TRUSTWORTHY DATA

DATA SHARING, KEY TO OUR STRATEGIC AUTONOMY

Olivier Dion and Arno Pons

THINK-TANK DIGITAL NEW DEAL

october 2022

DIGITAL NEW DEAL "TRUSTWORTHY DIGITAL COLLECTION"

“

EITHER WE SHARE DATA
AMONGST US,
OR IT REMAINS
CENTRALIZED BY BIGTECH.

SUMMARY

INTRODUCTION 3

I. THE DATA STRATEGY, A GERMANO-EUROPEAN IMPULSE
1.1. GENESIS OF A PROJECT INITIATED IN GERMANY 6
1.2. EUROPEANIZING THE AMBITION VIA GAIA-X 7

II. TRUST, THE PILLAR OF OUR STRATEGIC AUTONOMY
2.1. ESSENTIAL COMPONENTS OF THE TRUSTWORTHY ECOSYSTEM 12
2.2. ABOUT CLOUD ? 13
2.3. ABOUT AI ? 15
2.4. ABOUT DATA ? 16
2.5. VALUE CREATION WITH TRUSTWORTHY SHARING 17

III. OCCUPY THE STRATEGIC DATA SPACE FIELD
3.1. EUROPEAN CONTEXT 20
3.2. DATA SPACE CHARACTERISTICS 21
3.3. BENEFITS & USE CASES 22
3.4. DIFFERENT TYPES OF PLAYERS IN DATA SPACES 24
3.5. A STRONG NEED FOR GOVERNANCE 27

IV. DATA SHARING INFRATECH
4.1. DECENTRALIZING THE INFRASTRUCTURE WITH AN OVERLAY OF INTEROPERABILITY 33
4.2. INVOLVEMENT IN THE STANDARDIZATION EFFORT 35
4.3. BUILDING DATA SHARING COMMONS 36
4.4. THE INFRATECH ECOSYSTEM AS A NECESSITY 37

V. PROPOSITIONS FOR A NATIONAL PROGRAM 41

CONCLUSION 42
CONTRIBUTORS / ACKNOWLEDGEMENTS 43

INTRODUCTION

The European Union has always defined itself as an area of freedom of movement: free movement of goods, people, services, capital, and since 2018 data. This fifth freedom offers us the means to reconnect with the origins of digital technology, "sharing", to enable us to position ourselves on an equal footing with the American and Chinese giants. We have no real choice. Data has become the raw material of the 21st century economy: any service or product, in any field, if it wants to be competitive, must have a digital dimension based on the analysis or exploitation of data. Access to this data is therefore an enormous strategic power. Today, on the one hand, we have foreign giants that control gigantic pools of data, and on the other hand, a multitude of local actors, both public and private, that control small amounts of data, personal and non-personal, often unfit for use as they are. Either we share the data among ourselves, by equipping our players with a software infrastructure to share and exploit it, or it will continue to be centralized by a handful of giants, constantly capturing the value at our expense. It is this gamble that regulations are securing (GDPR yesterday, Data Governance Act and Data Act today), and that data sharing and data space initiatives (Gaia-X[1]) are deploying.

> *"We are defining today a truly European approach to data sharing. Our new regulation will enable trust and facilitate the flow of data across sectors and Member States while putting all those who generate data in the driving seat. With the ever-growing role of industrial data in our economy, Europe needs an open yet sovereign Single Market for data. Flanked by the right investments and **key infrastructures**, our regulation will help Europe become the world's number one data continent."*
>
> **Thierry Breton**, *Commissioner for Internal Market, November 2020*

This is a historical challenge for Europe. We have little time to win the crucial standardization battle against American and Chinese Big Tech, by setting up common standards (data models, legal and business models, etc.) for each sector,

[1] https://www.gaia-x.eu/

in particular within the Gaia-X framework, common standards (data, legal and business models interoperability, identity, etc.) in order not to depend on external technological solutions, and to be able to take an offensive approach to the issue of innovation.

Behind the question of a "sovereign data sharing infrastructure" lies the ability to develop new and innovative business models based on the opportunities created by this paradigm shift, in particular by resonating with the potential of Web3[2]. **This is a major challenge for the European Union, since it involves the construction of a decentralized Internet, in line with European values, as an alternative to the platforms of the Web giants.** This battle must absolutely be won by Europeans. And if possible, by those who defend the same definition of strategic autonomy as we do, especially since data is hosted in the information systems of European companies and administrations. This strategy must aim for independence by principle with respect to non-European actors, and a framework of trust by default, when we choose to rely on their technologies. This is why France must give itself the means to catch up with Germany, in order to rebalance and strengthen the Franco-German couple, and on the other hand to promote its political vision in the European context. Let there be no doubt about it, if we Europeans do not manage to guarantee our collective capacity to make the most of the data of our industries and services, others will take care of it, perhaps in a very qualitative manner, but to the exclusive benefit of their business models, and to the detriment of the European economy, growth and independence.

