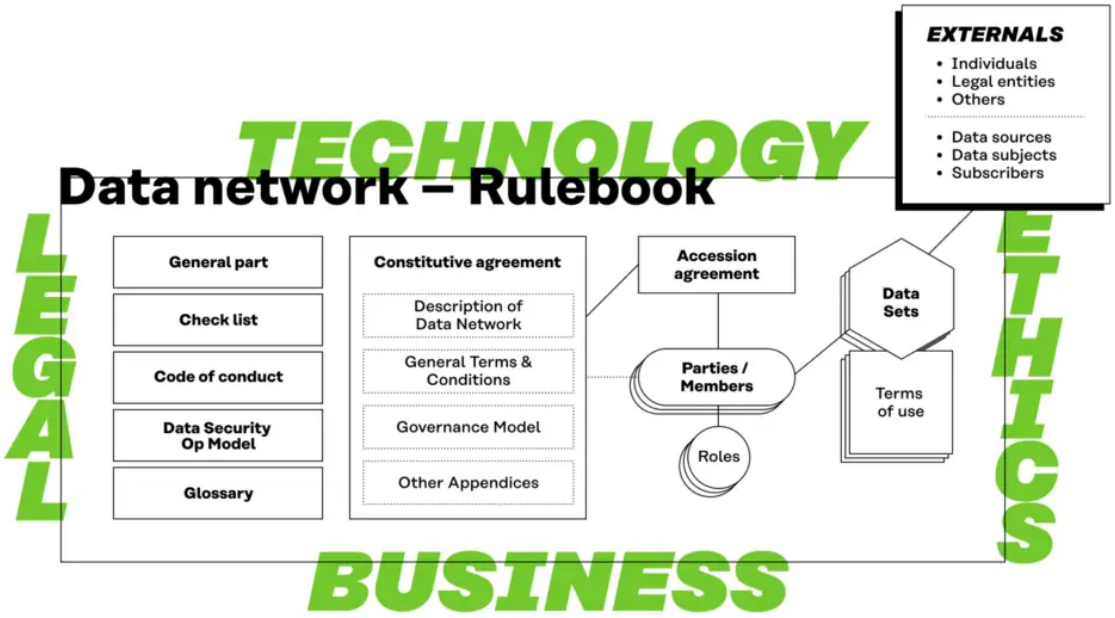
There are many benefits to sharing data. It may allow data users to access data for research purposes or for the development of their products and services. Sharing data may also allow data providers to improve their products or services and supporting the development of added value or services by third parties. The existence of rich ecosystems that create new products and services may become very attractive to users.

An increase in the number of service users, in turn, encourages new product and service developers and users to join the data ecosystem. This network effect may increase the value of a specific service and even the entire ecosystem. Furthermore, sharing data may lower the transaction costs of gathering data and allow data providers to combine their databases with minimal organisational changes.

Data networks that adopt the Rulebook Template must be fair, balanced and lawful in their processing of data. They must also be just and impartial toward their members and ensure that the rights of third parties are not infringed. Personal data must be processed in accordance with European and applicable national data protection regulations.

Data networks identify and manage risks associated with the sharing and processing of data while ensuring the exploitation of new possibilities that data offers. This includes also ensuring compliance with relevant competition legislation and that the data network will not have a negative impact on market competition and consumers. Provisions restricting access to the network are especially important to take into account in this kind of assessment.

Figure 1. The relations between different parts of a rulebook implemented from this Template.



It is possible to significantly improve commercial businesses and public services by better availing data. Sharing data across organization borders multiplies these opportunities. However, there are lots of obstacles that prevent cross-organizational data sharing. They include

- the lack of technical and semantic interoperability;
- inability to adequately identify different actors;
- the lack of data quality;
- cultural and attitudinal problems; difficulties in understanding the benefits from data sharing;
- risks related to losing control of data and trade secrets, infringing others' rights, and data protection; inability to coordinate data ecosystems and get all entities excited and involved;
- inability to define success and show value for all entities in a data ecosystem;
- inability to create a common vision, mission, purpose and values;
- inability to identify roles for each entity.

The Rulebook Template aims at helping to remove these obstacles. It enables and improves fairer, easier and more secure data sharing within data networks. A rulebook based on this Template describes legal, business, technical, and governance models that the members of the data network use when sharing data

with each other. It takes with the greatest importance into consideration ethical principles and especially the requirements that arouse from individuals' privacy and data protection.

The General Terms of the Rulebook Template as well as most of Glossary, Code of Conduct, and Checklists in contract Annexes are the same for all the data networks that use the Fair Data Economy Rulebook model. Only the Specific Terms are written case by case. Therefore, it is easier and more cost-effective to create data networks and ecosystems, if the rulebooks of different data networks have substantially similar basis. It simplifies collaboration and data sharing even between data networks and makes it easier for an organization to participate in several data networks. The similar Rulebooks ensure fair, sustainable and ethical business within the data ecosystems, which in turn enables increasing know-how, trust and common market practises.

In order to be able to use its own and others' data, the organization needs to understand broadly the business, legal, technical, and ethical perspectives of data sharing. It should especially recognize in which roles it acts in the data network, which data processing and refining capabilities it needs to have, and what are the minimum requirements to participate in the data network. The four main roles of the actors within a data network are:

1. **Data Provider**: one or several sources that provide the network with data. (Note: in earlier versions of the Rulebook Template for Data Networks, the term "Data Source" is used instead of Data Provider.)
2. **Service Provider**: one or several data refiners that combine data streams, refine data, and provide them further. Provides services to End-Users or as a subcontractor to other Service Providers.
3. **End-User**: one or several individuals or organizations for which a Service Provider has developed its services. Consumes, utilizes and accesses the value that is created in the data ecosystem.
4. **Infrastructure Operator**: one or several actors that provide identity management, consent management, logging, or service management services for the data network.

As recognised in the EU Data Governance Act, due in the summer of 2022, providers of data intermediation services are expected to play a key role in the future of the European data economy. A data intermediary can provide services to facilitate the aggregation and exchange of substantial amounts of data. Data intermediaries offering services that connect the different actors have the potential

to contribute to the efficient pooling of data as well as to the facilitation of bilateral data sharing. Specialised data intermediaries that are independent from both parties that offer up data and parties who make use of that data can have a facilitating role in the emergence of new data-driven ecosystems independent from any single player with a significant degree of market power.

Data networks, as described in this rulebook, can also include actors who provide data intermediation services. However, it is essential to keep in mind that each of the members of the data network can operate in multiple roles at the same time or change roles over time. Service Providers and Infrastructure Operators are natural candidates for independent data intermediaries in accordance with the Data Governance Act. Nonetheless, it should be noted that they are not always independent, but Data Providers and End-Users may occasionally also provide services or operate the infrastructure.

Also, note that in a wider context, even Data Providers may get data from external sources and there can be external parties, Subscribers, that receive data from the data network in accordance with the Data Sets' Terms of Use although they are not Parties of the Constitutive Agreement. The starting point is that the rulebook is open and public, which is required by the transparency principle and the data protection legislation. However, the network-specific parts of a rulebook contain also confidential rules that are not disclosed outside the data network.

1.2 Quick Start Guide: How to start working on a rulebook for data sharing

How to Start?

If you want to initiate a rulebook-based data network, follow these steps:

1. Read through this introduction (Part I) to familiarize with the various dimensions of the rulebook model.
2. Go through the Rulebook Template section (Part II). Your starting point may depend on the competences of your core team participants: business issues, legal experts, technology experts, ethical experts. You will work on these iteratively and expand with more stakeholders as the content of your version of the Rulebook matures.
3. Fill in the Description of the Network, both the Business & Operational Part and the Technology & Security Part. Go through and answer the Rulebook Template's Checklists in Contract Annexes to define how your version of the Rulebook should be implemented.

4. Check if you want to add more terms into the Glossary or change the existing definitions.
5. Read carefully all the Contractual parts and decide, how you want to complete them, and which terms and conditions need to be changed in your case.
6. Ask the Founding Members to sign the Constitutive Agreement and start sharing data. New members may join the data network by signing an accession agreement. The data network is governed in accordance with the Governance Model.
7. Give feedback to us on what kind of changes and amendments you made to the Rulebook and how we could improve the Templates.

1.3 Context and key concepts

Rulebooks are an important tool for a fair data economy and data sharing in general. To get the most out of a rulebook-building process and its eventual implementation and use, it is important to understand the larger context of rulebooks.

A fair data economy proposes pursuing two goals at once: putting individuals in control of their data and maximising the use of data. A fair data economy can serve the interests of individuals, existing service providers and data re-users alike, based on data portability and consent. The societal benefits of permission-driven data sharing include economic growth, individual empowerment, and broad societal benefits. (A roadmap for a fair data eoconomy, Sitra)

In keeping with these goals, the Fair Data Economy Rulebook is a tool for governing data networks which are needed for maximising data use and for aligning them with the ethical considerations for putting individuals in control of personal data about themselves. This rulebook, consists of a guide (Part 1) and a set of templates (Part 2) that can be modified, adopted, and used for the needs of a specific data network.

Data sharing requires a certain number of rules – who can and should do what, with which data, and so on. Bilateral data sharing is relatively straightforward compared to multilateral or networked data sharing and its rules are commonly set out in a contract between the two parties which governs the terms of sharing. When a larger number of parties decide to share data between themselves, a more complex form of governance is appropriate.

A rulebook is a collection of documents that can be used together to govern a data network, by which we mean a multilateral data sharing arrangement. More

precisely, a data network is a network consisting of a set of more than two parties that share data among each other. The goal of data networks is sharing data between parties in a responsible and legally sound way so that all can benefit.

There are several important concepts related to, and partially overlapping with, data networks. A data space is "a federated data ecosystem within a certain application domain and based on shared policies and rules" (Design priniples for data spaces – Position paper). Specific data networks can be parts of these larger, domain-specific data spaces in which they are connected in some defined way with other networks or methods of data sharing.

