Contribution to the European Commission's Call for evidence on a European Data Union Strategy

Arcep – July 2025

Data is at the core of the development of AI. In the past five years, AI use-cases have developed tremendously, with conversational models at the forefront. This development took place thanks to accessible datasets, such as the Common Crawl, as well as significant investments into data annotation, cleaning and pre-processing by AI stakeholders. Today, large language models (LLMs) are able to solve college-grade problems and tasks, with performances rising every month; while LLM-based agents are replacing older virtual assistants. In the meantime, AI can empower industrial use cases (e.g., anomaly detection, molecule discovery, medical diagnosis assistance). In order for the European economy to reach its full potential thanks to AI and to reach the Digital Decade targets, access to large amounts of high-quality data is crucial[1].

Such views were first expressed by Arcep in its answer to the European Commission's public consultation on the competition in generative AI: alongside with cloud computing power, talents and capital, data is a critical input for the development of AI.

As of today, Arcep believes that the full potential of European data still remains untapped, as availability of data for use and reuse is still quite low: data sharing ecosystems are still in the process of emerging, while some data remains in the hands of few players, with a lack of effective control for businesses and individuals.

The 2020 European data strategy laid the foundation of a regulatory and investment framework to tackle these issues, and contribute to the emergence of an interconnected single market for data. With the upcoming entry into application the Data Act (DA), completing the regulation laid down by the Data Governance Act (DGA) and sectoral regulations, their combined effect on data flow and the data economy should be visible in the short to medium term, fostering innovation and ensuring fair competition in data economy, echoing Arcep's regulatory approach and ambition.

In light of the ever-growing needs for data to develop and reap the benefits of AI, Arcep warmly welcomes the initiative of the Commission for a new European data union strategy, notably aiming at enabling scaled-up data use and availability. The principles already laid down by the DGA and the DA must be implemented: they provide a stable and predictable framework for data exchanges. Based on our experience as regulator of the DGA, we believe that minor adjustments may be necessary to ensure the economic viability of data intermediation service providers. Arcep is also in favour of simplification measures for businesses, notably through legal consolidation but without undermining the objectives of predictable regulation of data sharing.

[1] As highlighted by the European Commission's Communication on boosting startups and innovation in trustworthy artificial intelligence, as well as the EU AI Strategy (2018).

1 The regulatory framework for data access and data sharing, designed to free data, notably for artificial intelligence, is starting to take effect

Arcep is the competent authority for the enforcement of the provisions related to data intermediation services providers of the DGA in France, and has been given a mandate to regulate cloud services in anticipation of the Data Act's entry into force. This allowed Arcep to gather significant feedback from the cloud and data ecosystems on remaining barriers to data access and data sharing.

Stakeholders have reported that barriers to an efficient and open European data market still persist to this day: first, **there is a continuing lack of trust in data sharing agreements and on the potential illegitimate use of data**, second, **rights to access data are scarce** and third, **there is uncertainty about the value of data**, which hardly incentivises data sharing. Furthermore, concerns about digital sovereignty and international data transfers can also hinder trust and in turn, data reuse.

The previous data strategy, adopted in 2020, set out to remedy the previously mentioned barriers:

- with the DGA's Data Intermediation Service Provider (DISP) status as a means to increase trust in the data economy through dedicated neutral actors tasked with the sharing of data;
- by providing a governance framework and common technical tools with data spaces, meant to accelerate data exchanges, and allow for AI solutions to emerge;
- with the Data Act's access obligations aiming to increase data availability and circulation, and the effective interoperability it will provide through standardisation.

Most players think that the Data Act's imminent entry into application will allow many European businesses to better access, share and monetise their data.

First, the Data Act will empower users with more choice and allow European SMEs to compete with large actors for the provision of related services.

Second, it will protect European companies, including SMEs and start-ups, from unfair contractual practices with regards to data sharing.

Third, it will increase trust and interoperability within data spaces through standardisation.

