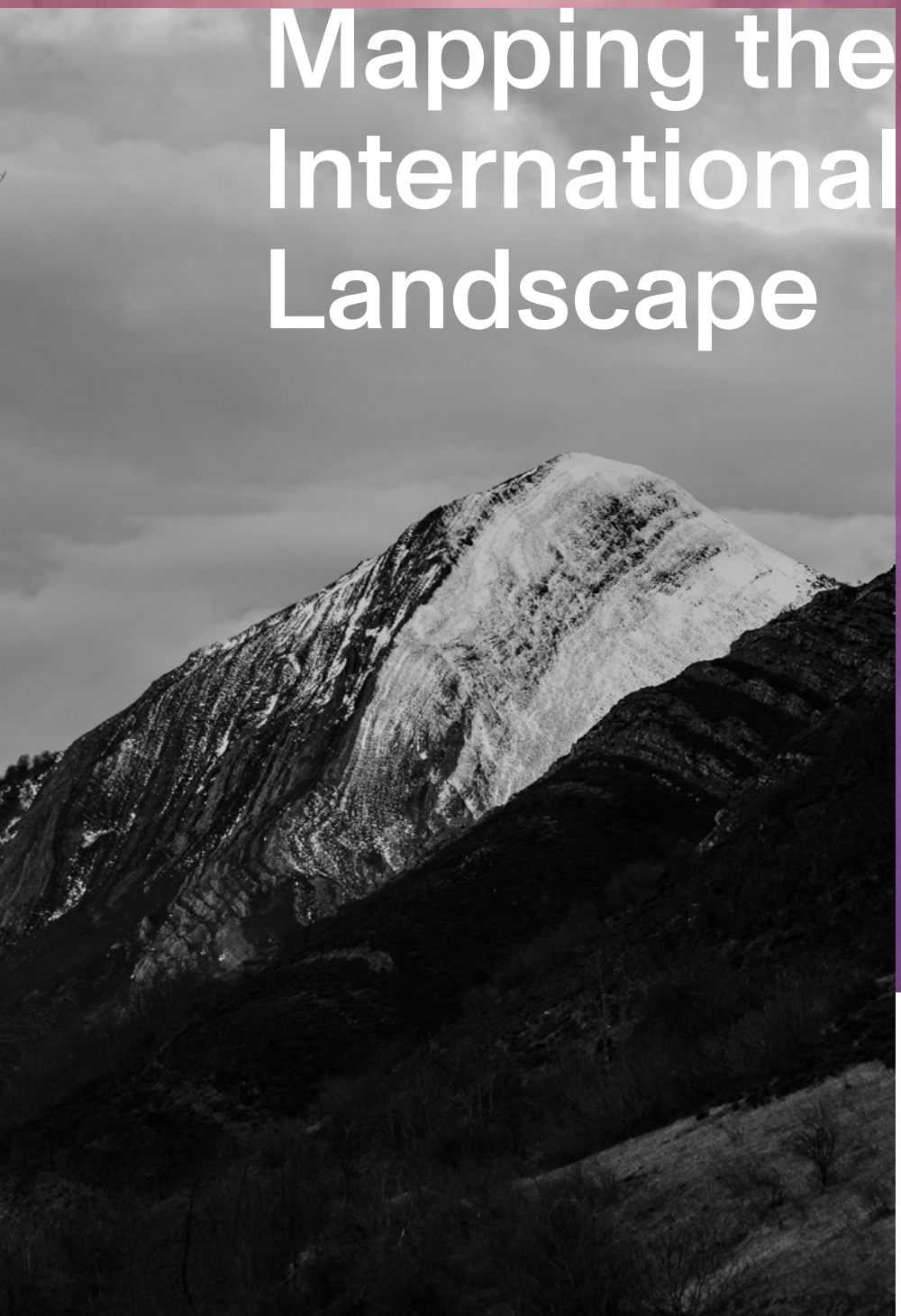
Labels and Certifications for the Digital World

Mapping the International Landscape



SWISS DIGITAL INITIATIVE

EXECUTIVE SUMMARY

The Swiss Digital Initiative (SDI) in Geneva aims to advance practice-oriented ethics in the digital age through specific projects. Launched in September 2019 by digitalswitzerland and under the patronage of Federal Councillor Ueli Maurer, it brings together academia, government, civil society and business to find solutions to strengthen trust in digital technologies and in the actors involved in ongoing digital transformation. A first landmark project of the SDI is the development of a Digital Trust Label. This label is intended to represent a mark of confidence that a service provider takes its promise of meeting consumer expectations seriously.

This SDI's label project is one among many other initiatives that have recently been pursued worldwide with the ambition to deliver seals, labels, and similar certifications. The present report aims at providing an overview of these projects in order to draw a picture of the diversity of existing types of initiatives and actors, their aim, their focus, and their way of functioning. It discusses success factors and critical aspects for developing a Label in the digital realm. This should allow lessons learned and emphasizes what role Switzerland, as hub for digital governance, could play.

With this aim, the present report opens with a thematic overview outlining issues of digital ethics in general, before delving further into details and nuances of the various label initiatives. At its core, it establishes a mapping of the most important initiatives and examines a relevant selection of them in detail, with a closer look on success factors as well as on similarities and differences compared to the Digital Trust Label. 12 projects from the map that are especially relevant for Switzerland have been analyzed in detail:

1 **Data Ethics Framework** (AI Ethics Impact Group, Bertelsmann Stiftung)
2 **The Digital Standard** (Collective effort: Consumer Reports; Disconnect; Ranking Digital Rights, The Cyber Independent Testing Lab)
3 **The Ethics Certification Program for Autonomous and Intelligent Systems** (ECPAIS), (Institute of Electrical and Electronics Engineers IEEE)
4 **Fair Data Economy Score** (part of Human Driven Data Economy IHAN, Sitra)
5 **Trustmark for the Internet** (EU Next Generation Internet Initiative)
6 **Trustable Technology Mark** (Mozilla OpenIoT Studio + ThingsCon)
7 **A Trustworthy Tech Mark** (doteveryone)
8 **D-seal, Seal for Data Ethics and IT Security** (Public Private Partnership)
9 **AI Certification** (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS)
10 **Independent Audit of AI Systems** (For Humanity)
11 **Label Numérique Responsable** (NR) (Institut Numérique Responsable)
12 **Apple's App Privacy Label**

Based on this analysis, the report concludes with recommendations addressed to the Swiss Federal Government on how to identify promising approaches. An integral part of this report is an Annex that provides, in the form of a living document, an easily accessible overview of the projects identified.

The report highlights a number of criteria a label should fulfill in order to have a chance of being successful. This includes that

- the label has to be known by its target users,
- it should be supported by a strong and well known organization,
- it has to convey a general message, with details and complexity being handled in the background,
- the governance of the labeling body has to be legitimate,
- the way the label organization is funded needs to be transparent and understandable for outsiders.

The present report identifies a highly dynamic and rapidly evolving ecosystem of labeling projects, in which, so far, no single initiative has gained sufficient traction to establish itself as a regional or international standard. In this environment, Switzerland can play a crucial role for supporting the development of successful labeling initiatives by offering its networks, resources and expertise to reinforce formal and informal coordination among the different initiatives. Also, the country could contribute to improving inclusivity and diversity within these initiatives, and it could support the general vision and integrating promising initiatives into existing multilateral efforts to tackle ethical issues and digital technologies.

1
INTRODRUCTION

Ethics for the digital realm is currently a burning issue both in national and international contexts. Numerous actors have been developing declarations, charters, codes of conduct or checklists to define which ethical standards should be respected when it comes to digital technologies. At the same time, projects have emerged that want to go one step further in trying to operationalise these standards. Their main focus is not the identification of principles, but their application in practice.

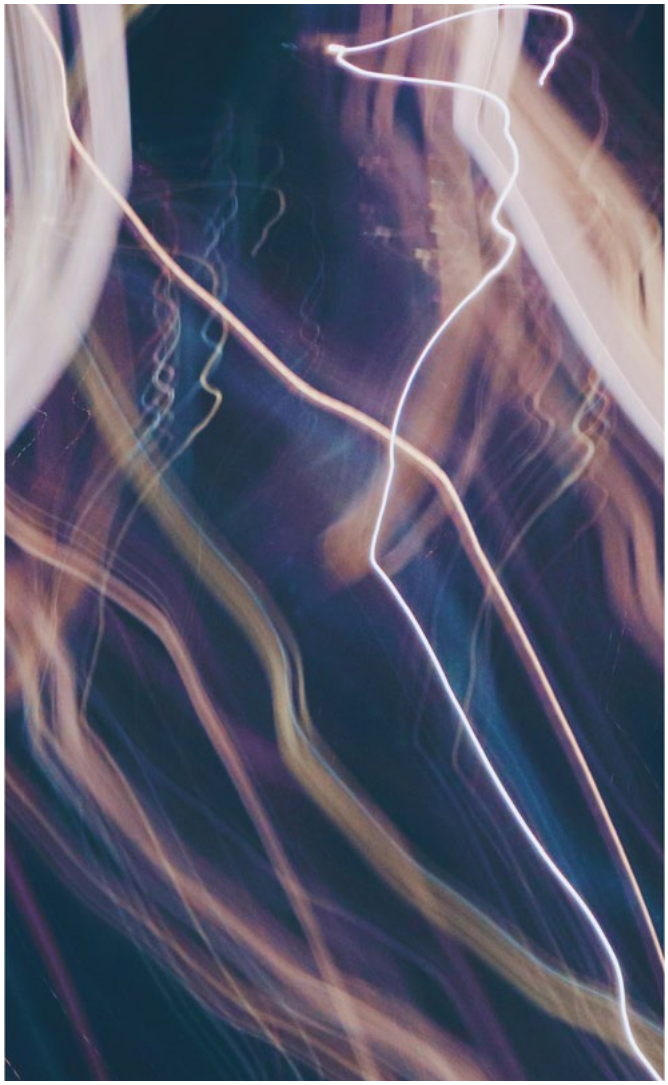
In this context, several initiatives have chosen to define benchmarks used as the basis for delivering labels, seals and similar certifications. The following report aims at creating an overview of these initiatives that are trying to operationalize ethical principles into practice through the means of creating labels, audit, certification or similar frameworks.

This mapping allows an understanding of the diversity of existing types of initiatives and actors, their aim, where their focus lies, and how these initiatives function. Many initiatives are still in their early development while others have been launched, yet failed to gain widespread acceptance and use. In the meantime, some have even been abandoned. This report illustrates the challenges these attempts faced and aims at identifying core factors of success. While a plethora of initiatives exists and new ones join the landscape seemingly every week, a meaningful contribution to this ecosystem may lie in their coordination.

Not all initiatives listed in the overview presented in this report (including its Annex) may be motivated primarily by the ambition to put ethics into practice. Some of them are at the first place oriented towards promoting consumers' data security or data protection, while others are mainly concerned with contributing to enhanced cybersecurity. However, the current international discourse on ethics in digital technologies shows that these objectives - for instance through criteria such as privacy protection, security, or transparency - rank among the predominant aspects referred to while defining cornerstones of ethically sound products or services for the digital realm. Thus, although their scope and ambition may touch upon specific relevant aspects, it is suggested for the present purpose to include these initiatives with regard to their contribution to informing and shaping ethical standards for the digital field.