Data ecosystems, on the other hand, are complex and interrelated systems of data, including but not limited to data spaces and data networks that may be governed by rulebooks.

Finally, there exist a number of terms to describe either the *roles* or *entities* that are parties to data sharing arrangements. In this rulebook, we use 'data provider' for any natural or legal person who provides (or permits the provision of) data for other parties to the data network. Other terms used in other contexts for this role are 'data source', 'data holder', 'data rights holder'; or 'data subject' if speaking exclusively of natural persons. Further, we use 'data source' for the specific *technical system* that contains or generates the data that is shared (e.g. a sensor).

Note that the agreement templates in this rulebook include additional, legally binding definitions of some terms that are used in the agreement. If the above and the definitions in the contract definitons are in conflict, the definitions in the agreement prevail legally.

2. Contractual Framework

2.1 Introduction

The Contractual Framework defines of the Rulebook template consists of the following parts:

Constitutive Agreement

- General Terms and Conditions
- Governance Model
- Description of the Data Network
  - Business and Operational Annex
  - Technology and Security Annex

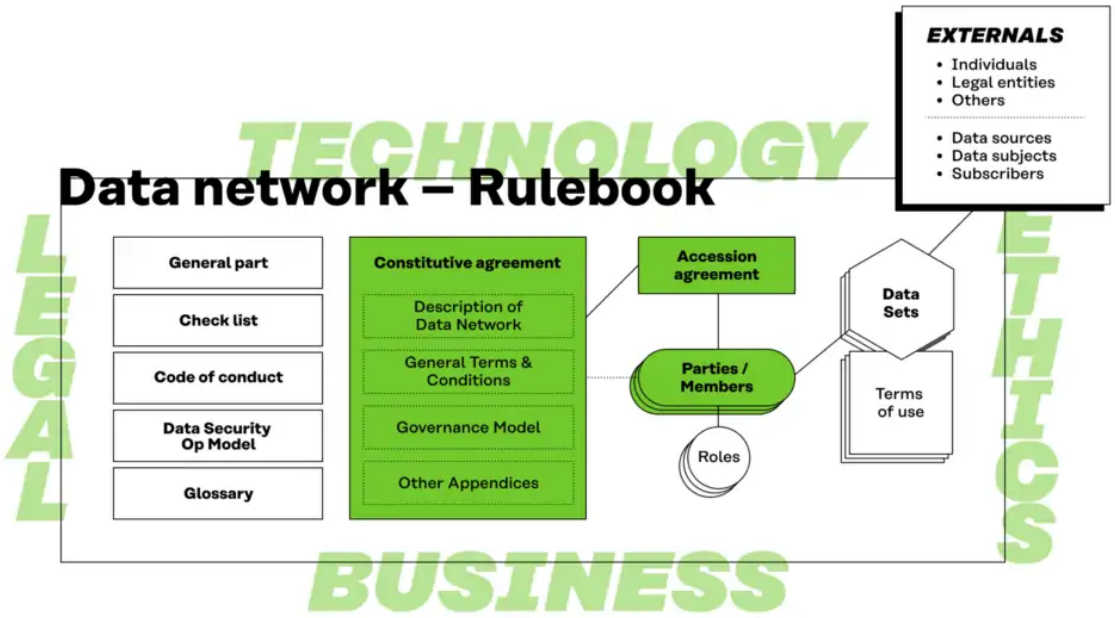
- Accession Agreement

Dataset Terms of Use

- Description of the Data Network
  - Business Part
  - Technology Part

The members of the Data Network are parties to the Constitutive Agreement either directly (the Founding Members) or through an Accession Agreement.

Figure 2. Contractual Framework's position in the entirety of the Rulebook.



Data Networks are established under the Constitutive Agreement, which is concluded by and between the Founding Members of the relevant Data Network. The General Terms and Conditions are included as an Appendix to the Constitutive Agreement.

Although the intention behind the General Terms and Conditions is to have them serve as a one size fits all baseline solution for various Data Networks, the reality is that each Data Network will require specific modifications to be made to the General Terms and Conditions. For this purpose, the template Constitutive Agreement includes a designated section for derogations from the General Terms and Conditions, which the Founding Members should review and amend to ensure that the contractual framework suits their Data Network. As such, the final contents of different Data Networks' Constitutive Agreements and their Appendices are expected to differ to a material degree.

We recommend that the Founding Members do not amend the General Terms and Conditions document itself but rather include any relevant amendments as derogations from the Constitutive Agreement. This will enable the Members to easily identify which amendments have been made without the need to compare the original General Terms and Conditions document to the amended version.

The Founding Members may allow new Members to join the Data Network under an Accession Agreement. Where the Data Network is established to allow for this kind of access, the Founding Members should describe the applicable accession criteria for new Members in the Constitutive Agreement. Furthermore, the Founding Members should consider whether they should define the criteria and process for accepting new Members to the Data Network in the Governance Model Appendix, together with other governance framework related matters that must be taken into consideration during the life cycle of the Data Network.

The Governance Model Appendix assumes that each Member nominates a representative to serve on the Steering Committee. The Steering Committee's mandate has been defined in a relatively broad manner to facilitate collaboration between the members and to organise the administration of the Data Network appropriately on a strategic level. This includes, e.g., a mandate to amend the Constitutive Agreement by a qualified majority of the Steering Committee representatives.

The purpose of the General Terms and Conditions is to serve as a tool during the operational phase of a Data Network. On the one hand, establishing a Data Network may involve material joint project investments by the Founding Members, while on the other hand, establishing a Data Network could require the Members to carry out individual actions. Any such potential project agreement by and between the Founding Members must be concluded separately and, where the Founding Members are open to welcoming new Members to the Data Network at a later stage, their contribution to cover the project costs should be agreed in the Constitutive Agreement and in any Accession Agreements.

In addition, the Members should also define any fixed term commitments for sharing the Data within the Data Network, e.g., where the Members seek to recover any investments, they have made for the purposes of establishing the Network or, alternatively, where they require reciprocity while sharing the Data.

The purpose of the template Dataset Terms of Use is to provide a template for the Data Providers to define the detailed terms and conditions that apply to the Dataset(s) that the respective Data Provider makes available within the Data Network. Where the Data Provider allows redistribution of the Data to any Third

Parties, the Data Provider should also define any applicable Dataset specific terms and conditions in the Dataset Terms of Use that the Members should include in their agreement with such Third Parties regarding the redistribution of the Data thereto.

By using the General Terms and Conditions, the parties agree to comply with them, unless the parties expressly decide to derogate from the General Terms and Conditions in the Constitutive Agreement. The Dataset Terms of Use, on the other hand, are supposed to be defined separately for each Dataset by the relevant Data Provider that makes the Data available to the Data Network.

The roles identified under the General Terms and Conditions for the Members of the Data Network include:

- **Data Provider** (makes data available within the Network);
- **Service Provider** (processes Data to provide related services and redistributes the Data, such as anonymisation, pseudonymisation or combination of Data);
- **End User** (uses the Data in its business); and
- **Operator** (provides services to facilitate the operation of the Network, such as provision of APIs, management of identities, connections and/or contracts).

In addition, **Third Party End User** has been identified as a role for any Third Parties who receive Data from Service Providers where the respective Data Provider has permitted such transmission of the Data.

It should be noted that individual parties may assume several roles within a specific Data Network and that, on the other hand, Data Networks may not necessarily require all roles. For example, the roles of Operator or even Service Provider may not be relevant if the parties exchange Data among themselves and use the Data in their respective businesses. On the other hand, the Data may pass through several Service Providers in certain Data Networks before End Users or Third-Party End Users receive and use it.

2.2 Premises

Both the Data shared in various Data Networks and the terms and conditions that apply to the Data may vary significantly. As it is not feasible to define a comprehensive library of terms and conditions that would cover all possible scenarios, provided in this rulebook template is a simple set of premises the authors assumed for the template contractual framework.

- the Data Provider may decide, separately for each Dataset, the Parties who are granted access to the Data;
- unless otherwise defined by the Data Provider in the Dataset Terms of Use or agreed by the Members, the Data Provider grants the right to use the Data free of charge;
- the provision of Data within the Data Network does not constitute a transfer of Intellectual Property Rights;
- the Data can be redistributed only to the Members of the Network, but Data Providers may allow redistribution of the Data to Third Party End Users under the applicable Dataset Terms of Use;
- the members are entitled to redistribute Derived Materials to third Parties, subject to possible additional requirements related to Intellectual Property Rights, and Confidential Information;
- where the Data involves Personal Data, by default the data recipient becomes a data controller;
- the Data Provider indemnifies other Parties against claims that its Data, which is subject to any fees, infringes Intellectual Property Rights or Confidential Information in the country of the Data Provider;
- the Members are entitled to use the Data after the termination of the Constitutive Agreement, in which case the Constitutive Agreement survives the termination, except for where the Constitutive Agreement is terminated as a result of Party's material breach; and
- the Data Provider is entitled to carry out audits related to its Data.

