THE TRINITY OF TRUSTED CLOUD, DATA AND AI AS GATEWAY TO EU'S COMPETITIVENESS

Within the Gaia-X Summit 2023, organised under the auspices of the Spanish Presidency of the Council of the European Union, the Spanish State Secretariat for Digitalization and Artificial Intelligence, alongside the Spanish National Data Office, puts forward this statement as an informal call-to-action for a more coherent and coordinated response within the areas of cloud, data and Artificial Intelligence across the different initiatives, programmes, and projects that are taking place both at the European Union and Member States level. This call for action is open for adherence to those stakeholders within the aforementioned disciplines.

It is well known that the 4th Industrial Revolution has made data a critical asset. This is because the consolidation of heterogeneous data sources permits better modelling of relationships, to improve processes, and to innovate and create new business models. A key feature of data is that it is a reusable asset, that can be capitalized by many actors at the same time without any loss of value. This peculiar characteristic underpins the global data economy, allowing to benefit from the opportunities of exchange and reuse of data.

Data exchange provides two relevant benefits: on the one hand, they increase the general availability of data (both in terms of volume and granularity); and on the other, they serve to produce higher quality datasets. This is precisely the value proposition of a flexible, dynamic and highly scalable data economy: access to more and better data (from heterogenous yet complementary domains) improves the quality of products and services within value chains. For this reason, the European Commission's European Strategy for Data seeks to efficiently calibrate the data markets dynamics alongside the tech capabilities of the cloud, whose strength lies in providing flexibility. This way one can more easily achieve interoperability across IT systems, a key element to developing distributed data ecosystems.

However, governing this type of ecosystems is not exempt of challenges, one of them being how to articulate trust across participants and resources. For this very reason, the Gaia-X initiative emerges in 2019 to generate a conceptual and architectural model, including standards, for IaaS/ PaaS/ SaaS offerings that provide traceability and transparency in their operation, as well as composability and reversibility functionalities (often simplified as '*openess*'). We dub the near-future federated cloud services running in Europe under this model as *TrustedFedCloud*, as it underpins the regulatory compliance efforts needed to enable a sovereign and resilient data economy.

From a functional and technical perspective, this infrastructure also represents the optimal choice over which to deploy data spaces, the novel

paradigm by which the EU aims to achieve an efficient data exchange at scale. This is due to the federated nature of these cloud services being a facilitator of the dynamic interconnection across participants. If cloud has proven a breakthrough in terms of flexibility and effortlessness to users, a federated Cloud furnishes easy and seamless scalability, far more reaching that any monolithic offering. Additionally, its transparency and compliance-ready capabilities provide operational guarantees, key to business-continuity and digital sovereignty goals.

Thus, if we seek for a flexible, dynamic, and highly scalable data economy, the TrustedFedCloud is the optimal underlying model to build upon when defining business data exchange requirements and rules. This concept is not yet available in practice, but significant strides have already been made, and should remain a heavy focus for the European ecosystem. Among other reasons, and with data being the most important raw ingredient of the new and disruptive wave of artificial intelligence, there is a clear need for improved data quality and data governance in both the development and operationalisation of AI products.

This is pegged to recent advances in artificial intelligence being driven by Deep Neural Networks. These are large computational models trained on enormous datasets, representing a competing approach to AI from that of traditional symbolic or deterministic models. There are 3 main ingredients behind the recent success of DNNs: algorithms, computational capacity, and access to large amounts of data. This allows us to define the following complementary paradigms for neural network-based AI development:

- **A neural network paradigm based on experimentation.** Since most of the algorithms employed commercially today are based on running the classical "backpropagation technique" over very complex architectures made up of different types of "neural layers", this scenario is based on the heuristic ideation of new architectural patterns. These get trained over the same datasets, retaining those that work best in practice. This trial-and-error model therefore implies heavy experimentation efforts, alongside the necessary funding.
- **A neural network paradigm based on computational capacity.** Once the aforementioned optimal architectures are found, a second paradigm for AI advancement relies on capitalising on them by expanding the size of their parameters, i.e., making the networks bigger. This will typically also involve using larger datasets to train over.

  This race for larger models and more data to feed them, implies a pressing need for computing power. This has resulted in larger as well as dedicated data centres, also coupled with innovative parallel computing architectures, when possible. This has driven a booming market for both the building of new data centres as well as the design and manufacture of larger, faster and more efficient chips.