The present report is structured in the following way: It firstly starts with a thematic overview outlining issues of digital ethics in general, then delves further into details and nuances of the various label initiatives. Secondly, the report proposes a mapping of the most important initiatives worldwide and a detailed presentation of a selection of these initiatives. This report thirdly concludes with an analysis of those projects examined in more detail as well as with recommendations addressed to the Swiss Federal Government on how to identify initiatives with a high potential of developing successfully and to support promising approaches. An integral part of this report is an Annex that provides an overview of all projects identified. The list of projects presented in this Annex is to be considered as a living document, as the ongoing activities in the field covered by the present report will continue to evolve and will remain very diverse.

The timing of this mapping report seems highly appropriate. The field is very dynamic and changing fast. However, the ecosystem is still not stable and no initiative could gain sufficient traction to establish itself as a regional or international standard. This is the opportunity which Switzerland, as hub for the governance of digital matters, could seize in bringing this ecosystem and particularly promising initiatives to the next level of maturity.



2
SETTING THE SCENE

This report focuses on initiatives and projects that propose some kind of auditing and/or certification mechanism for promoting as well as certifying responsible and ethical digital tools and services, in particular those initiatives taking the form of a label. The term "label" will be used to include all projects defining a certain standard and giving a certification upon achievement of this standard.

Before examining such label initiatives, it is necessary to get a better overview of the overall landscape of relevant projects with the ambition to put ethics into digital technologies and practices. For that, the present report will move from the overall picture of ethics in digital matters to the trust focused labels.

a.
Digital technologies and ethics

Ethics in digital technologies and practices is the buzzword of the post Cambridge Analytica era. As diverse scandals bearing upon digital technologies have been made public, the requirement to consider ethics an integral part of the way digital technologies are conceived, designed and used has gained traction. Ethical considerations and responsible innovation have become an integral part of the discourse on technology.[1] This bears upon companies, public institutions, law makers, civil society at both national and international levels - who are all at the same time confronted with the expectation to avoid mere "ethics washing".

There exist over a hundred different ethics guidelines, particularly regarding Artificial Intelligence systems[2]. Several countries have included in their AI strategy basic ethical considerations and ideas of implementation. Literature suggests that these guidelines, principle statements, declarations and policy papers show a significant degree of convergence on a number of ethical principles that characterise ethically sound AI systems, while being very unequivocal on the way they need to be interpreted and applied[3]. According to Jobin/Ienca/Vayena (2019), for instance, 11 values/principles that are repeatedly mentioned in these documents can be highlighted, namely transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity[4]. However, only five of these are to be found at least in 50% of the documents assessed, with not a single one being mentioned in all of them. In addition, the exact definition of these values/principles remain very disputed and the implementation in practice unclear.

Among these diverse documents dealing with ethical issues in the field of digital technologies, the notions of trust and trustworthiness often occupy a particularly prominent role. Raised from the perspective of users and/or citizens, digital technologies are required to become trustworthy. In order to be deemed as such, however, the way digital technologies are conceived, designed, and applied through particular digital services must comply with a set of ethical and/or legal criteria. The exact list of criteria making up "trustworthiness" for digital services is the object of ongoing discussions. The list of initiatives dealing with digital trust presented in

1 Martin K, Shilton K, Smith J (2019) Business and the Ethical Implications of Technology: Introduction to the Symposium. Journal of Business Ethics: 160. 307-17. doi.org/10.1007/s10551-019-04213-9.

Brom FWA, Chaturvedi S, Miltos L, Zhang W (2015) Institutionalizing Ethical Debates in Science, Technology and Innovation Policy : A Comparison or Europe, India, and China. In: Ladikas M., Chaturvedi S., Zhao Y., Stemerding D. (eds) Science and Technology Governance and Ethics. Springer, cham. doi.org/10.1007/978-3-319-14693-5_2.

Mehlich J, (2017) "Is, ought, should" – scientists' role in discourse on the ethical and social implications of science and technology. Palgrave Commun. doi.org/10.1057/palcomms.2017.6.

2 See the overview prepared by OECD AI Policy Observatory https://oecd.ai/dashboards?selectedTab=countries or by Algorithm Watch https://inventory.algorithmwatch.org/.

3 Jobin A, Ienca M et al. (2019) „The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence 1(9): 389–99.

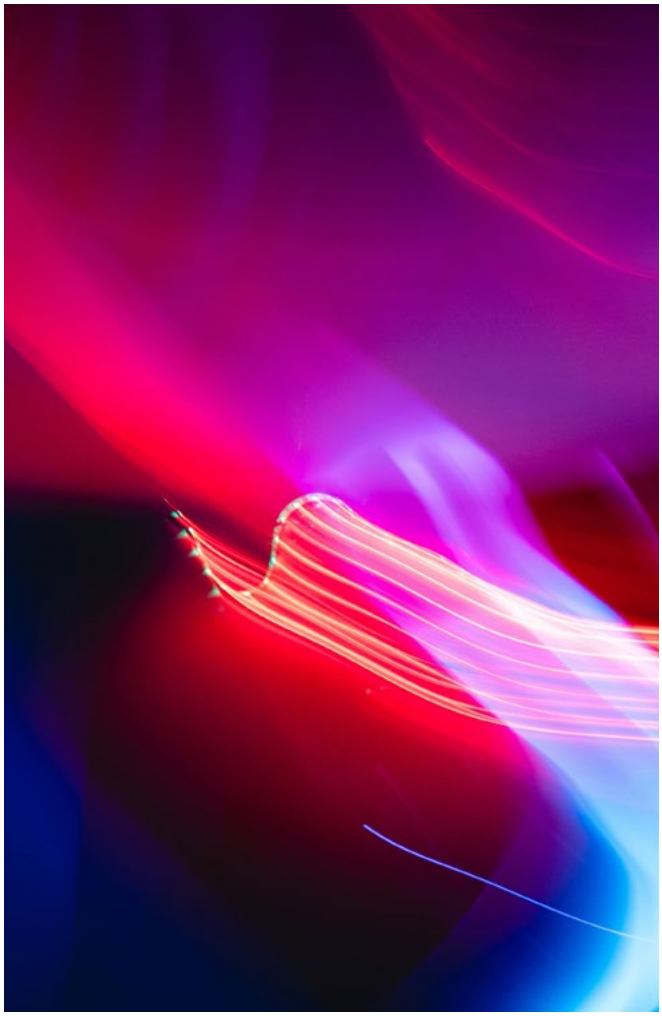
Floridi L, Cowls J (2019) „A Unified Framework of Five Principles for AI in Society." Harvard Data Science Review 1(1): 1-15

4 See Jobin/Ienca/Vayena 2019, 395

this report is a good illustration of these current debates. In short, for users and citizens, trust and trustworthiness are employed to crystallise both the process which digital technologies' design should go through and the features which the final product/service should display. As a consequence, many initiatives aim at providing guiding resources for tech developers to embed ethical considerations in the design and development process from the beginning.

This intense activity around ethical guidelines can be seen as part of a debate about the need for regulation in tackling the conception of digital technologies and the way they are used.[5] The US Artificial Intelligence Initiative[6] emphasises for example the importance to foster trust in the technology yet underlines the importance of voluntary measures and warns of barriers to AI innovation[7]. The tech sector, however, is keen on guidance, which causes a rift between the discourse on innovation in Washington DC and Silicon Valley. A number of high ranking tech entrepreneurs have sounded the alarm on the negative impacts technology has on society and called for changes how tech giants work and for more government regulation. In an opinion piece published in 2019, the president of Microsoft, Brad Smith, writes that "the greatest risk facing technology firms isn't overregulation – it's that government won't do enough, swiftly enough".[8]

More specifically, the drafting and increased consideration of these ethical resources should be set in the context of legislative evolutions. Most importantly, data protection law has been a key policy field to operationalise improvements in relation to digital technologies. Two of the most important legal frameworks are worth mentioning, particularly with regard to privacy. The EU's General Data Protection Regulation (GDPR)[9] and the California Consumer Privacy Act (CCPA)[10] aim to give users control over their personal data and reinforce their capacity to interact safely with digital tools. The GDPR is remarkable as it affects all companies that offer services or impact persons in the EU, regardless of where the company is based. Following both the GDPR and the CCPA, companies need to have clear and transparent policies on data collection and inform users whether and which data is being collected. The two legislations are fairly consistent when it comes to consumer rights such as the right to have one's data deleted, right to be informed, right of access, but differ when it comes to the right to opt-out. Through the importance of the European and Californian (respectively US) market, the two legislations have considerable power to create and influence regulatory standards around the world.



5 The more general context of this discussion bears upon the desirability of standardisation for technical products and its impact on innovation. See a German study arguing that standardisation brings benefit for the entire economy and is considered a catalyst for innovation for which there is high demand among tech developers and executives. Blind K, Jungmittag A, Mangelsdorf A (2011) "Gesamtwirtschaftliche Nutzen der Normung" DIN, available at: https://www.din.de/resource/blob/79542/946e70a818ebdaacce-9705652a052b25/gesamtwirtschaftlicher-nutzen-der-normung-data.pdf. For an overview on issues regarding labelling projects see also NGI Forward (2020): Report: Digital Trustmarks, https://ngi.eu/wp-content/uploads/sites/48/2020/01/NGI-Forward-Digital-Trustmarks.pdf (see also Initiative No. 5 described in detail below).

6 Exec. Order No 13,859 of Feb. 11, 2019, Maintaining American Leadership in Artificial Intelligence, 84 Fed. Reg. 3967 (Feb. 14, 2019), available at: https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

7 Memorandum M-21-06 of Nov. 17, 2020, Guidance for Regulation of Artificial Intelligence Applications. Executive Office of the President, Office of Management and Budget, available at: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

8 Smith B, Browne CA (2019) "Tech Firms Need More Regulation" The Atlantic. September 9. https://www.theatlantic.com/ideas/archive/2019/09/please-regulate-us/597613/.

Levinson-King R (2019) "Tech entrepreneurs call for more government regulation" BBC News. September 19. https://www.bbc.com/news/technology-49719946.

9 European Parliament and Council of European Union (2016) Regulation (EU) 2016/679. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.

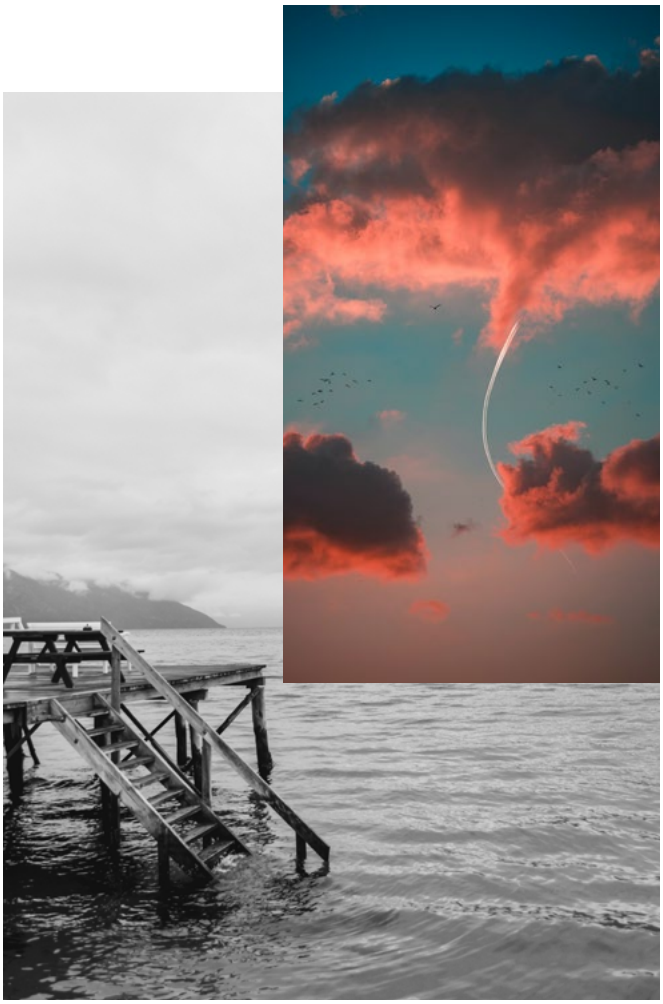
10 Goldman E (2018) An Introduction to the California Consumer Privacy Act (CCPA) International Association of Privacy Professionals IAPP. Available at: https://iapp.org/media/pdf/resource_center/Intro_to_CCPA.pdf.