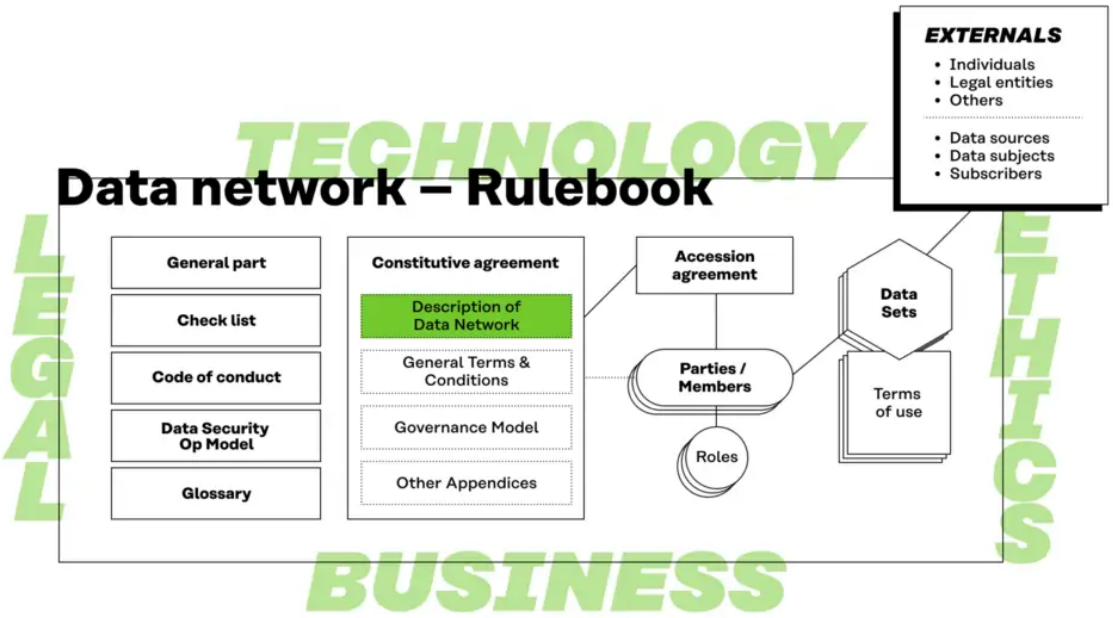
Process-wise, the Members need to carefully analyse their needs and objectives against the principles above. If needed, the Members of the relevant Data Network may wish to amend these principles on a case-by-case basis either at the level of the Data Network by indicating any necessary derogations from the General Terms and Conditions in the Constitutive Agreement and/or by defining a more detailed template for the Data Network specific Dataset Terms of Use.

In addition, each Data Provider should define, within the framework established for their respective Data Network, the terms and conditions that apply to their Data. Furthermore, more detailed conditions may be added in order to accommodate for different and more multi-faceted business models and e.g., framework for processing of Personal Data. The Members of the Data Network may also need to add a mechanism that facilitates transfer of data also to third parties.

3. How to describe your Data Network?

This part of the Constitutive Agreement describes the Data Network. It consists of two subparts: Business and Operations and Technical and Security.

Figure 3. Description's position within the entirety of the Rulebook.



3.1 Business and Operations Perspective

3.1.1 Introduction

Business and Operations collects the business and operational decisions raised via the rulebook checklist questions and answers and otherwise during the ecosystem design.

The Business and Operations part is divided into two main sub-parts; data ecosystem canvas and accompanying questions define the high-level summary and structure for the business aspects of the data network, and subsequent chapters provide the more in-depth business design information that does not fit in the canvas itself. Further design related documents (contracts, presentations, etc.) can be added to provide additional level of detail for the data network, if available and needed.

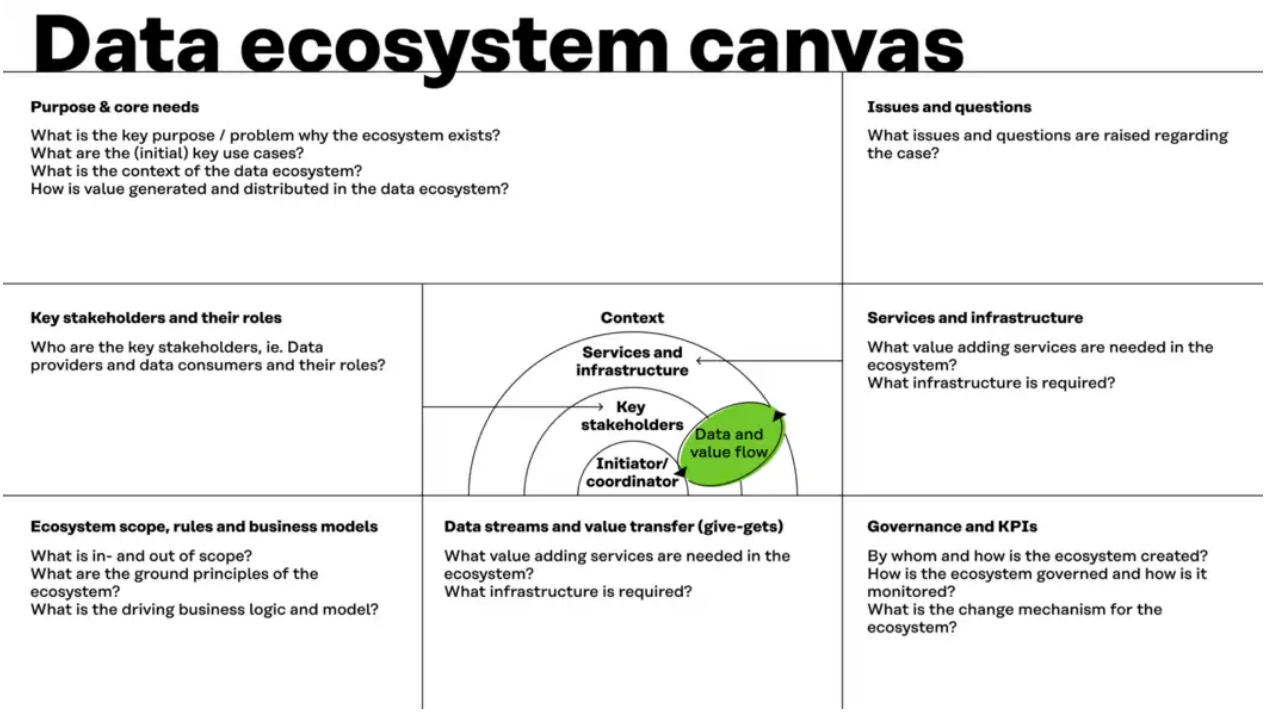
3.1.2 Data Ecosystem Canvas

The following Data Ecosystem Canvas is a high-level summary of the business design for the data network. Business and operational rulebook checklist questions (section 4.6.1.3) help in defining the content in each of the canvas fields.

The answers cover the following categories:

- **Purpose and goals:** Key raison d'être for the data network.
- **Roles and responsibilities:** Naming the key stakeholders, their roles and responsibilities.
- **Business logic and data value:** Value exchange and business models that are made possible by the data ecosystem.
- **Governance:** Mechanisms to make decisions at the level of data network. administer and monitor its operations.
- **Data services and infrastructure:** Required value adding components required of the data network.

Figure 4. Data Ecosystem Canvas.



3.2 Technical and Security Perspective

3.2.1 Introduction

Technical and security template collects the technical and security decisions during the infrastructure and system design for the data ecosystem. The checklist assists to identify and define common technical needs and their main characteristics.

The template is currently mostly a placeholder but provides topics that should be covered and discussed while designing the technical solution. The template acts as a master document for infrastructure specifications as well as for the work split between participants' systems and the common infrastructure. Further technical

design related can be added or linked to provide additional level of detail for the data network, if available and needed.

3.2.2 Overview of the technical solution

The prerequisite of the technical design is that we have at least an initial understanding of the common needs, parties and roles involved, as well as requirements of the common solution. Checklist tool is aimed in supporting in the process as well as in identifying more detailed technical requirements. When the parties have reached an initial and common understanding of needs, parties and common core functionality, document also a short introduction for the technical part.

3.2.3 Core security questions

A prerequisite for the operation of a data network is a common understanding and documented state of mind on the (security) risks, practices and solutions related to the data network. What security perspectives does data sharing bring to the data network, such as existing corporate information security standards and policies? Some of aspects to consider are:

- What is the common technical solution basis related to security and privacy?
- How is data security ensured for the shared data throughout the data network?
- What security and privacy features are needed in the common solution as well as at the participants and how will these be implemented and managed? What common activities are needed?
- What are the references and standards to be used for security and in the data network?

Not all security-related issues are necessarily agreed jointly, but each party to the data network is responsible for ensuring that security is implemented and that the conditions and requirements for the other parties and the network are met.

This solution is not only static, but the operating model must also take into account issues related to security monitoring and improvement. The following chapter discussed the security aspects more in detail.

4. Data security operating model for data networks

4.1 General

These instructions set out the principles for developing data security operating model for data networks, in accordance with which the Data Network takes care of secure processing of data. Data security needs may vary considerably in different networks, so it is important to modify the operating model on a case-by-case basis and update it with sufficient frequency. The more valuable the data shared or the more damage data security breaches may cause, the more the network should invest in high-quality data security.

In data sharing networks, trust between the parties involved is the key for success. In this respect, an adequate level of data security is an essential key element. The network rules must be drawn up in such a way that they support sufficient trust. The members of the network may to some extent share responsibilities between them, but not outsource them entirely.

In this context, it should also be noted that the skills and willingness to bear risks may differ between the parties. In this respect, too, it is important that the data security operating model is based on jointly agreed views on what levels of risk or residual risks after risk prevention measures are acceptable and how they are to be shared.

Data security must be taken into account from the outset, already when planning the operations, and it must be understood as a continuous activity throughout the life cycle of data sharing. Just as personal data protection is required to be included by design and by default (Article 25 of the GDPR), it may also be impossible or at least very expensive to add data security afterwards.