More generally, data sharing obligations have encouraged, and will continue to encourage, data flow in the EU.

Public sector data and sectoral data can benefit from such obligations: the Open Data Directive has increased the availability of high-value datasets, local and regional mobility authorities are using data regulations to make available the data needed to organise intermodal transport services and for further innovative reuse, and the upcoming European Health Data Space will allow start-ups to develop innovative health solutions. **Such sharing obligations and access rights will broaden the scope of accessible data and put larger volumes of high-quality data in circulation. Therefore, Arcep believes that the effectivity of data access rights, in particular those from the Data Act, will be crucial to train innovative AI models, including in the industrial sector.**

In turn, these will amplify the need for trusted governance frameworks, such as data spaces, and neutral intermediaries, notably in the form of DISPs. To this end, **Arcep supports the creation of new data spaces and more broadly data sharing infrastructures when there is a relevant need**, and acknowledges the importance of **strengthening both the governance and interconnection of existing data spaces**. Both elements are essential to ensure the effectiveness of this approach. On that note, Arcep welcomes the efforts that have been made by the European Commission and stakeholders to have European data spaces emerge.

Overall, Arcep believes that the framework laid down by the previous European data strategy constitutes a solid foundation, whose efficient enforcement can unlock massive amount of data, and ensure a trustworthy environment for data sharing, strengthening the European position in AI.

2 New measures could be taken to foster data circulation and secure new data access rights, unlocking more data for AI and the EU economy

Leveraging the DGA framework and DISPs to set up AI Data Labs.

To address the specific needs of AI developers, a trustworthy and interoperable data governance, notably through interconnected data spaces – e.g. thanks to interoperable DISPs – could offer access to data from various sectors, and promote trust and cooperation by ensuring security or integrity of shared data, as well as the respect of sectorial rules. Furthermore, the needs of AI developers often go beyond accessing data, and also encompass data cleaning, enrichment, pseudonymisation or anonymisation, or the production of synthetic data. Subject to minor clarifications in the DGA, notably on the possibility for DISPs to enrich or improve datasets, these data governance structures could constitute the blueprint for the Data Labs proposed by the Commission in its AI Continent Action Plan.

Incentivising the development of interfaces to ease DISPs and other stakeholders' access to data.

Some data intermediaries report that they are sometimes confronted with technical difficulties when accessing data on behalf of users, e.g. when exercising portability rights, notably in the absence of interfaces. Such difficulties can hinder data flow, and prevent data intermediaries from realising their full potential in a neutral, secure and trusted manner. As such, for major data sources, being private or public, the development and provision of data access interfaces, such as APIs could be incentivised.

Extending existing cloud services portability rights to on-premises software to free more data.

Low portability can hinder the possibility for businesses to develop or use new services with their data. In particular, the traditional software sector sometimes still suffers from unjustified barriers to switching provider and porting data. This is believed by some stakeholders to be detrimental, as in many sectors (medicine, agriculture, etc.), cloud adoption is still ongoing, or professionals might prefer on-premises solutions. Not unlike cloud solutions, the barriers customers face when they try to switch or port their data can be economical, contractual, or technical. With the Data Act applying to all cloud services, the portability right could be extended to on-premises software, allowing more data to flow.

Helping compagnies assess the value of their data to further incentivise data sharing.

In the general case, some companies report that they primarily invest in raw data, either collected or bought, in order to transform it into high-quality datasets available for internal use cases or AI training. However, assessing the value of high-quality datasets can prove challenging, especially for SMEs and start-ups. This causes reluctancy over data being sold or shared, as the value created thanks to data pooling from various sources can also prove hard to assess. Outside of the European Union, some measures were taken to incentivise companies to better evaluate the value of its data[2]. In light of these evolutions, the development of such options in Europe could help companies engage in innovative projects requiring joint uses of scattered data or the production of new high-quality data.