In December 2020, the European Commission has presented a set of measures in the "Digital Services Act"[11]. The proclaimed objective is to ensure a safe and accountable online environment. The proposed regulation is built upon a distinction among relevant digital actors (ranging from intermediary services to very large online platforms reaching more than 10% of the EU population) and a set of crescendo obligations mirroring their potential for detrimental impact on their clients and on the broader public. Hence the obligations that different digital actors must comply with are proportional to their role, size and impact within the online ecosystem.

As part of the ambition to make the EU fit for the digital age, the European Commission has proposed a further regulation called the "Digital Markets Act"[12]. This regulation aims at making sure that large online platforms qualifying as "gatekeepers" will be under stricter control in order to ensure fair competition. The objective is also to make the EU a place for innovation and to give consumers the opportunity to get more and better services and products to choose from.

Generally, it is interesting to note that these legislative efforts take consumers' rights, and more broadly international human rights, as normative backbone[13]. Legislation is conceived as a way to protect human rights of individuals, acting either as consumers/users of digital services or as citizens impacted in their political self-determination by digital services. Human rights offer a broadly shared set of rights and duties for digital technologies, particularly with concerns about privacy, personal integrity, equal access, non-discrimination, freedom of expression and safety[14]. As exemplified by the EU legislation, this human rights focus can be combined with a competition law approach (securing fair conditions of competition).

Overall, the relations between ethics and legal frameworks regulating the fields of technology is best conceived as a relation of mutual enrichment. Ethical norms, often considered under the heading of "soft law" (as opposed to hard law), can be the forerunners of legislative acts. They offer a more informal forum for discussion and negotiations across diverse fields and with a diversity of actors taking part in the conversation. As such, they can exercise a certain pressure on the way issues are framed and how they are addressed by national, regional or international political decision-makers.



b.
The function of labels

There are a great number of labels and standards for the digital world in the pipelines of civil society organisations, government agencies, research institutions, private actors and professional associations predominantly in Europe and North America, a list of which will follow in the next chapter. These initiatives, while diverse in application, actor and target, all broadly have a common goal: to foster ethical digital services by providing a benchmark to evaluate them. They primarily function through information, either offering guidance to technology designers and companies, with regard to which ethical considerations need to be taken into account, or informing consumers on what to look out for when choosing products and services.

11 15th December 2020, Proposal for a Regulation on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. COM(2020) 825 final.

12 15th December 2020, Proposal for a Regulation on contestable and fair markets in the digital sector (Digital Markets Act). COM/2020/842 final.

13 For a mapping and visualisation of rights-based approach to AI, Fjeld, J., Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Harvard, Berkman Klein Center for Internet & Society.

14 For an overview, Wagner, B., M. C. Kettemann, et al. (Ed.) (2019). Research Handbook on Human Rights and Digital Technology : Global Politics, Law and International Relations. Research handbooks in human rights. Cheltenham, UK ; Northampton, MA, Edward Elgar Publishing.

In short, the different initiatives pushing the idea of a label (and the similar ideas gathered under this umbrella term) broadly share a similar vision about the function of the label. A label is a way of publicly ensuring trustworthiness of a digital service. If the label is widely used among companies creating a digital service and has gained recognition and acceptance among users, it is highly useful to communicate important information about a service. Increased transparency empowers users through information and heightens digital literacy. It therefore allows users to make informed choices in a free and competitive market. As a consequence, a label has potential to level the playing field and make ethical behaviour a competitive advantage. However, a label also has the downside of raising the barrier of entry, as administrative costs of implementing label criteria, and assessing these, can prove costly for companies with limited resources (such as SME and startups).

Labels have an essential relation with trust: they represent, sometimes in one single image/picture, the values and principles the label stands for. Beyond their content, these labels are trusted because the organisations that issue and control them are trustworthy. While consumers do not know the exact processes nor technical details a product or company goes through in order to be granted the right to display the label, there is a sense of trust that it is gained through transparent, rigorous processes by an organisation independent of the producer and audited regularly.

In that sense, it is key to understand that a label which is about trust in digital services has a double relation with trust: its content attempts at grasping what trust is about for users working with digital services, and the mechanisms of audit and control of the label also need to be trustworthy. There must be trust in the integrity of the label through safeguards against conflicts of interests.

If the label has been created by independent experts and is based on scientific analysis, the trust a label receives from consumers is high regardless of the product type or the country.[15] As e-commerce shopping continuously increased throughout Europe, consumer security and protection concerns gained traction. As a successful example, the label Trusted Shops is considered Europe's leading trustmark as it guarantees consumers that they are safe from scammers when purchasing online. Online retailers benefit from the increased trust and safety consumers feel which boosts conversion rates and sales. As a major difference to the trust label to be issued by the SDI, this certification applies to the entire online shop in question, whereas the SDI label shall apply to a particular digital product or service deployed by the individual company.

With respect to trust in digital technologies, the media have early been put under the pressure to distinguish distinct types of news. The fake-news debate is the most well-known example of the trust-based challenges bearing upon the media. The US based "Trust Project" is a perfect example of the attempt to provide news' consumers with information about the news they get[16]. In that sense, it works like a label wanting to empower citizens to make better decisions. Despite its societal relevance, this specific media debate is not part of this report because it primarily bears upon an assessment of content or producers of content.

Indeed, labels are not reserved to digital services, quite on the contrary. Labels are part of our everyday life when deciding which food we consume based on nutritious values uniformly colour coded – or which products to buy based on informations related to the sustainability performance of the products. In terms of international trade in general, for instance, consider "Fairtrade" standards and seals. And in the field of electronics, labels that qualify the energy consumption of a particular device help achieve informed consumer choices. All these labels offer information that is understood at first glance, known and accepted by most people. Also, they are audited for producers to whom the label applies, and they are issued by trusted label owners, whose reputation is key for the label acceptance. This is the case of organic food labels: it is a market signal of quality that allows producers to sell their product at higher prices than conventional products – and successfully so – for implementing higher ethical standards.

15 Rupprecht CD, Fujiyoshi L, McGreevy SR, Tayasu I (2020) Trust me? Consumer trust in expert information on food product labels. Food and Chemical Toxicology.

16 https://thetrustproject.org.

When comparing the analogous world of labelling to the digital realm, key differences appear clearly. Firstly, labels on digital services must account for different realities. Most importantly, digital services, by their very nature, can represent a threat with respect to personal data. By contrast to labels on food or devices, a label on digital services needs to take into account the data-based flows of information between the user, the service and the companies involved. Data that is gathered on an individual is more intimate and contextual: data taken from one aspect of a person's life is by no means representative of the person or their social group, yet is often used in that way. Taking the example of training data for algorithms, biased and incomplete data sets can reinforce discrimination and exclusion with great risks for the wellbeing of individuals – especially when used in the distribution of important resources such as credit worthiness, welfare benefits or housing. Any label dealing with digital services must make room for data rules.

Secondly, a major challenge in the digital realm is the fast pace of continuously evolving technology. How can a label keep up with a digital service being updated several times a week? The trade-off is between a label that is specific enough to account for the technical features of digital services (facing the problem of updates) and a broader label which focuses on more general principles (less affected by the problem of updates).

Thirdly, labels dealing with digital services face the issue of territoriality more directly. Organic labels need to consider complex chains of production, transport, delivery. Yet, the physical movement of goods (eg. fruits) makes it easier to follow the different steps and address potential problems. For digital services on the other hand, the data-based processes might produce complex situations when it comes to identify and assess all the actors involved and the corresponding responsibility. This challenge has an important legal dimension when it comes to identifying which norms and regulations apply to the specific service.

c.
Intermediary conclusions

Overall, the context in which the current initiatives and projects are being developed entails the following key points:

- There has been an intense activity of creation of ethical resources dealing with digital challenges. These resources have taken the form of charters, declarations, codes of conduct or public policy strategy documents.

- The interest for ethical values and principles is paralleled by legislative initiatives which directly address the challenges raised by digital technologies. Adaptations in data protection law, privacy law but also consumer protection law and competition law are key elements of these legislative efforts. The EU assumes an important leadership on these challenges.

- Several actors have looked for ways to go beyond the identification and proclamation of these ethical resources. They have been looking for methods to operationalise these resources. The idea of labels encompassing several ethical requirements is one way to proceed with operationalisation.

3
MAPPING EXISTING INITIATIVES

The field of initiatives dealing with ethical challenges of digital technologies is highly dynamic and fast growing. This report entails two overviews. In the Annex, readers can find the complete list of initiatives identified during our research. In the following chapter, an overview of the initiatives deemed most relevant for the purposes of this report is presented, with 12 initiatives being examined in more detail.

The initiatives were found through search of meta-studies, conference presentations on the topic, news articles and websites of organisations dedicated to promoting ethics in digital technologies. We also took advantage of the co-development process we organised for the SDI in Summer-Fall 2020 in which we contacted more than 100 organisations to comment on a first draft for a Digital Trust Label[17].

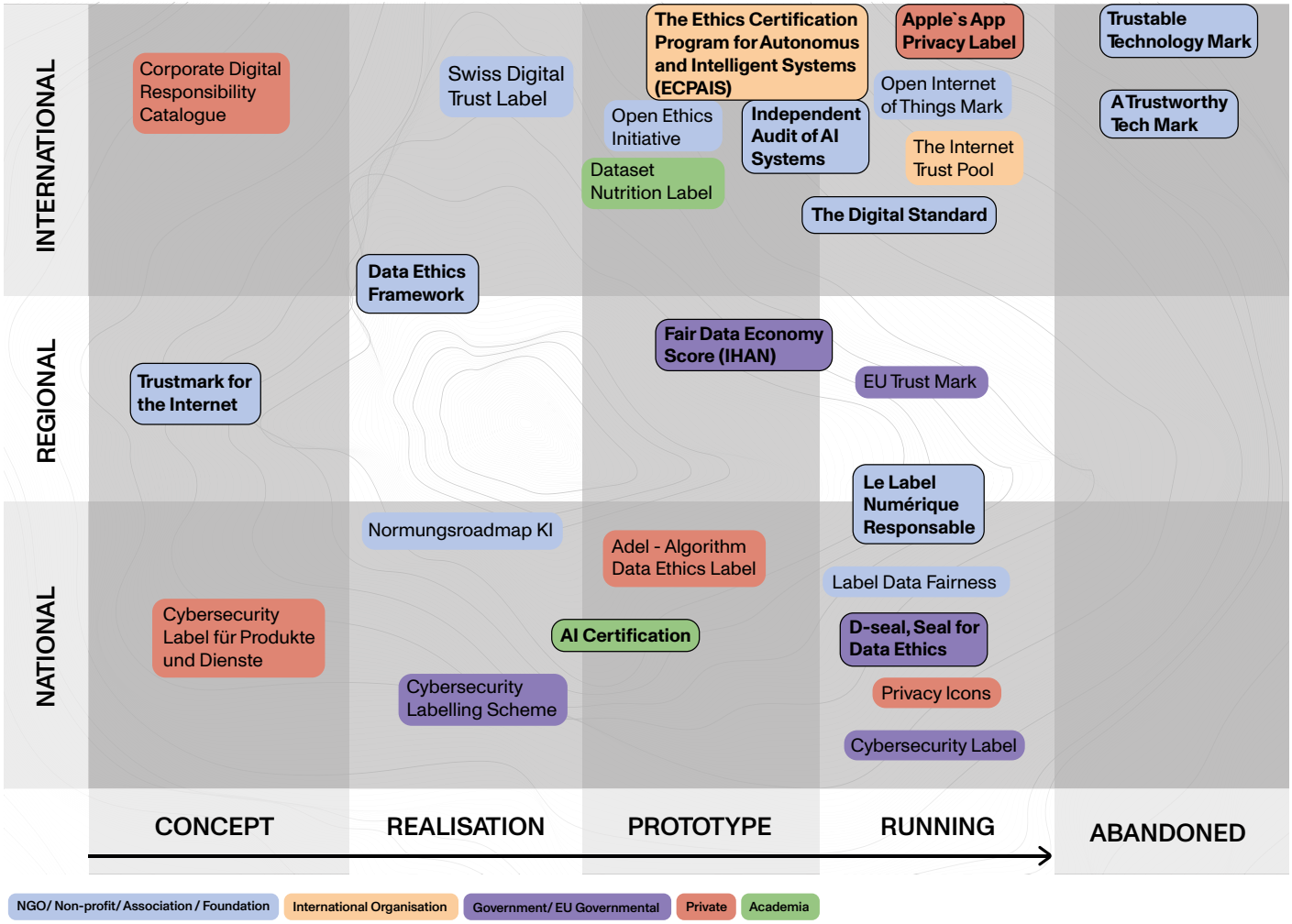
Criteria for inclusion in the broader list (see Annex) are the following:

- Formulating normative requirements for digital services or products, that is formulating a proposition on how digital technologies should be designed/implemented/used.
- Entailing considerations about ethics, preferably having ethical values/principles as main focus. Initiatives that focus exclusively on technical standards were generally left out, unless the technical dimension was clearly approached in a values-based perspective. Similarly, declarations or guidelines dealing exclusively with ethical values/principles - without link to operationalisation on digital services/products - have also been left out.
- The form of the initiative was left open: recommendation, policy framework, regulative proposal, standard or label.

Within all the initiatives listed in the Annex, we made a second round of selection to map the most relevant initiatives. The mapping focuses on the initiatives formulating/proposing a label which is applied to a digital service/product.

The mapping below shows the diversity in initiatives with regard to actors involved, level of maturity as well as the desired scope of the initiative. These three points of information offer an easily understandable overview of the complex matrix of initiatives.

Initiatives in bold are presented in more detail on the following pages.

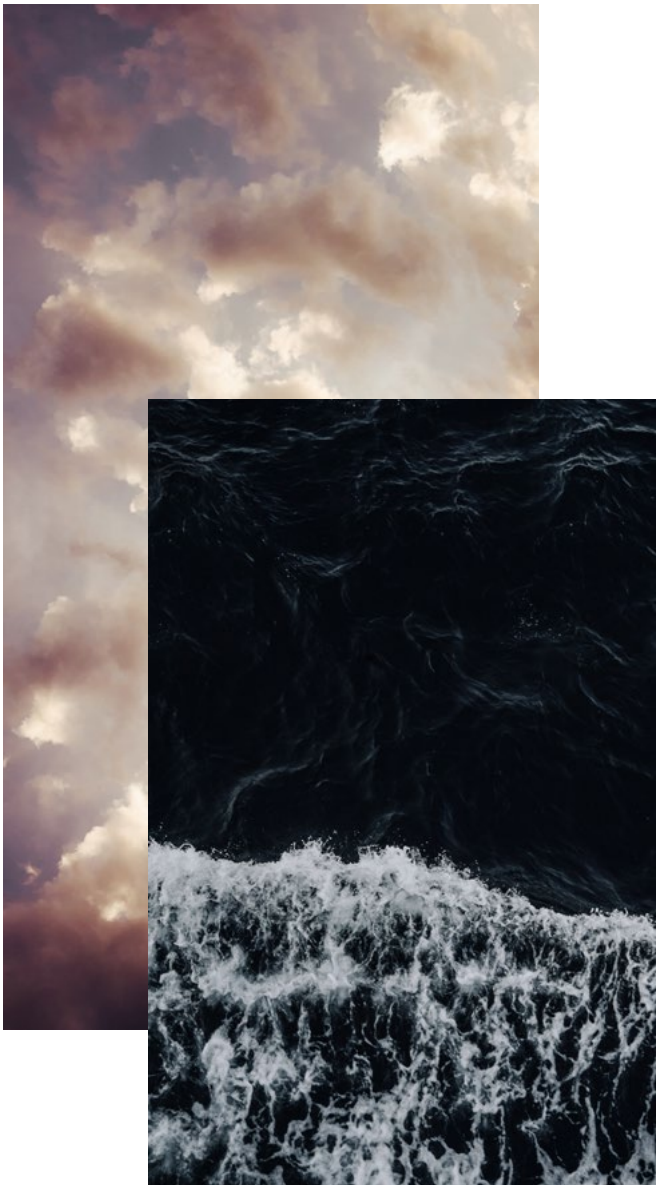


17 A report of this co-development process can be found here: https://a.storyblok.com/f/72700/x/389353b97c/sdi_report_8.pdf.

4
PARTICULARLY RELEVANT INITIATIVES FOR SDI AND SWITZERLAND

The following section presents 12 projects from the map that are especially relevant for Switzerland. These initiatives were chosen to offer a broad overview of existing labels and similar initiatives internationally, to compare and shed light on the diversity of objectives, focus, and to better understand the corresponding challenges and the success factors of different labels. The initiatives were chosen as they fulfil one or several of the following criteria:

- The organisation behind the initiative is a major actor.
- The project is unique and innovative or otherwise of particular interest.
- There are strong similarities with the Digital Trust Label.
- The initiative contains lessons about the development process and the values behind.
- Examining the initiative allows to draw conclusions about a successful implementation, as well as the challenges and success factors of such projects that may inform future strategy of the SDI and Switzerland.



The following projects are presented in detail:

1 **Data Ethics Framework** (AI Ethics Impact Group, Bertelsmann Stiftung)
2 **The Digital Standard** (Collective effort: Consumer Reports; Disconnect; Ranking Digital Rights, The Cyber Independent Testing Lab)
3 **The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)**, (Institute of Electrical and Electronics Engineers IEEE)
4 **Fair Data Economy Score** (part of Human Driven Data Economy IHAN, Sitra)
5 **Trustmark for the Internet** (EU Next Generation Internet Initiative)
6 **Trustable Technology Mark** (Mozilla OpenIoT Studio + ThingsCon)
7 **A Trustworthy Tech Mark** (doteveryone)
8 **D-seal, Seal for Data Ethics and IT Security** (Public Private Partnership)
9 **AI Certification** (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS)
10 **Independent Audit of AI Systems** (For Humanity)
11 **Label Numérique Responsable (NR)** (Institut Numérique Responsable)
12 **Apple's App Privacy Label**

01

Data Ethics Framework

AI Ethics Impact Group, Bertelsmann Stiftung

Brief Description

This initiative represents a framework to operationalise ethics in practice through a proposal of what an AI ethics label could look like based on six values: justice, environmental sustainability, accountability, transparency, privacy, and reliability, based on contemporary discourse and operability.
It includes a risk matrix:

- to assss the risk of AI use in specific domains according to ethical challenges of AI;
- to help classify an application in context. Categorisation of AI systems into 5 classes thus indicating the level of regulation required.

0 = no considerations of AI ethics needed
4 = no algorithm should be used to take decisions
Goal: Control, oversight and comparability of different AI systems.
Operationalisation & measurement through a VCIO model: values, criteria, indicators and observables
The proposed label visualisation is inspired by energy efficiency labels as it is easily understood and recognised by consumers.

Target Groups

Organisations developing and using AI, policymakers and standard setting bodies

Regional Scope

EU (and global): specifically designed to enforce European values and protect citizens in Europe. The authors see roles for European standards developing organisations and European policymakers, yet also potential for actors worldwide adapting and operationalising the framework.

Product of focus

Products and services using AI systems

Self-assessment or third party

The authors aim at assessment through third party and regulatory bodies. However, concrete implementation is not outlined in the framework, but left at the discretion of EU standardisation committees and other stakeholders.

Development process

Little public information available on the development process itself.

Framework created by an interdisciplinary consortium of experts and academia:

- HLRS: High performance computing Center | Stuttgart
- Karlsruhe Institute of Technology
- Technische Universität Darmstadt
- Technische Universität Kaiserslautern: Algorithm Accountability Lab
- Universität Tübingen: International Center for Ethics in the Sciences and Humanities
- iRights.Lab (independent Think Tank)

Actor type

Foundation

Country of origin

Germany

Level of maturity

Realisation (2)
Next Steps: With completion of the report proposing how a label could be operationalised, the initiators invite the EU standardisation committee to come together to refine, complete and further operationalise the label proposal. The European Commission is invited to develop regulation to give teeth to the standards outlined in the document.

The authors see their future role in supporting standardisation and policy actors, initiating networks and activities, raising awareness and refining conceptual ideas.

Success factors and challenges

Legitimacy through the institutions involved, who are trusted to act in the interest and for the wellbeing of end users and society without financial motivation.

A multimethod approach was used to develop the framework, which allows to solve the issue of the inherent context dependency and the socio-technical nature of AI systems, as well as the diverse interests of different stakeholders (end users, policymakers, companies).

Clear goal of being primarily implemented by a strong and legitimate body: European standardisation committees supported by the European Commission.

Challenge: The European Commission invited to further develop the label, yet the question is open as to how authors can assure that the document will be developed to become an implemented label.

Comparison to the Digital Trust Label

Similarities:

- Human centred for citizens and consumers to make informed choices
- Ethics based
- visualised label

Differences:

- This label focuses on one technology - AI - and addresses organisations developing and using AI as well as policymakers and standard setting bodies.
- It primarily has the aim to be implemented by policymakers and standardisation bodies.
- The framework explains the "how to", but there is no concrete plan to operationalise the approach. Instead, other stakeholders are invited to implement.

02

The Digital Standard

Collective effort: Consumer Reports; Disconnect; Ranking Digital Rights, The Cyber Independent Testing Lab

Brief Description

This initiative represents a framework to operationalise ethics in practice by setting standards and criteria for testing and rating software and Smart devices on the basis of privacy, security and data practices.

Framework content:

- It is human centred and consumer-oriented to evaluate how t echnologies respect consumer's interests and needs through four user values: Security, privacy, ownership, governance. Each category is made up of several indicators which further include variables that operationalise the evaluation, same as the SDI label.
- It defines state of the art practices from the design stage on, within electronic software industries.
- It creates digital privacy and security standards to help guide the future design of consumer software, digital platforms and services, and Internet-connected products.
- It offers advanced tools to create tailor-made tests for specific devices.

Target Groups

The framework offers information to companies on state of the art practices and how to implement these.

Regional Scope

USA

Product of focus

Electronic software, hardware and services

Self-assessment or third party

- Voluntary, self-assessment by companies
- Third party assessment as case studies of specific technologies by Consumer Reports, primary organisation behind the Digital Standard (Connected Cameras, Mental Health Apps etc.). Other organisations are welcome to contribute a case study.

Development process

An open source, living document and collaborative effort (available on GitHub) co-created by product developers, researchers and technologists. The Digital Standard is created and managed by the NGO Consumer Reports together with Disconnect, Cyber Independent Testing Lab and the New America Foundation (Ranking Digital Rights project).