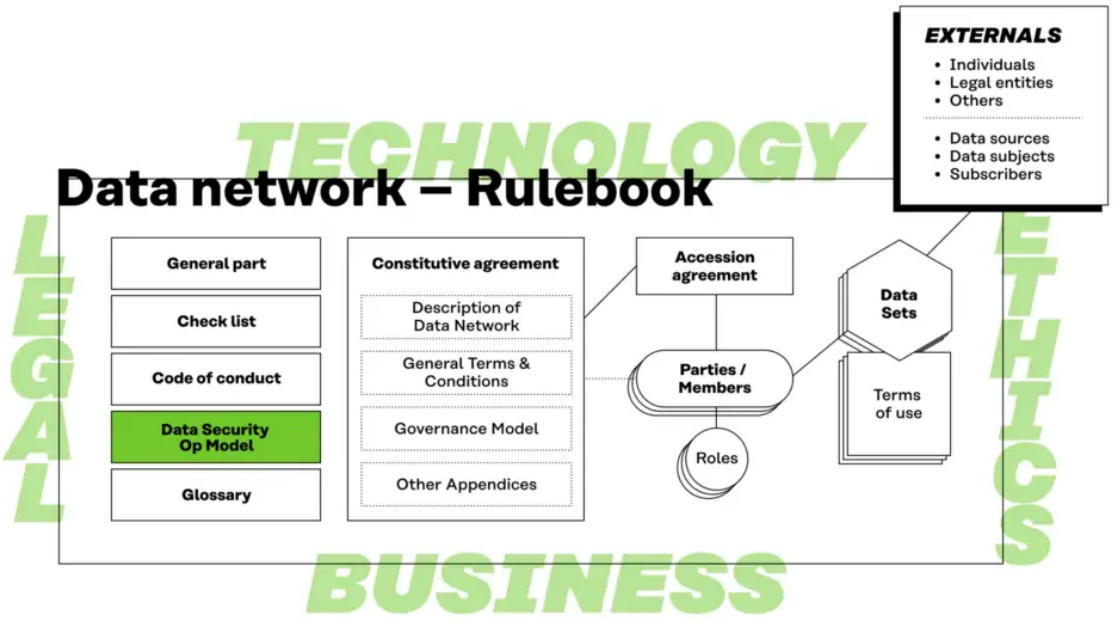
Unless specifically agreed otherwise, the data security operating model is not part of the aforementioned Constitutive Agreement of the Data Network. It is therefore important that the rights and obligations of significance to data security which the members of the network are expected to commit to are included separately in the agreements. Attention should be paid to this when reviewing the security questions of the check list below and when building a data security operating model for the Data Network.

The guidelines for the data security operating model for data networks were created on the initiative of the Digipool of the National Emergency Supply Organisation (Technology Industries of Finland) and with the support of the National Emergency Supply Agency as part of Sitra's Rulebook for a Fair Data Economy.

The guidelines were prepared by experts Olli Pitkänen, Sami Jokela and Marko Turpeinen from 1001 Lakes Oy and Antti Nyqvist, Pool Secretary of Digipool. The

workshop organised for developing the guidelines involved members of the Digipool, companies from many different sectors and representatives of universities and organisations.

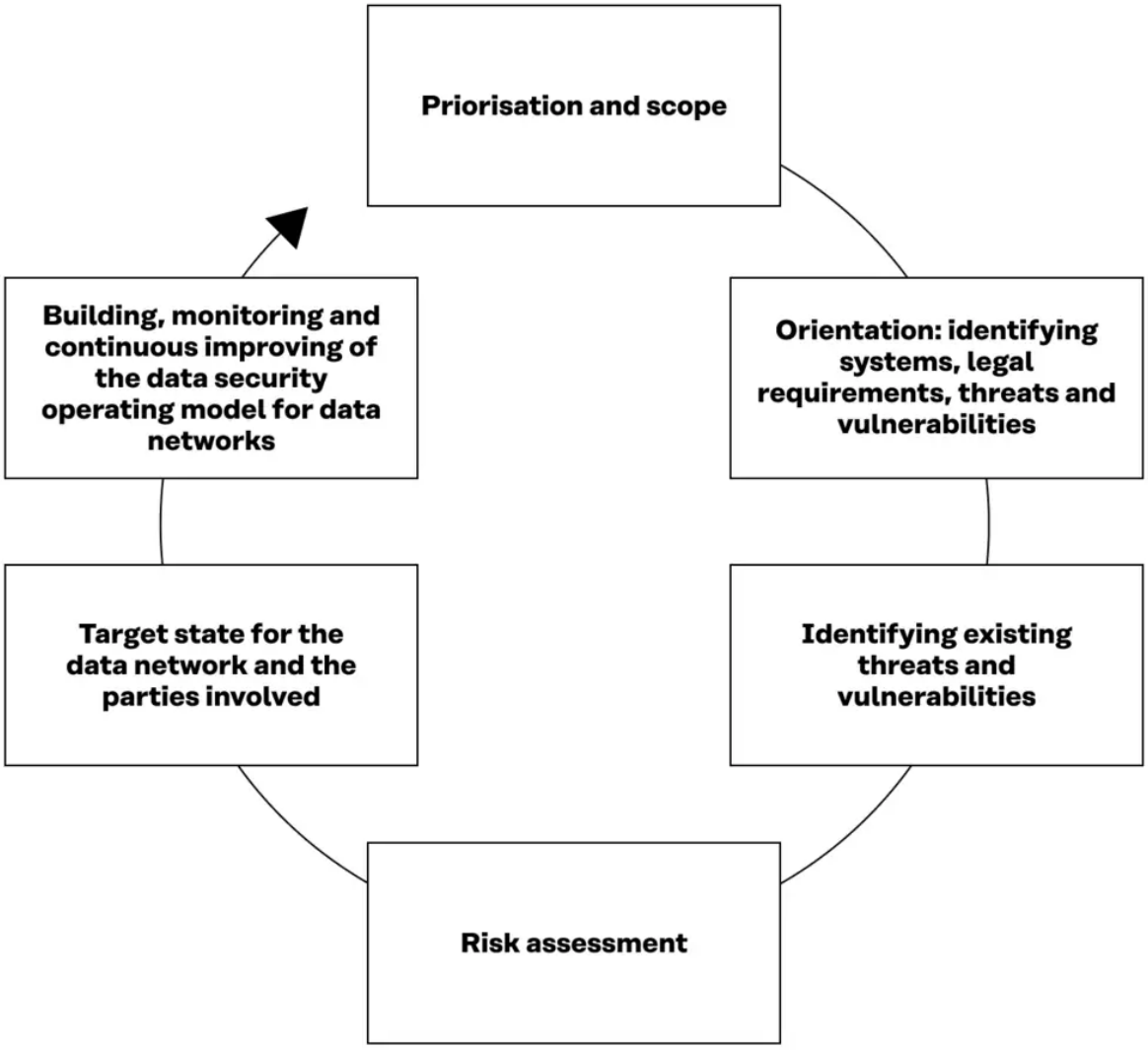
Figure 5. Data security operating model in the rulebook as a whole.



4.2 Data security process

The following diagram shows in a simplified format the development process of the data security operating model for data networks, or, in short, the data security process, described below.

Figure 6. Development process of the data security operating model for data networks, or, i.e., the data security process.



4.3 Prioritisation and scope

The Founding Members of the Data Network should appoint a working committee involving representatives of all Founding Members to prepare a data security operating model. The first task of the working committee is to define the operating environment. As far as the data network is concerned, this can be done simply by referring to the items in the Business Part of the rulebook of relevance for data security, if they have already been drawn up.

After this, you need to limit what is included in the scope of the data security covered by the rulebook. The essential point is to pay attention to data sharing and network-like activities. In this context, it may not be necessary to resolve other security issues, which may be important as such.

The working committee must also take a stand on what is the object of protection: the data covered by the rulebook in accordance with the Constitutive Agreement

within the contractual framework of the rulebook, or something else?

4.4 Taking account of data protection

If the data contains any personal data, the data protection legislation shall be applied. The key legislation is the EU's General Data Protection Regulation (GDPR), which applies to all processing of personal data. In addition, depending on the situation, national data protection provisions and provisions covering special sectors may also apply. For example, in Finland, depending on the type of data concerned, the Act on the Protection of Privacy in Working Life (759/2004) must be taken into account in matters relating to technical supervision of employees and the Information Society Code (917/2014) as regards messages transmitted and added value services.

For more information

For more information, see Korpisaari, Pitkänen, Warma-Lehtinen: Uusi tietosuojalainsäädäntö. (In Finnish; New Data Security Legislation) Alma Talent, Helsinki 2018.

In the GDPR, personal data refers to all information related to an identified or identifiable individual, i.e., **a data subject**. As the definition of personal data is extensive, it is generally safe to assume that the data may include personal data, even when this is not obvious. Only if the processor can be totally certain that the data does not contain any personal data, the data protection legislation can be ignored.

Article 24 of the GDPR is a general provision on the responsibility and liability of the controller. This article lays down what kind of measures the controller must implement to ensure and to be able to demonstrate that the processing of personal data is performed as required by the Regulation. The provision includes both an obligation to act carefully and an obligation to demonstrate what kind of measures have been implemented to ensure lawful processing.

Article 32 contains specific provisions on data security. In accordance with the article, the controller and the processor of personal data shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk.

Ensuring that data protection is properly managed often provides significant benefits and competitive advantage for the controller, while, in the reverse case, neglecting data protection may become very expensive. The implementation, recovery and damage related to data protection may have an impact on the amount of any administrative fines imposed in accordance with Article 83 of the GDPR.

In data networks, particular account must be taken of the fact that, in principle, personal data may only be collected for a specific, explicit and legitimate purpose and they must not be further processed in a manner incompatible with those purposes. This may be a problem if different parties have different needs for the processing of personal data and if those needs change over time. Therefore, attention should be paid to the purpose of use of personal data already when setting up a data network.

It is good if, as a rule, the rulebook is prepared in such a way that the intended purpose of use of the data specified in it is sufficiently accurate and explicit as the legitimate purpose of the processing of personal data and the purpose of the processing does not need to be specified separately as regards personal data. If there are many types of data to be shared, it may not be possible to define the specific purpose of use in detail for all data concerned. Instead, attention must be paid to the matter separately with regard to different types of personal data.

