

Summary

- CSC supports the data strategy and its objectives. Developing data economy is crucial for Europe's competitiveness and contributes significantly to all the strategic priorities of the European Commission, including the Green Deal.
- Sufficient funding for digital priorities must be secured in the EU budget. The funding levels of Horizon Europe, Digital Europe programme, CEF2 Digital and Erasmus+ should be as close to the Commission's original May 2018 MFF proposal as possible.
- In building data infrastructures, existing infrastructures and work must be leveraged, to ensure interoperability and continuity.
- High-performance computing, data management, AI and connectivity networks must be developed in convergence to create synergetic ecosystems.
- Sustainable business models for data infrastructures must be developed, to ensure their continuation after the initial funding phase.
- Data policies such as FAIR principles must be promoted, and data movement across borders and sectors must be enabled.
- The European Interoperability Framework must be applied when developing data spaces.
- Legal barriers for the free movement of data must be identified and removed. For example, the DSM directive on copyright must be revised to enable a European data economy.
- Special attention must be given to better usage of health data. Its secondary use in research has massive potential, as well as better cross-border exchange of health information.
- In decisions regarding placement of data centers, smart specialisation strategies and incentives for environmental sustainability and carbon-neutrality must be applied, to ensure cost-efficiency at European level, as well as contributing to the targets of the Green Deal.
- Data skills are crucial for Europe's competitiveness. EU needs urgent investments in broad skills and competence development across all sectors and levels, including teachers.

CSC sees the European strategy for data as a welcomed and instrumental part of the Commission's strategic priority "a Europe fit for the digital age". Data economy and digital technologies have an enormous potential to transform the economy and society for the better. Data is one of the key elements in sustainable economic growth, and therefore its efficient sharing and usage is crucial also for the objectives of the European Green Deal. Digitalisation, research, innovations and the targets of the Green Deal are tightly intertwined and must be seen as an entity. Digitalisation is not mainly about technology, but it is an entirely new paradigm that requires change in policies, organisational structures, processes and collaboration. It is important to recognize the ways in which data brings added value to the society.



CSC supports the idea of creating specific, high-value data spaces with the goal of using them as an interoperable European data space. It is critical to ensure, that the different sectors will not become separate silos. Data is cross-cutting by nature, and the free flow of data between sectors and borders is a key element in building a data based economy.

In order to compete on a global scale, EU needs to strengthen its own data management capabilities and skills. Other key capacities, such as world-class high-performance computing capacity, reliable networking connections, and skills development must be developed coherently as a part of the data infrastructures. To reap the full benefits of data, it needs to be combined with computing and analysis capacities. The EU must also invest in the strengthening of high-speed digital connections, not only intra-EU, but also between Europe and strategically important regions such as North America and Asia.

This basic capacity building requires sufficient funding in order to take critical steps forward. Thus, the proposed funding level of Digital Europe Programme, Horizon Europe and Connecting Europe Facility must be secured in the negotiations of the multiannual financial framework. Sustainable business models coupled with life-cycle management must be developed for the data infrastructures, in order to ensure their continuation.

In the implementation phase, it is crucial to ensure that the data spaces truly become interoperable in the future. The data spaces should be built on top of the existing infrastructures and initiatives, such as European Open Science Cloud¹. Other good examples of work that has been done to create functioning infrastructure for data are initiatives such as Public Sector Information directive and INSPIRE directive. The idea of federated data infrastructures² should be implemented wherever it is found suitable – thus, when creating European infrastructures, it should be preferred to renew and connect existing ones and make them interoperable, instead of building completely new ones in areas where there already are functioning data infrastructures.

In addition to data infrastructures, a truly functional data economy needs data policies and practices that enhance the capture, management and usage of data. The Commission must promote policies that support consistency with the FAIR principles (findable, accessible, interoperable and re-usable).³ Especially important is the interoperability of data. Interoperability and movement of data must be ensured between sectors (research, business and public) and between the data spaces. Interoperability must be systematically promoted in legal, organisational, semantic and technical level, according to the European Interoperability Framework.⁴

From the policy point of view, legal interoperability is of utmost importance. CSC supports the Commission's view that increasing Europe's share of data economy requires an attractive policy