The organisations involved continuously update the framework on a quarterly basis with input received from experts and civil society. Anyone is welcome to contact the Digital Standard with ideas of improvement.
First case studies using the Digital Standard to assess technology started in 2019 and are ongoing.

Actor type

NGO

Country of origin

USA

Level of maturity

Prototype already running but continuously updated (3-4)

Success factors and challenges

- Community led and living document actively seeking comments assures the inclusion of a diversity of insights or different stake holders as well as societal groups. It also solves the issue of constantly updating technology and changing social attitudes thanks to frequent periodical reviews.
- Picked up and promoted by several trusted and strong actors who promote the framework. The open development process that allows many important stakeholders to participate, helps strengthen trust in the Digital Standard and also distribution and education about it. Organisations that have participated in its development are likely to promote it within their network.
- The initiating organisations are major, well established players in protecting consumer and digital rights, bolstering legitimacy and trustability of the framework.
- The operationalisation criteria and variables are publicly accessible and in a language easily understandable to laypersons.

Comparison to the Digital Trust Label

Similarities:

- Criteria of operationalisation, especially security, are very similar to the SDI. However, The Digital Standard goes further by adding criteria of personal safety, testing if the company helps users protect themselves from grief, abuse and harassment.
- Goal: empower people to make better and informed choices about the products they use.

Differences:

- Not a label but a framework.
- Based on self-assessment by companies, not checked by third parties. The Digital Standard includes the possibility that third party organisations (such as consumer protection organisations) assess products in the form of case studies.
- Complete testing procedures are developed as part of the framework.
- The framework is freely and publicly available for organisations and individuals interested in the assessment of digital services.
- Living, open-source consensus based, content fully available on website inviting commentary.
- Focus specifically on human rights, freedom of expression as well as governance and/or business model.

03

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) Institute of Electrical and Electronics Engineers IEEE

Brief Description

IEEE is one of the world's largest technical professional organisations with the objective to advance technology for the benefit of humanity. IEEE develops standards rooted in consensus, due process, openness, right to appeal and balance.

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) creates specification and metrics for certification and marking processes: a number of labels and standards.
Core values: trust, growth and nurturing, partnership, service to humanity, integrity in action

The ECPAIS program includes the development of three labels treating the following issues in Autonomous and Intelligent Systems (AIS) separately:

- Transparency
- Accountability
- Algorithmic Bias

Goal:

- Establish societal and policy guidelines for autonomous and intelligent systems to remain human-centric, serving humanity's values and ethical principles.
- Prioritise human wellbeing with AIS.

Target Groups

Organisations, businesses

Regional Scope

International

Product of focus

Autonomous and Intelligent Systems (AIS), products and services (e.g. smart homes, companion robots, autonomous vehicles etc.)

Self-assessment or third party

Self-assessment and third party

Development process

First stage of development (currently ongoing):
work done by experts/members IEEE

Second stage (if going into a normal IEEE standardisation process):
Global, open and inclusive, according to IEEE procedures

Actor type

International organisation, technical professional association

Country of origin

International (USA based)

Level of maturity

Prototype is running in the form of several case studies with public institutions in the EU (3-4)

Success factors and challenges

- Major international organisation of technical professionals with truly international reach developing the project, following set standards: trustworthy organisation, global reach, technically sound.
- Process is open for specialists, academic institutions and government organisations involved in policy and regulations.

Comparison to the Digital Trust Label

Similarities:

- Focus human centered, values-driven design
- Similar core values and goal:
  - Foster trust and empower people through transparent information to make informed choices.
  - Ambition to "easily and visually communicate to consumers and citizens" whether the certified services "are deemed 'safe' or 'trusted' by a globally recognized body of experts".
  - Reinforce an ecosystem of responsible creation and use of digital tools.

Differences:

- Technology based, focus on AI systems
- Is not all-encompassing but ethical considerations are topic specific.

04

Fair Data Economy Score

part of Human Driven Data Economy IHAN, Sitra

SWISS DIGITAL INITIATIVE

Portrait of Initiatives

Brief Description

A criteria framework and maturity model guides organisations to evaluate their fair and sustainable use of data – such as the collection, sharing and use of data – to contribute to an ecosystem of fair data.

The tool is divided into several stages with detailed steps to follow and further recommended actions to take. These criteria were developed to become the first version of the Fair Data Economy Score tool to measure the maturity of a company's data use. It aims to encourage approaches to data use that are open, transparent and above all trust-promoting.

The model is based on the European Union's data economy principles: trust, access, human-centricity, innovation, competence and sharing. Fairness is measured through six dimensions each containing 10 questions in which ethical aspects are embedded (1. data architecture and technology, 2. data management and capabilities, 3. the organisation's values, culture and competence, 4. data -drive services, 5. value creation and profit performance, 6. operating as part of an ecosystem).

The maturity model is one tool of a larger project: The Human Driven Data Economy which establishes an ecosystem for a fair and functioning data economy through various initiatives, activities and projects. Having shared rules, tools and sharing data within a fair and trustworthy ecosystem shall provide European companies with a competitive advantage:

- Support new internet standards for data productising, portability and interoperability to boost the emergence of global data markets.
- Where data flows more freely, available to operators of all sizes, data shared fairly between different sized companies and bodies of the public sector.

The model aims to create a method for data exchange and set up European level rules and guidelines for ethical use of data.
The model is built on trust, a human centric approach, community, and sustainable growth, see their Q&A

Target Groups

Organisations, Companies

Regional Scope

Europe

Product of focus

Data use within organisations

Self-assessment or third party

Primarily self-evaluation, but also third party

Actor type

Public foundation, independent fund and think-tank

Country of origin

Finland

Development process

The criteria of the maturity model were further developed in autumn 2020 in co-creation between Sitra and several external finnish and international experts. They provided insights on given topics through workshops and case-interviews.
The IHAN project as a whole is the fruit of international cooperation and open access as well as open data. Developers can now log into the Beta version, test it and add their own data sets for others to use.

Level of maturity

Prototype (3)
The first version of the data economy criteria and maturity model was accomplished in late 2020. Version 2.0 will be tested online via a fully functional developer portal accessible for the whole internet community and through workshops from February to May 2021. Three workshops are planned, first with NGOs, then companies and lastly with regulators, decision makers and interest groups.

The overarching IHAN programme started in April 2018 and will end in 2021. A first version of the IHAN fair data economy infrastructure testbed (see: rulebook to operate in the ecosystem) was released in June 2020 (with version 1.2 released in January 2021).

Success factors and challenges

- Sitra, the initiating organisation is an experienced organisation, running for several years already.
- The organisation is under supervision of the finnish parliament. Sitra is well funded and has strong political support both nationally and transnationally, aligning with the EU strategy for data.
- Diverse stakeholders representing civil society, companies and regulators and interest groups involved and actively participating at every stage of development of the different tools. This enables the tools to be useful, considering diverse perspectives and answers to real needs. This bolsters its legitimacy and credibility.
- Diversity of initiatives to create a wholesome system regarding data.

Comparison to the Digital Trust Label

Similarities:

- Strong collaboration with industry
- Similar goal to create wide range change within the digital industry for ethical and human centered data use.

Differences:

- The project is part of a larger endeavour to establish best practices and a complex infrastructure for an industrial ecosystem to develop.
- At the moment there is no corresponding label for the consumer but as a primary aim a tool for self-assessment by businesses.

05

Trustmark for the Internet

EU Next Generation Internet Initiative

Brief Description

This project entails a series of labels under one umbrella trustmark. It is funded by the EU's Horizon 2020 research and innovation programme.

Aim: build a more democratic, resilient and inclusive human-centric future internet that is safe and protects users.

- The idea is to create a Next Generation Internet trustmark to support the development and use of responsible technology and software. It highlights the importance of having specific labels as building blocks for a more comprehensive "umbrella" trustmark that covers the full digital experience that would be easy to understand and recognised by users.
- Criteria that may be included: (Cyber)security, privacy, best data practices, AI ethics such as bias and inclusive representation, transparency, accountability and environmental sustainability of the internet itself.
- Visualisation: traffic light system and indicating where further information can be found through a "SmartLabel" QR code. This online repository would account for the fact that digital products update regularly and change significantly over their lifecycle.
- The trustmark's content is built upon GDPR and will include future legislation, but aims to go beyond a minimum standard by complementing the law.

Target Groups

Companies

Regional Scope

EU

Product of focus

A comprehensive label for the entire digital realm of the internet. Several technology specific labels for all technology and software related to the internet (AI systems, digital services, IoT) would be created under the wing of a single "umbrella" label.

Self-assessment or third party

Third party

Actor type

EU program: a flagship initiative by the European Commission NGI Forward – strategy and policy arm of the Next Generation Internet

Country of origin

EU

Development process

External experts, civil society, industry and government come together in an International consortium: Nesta (UK), DELab, University of Warsaw (Poland), Edgeryders (Estonia), City of Amsterdam (Netherlands), Nesta Italia (Italy), Aarhus University (Denmark), Resonance Design (Belgium).

Based on research by Nesta and interviews with experts on responsible and ethical technology, the initial 2020 report outlines what this trustmark would look like.

The report states that a successful trustmark should be initiated and overseen by a large, transnational and publicly accountable organisation with enough capacity, resources and legitimacy..

The project started in January 2019 with research for the initial report published a year later in 2020. Horizon 2020 funding lasts 3 years.

Next steps: NGI Forward projects will gather stakeholders to co-develop and co-design the label. The European Commission is invited to take lead in facilitating and shaping the trustmark.

Level of maturity

concept (1) → final report

Success factors and challenges

It is too early to judge the success of the project as it is still only in the conception phase. Challenges ahead may be:

- its further development in a comprehensive way, yet still fit for purpose
- encouraging companies to participate
- the trustmark's governance
- generate trust, recognition and legitimacy among end users
- making it meaningful, yet accessible for end users

Comparison to the Digital Trust Label

Similarities:

- Goal, generate trust and legitimacy for end users

Differences:

- Inherently transnational project with developing organisations from across Europe.
- Umbrella label, comprising the entire realm of digital services and devices, therefore potentially bigger field of application.

06

A Trustmark for IoT

Mozilla OpenIoT Studio + ThingsCon

Brief Description

This project represents a prototype trust mark (label) aimed at users of IoT, to indicate excellence of a company and offer alternatives to basic certification schemes that have very low requirements.

- Aim: empower users to make informed decisions and enable companies to prove that their connected products are trustworthy.
- Hard to earn, easy to apply to and free to use.
- No certification in the legal sense, nor third party audit. Companies answer specific criteria questions, which are then reviewed by independent experts at ThingsCon. If they fulfill certain conditions, demonstrating a strong commitment to build exceptionally responsible technology they receive the right to display the label for free.

Target Groups

Companies

Regional Scope

International

Product of focus

Entire companies that work with IoT, all products of a company that passes the audit receive the label

Self-assessment or third party

Self-assessment with third party control

Development process

Supported by a Mozilla Fellowship, an initial research and report on the potentials and challenges of a trustmark for IoT was published in 2017. It is open source, including a large number of experts, and "living" through continuous input and adaptation.

Actor type

NPO, Foundation

Country of origin

USA, Germany

Level of maturity

Abandoned at prototype level in July 2020. However, ThingsCon are convinced that the trustmark/label model is strong and appropriate to strengthen digital rights of users.