If a rational, commonly used set of criteria for the anonymisation of data can be found, it can be used to exempt such data from the obligations of data protection legislation, as anonymised data can no longer be linked to any individual person.

For more information

For more information, see, e.g., The Office of the Data Protection Ombudsman: Pseudonymised and anonymised data

It may also be possible to use the data network to fulfil the obligation to provide information laid down in the Act and to report data security breaches, as long as the relevant obligations have been included in the Constitutive Agreement of the rulebook. For example, it may be agreed that data subjects may submit requests for information to any party in the network or to the most visible party from the perspective of the data subjects, which forwards it to the correct controller, or that information is centrally provided by one party of the network on behalf of all parties. However, it should be noted that such an arrangement must not undermine the rights of data subjects and that they always have the right to deal with the correct controller.

4.5 Orientation: identifying system, legal requirements, threats and vulnerabilities

When all the items covered by data security in accordance with the rulebook have been prioritised and specified, the objectives of data security must be defined and the management methods by which they can be achieved must be reviewed. We will return to how the management methods are defined in further detail below.

The Founding Members must **define the objectives of the entire Data Network** with regard to data security, and individual parties must **define their own objectives** with regard to the Data Network. Depending on the type of data being shared and the sector in which the parties operate, it may be necessary to apply special legislation, the requirements of which must be taken into account. In total, such special provisions can be found in hundreds of sector-specific regulations. For example, in Finland, the Act on the Electronic Processing of Client Data in Healthcare and Social Welfare contains more detailed provisions on the processing of customers' personal data.

Once the objectives have been defined, it is possible to identify which data security management methods are suitable for the data network concerned. A more detailed definition of the management methods takes place at later stages of the process. Data security management methods can be divided into three categories: administrative, technical and physical means.

Administrative means are related to risk management, data security documentation, personal security and training. Technical means are implemented using either software or hardware. These may include centralised log management, firewalls, intrusion detection systems (IDS), intrusion prevention systems (IPS), and identity and access management systems. Physical means of data security management, on the other hand, refer to the protection of resources and personnel by means such as fences, locks, lighting and camera surveillance.

It is important to take into account all different means of data security management to ensure that the Data Network can reach a sufficient level of data security. Multiple layers of control should be created around the object being protected to achieve multi-layer protection.

It should be emphasised that, at this stage, the various data security management methods available are just being explored in general. It is advisable for the Data Network and the organisations involved to select the management methods best suited for their specific circumstances only after carrying out the security risk assessment process described below.

4.6 Overview of security threats

In the traditional local deployment model of applications, each organisation's sensitive data remains within the organisation and is subject to its physical, logical and personnel security and access control policies. However, the cloud model commonly used by data networks stores the data outside the organisation's boundaries. For this reason, additional security checks must be carried out to ensure data security and to prevent data security breaches caused by data security flaws or malicious employees.

Information on current data security threats can be found, for example, on the websites of the European Union Agency for Cybersecurity (ENISA) and national information security authorities.

Cloud computing may employ sets of hardware and software from a variety of computer networks around the world. It enables more cost-effective and quicker sharing of data. This has been identified by criminals who exploit viruses and other malware for purposes such as attempting to steal sensitive information, disrupt services, or damage the cloud computing networks of enterprises.

According to IBM Security, globally, the average total cost of each data breach amounts to USD 4.24 million. In Scandinavia, the average total cost per data breach is slightly lower, USD 2.67 million. It took an average of 287 days to identify and mitigate a data security breach. The longer it takes to identify the violation, the higher the cost. The cost of a data breach with a life cycle of less than 200 days were nearly one third lower than those of a data breach that lasted more than 200 days.

In the following, some important trends with an impact on data security:

1. The interdependency between societal processes and information systems increases

2 New interdependencies between organisations and the state emerge

3 Information security issues become more international

4 Needs to manage private or confidential information and public appearances in ICT environments increase

5 Protection of personal data becomes a considerable political issue

6 It becomes increasingly difficult to ensure the correctness of information

7 The correctness of information becomes increasingly important

8 Data gathering increases

9 Data combination from different sources increases

10 Traceability of persons and goods increases

11 Malicious action against information systems increases

12 Quality and security issues are increasingly taken into account in software development

13 Automation/autonomous systems are increasingly employed to effect security

14 Availability of information increases as the public information resources are opened

15 Commercial interests drive actors to restrict access to proprietary information resources

16 Governance of access to information resources in organisations becomes more difficult

In the following, information security threats are discussed especially from the perspective of how they affect data sharing in data networks.

4.6.1 Using data in a manner incompatible with the purpose

Data may also be used for purposes not agreed in the Constitutive Agreement or in the Dataset Terms of Use or not taken into account in the rulebook. This may bring up issues that the Data Provider has not taken into account but to which it has committed itself. Typically, this involves harm suffered (e.g., data reveals some matters that the party does not want to share) or some other party building business from this without sufficient compensation to the Data Provider. Another subset of this is building and using models to replace data.

4.6.2 Data leaks

Data ending up with a wrong party, deliberately or unintentionally. This may occur either through an IT error (close to traditional data security threats) or by a party transmitting the information, its subset or a data model built from the data to a third party. In data leaks, mistakes made by users and incompetence often play an essential role. These can be reduced by means of education and training, as described below. Data leaks may also take place in such a way that data has been used for a purpose such as teaching artificial intelligence, with some confidential information having remained in the AI model, which may thus leak to outsiders through the model.

4.6.3 Responsibility for data

Data integrity. The data does not correspond to what has been agreed, is incorrect, or has been changed. This essentially involves liability issues: who is responsible if the data does not correspond to what has been agreed or assumed, or if it has been changed. Responsibility and liability issues are discussed more extensively in the Contractual Framework of the rulebook, especially in sections 3 Role-specific responsibilities, 5 General responsibilities and 11 Liability, and potential exceptions and specifications concerning them in the Constitutive Agreement and the Data Set Terms of Use.

In addition to the above, data liability issues include unforeseen responsibilities, such as could new surprising responsibilities related to data, for example, regarding the processing of personal data, emerge through developing interpretations made by authorities, as the way data protection legislation is interpreted is just finding its form. By definition, it is difficult to take such eventualities into account, but to a certain extent it is possible to take a stand on who will bear responsibility for them in the agreements.

4.6.4 Accidental sharing of data

Not all data loss events are the work of sophisticated cybercriminals. In fact, a shocking number of data breaches are caused by a company's own employees who accidentally share, misplace or mishandle sensitive data. It is common for employees to share, grant access to, lose or misprocess valuable information, either accidentally or because they are unaware of security practices. It is also regrettably common for employees to deal carelessly with devices on which data is stored, or to forget them in public places or to store them without proper protection. For example, according to a report published by Shred-it in 2018, 40% of data security breaches are caused by such behaviours. This major problem can be addressed by training employees and other users, but also by other measures, such as data loss prevention (DLP) technology and improved access controls.

Example 1

The electronic notebook of a service company employee was stolen. The stolen notebook contained the names, last names, genders, addresses, and dates of birth of more than 100,000 customers. Afterwards, it was not possible to fully determine the content of all data that had been lost. Access to the hard disk of the device was not protected by any password. Some of the data could be restored from daily backups.

Example 2

The HR department of a public administration office sent an e-mail about future training courses to those registered as jobseekers. The email was accidentally attached with a document containing the personal data of all these jobseekers (name, email address, postal address, personal identity code). The number of persons affected was over 60 000.

People do make mistakes, and mitigating the risks associated with those errors is critical for protecting data privacy. Company data is one of the most valuable assets that any business controls, and it should be protected accordingly. To put it simply, data access should be a system that minimizes exposure and reduces the risk of accidental or malicious misuse.

Another security threat related to people is the limited opportunities of the organisation's cybersecurity team to monitor all risks. System administrators are often overworked when trying to protect sensitive information. This leaves companies exposed, and it should increase the impetus to implement automation

wherever and whenever possible, without forgetting that automation itself various different risks.

4.6.5 Phishing and social engineering

Social engineering attacks are a primary vector used by attackers to access sensitive data. They involve manipulating or tricking individuals into providing private information or access to privileged accounts.

Phishing is a common form of social engineering. It involves messages that appear to be from a trusted source, but in fact are sent by an attacker. When victims comply, for example by providing private information or clicking a malicious link, attackers can compromise their device or gain access to a corporate network.

Phishing emails are on the rise. At the same time, new technology and increased information accessibility are making these attacks more sophisticated, increasing the likelihood that hackers will successfully infiltrate your IT systems. Despite every business' best efforts, these malicious messages may make their way into employees' inboxes. Managing this traffic and equipping employees with tools, education and training to defend against these threats will be critical.

Email addresses and passwords are in particularly high demand by cybercriminals. This is the primary data stolen in data breaches. Since this information can be used to deploy other, more diverse attacks, every company needs to be aware of how their data could be used against them.

Example 3

A telecommunications company's customer service centre receives a call from a person who presents himself as a customer. He asks the company to change the email address to which the customer's billing information is sent. The contact centre employee confirms the customer's identity by requesting certain personal information as required by company procedures. The caller provides the requested tax number and postal address of the customer correctly (because he has access to this information). After confirmation, the operator makes the requested change and from then on, the billing information is sent to the new email address. The

procedure does not require a notification to the former email address. The next month, the legitimate customer contacts the company and inquires why he does not receive his billing data to his email address and denies having requested his email address to be changed. The company later notices that the information has been sent to a wrong user and cancels the change.

4.6.6 Insider threats

Insider threats are employees who inadvertently or intentionally threaten the security of an organisation's data. There are three types of insider threats:

- Non-malicious insider – these are users that can cause harm accidentally, via negligence or incompetence.
- Malicious insider – these are users who actively attempt to steal data or cause harm to the organization for personal gain.
- Compromised insider – these are users who are not aware that their accounts or credentials were compromised by an external attacker. The attacker can then perform malicious activity, pretending to be a legitimate user.