¹ <https://ec.europa.eu/digital-single-market/en/european-open-science-cloud>

² <https://www.knowledge-exchange.info/event/federated-rd-infrastructure>

³ <http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>

⁴ https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf



environment. Before introducing any new legislation, the implications on data economy must be carefully analysed. Existing legislation must be screened and any legal barriers to the movement and reuse of data should be removed unless found absolutely necessary to protect personal integrity and privacy. Currently, not all legislation is contributing to the proper functioning of the data economy. Some articles of the Digital Single Market ('copyright') directive are clearly harmful for the development of the data economy and AI. For example, whereas the article 3 of the copyright directive gives a mandatory exception for text and data mining (TDM) for research purposes, the article 4 allows rightholders to opt out of the exemption if the TDM is done for commercial purposes. This puts commercial users of TDM, such as AI developers, start-ups and other innovators, at a competitive disadvantage compared to the United States, where TDM is deemed fair use. CSC asks the Commission to promptly revise the copyright directive and especially its article 4, giving a full exemption also for commercial users of TDM.

Another important aspect is semantic interoperability. Instructions, data policies, quality of metadata and services that support data management all affect findability and reusability of data. Common data models, practices, standards and skills are needed to make data move between the member states, actors and sectors. Technological solutions and planning must be based on end-users' perspective. A good way to incorporate end-users' perspective is co-design model in service development. In developing tools and policies for interoperability, it is important to take into account the global aspect, ensuring that all policies comply with global standards and practises. For example, the tools developed for data management in the global Research Data Alliance must be taken into account and leveraged wherever possible. In this way, the EU can avoid creating non-interoperable layers and duplicate work.

There are a few key thematic areas where we see especially great potential in better data flow. First of all, openness of public sector data should still be increased. Data that is collected by publicly funded actors must create benefit for the whole society, which is why the data must be open and available for everyone to use. In this way, the public sector data can be used for building sustainable public data registries, which will enable stronger data based decision-making and policy implementation across all sectors of society. This will make Europe a strong global player, as the European data can be used for building European competences. CSC supports the national implementation of the Open Data Directive and urges the Commission to ensure that the directive is properly implemented in all member states.

Health data is another type of data that has a very high potential to work for the common good. Better movement and usage of health data could improve the health of Europeans and contribute to the competitiveness of European industry. Covid-19 crisis has only highlighted the importance of anonymised and aggregated health data. However, the situation has also showed the limits of data sharing between the member states. In addition to better cross-border exchange of health data, its secondary use for research purposes must be systematically promoted. Research use of health data must be done in a secure manner, protecting individuals' rights to privacy. A good example of how to implement the secondary use of health data in legislation can be found in Finland.

A third crucial factor for the development of data economy, after proper data infrastructures and an attractive policy environment, is skills development. Europe needs both more data scientists and better general knowledge of data's role in modern society. The role of data in research is growing, which means

that data science should be better incorporated in all university programmes, as well as already in earlier stages of education. Hundreds of thousands of Europeans will need reskilling in the coming years. Therefore, life-long learning should be developed and made easier for everyone. Understanding the basics of data science would contribute to the needs of the future work life, but citizens also need to have certain basic skills to benefit from the data economy. Here the digitalisation of education can act as an enabler; one example being the Elements of AI MOOC provided by the University of Helsinki and Reaktor⁵. The needs of data economy and digitalisation of education must play a major role in the new Erasmus+ programme.

Finally, in order to reach the ambitious targets of the EU for climate neutrality, measures must be taken also in the digitalisation and data area. In addition, regarding the current financial challenges that EU is facing, it is crucial to consider the cost-efficiency of investments. Thus, smart specialisation strategies must be applied: on a European scale, common, pan-European structures would be located in regions where they bring most added-value with lowest cost for European taxpayers and with lowest possible environmental burden. For example, in choosing locations for data centers, which by default consume vast amounts of energy, the criteria must support eco-, energy- and cost-efficiency and cost-efficiency. For data centers, energy costs are the largest run-time cost element (40–60%) that depend primarily on energy prices, secondarily on nature and climate conditions, and thirdly on energy effective technologies. The variation of energy prices within the EU is big: the highest price might be even three times as high as the lowest.

All in all, the EU needs to take measures in a broad range of areas and levels, in order to build a globally competitive data economy. Europe must build its own competence and a strong infrastructure base for data, and support this with a regulatory framework that enables a flourishing data economy.

⁵ <https://course.elementsofai.com/>