Success factors and challenges

Reasons the project was abandoned:

- Funding: failure of finding a sustainable business model while being open source and free.
- Outreach too limited: while there is great demand on the topic of trustable/responsible technology among policymakers and consumers, ThingsCon failed at getting enough traction among companies.
- Difficulty to enforce the label when labeled digital services were sold to another company. High responsibility toward consumers with potential backlash.
- The bar might have been too high. Also, unclarity within companies which team was best suited to answer the respective questions, which might be a red flag in itself.
- The organisation was too small to carry through with the project. However, ThingsCon is now involved with the NGI Forward initiative "Trustmark for the Internet".

Comparison to the Digital Trust Label

Similarities:

- Aim to strengthen trust in end users and contribute to higher standards among technology providers.

Differences:

- Only applies to IoT
- Not a label per se, but support and quality control resources available for companies to use throughout the design process.
- Developing organisation was small, with limited reach and resources.

07

A Trustworthy Tech Mark

doteveryone

Brief Description

Label

- Create trustworthiness of the whole system: not only one facet of technology, but the entire way products and services are built, maintained, supported and used. Change in tech companies' behaviour regardless of product sector or team skills.
- Takes up an interconnected understanding when defining core values (fair, ethical and sustainable) based on how people describe their concerns about technologies.
- Create a specific checklist of requirements that works for a variety of products, services and systems. Values-based approach, encompassing standards and best practices which exist or are emerging within specific technology fields, such as unbiased algorithms, privacy and security.

Target Groups

Companies

Regional Scope

International

Product of focus

Digital services and products in general

Self-assessment or third party

Third party

Development process

First announced in 2017, abandoned in Sept. 2019

Actor type

Think Tank

Country of origin

UK

Level of maturity

Abandoned at a concept stage:

- doteveryone describes the specific challenges of making a trustmark work for digital technologies in a comprehensive article. Reasons why the project was abandoned are summarised under "success factors and challenges".
- doteveryone have instead developed a set of tools to help organisations be more responsible. In May 2020 doteveryone disbanded, and major parts of their work was taken over by the Ada Lovelace Institute and Open Data Institute.

Success factors and challenges

Reasons for abandoning the project according to doteveryone:

- Digital services are too complex and fast changing as are societal attitudes towards them. According to doteveryone "digital services aren't bananas", this makes it hard to set common standards and evaluate these.
- Developing a trust mark is a multi-year investment:
  - a brand has to be developed in which users trust
  - setting standards and creating a sustainable audit system.
- doteveryone considered itself too small an organisation to achieve this.
- Consumer's choice is already limited and confusing. Helping tech developers build trustworthy technology was considered more important.

Comparison to the Digital Trust Label

Similarities:

- Aim to strengthen trust in end users and contribute to higher standards among technology providers.

Differences:

- Adopted a systems-approach rather than a features-approach
- Focus on a values-based approach, and business and product choices at a higher level.

08

D-seal, Seal for Data Ethics

Danish Ministery for Industry, Business and Financial Affairs

Brief Description

The label was launched as part of several initiatives of the Danish National Strategy for AI with the aim to make the public and private sector use data and AI more responsibly, assure transparency and provide data ethical guidelines. It is funded by the Danish Industry Foundation through 18 million DKK (approx. 2.6 million CHF).

To obtain the label, a company must meet criteria defined for their business type, which include IT-security, privacy and data ethics. They deal with questions regarding the relation between society, suppliers, management, organisation and operations and consumers.

The D-seal's criteria are not openly accessible in English, but are based upon several frameworks from European and national councils, committees and task forces.

The label is aimed at all sizes of companies and types. However, it does not apply a "one size fits all" labelling programme but uses a clearly defined risk-based approach, where criteria are differentiated based on an initial risk-profiling of the company in question.

It offers the opportunity for companies to show that they treat user data ethically and their AI products are secure and trustworthy.

Target Groups

Companies

Regional Scope

National; with hope to serve as a model for similar initiatives in other countries

Product of focus

Company as a whole, Services, IoT, Devices, Mobile Apps

Self-assessment or third party

Self-assessment

Development process

Closed process through an advisory board of 20 experts including the following organisations: The confederation of Danish Industry, the Danish Chamber of Commerce and SMEdenmark, Danish Business Authority and the Danish Consumer Council.

The label is endorsed by experts within IT-security, privacy and data ethics.

Actor type

Public Private Partnership

Country of origin

Denmark

Level of maturity

Running (4)
Planned commercial launch in spring 2021. Currently the D-seal criteria are being tested by companies.

Success factors and challenges

- Backing and endorsement by several key industry players make it likely to be accepted by companies.
- Public private partnership makes it likely to be more easily implementable and the public arm and the participation of the Danish Consumer Council makes it more trustworthy to end users.
- In terms of challenges, the closed nature of the development process and the involvement of predominantly business actors may make the label less credible for end users and might run the danger of setting the bar too low to truly be useful.

Comparison to the Digital Trust Label

Differences:

- Stronger emphasis on technical issues
- Focus on personal data security and privacy
- Ambition to remain predominantly a national label but inspire similar projects abroad
- Industry driven, attached to a governmental department
- Closed development process with no external organisation that assessed the label. Lack of stakeholder diversity.
- Applies to an entire company and all their products.

09 AI Certification

Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS

Brief Description

Certification system for trustworthy use of AI

Inspired by the EU High-Level Expert Group's recommendation on trustworthy AI, this project offers a certification process and inspection catalogue to assess AI applications.

- The certification examines technical issues as well as ethical and legal aspects.
- Aim:
  - Identify regulatory gaps, make these transparent and close these gaps through voluntary control criteria.
  - Build in checks and controls at the design stage of an AI application.
  - Contribute to the creation of "Made in Europe" standards: fairness, transparency, autonomy, control, data protection, safety, security, reliability.

Target Groups

Companies

Regional Scope

National

Product of focus

AI Systems

Self-assessment or third party

Third party

Development process

The development process relied on experts from the fields of machine learning, law, philosophy, ethics, and IT security. The basic principles for technically reliable and ethically responsible artificial intelligence will be developed in an openly organised process that involves a wide range of stakeholders from business, research, and society. Certification process together with the Fraunhofer Institute and Germany's Federal Office for Information Security (BSI). The inspection catalogue is a "living document" continuously updated.

Involved are: Fraunhofer IAIS, Federal Office for Information Security (BSI), the universities Bonn and Köln and RWTH Aachen, Deutsches Institut für Normung DIN, several DAX-30 companies from a diversity of industries.

Actor type

Academia and Government

Country of origin

Germany (regional Government Nordrhein-Westfalen)

Level of maturity

Prototype (3)
White paper (2019) outlining the certification process, inspection catalogue published (2020) now begins certification of AI applications through the organisations responsible for the project.

Success factors and challenges

- Public private synergies with backing from industry stakeholders can favour acceptance among companies. Academia and government as leading the process and content development strengthens the project's legitimacy and trustworthiness.
- The certification catalogue as a living document with frequent updates solves the issue of fast changing technology.

Comparison to the Digital Trust Label

Similarities:

- Development process including academia, companies, and society.

Differences:

- Originates from a regional/subnational government with the aim to be applied nationally.
- Focus on AI.
- Not a consumer facing label but industry oriented certification system

Independent Audit of AI Systems IAAIS

For Humanity

Brief Description

For Humanity is a New York based charity examining and aiming to mitigate risks related to AI and Automation in areas of trust, ethics, bias, privacy and cybersecurity at a corporate and public-policy level. Main points of their framework for auditing AI systems are:

- Five key areas: privacy, bias, ethics, trust and cybersecurity.
- The five audit silos are assessed by fellows of For Humanity.
- The framework audits entire companies, but specific services and products using AI and autonomous systems can also be awarded the "seal of approval".
- Goal:
  - To build an infrastructure of trust.
  - To make the creation of safe and responsible AI and automation profitable, while dangerous and irresponsible AI systems shall become costly.
- Audit rules are: Implementable, binary (compliant/ non-compliant); open source and Iterated; unambiguous; consensus-driven, measurable.
- The audit would be free, yet if the audit has been successful a fee needs to be paid for the right to display the logo (see a Q&A here).

Target Groups

Companies are audited and labeled for end user information

Regional Scope

Global aspiration, (launch initially planned in New York City)

Product of focus

Company as a whole, digital services and products using AI Systems

Self-assessment or third party

Third party; audit by For Humanity

Development process

Global open source and crowd sourced process of volunteers in collaboration with fellows at the For Humanity Center.

Actor type

Foundation

Country of origin

USA

Level of maturity

Prototype preparing for launch (3)

Success factors and challenges

- Strong expert knowledge with specialised researchers working on the project
- Crowdsourced and international

Comparison to the Digital Trust Label

Similarities:

- Trust focused

Differences:

- Very detailed with many different properties
- Properties are application specific
- Carrying organisation is rather small with limited resources

11

Le label Numérique Responsable (label NR) Institut du Numérique Responsable Suisse INR-CH

SWISS DIGITAL INITIATIVE

Portrait of Initiatives

Brief Description

The label NR comprises a Label and a charter (charte du Numérique Responsable).

The Institut Numérique Responsable in Switzerland was founded in 2020 in partnership with the Belgian Institute for Sustainable IT and INR in France who is the main initiator of the NR label.

Reducing negative impacts digitalisation may have on the economy, society and the environment, the opportunities digitalisation offers to reduce humanities impact on these areas and to create lasting values for responsible and inclusive innovation.

Based on an initial charta on digital responsibility, the label focuses strongly on creating ethical, regenerative and inclusive technology use, and to encourage organisations to take up their social responsibility and contribute to sustainable technological development.

Five core values outlined in the charta:

- Biodiversity, environmental sustainability, carbon-neutrality: limit the impact of digital tools on the environment
- Social inclusion: development of technology and digital services accessible to all, inclusive and long-lasting
- Transparency, trustworthiness, ethics and responsibility
- Create digital tools that are measurable and readable: (e.g. respecting ethical norms when collecting, analysing and sharing data)
- Favor the emergence of new behaviours and values: social innovations in the definition of new digital systems and services.

Labelling process:

The label process is managed by the french agency LUCIE. Organisations have to go through six stages, accompanied by the agency LUCIE.

1. The first step entails participate in an education course on digital responsibility as an organisation, to then propose an action plan on how to implement changes within the organisation responsible and ethical digital
2. An independent labelisation committee then evaluates the action plan and assesses whether or not to grant the label.
3. 18 months later, organisations having obtained the label will be audited by external organisations (SGS and Bureau Veritas) on the implementation of the commitments outlined in the action plan.

Organisations pay a licensing fee for the label as well as for the audit by external partners. The label aims to be accessible and therefore costs depend on company size (between €1000-6000).

Target Groups

Private and public organisations

Regional Scope

France and Switzerland

Product of focus

Companies as a whole go through a process to assure responsible behaviour in the digital realm

Actor type

NPO

Country of origin

France

Self-assessment or third party

Self-assessment followed by third party evaluation and label attribution

Development process

Launched in France in 2019 by INR in partnership with Agence Lucie and the support of the following organisations: Agence de l'Environnement et de la Maîtrise de l'Energie (FR), Ministère de la Transition Écologique et Solidaire (FR), WWF, SGS, Fing, France Digitale, France iT, IDDRI

Level of maturity

Running (4)

Success factors and challenges

- Backing by important organisations that reinforce the label's trustworthiness.
- Demanding labelling process enhances credibility.
- Independent labellisation committees hold the power to grant the label, assuring the label's independence.
- The labellisation process is lengthy and involves several actors. This should ensure the implementation of a lasting and coherent digital responsibility policy within the organisation.

Comparison to the Digital Trust Label

Differences:

- Strong focus on environmental impact, biodiversity and sustainable development including consideration for example on CO_2 emissions.
- While it is called a label, organisations applying for it undergo a lengthy process in close exchange with the label supporting consulting agency with the aim to create overall sustainable change within the organisation as a whole.
- A sustainable business model for the cerfication process is well established including price discrimination according to organisation revenue.