When companies consider their cybersecurity risks, malicious outsiders are typically on top of their mind. Indeed, cybercriminals play a prominent role in some data heists, but company employees promulgate many others.

Verizon's 2019 Insider Threat Report found that 57% of database breaches include insider threats and the majority, 61%, of those employees are not in leadership positions when they compromise customer data.

One form of insider threat is bribery. Company data and intellectual property are both incredibly valuable and, in some cases, employees can be bribed into revealing this information. For example, in 2018, Amazon accused several employees of participating in a bribery scheme that compromised customer data, and in 2019, it was discovered that AT&T employees received bribes to plant malware on the company network. Of course, bribery isn't the most accessible way to perpetuate a data scheme, but especially for companies whose value resides in their intellectual property, it can be a serious data security concern.

Example 4

During the period of notice, a company employee copies business information from the company database. The employee is entitled to access this data only to perform his duties. Months later, after having resigned from the job, he uses the information thus obtained (basic contact information) to enter new processing data. As data controller, he intends to contact the company's customers to attract them to his new business.

4.6.7 Ransomeware attacts

Ransomware is malware that infects corporate devices and encrypts data, making it useless without the decryption key. Attackers display a ransom message asking for payment to release the key, but in many cases, even paying the ransom is ineffective and the data is lost.

The cost of ransomware attacks more than doubled in 2019, and this trend is likely to continue well into the future. Many ransomware attacks begin at the employee level as phishing scams and other malicious communications invite these devastating attacks.

Many types of ransomwares can spread rapidly and infect large parts of a corporate network. If an organisation does not maintain regular backups, or if the ransomware manages to infect the backup servers, there may be no way to recover.

Example 5

A ransomware attack is targeted against a small industrial company's computer systems, and the data stored in them is encrypted. The company itself had properly encrypted the data in advance, so all the information accessed by the ransomware

program is already encrypted, and the decryption key is not compromised in the attack. Therefore, the attacker only has access to encrypted data. The company uses the expertise of an external cyber security company to investigate the case. All logs of outgoing emails and other data streams are available. After analysing the logs and the data collected by the company's observation systems, an internal investigation supported by the cybersecurity company establishes with certainty that the attacker only encrypted the data without having access to its content. The personal data affected by the breach relates to the company's customers and employees, a total of a few dozens of persons. The backup copy is easily accessible, and the data is restored a few hours after the attack. The data security breach has no consequences for the day-to-day operations of the company. There is no delay in employee payments or processing of customer requests.

Example 6

The server of a public transport company is exposed to a ransomware attack and the attacker encrypts the data on the server. An internal investigation reveals that the attacker not only encrypted the information but also retrieved some of it for himself. The compromised data includes the personal data of customers and employees, and thousands of people using the company's services (e.g., buying tickets online). In addition to basic personal data, the data breach involves ID card numbers and financial information, such as credit card information. There is a backup database, but the attacker encrypts that as well.

4.6.8 Data loss in the cloud

Many organizations are moving data to the cloud to facilitate easier sharing and collaboration. However, when data moves to the cloud, it is more difficult to control and prevent data loss. Users access data from personal devices and over unsecured networks. It is all too easy to share a file with unauthorized parties, either accidentally or maliciously.

4.6.9 Bad password hygiene

The study by Thomas et al concluded that 1.5% of all login information on the internet is vulnerable to credential stuffing attacks that use stolen information to inflict further attacks on a company's IT network. Many login credentials are compromised in previous data breaches, and with many people using redundant or easy-to-guess passwords, that information can be used to access company data even when the networks are secure.

Therefore, best practices like requiring routinely updated passwords are a simple but consequential way to address this preventable threat (see the instructions by Traficom).

4.7 Identifying existing threats and vulnerabilities

Once the data security objectives have been defined and the management methods identified, it is advisable to **assess and document the current situation**. In this respect, the situation is significantly different if data is already shared between the parties in some way compared to the situation in which data sharing is only about to begin.

How has data security been currently ensured to the extent that data is already being shared between the parties? What will change from a data security perspective when data is shared in accordance with the Technical Part and Business Part of the rulebook above?

In this matter, it is also worth taking into account the legal requirements and the extent to which international regulations, such as those at the EU level, and national provisions in different countries must be taken into account. As a rule, the regulations on data sharing in the EU are highly harmonised. For example, the General Data Protection Regulation (GDPR), as well as all other EU regulations enforced in the form of a regulation, are in force as such in all EU Member States and the EEA.

On the other hand, there may be significant differences in how EU directives have been implemented in the legislation of different Member States, for example. If data is also shared outside the EU and the EEA, you must be prepared for the

possibility that legislation can be completely different or, in particular, that the transfer of personal data may not even be permitted without special arrangements.

The key question is whether the data is shared directly between the parties or whether the data is stored somewhere in a central repository from where each party can retrieve it. This also has an essential impact on assigning the data security management requirements to the right parties. One option can be using an escrow type operator, a reliable third party, which will manage certain obligations against a fee.

4.8 Risk assessment

As described above, the Data Network and its participants must identify **security threats and vulnerabilities**. Next, they **assess the severity** of the risks by determining the likelihood of each threat and the extent of the damage in the event that the threat materialises.

In data networks, assessing risks is primarily the responsibility of the Founding Members, but in practice, they should entrust it to, for example, the information security working committee described above. Here again, particular attention should be paid to the perspective of the entire data network.

The severity of risks may appear quite different when viewed from the point of view of the entire network than when viewed from the perspective of an individual party. Still, every party involved is responsible for the risk assessment from its own perspective.

The severity of the risk is the product of likelihood and the extent of damage. Since it is all but easy to assign a numerical value for the likelihood or the extent of damage, in practice, it is often best to use some kind of classification, such as a three-step approach, when assessing them, and then place the risks in a matrix where one axis represents the likelihood and the other the extent of damage. This makes it quite easy to see which risks are the most serious and require most attention.

Risk management refers to processes and practices that can be used to identify, assess and mitigate risks and to assign responsibilities for these measures. It makes sense for the Data Network to define common operating methods for risk management insofar as they concern the operation of the network. The data security operating model should include adequate definitions of all areas of risk management that support the risk management of the entire network.

4.9 Target state for the Data Network and the parties involved

Once the risks have been assessed, the Founding Members can agree on the **target state of information security** for the entire Data Network and its individual members. Regarding the extent to which the parties want to make the target state legally binding, it is necessary to include these in the binding terms and conditions of the Constitutive Agreement, the Accession Agreement or the Dataset Terms of Use.

The network must have a common intent regarding the risk level and the extent of the measures. This may require a considerable number of negotiations and coordination of objectives, especially if the levels of knowledge, expertise and security requirements vary among the network participants. Depending on the nature of the Data Network, the target state is documented by describing, for example, what the network aims to achieve in the following areas:

1. Data security policies
2. Personnel safety and security
3. Management of the data to be protected
4. Access control
5. Encryption
6. Physical and environmental safety
7. Operational safety
8. Communication security
9. Acquisition, development and maintenance of systems
10. Relations with suppliers
11. Security incident management
12. Data security aspects related to business continuity management
13. Compliance

4.10 Builging, monitoring and continuous improving of the data security operating model for data networks

Once the target state has been defined, **a data security operating model** to support the achievement, maintenance and continuous improving of the target state will be built. Depending on the data network, it includes topics such as:

1 Organisation of data security

Depending on the size, operation and purpose of the data network, it may be appropriate to establish a steering committee subgroup in accordance with the Governance Model to ensure information security and the achievement, maintenance and development of its target state, or to assign these tasks to the steering committee or to assign responsibility for them to certain appointed persons, for example. A definition of this binding on the parties should usually be included in the Governance Model or the Constitutive and/or Accession Agreement.

As part of organising data security, responsibilities for who reacts if any risks materialise and in what manner should be defined. This includes not only immediate measures to limit and rectify damages, but also communication and learning from what has happened.

The GDPR usually obliges parties involved to report personal data security breaches to the Data Protection Ombudsman and, in certain cases, the persons affected as well. As a rule, a notification to the Data Protection Ombudsman must be submitted no later than 72 hours after the data security breach has been discovered. This is such a short time that it is a good idea to plan in advance how to react to a data security breach and report it, so that it is not necessary to start considering it only after the breach has occurred.

2 Management's commitment to data security. Who/what body has leadership over the Data Network?

Ensuring the management's commitment is essential for the data security operating model. Therefore, the Founding Members must identify the person or party who is in a leadership position regarding data security in the Data Network and ensure their commitment to achieving the target state.

3 How is data security measured?

The Data Network should agree on indicators for measuring data security. What indicators are used is essentially dependent on the type of data network setup. The indicators should be defined in such a way that they are related to the objectives set above, that the parties can influence them through their own measures and that they are particularly significant in terms of the risks assessed as most serious above.

4 What kind of audits and reviews are carried out in the Data Network?

In order to build trust, the data security process must be sufficiently transparent. In particular, it is worth investing in transparent risk management of the network.

To ensure data security, it can also be agreed that the parties have the right to audit and review each other's systems and facilities. If such audits and reviews are deemed necessary, a term should be included in the Constitutiv Agreement, the Accession Agreement, the Dataset Terms of Use or the Governance Model that obligates parties who may be subject to such measures to allow access to their systems and facilities. The term must be so unambiguous and clear that there is no doubt about who it concerns, what the audit or review may focus on, who can perform it, who is responsible for the costs arising from it and what kind of confidentiality obligations the party performing it has.