  However, this paradigm remains very polluting, as heavily-loaded data centres often require large amounts of water for cooling, and with these matrix machines having significantly increased their CO2 footprint, following the recent popularization of chatbots.
- **A neural network paradigm based on high-quality data.** As an alternative to

continuously building larger and/or more exotic models, we vouch for a "European AI way" underpinning its success factor on the curation of higher quality data. Since data is the main ingredient behind neural network training, we posit that having access to better-parameterised datasets represents a feasible option that would not only save valuable time for researchers, but also avoid the environmental degradation resulting from the trial & error approach over progressively bigger architectures.

Furthermore, working with data rightly governed (according to European regulatory frameworks and industry's best practices) also represents a major value proposition in ensuring a compliant deployment of these advanced AI models.

We are nonetheless tasked with outperforming its many challenges. This is because in a future made up of large and heterogeneous volumes of datasets, these will be hiper-distributed, often even at the edge. This implies that to reach the full potential of the European Data Economy, critical mass will have to be achieved by surmounting the scales naturally found within the context of data exchange across many and differing stakeholders. One can find that the confederated and thematic nature of these stakeholder networks leads to a large degree of dispersion, with different initiatives and projects taking place across individual sectors, each often bounded by their own standards, specific datasets, technologies and business and regulatory frameworks, thus leading to the creation of highly specialized siloes.

In this sense, we are very optimistic about the initiatives, programmes and projects already in-flight at the European-level, Member State-level and even industry-driven. There are too many to list exhaustively, but some of these include the aforementioned Gaia-X, the new efficient and ultrafast cloud services of the IPCEI-CIS, the European Commission-driven data spaces, the smart middleware for cloud-to-edge federation, the cloud marketplace and rulebook, the legal structural framework provided by EDICs for the sustainability of multi-country digital projects, and even the plans for edge computing and the APIfication of networks recently announced.

While there is no doubt of the merits that a vibrant and pervasive data economy can bring to our Union, and positive steps forward that have been taken, we nonetheless recognise that a cohesive coordination of the European ecosystem needs to be more widespread and effective. This effort shall drive the practical implementation of the concepts mentioned above by way of real industrial projects, monitor their deployment and efficient operation, and advance the standardisation of trustworthy cloud and data labels. The latter integrate governance rules, technical concepts, as well as the fundamental principles of data spaces, and are conducive to achieving high levels of cybersecurity, data protection, algorithmic transparency, and infrastructure portability and flexibility, all pillars of the European vision that we aim for.

For the Single Market for Data to become a reality, the processes of data exchange ought to follow a common governance framework. And while the novel set of digital and data legislation such as the Data Act or the Data Governance Act provides a northern star and is quickly becoming a global reference (with expected additions in the near future), there is still the need to articulate them in practical terms, for only in that way they will drive benefits like traceability and trust, and empower the collectivisation of the latent value in data.

In short, Europe has the capacity to become a world leader in a data-driven future based on the reliable and trustworthy processing of data, with data spaces in key industry verticals as input to an AI that will serve for international competitiveness and achieving productivity gains, all while fulfilling the European values and rights that define us, including the fostering of open strategic autonomy efforts.

"The distributed and heterogenous nature of the European Digital Economy can be fostered optimally by way of collaborative, interoperable and coordinated actions."

Signed by:

Carme Artigas
Secretary of State for Digitalisation and AI, Ministry of Economy and Digitalisation

Alberto Palomo
Chief Data Officer of the Government of Spain

Supported by:

Ulrich Ahle,
CEO, Gaia-X Aisbl

Ana García Robles
Secretary General, BDVA (Data, AI, and Robotics aisbl)

Lars Nagel
CEO, International Data Spaces Association

Prof. Dr. Boris Otto
Co-ordinator of the EU Data Spaces Support Centre and Director of Fraunhofer ISST

Daniel Sáez Domingo
President Gaia-X Hub Spain

Laura Olcina Puerto
President of the Federation of Technological Centres of Spain, FEDIT and President of the Executive Board of AI4ES

Pablo Coca Valdés
Executive Board, Red Cervera Cel.IA

Dolores Ordóñez
President of PLANETIC, Spanish Platform for the Adoption and Dissemination of ICT

Luis Pardo
General Director
AMETIC