12 App Privacy Labels

Apple

Brief Description

This initiative entails a series of easy-to-read labels for end users. Different icons indicate at a glance details on the privacy and data practices of an application in Apple's App Store. The indicators only show specific factual information (data type the app collects, whether the data is linked to the user or used to track them). It does not entail complex criteria to evaluate an app.

Apple requires App developers to provide information about some of their app's data collection practices on the product page in the App Store. Apple will not remove Apps that do not disclose this information, but Apps cannot be added to the Store nor be updated without this privacy information. These labels are among several policy initiatives by Apple that aim at offering users higher levels of privacy.

The move has been severely criticised by Facebook, whose Apps such as Whatsapp or Instagram need to show a very long list of labels, the Facebook App itself seemingly checking every single box of Apple's data gathering categories.

Consumer Reports offers a guide for users on how to read and to navigate the different labels.

Target Groups

App developers/companies

Regional Scope

International

Product of focus

Mobile apps featured in the App Store

Self-assessment or third party

Self-Assessment with random audit control by Apple developers. There is political concern that app developers may not be truthful and the US House of Congress has inquired as to how and when these audits for label accuracy happen.

Development process

Development internal to Apple

Level of maturity

Running (stage 4)
First announced in June 2020; Launched December 14, 2020

Actor type

Private

Country of origin

USA

Success factors and challenges

- Power of the organisation: Apple is under antitrust investigation for unfair market behaviour and the power it holds over apps through the App Store. Due to the wide use of Apple products throughout the world, and Apple's predominant market presence, most app developers are forced to apply the labels.
- If developers lie, they risk being banned from the App Store and face penalisation from the US Federal Trade Commission. These can be powerful incentives to stay honest.
- The labels offer transparency to users and create a strong precedent in the market. This initiative signals that user privacy and responsible data handling can become a competitive market advantage, when differences in an app's data policy are made visible.

- The initiative relies on self-declaration and therefore the honesty of app developers.
- The labels still need further work when applied in practice. Many apps have proven not to be truthful about their privacy policies (see Washington Post). This can create a false sense of security in users and may cause more harm than good.
- While the labels give a good overview of how much data an app collects on a user, the labels do not show what the data is being used for. This can be considered important information for empowering end users.
- A label must be as easy and clear as possible. Yet, the icon labels can get complicated for apps that collect a lot of data.
- Apple's labels do not put the information into context, so it is very difficult for the user to evaluate the product. Many users lack necessary levels of technological literacy to understand what the labels actually mean.

Being created by a major actor, the App Privacy Labels have undergone thorough criticism. The general spirit of the initiative however, has been widely applauded within the tech community as a step in the right direction. Hopes are high that it will help reform company behaviour. Forcing developers to more transparency can further help privacy experts and regulators to see if practices actually violate an app's privacy policies. The labels shift the burden of finding and translating information of an app's data handling away from users and places it on developers.

Comparison to the Digital Trust Label

Similarities:

- High reliance on empowering users through information and strengthening their capacity for taking well-informed decisions.

Differences:

- The App Privacy labels only show specific information. The products have not undergone an in-depth evaluation of their ethical merit and the labels do not offer any guarantee of quality, as opposed to the SDI label's nature and purpose.
- The Apple labels were developed purely internally, without the involvement of several stakeholder groups or government involvement.

5
ANALYSIS AND RECOMMENDATIONS

ANALYSIS

The analysis of the initiatives described above can be summarised as follows:

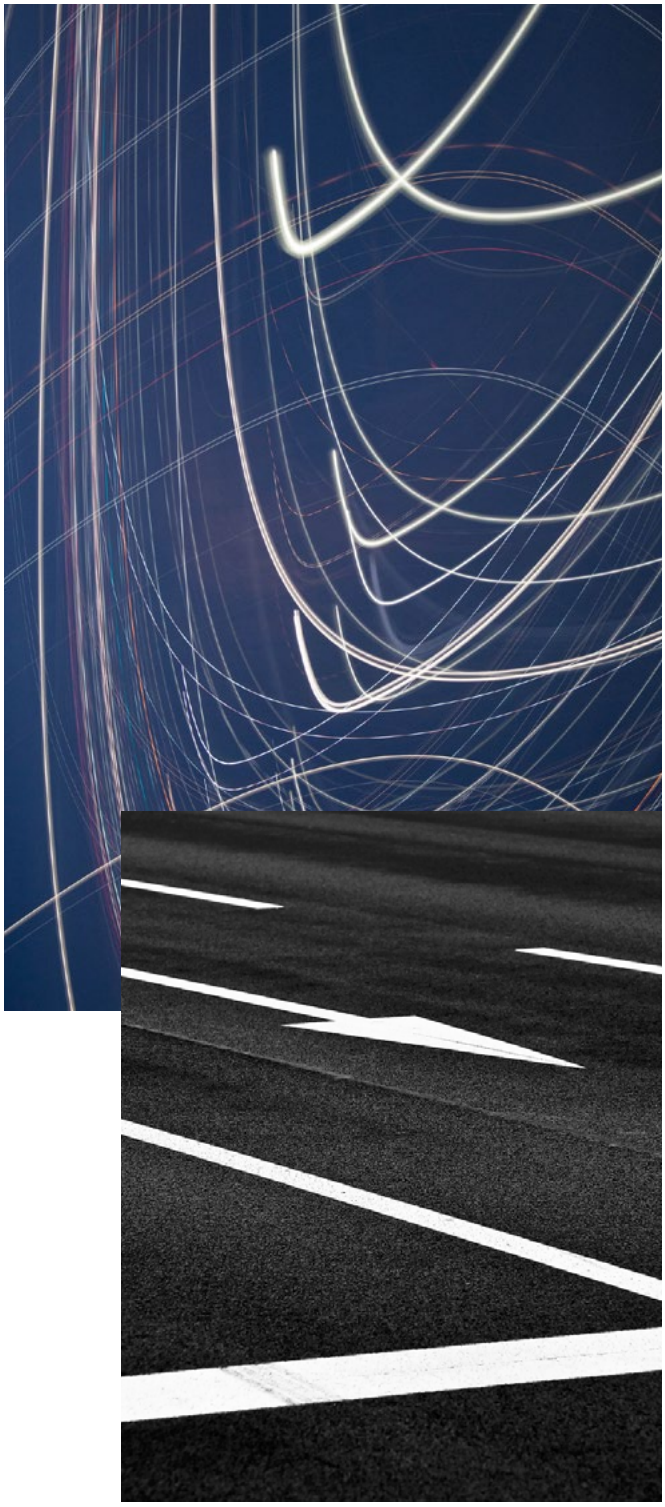
- Common objective: in their basic structure, all label initiatives share a common objective: providing audited and certified information about the features of a specific service or a company in order to enable users to make well-founded and well-informed decisions. The methods to achieve this objective are fairly diverse.

- Legitimacy is key: the content of the label is important, but more crucial is the governance structure surrounding the label (design, development, control, audit, certification of the label). The process foreseen to get the label, mainly the audit mechanism, is to be assessed with respect to its efficiency (incl. costs) but also with respect to its legitimacy. As a minimum, it is about transparent information on the governance structure of the label. Going further, it is about good practices in putting virtuous incentives in place and making sure that checks-and-balances guarantee the high-quality of the labeling process.

- Learning from project failures: as described above, some initiatives failed to reach their objectives and were abandoned. Their experiences show the numerous challenges on the way to a successful label, which must be taken into account. These challenges can be:

  1) Challenges linked to the nature of digital services: most digital services rely on a complex architecture and evolve in very dynamic environments. They develop and change sometimes rapidly. This is also true for users' and consumers' attitudes towards them. Hence, for organisations aiming at certifying particular services, it may represent a matter of complexity to track these changes and keep the pace of public expectations.
  2) Challenges linked to the business model of the labeling process (funding the project, without setting bad incentives in place): some organisations proved to consider themselves as too small and lacking adequate resources for providing the necessary infrastructure and services beyond label development.
  3) Challenges linked to the missing traction among actors behind initiatives, companies and consumers/users: developing a trustmark is a multi-year investment, where a brand has to be developed in which users trust, setting standards and creating a sustainable audit system. This entails continuous efforts to keep close ties with potential users and companies interested in having products labeled. Such multi-year efforts pertain in particular to building up the initiative's own reputation and thus require adequate long-term oriented governance structures and funding.

- International competition: the diversity of the more than 50 identified initiatives shows that the topic is of great interest to public institutions (incl. Academic institutions), but also private actors and NGO. The capacity to define what is relevant when it comes to trustworthy digital services is the object of an international competition about leveling the digital playing field. However, The fact that none of the initiatives has yet been able to implement an internationally successful label for digital services or organisations shows the complexity of the topic.



RECOMMENDATIONS

Towards a successful label

A label must fulfill the following criteria if it wants to have a chance at being successful. These criteria are no guarantee, but they are mandatory requirements:

- The label has to be known by its target-users. End-users of the label can be consumers or companies, both are considered users in a B2B or B2C constellation. Users must be able to recognise the label and associate it with a specific quality/ambition. This means that its name/visualisation must be easily remembered and that it must convey part of the quality/ambition of the label. This also means that one needs to think big to make the label known. This requires important financial resources and know-how on a regional or international scale (e.g. to organise an awareness-campaign).

- The label should be supported by a strong and well-known organisation. This means that its reputation and communication capacities prove for its credibility and trustworthiness, or that the label must be promoted by a coalition of smaller actors which have a strong presence in specific geographical areas.

- The label has to convey a general message. Buying a piece of food with the Fairtrade label gives consumers a sense that what they buy is fairer than goods without the label. Firstly, a label should give a general impression (more fairness e.g.). It does not deliver specific technical knowledge. A tiny minority of food-buyers actually know which fairness criteria are being assessed. But the important message is the general one: buying this piece of food is a seal of quality with respect to social and environmental conditions of production. The details and complexity are handled in the background. A mechanism is possible which makes additional information available on demand, but not in the primary design of the label. Secondly, the label is not a guarantee of perfection. A label is more often relational: it signals that X is better/fairer/more just than other options. It gives relevant information to the consumer who is in the position to make a choice. This point is crucial in digital matters, as often the problem resides in the lack of competition among service-providers (important actors exercising a monopoly). For that type of situation, the label cannot contribute much to making the position of users better.

- The governance of the labeling body has to be legitimate. The labeling process has to be perfectly clear and understandable to outsiders. It must be possible to assess whether the commercial and financial incentives set in place are compatible with the promises made by the label. Key steps of the labeling process are the audit by a third-party (which third-party is acceptable, under which conditions, for which price) and the ongoing assessment of the fulfilment of the criteria foreseen by the label (under which conditions must the service provider be assessed again)..

- The way the label organisation is funded needs to be transparent and understandable for outsiders. Profit-oriented labels are in tension with the broader goal pursued by the label (e.g. ethics in the digital realm or fairness in production of agricultural products). For users, it is not clear whether a label pursues this general goal or profit. This does not exclude that the label organisation raises fees in order to 1) fund the further development of the label and/or to 2) invest in other projects aiming at the same broader goal. But the purpose of the funding is to be understandable and consistent. This consistency is part of the legitimacy required for a successful label.