5 Documentation of the Data Network's data security model

The data security operating model for the Data Network will be incorporated into the Data Network's rulebook. As a rule, it is a similar document to the Code of Conduct, which describes the mutual understanding between the parties, but is not a contractually binding part of the Constitutive Agreement. Therefore, all obligations that should be contractually binding on the parties must be recorded separately in the relevant sections of the Constitutive Agreement or its annexes. It is also possible to include the entire data security operating model as a binding annex to the Constitutive Agreement. However, this may make the operating model too rigid when it should be possible to reform and maintain it in a flexible manner.

As regards documentation, it is also worth noting that Article 30 of the GDPR requires that a record of processing activities of personal data be maintained. This must also be made available to the supervisory authority on request, allowing them to assess the lawfulness of the processing activities. Furthermore, in accordance with Article 5, the controller has a general obligation to be able to demonstrate compliance with the Regulation (*accountability*) at all times.

6 Continuous improving: how to ensure data security continuity and process development?

As the Data Network develops, operations change and new parties bring new needs to the network, it is important to ensure that data security remains at the desired level and that the data security process of the Data Network is continuously improved. It is advisable to enter separately in the Governance Model how the data security operating model is reviewed and updated whenever necessary and at agreed intervals.

7 Changes in the purpose of use of data and their management

The purpose of use of data is defined in the Constitutive Agreement or the Data Set Terms of Use. This is essential for the implementation of 1) the purpose limitation of data protection and 2) data access rights.

The purpose limitation of data protection in accordance with the General Data Protection Regulation (**GDPR**) restricts the purposes for which the controller can use information collected on people. There are two important aspects in purpose limitation: personal data shall be collected for specified, explicit and legitimate purposes (purpose specification) and not further processed in a manner that is incompatible with those purposes (compatible use). The purpose limitation does not prohibit the data collected for one purpose from being processed for another purpose, as long as it is not incompatible with the original purpose. This principle is aimed at, on the one hand, protecting the legitimate expectations of data subjects with regard to the processing of personal data concerning them and, on the other hand, enabling further use of personal data within certain limits.

However, when it comes to the rights of use, the essential question is for what purposes the holder of the rights, such as the holder of copyrights, producer's rights or the right to data based on the protection of trade secrets, has allowed the use of the data. Normally, they do not have the rights of use for any other purposes.

In other words, the basic rule is that data may not be used for purposes other than those specified. However, the circumstances change, new needs may arise for the parties to the network regarding data, and new parties joining the network may bring with them needs that could not be anticipated. Therefore, there may be good grounds for changing the purpose of use. When such a need arises, it must be determined how it can be done in accordance with data protection legislation, such as by asking people for their consent to process their personal data for a new purpose. On the other hand, it is usually necessary to negotiate about the terms of use that can be reformulated with the holder of the rights if the holder has the right of veto on the new purpose and it does not fall within the scope of the old definition.

4.11 Sources used for the data security operating model

EDPB Guidelines 01/2021 on Examples regarding Data Breach Notification

Forbes: 10 Data Security Risks That Could Impact Your Company In 2020.

IBM Security: Cost of a Data Breach Report 2021

Imperva: Data Security.

Päivi Korpisaari, Olli Pitkänen, Eija Warma-Lehtinen: Uusi tietosuojalainsäädäntö. (in Finnish; New Data Security Legislation) Alma Talent, Helsinki 2018

Olli Pitkänen, Risto Sarvas, Asko Lehmuskallio, Miska Simanainen, Vesa Kantola, Mika Rautila, Arto Juhola, Heikki Pentikäinen, Ossi Kuittinen. Future Information Security Trends. Kasi Research Project, Tekes Safety and Security Research Program, Final Report, March 11, 2011

Olli Pitkänen: Tietosuojasäädösten muutostarve (in Finnish; Need to change national data protection statutes), Prime Minister's Office, 41/2017

Olli Pitkänen, Päivi Korpisaari, Rauno Korhonen: Miten kansallista lainsäädäntöämme pitää muuttaa EU:n yleisen tietosuoja-asetuksen vuoksi? (In Finnish; How should Finland's national legislation be amended due to the EU's General Data Protection Regulation?) In Korpisaari (ed.) Yearbook of Communication Law 2016, Forum Iuris, 2017

Subashini, Subashini, and Veeraruna Kavitha. "A survey on security issues in service delivery models of cloud computing." Journal of network and computer applications 34.1 (2011): 1–11

M. Swathy Akshaya and G. Padmavathi. "Taxonomy of Security Attacks and Risk Assessment of Cloud Computing" in J. D. Peter et al. (eds.), Advances in Big Data and Cloud Computing, Advances in Intelligent Systems and Computing, 2019.

Kurt Thomas, Jennifer Pullman, Kevin Yeo, Ananth Raghunathan, Patrick Gage Kelley, Luca Invernizzi, Borbala Benko, Tadek Pietraszek, Sarvar Patel, Dan Boneh and Elie Bursztein. Protecting accounts from credential stuffing with password breach alerting. Proceedings of the USENIX Security Symposium 2019

5. Ethical principles: Shared values of the Data Network

Ethical model assist in surveying and improving the ethicality of the data network, and activities and organisations that are part of it. The incorporated Code of Conduct is not a standardized ethical code for all cases as the organisations and networks – that want to use this – most likely are very different depending on context and various other reasons.

Therefore, as different data networks using this rulebook may differ, case by case, this code cannot be seen as sufficient but necessary list of issues needed to be taken care of. This means that more detailed and specific codes or other ethical guides should be considered and reviewed – based on demands of specific data network or organisation that implements this code of conduct.

The demands for ethical codes are different based on what kind of information is used and also the target of use may raise different ethical concerns. As an example, whilst using health information the medical codes of ethics need to be considered. As against, when data that are processed within a data network does not include any personal data, the focus should be in respecting e.g., intellectual property rights.

Ethical code should not be seen as a way to restrict the actors of a data network but instead as a set of commonly acceptable norms that make cooperation between members more convenient by setting the direction for more detailed rules defined by implementing organisations. The code is not an obstacle. Like laws, it helps to create trust in a data network, which is needed for gaining real benefits and new business opportunities. This code of conduct is based on the respect between different stakeholders, transparent communication and ambition to seek the values that are commonly acceptable.

The code itself does not have intrinsic but instrumental value. Most important is that organisation has a real goal to improve its processes and policies to be more ethical. The mere mechanical following of codes is better than not following those. However, the aim should be changing the culture of the organisation to such that it put ethicality in everyday actions, so the change comes from inside – not from outside. In that case, change is durable and will help the organisation to meet demands that society justifiably sets for them.

Acting ethically is not a mere cost but possibility for a resilient business.

Figure 7. Code of conduct in the rulebook model.

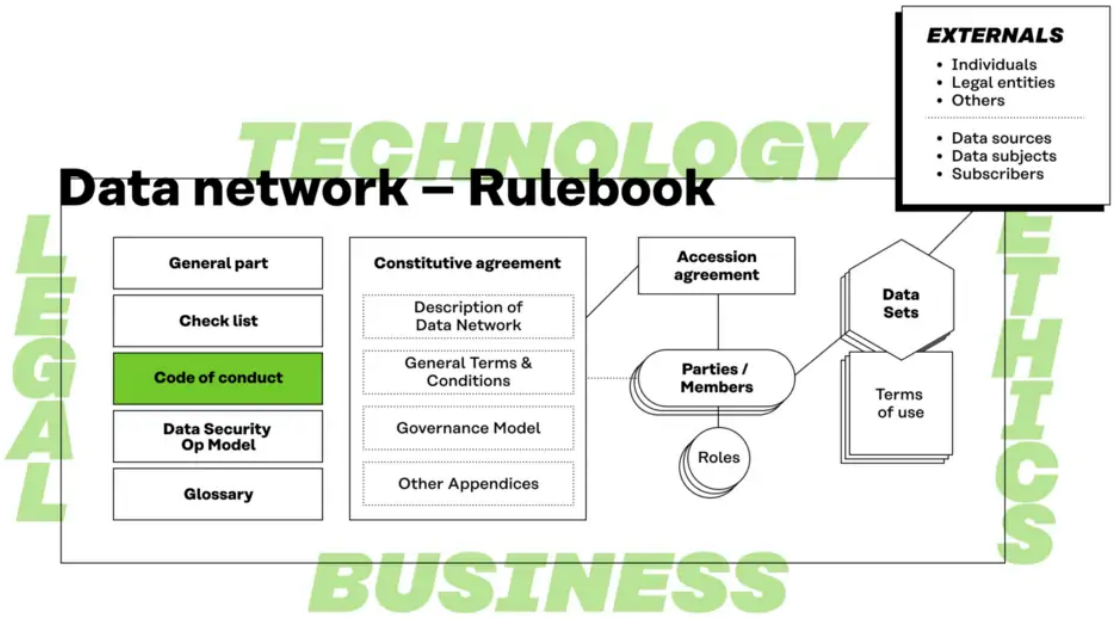


Image: Topias Dean / Sitra

There exists a consensus amongst normative theorists of cultural pluralism that dialogue is the key for securing just relation between different groups. Discourse ethics is an applicable tool to bring different views under constructive communication and thus to facilitate a more transparent and rational discourse. Discourse ethics provides a mechanism for considering different moral views and intuitions of different stakeholders.

However, this code of conduct does not focus on ethical theories. Instead, the purpose of this code is to approach the topic from the practitioners' point of view and to provide conceptual and analytical tools for assessing reasons on the basis of which the question "what should we do?" can be answered. This is done by presenting the values seen as important for data economy as well as by offering the maturity model (see next section) which can be employed in the analysis of the state of affairs in an organisation and consequently in the search of ways of improvement and development.

The following values have been found to be important in the research conducted during the IHAN project. In order the aim of fair data use to be achieved these values should be noted and respected in everyday practises.

Sources and further material on ethics:

James, M. R. (2003). Communicative Action, Strategic Action, and Inter-Group Dialogue. European Journal of Political Theory, 2(2), 157–182.