[2] E.g. by allowing them to include data in their balance sheets, or by setting up third parties in charge of assessing their value and possible scenarios for distributing the value generated by data sharing among contributors. See for example *China's Data Strategy* - European Union Institute for Security Studies.

Ensuring contestability in the AI market through FRAND access to usage data from large players.

Arcep believes that open markets and contestability are key to contribute to the availability of European alternatives for digitalisation and to the EU digital strategic autonomy. In this regard, uneven access to data linked to the development of generative AI could constitute a new source of market foreclosure**.** General-purpose AI services can be either integrated into search engines or used as stand-alone applications. These services are continuously improved thanks to user-provided data and feedback – the more data, the better. This further reinforces the concentration in the search engine market, and the same concentration dynamics that were witnessed with search engines may manifest in the stand-alone AI market, leading to a "winner takes all" situation. To ensure contestability in the general-purpose AI sector, FRAND conditions could be used to avoid concentration of such data.

3 Regulatory consolidation and streamlining could improve regulatory clarity, while maintaining stability in the recent data framework

In light of the many regulations applying to data, being general or sector-specific, consolidation of existing regulations could improve regulatory clarity. Some regulations, such as the DGA and the Free-Flow of Non-Personal Data Regulation (FFDR), are lesser-known, and the rules they contain could gain visibility through consolidation, for instance with the Data Act. This could also reduce regulatory uncertainty, highlighted by some stakeholders, linked to the interplay between these regulations, and avoid overlapping rules. Such consolidation also calls for a clear and ambitious governance, including cooperation mechanisms, and the existing EDIB could see its mandate and composition adapted (e.g. a two-layer governance based on a strategic level and a technical level for operational matters, and an institutionalised liaison with BEREC on cloud and AI topics in the context of the Data Act).

Through its experience in the telecommunication sector, Arcep believes that regulatory stability is key, particularly to stimulate investments. For data sharing and access, the DGA is still recent and hard to evaluate thoroughly, while the Data Act will soon enter into application. The combined effects of the DGA and the Data Act are yet to come, both for the industry and public bodies. Therefore, we think that only cautious and targeted adjustments should be considered with regard to these texts.

In view of the challenges raised by data sharing and access for AI, and the safeguards needed in this context, we consider that the underlying objectives of the DGA, both of fostering trusted data sharing and avoiding conflicts of interest through market design and regulation, are still relevant. As such, **Arcep considers that maintaining the mandatory regime for data intermediation services could prove beneficial,** notably to give more time for the data sharing market to develop, while sustaining a trustworthy environment for data exchange and fostering innovation. Furthermore, this would avoid the concerns raised by DISPs of potential competition distortion between regulated and unregulated data intermediaries that might arise from a voluntary regime. **In the meantime, labelled DISPs and others could be better demarcated**, for example through an improved publicity for labelled DISPs, to provide a clearer signal of choice to the market.

Furthermore, to reduce administrative burden, some obligations applying to DISPs could be streamlined with other regulations: notably, the security of data could be covered by a horizontal framework, such as the NIS 2 directive.

In addition, further clarifications to the obligations to give legal certainty could be considered because stakeholders mentioned the difficulties for DISPs to emerge and be economically viable due to the current regime being too restrictive. In particular, the perimeter of supplementary services that DISPs could provide could be clarified, and explicitly include services meant to improve data usability, such as data analysis, data curation, pre-processing or assessing data value, as long as they are performed on data meant to be shared through intermediation services, at the explicit request of the

data holder, or on data acquired through such services, at the explicit request of its data user. This could in turn favour the pro-innovation objective of the DGA, while maintaining the regulatory focus on potential conflicts of interest prejudicial to the trust of the intermediated parties.

Finally, regulatory sandboxes, more suited to market testing phases, could notably support SMEs and start-ups, given that the development of innovative data sharing services can involve significant risks that are hard to predict, while market outlooks and economic viability are uncertain.