A successful label might be a transition

It is key to understand that even a successful label remains connected to a specific time and a given context, and that it can contribute to inspiring further development both of the label itself and regulatory frameworks with which it interacts. A label, therefore, is very likely to be part of a transition. If compared to soft law, the functions of a label could be described as the following. The label crystallises the expectations of users and companies at a certain time X. It provides an object of debate and negotiations for these expectations among different stakeholders. As for soft law, it means that a successful label could be the forerunner of distinct types of subsequent measures. Firstly, a national or regional label can be geographically extended through new partnerships; its scope widens, which could cause the label's content to evolve. Secondly, a label can be used as inspiration and resource for an international standardisation process led by ISO. This requires support by a national point of contact (e.g. Swiss Association for Standardisation). Thirdly, a label can be used as inspiration and resource for regulatory or legislative projects at a national, regional or international level.

Lessons learnt for Switzerland

Switzerland could play an interesting role if it aims at achieving the broad vision shared by the majority of initiatives addressed in this report: go beyond declarations about ethical values and principles and focus on how to make them become a tangible reality in society. The role which Switzerland could play needs to evolve with the dynamic situation described in this report.

In the current phase of development, several initiatives are emerging. An important number of them will be abandoned, while some of them might come out as regional or international initiatives strong enough to be recognised by consumers. For a country like Switzerland, the time is not rife to choose and support a single specific initiative. It should rather make sure that the current phase of intense activities does not last too long and that it remains productive with respect to the overall vision pursued. Switzerland has a reputation of neutrality and is in high esteem for its expertise in mediation. Such a position is particularly needed in the face of the growing cleft between great powers in a race for digital and especially AI supremacy.

For now, Switzerland might:

- Offer its networks, resources and expertise to reinforce formal and informal coordination among the different initiatives. The objective is to play a constructive, co-development moderation role. Geneva, as hub for the governance of digital technologies, is best placed to fulfil this mission. It can be supported by scientific diplomacy actors (such as the EPFL, the Swiss Universities, GESDA), but also by important, transnational companies present in Switzerland. Concretely, this could mean:

  a) the organisation of a conference in Geneva inviting the most promising initiatives to connect with the ambition of creating/contributing to an ecosystem
  b) the invitation of key people of the different initiatives to follow a one-week "deep-dive" in the Geneva ecosystem (with the ambition to create a network of informal Swiss ambassadors)
  c) the support of specific organisation bringing the relevant expertise and scope of activities in realising this coordination effort
  d) enabling exchange of practices and use of synergies. Switzerland can be the platform, where the dots are connected and the initiatives can learn from each other.
  e) moving forward an own Swiss pilot or partner with like minded or promising initiatives to launch a pilot.

For Switzerland, the support of these initiatives in the form of a coordination-platform is a very interesting way to position itself at the heart of the debate on digital technologies and their global governance.

- Improve the inclusivity and diversity of the initiatives. For the time being, the vast majority of initiatives are situated in Europe and Northern America. There is firstly an interest in making these labels more efficient by securing the possibility of an international uptake. Many of these initiatives claim a global scope on paper, yet are clearly tied to Western concepts and tested on a Western public. This may hamper the likelihood that the initiative will be accepted or work outside of its region of origin. For Switzerland, this would require to activate networks making this inclusivity possible. That could be through organisations present in Switzerland and able and willing to contribute to enriching the development process of the labels (e.g. Confucius Institute in Geneva). Similarly, it could mean networks working across the globe who are able to open a conversation which proves more inclusive. We could think of the Swissnex network, the Switzerland Global Enterprise network, but also of privately led networks such as Seedstar (startups network in emerging markets). In addition, Switzerland could proactively invite promising organisations active in Asia, Africa and South-America to visit Geneva and get a "deep-dive" into the Geneva ecosystem.

For Switzerland, this support is at the junction of the interest to fulfil the broader vision of ethics and digital technologies, but also justice commitments to improve regional and global governance processes by making them more inclusive. This is the opportunity to show the capacity of Switzerland to combine economic, academic, political, civil society networks.

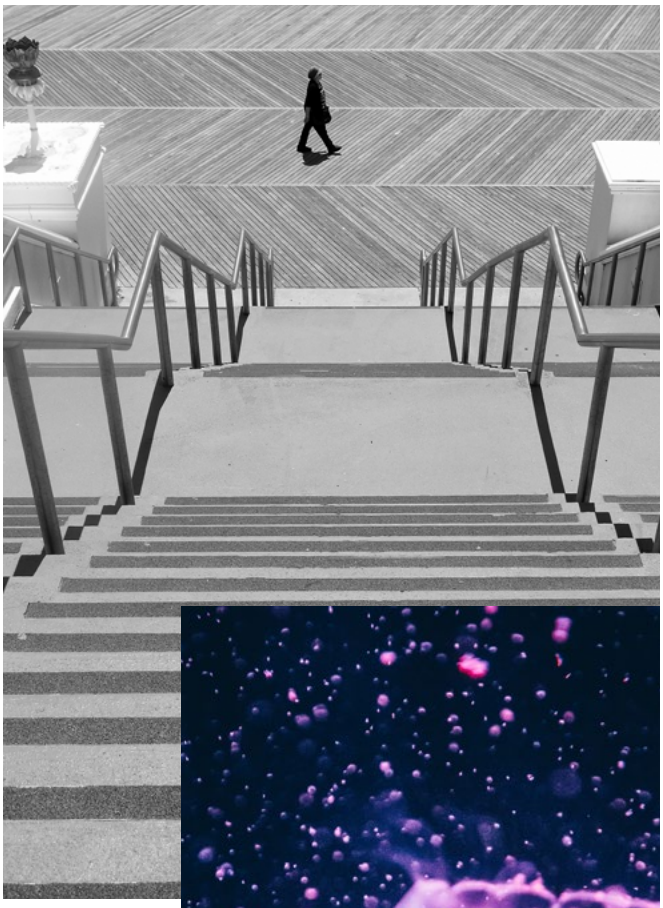
- Support the general vision and integrate promising initiatives into existing multilateral efforts to tackle ethical issues and digital technologies. Through its networks in the UN and the ecosystem gathered in Geneva, Switzerland could help connect general political ambitions at the UN-level with operationalisation initiatives. As explained in the previous analysis, the objective is to complement high-level declarations and commitments with concrete projects able to transform the way consumers and companies deal with digital services.

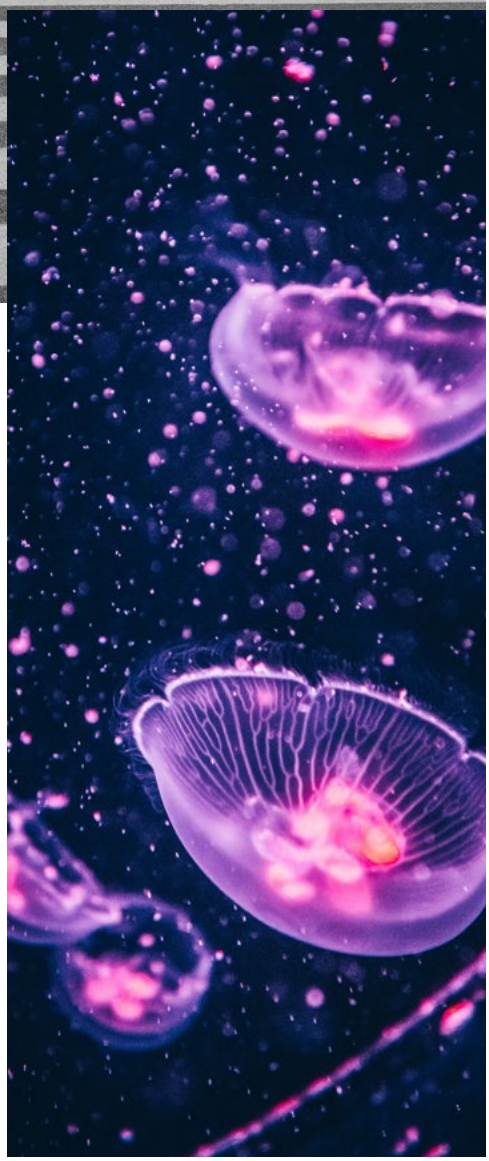
This support promotes the brand "Switzerland" as an ecosystem at the cutting edge of innovation, driving responsible governance, strengthening business while championing human rights in the digital realm.

In a mid-term perspective - when it comes to crystallise few promising initiatives - Switzerland is best placed to promote:

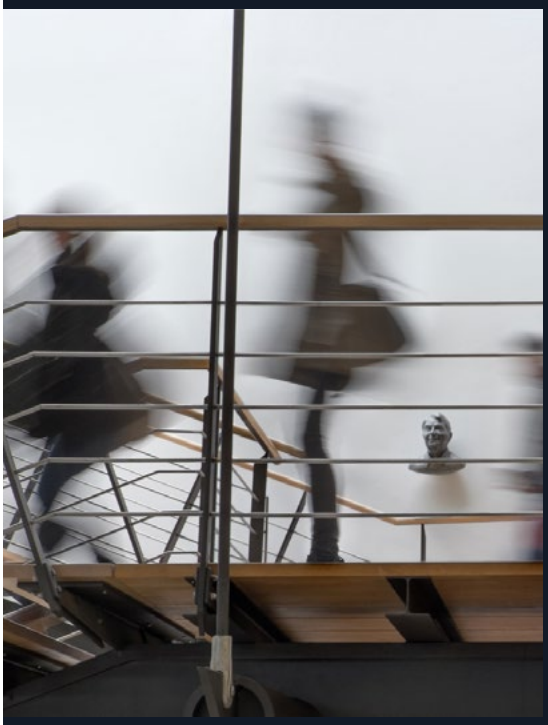
- A human-rights approach to labels for digital services: labels supported should be conceived in a way that improves well-being and capacity for autonomous choices.

- An inclusive approach, both from the types of actors involved (academia, business, civil society), and from the geographical distribution of the stakeholders.





The present report was elaborated by the Swiss Digital Initiative as part of the National Strategy for Protection of Switzerland against Cyber-Risks (NCS)



Authors

Swiss Digital Initiative
Niniane Paeffgen and Simon Perdrisat
c/o Campus Biotech
Chemin des Mines 9
1202 Geneva

ethix - Lab for Innovation Ethics
Zora Muriel Schmid and Johan Rochel
Zweierstrasse 100
8003 Zürich

Geneva, March 2021

ABOUT THE DIGITAL TRUST LABEL

The Swiss Digital Initiative (SDI) is working on developing the first Digital Trust Label that denotes the trustworthiness of a digital service in clear, visual and plain non-technical language.

The Digital Trust Label will be a combination of bio label and nutrition fact table for the digital world: it shows that mandatory criteria are fulfilled by a digital service, while at the same time giving users more information and transparency about four dimensions of the digital service: Security, Data Protection, Reliability of a service and Fair User Management (making transparent automated decision-making).

Eight test partners from the public and private sector are involved in the project: Axa, Booking.com, Canton Vaud, Credit Suisse, IBM Switzerland, SBB, Swiss Re and Swisscom.

Digital Trust cannot be defined by one actor alone, but can only be the result of the close collaboration of all relevant actors: academia, civil society, consumer protection, the private and public sector. This is why the SDI involved all relevant stakeholders in the development and made the criteria and development process as transparent as possible. The Label is understood as an ongoing and collaborative effort for strengthening transparency, trustworthiness and understandability of digital applications.

ABOUT THE SWISS DIGITAL INITIATIVE

The Swiss Digital Initiative aims to bridge the gap between principles and practice and to safeguard ethical standards in the digital world through specific projects. It brings together academia, government, civil society and business to find solutions to strengthen trust in digital technologies and in the actors involved in the ongoing digital transformation. The initiative has a global focus and is headquartered in Geneva, Switzerland. It was initiated by the association digitalswitzerland and under the patronage of Federal Councillor Ueli Maurer.