Stahl, B. C. (2012). Morality, ethics, and reflection: A categorization of normative IS research. Journal of the Association for Information Systems,13(8):636–65.

ACM code of ethics (PDF) is ethical code that gives insights for computing professionals and managers to ethical issues that should be taken account in practice.

Ethics Guidelines for Trustworthy AI by High-Level Expert Group on AI set up by the European Commission.

Data ethics canvas (PDF) provided by ODI (Open Data Institute) that focuses on helping identify and manage ethical issues of using data.

The ethics of Big Data (PDF): Balancing economic benefits and ethical questions of Big Data in the EU policy context

5.1 Accountability and Auditability

The members of the data network are responsible for what they do, and they must be able to give satisfactory reasons for it. This means that all actors are expected to follow the Rule Book of the data network and especially its contract. All the contracts also should follow the Code of Conduct and the Rulebook of this data network. The responsibility is towards members of the data network, but also the external stakeholders – e.g., individuals, whose personal data may be processed in the data network.

The operations within the data network must also be auditable, i.e., an auditor needs to be able to achieve a comprehensive examination of the data processing within the data network. Therefore, the members' records, logs and documents on data processing are well-organized and complete, their personnel are transparent in their dealings with the auditor, and the members have a good system of internal control, security and documentation in relation to data processing.

5.2 Avoid harm

All actors in the data network should avoid causing harm but instead focus on creating value (direct or indirect) for the whole data network and all the people that are affected by the actions of this data network.

5.3 Justified Processing of Personal Data

Personal data shall be processed on a fair and lawful basis, like for example on the basis of an informed consent of the individual, in accordance with a contract with the individual, a legal obligation, a vital interest of the individual, in the public

interest, or for the purposes of the legitimate interests, given that the interests and fundamental rights and freedoms of the individual are not threaten, in particular where the individual is a child.

5.4 Fairness, justice and equality

All actors in the data network should promote fairness, justice, and equality among individuals. Fairness means that everyone is treated with respect regardless of their socio-economical background or status. Likewise, the benefits (economical and others) must be balanced between all stakeholders in such a manner that individuals that are the source of data are not seen as mere exploitable resources.

To ensure fair use of their information, individuals are granted true possibilities to understand and control their personal data that are collected, transferred and otherwise processed in the data network.
The rules and the structure of the data network secure the benefits and rightful expectations of all the parties. This requires a balanced power structure in the data network and transparent consensus-oriented governance.

5.5 Human-centricity

People live in different environments, and they have personal lived experiences of their own life. They must be respected and empowered. This means that individuals have to be seen and treated as active actors with opportunities to make their own choices in the data network. They must be able to keep full and effective self-determination. Furthermore, their needs and wishes should be taken account instead of reducing them as objects or subjects.

5.6 Privacy

Privacy is one of the central issues in data economy. Therefore, privacy must be respected and protected. The data network is based on the use of information, which sets high demands for privacy as information can be sensitive and private. Thus, this means that personal data shall be processed lawfully, fairly and in a transparent manner in relation to the individuals.

Personal data shall be collected for specified, explicit and legitimate purposes and it shall not be processed further in a manner that is incompatible with those purposes. Only personal data, which are adequate, relevant and limited to what is necessary in relation to the expressed purposes, shall be processed.

Organizations do not collect personal information that they do not need. All the personal data that are processed have to be necessary for the specific use. The

members of the data network take reasonable measures to ensure that personal data are accurate and up to date. Personal data must not be stored longer than necessary for the purposes for which the personal data is processed.

To ensure the integrity and confidentiality of privacy, personal data must be processed in a manner that ensures appropriate security of the data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.

To preserve the integrity, confidentiality and availability of the data, organizations need to develop and implement an information security policy framework. When merging data sets, privacy should be guarded even more carefully than normally. Anonymization of data is recommendable, whenever feasible.

Any organization must also be accountable, i.e., it needs to be able to demonstrate its compliance with the principles mentioned above. Therefore, the processing of personal data must be planned and documented. There should exist clear, documented processes for data collection, storage, use, and distribution.

For collected data, there needs to be a clearly documented life cycle plan where the collection, archiving and possible erasing of data are described. The relevant parts of the life cycle plan are available to data providers and individuals related to the data.

5.7 Security

All the members of the data network are responsible that their collection, use, storage, sharing, and other processing of data are secure. This means that proper security solutions and processes are used and also that monitoring, patching, and reporting of security issues are properly designed.

Personal data on individuals must be properly secured and the risks to the rights and freedoms of individuals should be analysed. All the necessary technical, organisational and personal actions must be implemented to minimise security threats to individuals whose information is processed. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of individuals, all the members of the data network shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk.

Likewise, data breaches must be responded without delays. When the personal data breach is likely to result in a high risk to the rights and freedoms of natural

persons, the responsible member of the data network shall communicate the personal data breach to the data subject without undue delay.

5.8 Sustainability and Circular Economy

All the members of the data network are guided and incentivised to develop and deploy sustainable solutions in alignment with a more sustainable, circular economy. The members will implement the data network in a manner to make its operations more sustainable and circular, thus reducing its negative externalities on the environment, climate, and natural resources.

5.9 Transparency

The data network is based on co-operation and respect for information sources. Transparency is important to develop trust. The data shall be processed lawfully, fairly and in a transparent manner. Any information addressed to the public or to individuals must be concise, easily accessible, and easy to understand, and clear and plain language and, additionally, where appropriate, visualisation is used.

This does not mean that information is open to everybody without restriction. Instead, it means that all the members in the data network should know (when/if possible), what data are offered in the data network and by what requirements to promote transparency of network. To support real-time economy, the members of the data network do not unnecessarily detain data but share them as soon as possible.

The use of unnecessary legal jargon should be avoided. If an individual is asked to give a consent or to accept an agreement, it must be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, and using clear and plain language. Additionally, honest information should be provided to individuals for understanding what data regarding them is being collected and how it is being processed.

5.10 Continuous improvement

Ethical issues vary and different issues may come up case by case. Thus, ethical evaluation should be a continuous process in organisation and there should be institutional support for this. Therefore, the management of a network member should support the organization's employees by ensuring that they have real opportunities to uphold, promote, and respect the principles of the Code of Conduct.

Ethics is implemented in daily actions of individuals or it is not implemented at all, as only individual can make the moral decision. However, without institutional support for ethical decisions there is a higher risk of unethicality as individuals lack the needed autonomy for being able to make moral decisions.

5.11 Support for individuals

All the members of the data network should support individuals in (a) getting information about use of their personal information, (b) understanding information, practices, contracts, and their consequences, and (c) participating, contributing, and influencing in systems and practices when using personal information of those individuals. The aim is to ensure that if individuals need information or have justified demands for Information, they are given needed support (Guidelines, personal help etc.) in transparent ways. The focus should be in creating low-barrier way to look overuse of personal information for those individuals from whom it is directly collected, or other way received.

5.12 Communication

Appropriate communication is fundamental throughout the data network's life circle. It is essential for individuals, organizations, and the society as a whole. Each of them needs in addition to different contents and timing, also apposite communication channels and manners. The above-mentioned ethical principles are put into practice with communication. Furthermore, communication is the way to demonstrate the organization´s commitment to them. The management has a special responsibility to articulate, apply, and support the organization's culture and processes that reflect the principles of this Code of Conduct.

Publication details

Title
Rulebook for a fair data economy

Subtitle
Rulebook template for data networks

Place of publication
Helsinki, Finland

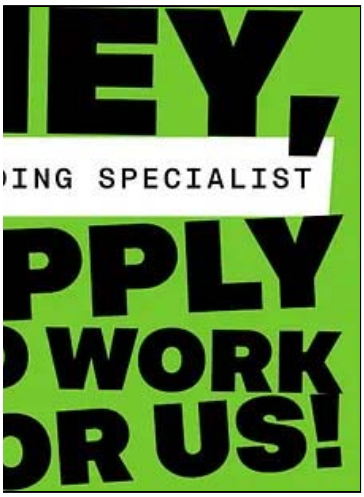
Year of publication
2019

Publisher
Sitra

SHARE

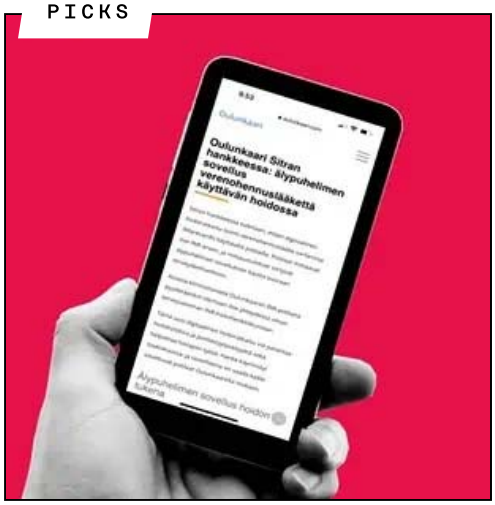
Recommended

Have some more.



system Specialist, Fair nomy

PICKS



Smartphone home measurements of medication effects help build the data economy

Hous curb ener solut

What's this about?

TOPIC

IHAN® project

With our IHAN® project, we laid the foundation for a fair data economy, in which successful digital services are based on trust and create value for everyone. The project ended in June 2021.

READ MORE

THEME

Fair Data Economy

We are building a human-driven, fair dat European value base. The use of data crea competitiveness and helps to develop soc interests of individuals, businesses and so

READ MO

CHANNELS

Twitter

Facebook

Linkedin

Youtube

Instagram

Slideshare

CONTACT US

The Finnish Innovation Fund Sitra
Itämerenkatu 11-13, PO Box 160,
00181 Helsinki
Telephone +358 294 618 991
Telefax +358 9 645 072
Email firstname.lastname@sitra.fi sitra@sitra.fi

Business ID 0202132-3