The objective of this position paper is to shed light on the debate on these complex issues. First of all, by reminding us what is at stake regarding data sharing for the full control of our technological and economical destiny, and then by explaining as precisely as possible the role of the data spaces in this strategy. For France, the challenge is to define a "French data strategy" in line with the European data strategy. Data must not become the great forgotten issue of our ambition for trustworthy digital technology. **We are convinced that only a strategist State can think and drive a coherent set of "Cloud, AI and Data" programs that will enable us to offer business players "end-to-end" solutions covering the entire digital trust value chain.**

[2] Web3 is a term used to sign the design of a decentralized web using blockchain technology - Wikipedia

GERMANY IS AIMING AT "DATA SOVEREIGNTY", WHEN FRANCE PREFERS TO THINK ABOUT "INFRASTRUCTURE SOVEREIGNTY".

I. THE DATA STRATEGY, A GERMANO-EUROPEAN IMPULSE

1.1 GENESIS OF A PROJECT INITIATED IN GERMANY

Once upon a time, there was Gaia-X... Let's take a moment to recall the genesis of this initiative to better understand our situation and to be able to outline the role that France can, and must, play.

Launched in February 2020, the European data strategy[3] is based on the notion of "data spaces": a novel concept in public policy. A data space[4] brings together public and private actors who want to share their data (personal and non-personal) to create value together, using a decentralized infrastructure and a common governance. **For many French players, this is a new thing they will discover with Gaia-X, at the end of 2019. For German players, it's an older story that began in 2015.**

The data space notion appeared in the academic research literature in the 2000s[5]. *"Data space systems offer services on data without requiring prior semantic integration. Unlike existing information integration systems, data space systems offer answers at best before semantic mappings are even provided to the system. Data spaces offer a "pay-as-you-go" approach to data management. Users (or administrators) of the system decide where and when it is worth investing more effort in identifying semantic relationships."*[6]

The data space is a catch-all concept that lends itself well to strategic considerations that seek to identify large data sets without characterizing them too much. The Germans used it to develop their "Industry 4.0"[7] project, which was designed to transform digitally their powerful industry[8]. The Fraunhofer- Gesellschaft[9], internationally renowned applied research institute, was charged with launching the initiative in October 2015. A few weeks earlier, the federal government had announced its support, and grants, for the Fraunhofer's work on the subject. On the same occasion, the creation of the Industrial Data Space Association (IDSA)[10] was also announced. The research is not only theoretical, by nature the Fraunhofer works closely with companies, and more than 60 use cases of the Industry data space had already been identified in about twenty companies.

Boris Otto, one of the researchers, published the first document on the Industry data space at the end of 2015 (followed in 2016 by a white paper describing the reference architecture of the German Industry data space). In January 2017, he became Director of Fraunhofer ISST, the institute's entity in charge of software and systems engineering, who will lead the initiative, and bring together a total of 12 Fraunhofer institutes.

Always supported by the federal government, Fraunhofer will further develop his proposal. The data space concept for industry remains a reference and will later be extended to many other activity areas. In 2016, the *'Industrial Data Space Association'* (IDSA) becomes *'International Data Space Association'*, marking the desire to extend the concept and to

[3] https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283
[4] "Data space" should not be confused with "Data hub" used by large companies. The latter do not have open competition governance, and are not connected with other initiatives within a European policy framework.
[5] https://en.wikipedia.org/wiki/Dataspaces
[6] A first tutorial on dataspaces by Michael Franklin, Alon Halevy, David Maier – August 2008; See also: "From databases to dataspaces: a new abstraction for information management" – December 2005
[7] https://fr.wikipedia.org/wiki/Industrie_4.0
[8] 'Overview of the industrial data space' - Oct 2015 - Fraunhofer- Boris Otto
[9] https://www.fraunhofer.de/en.html
[10] "The industrial data space: Architecture blueprint for digital sovereignty" - 2016:

internationalize it. Fraunhofer provides this to some European countries as well as to the Commission in the years 2016 and 2017[11]. However, the German Institute did not limit its promotion of IDSA inside Europe since it was presented at the *International Manufacturing Technology Show* (IMTS) in Chicago in September 2018.

Gradually, the European Commission will embrace the concept of data space to make it the core of its strategy. The subject will only be known to the Commission in 2016.

The German work of Fraunhofer is mentioned in a DG Connect working paper[14] as early as January 2017[15], but it is the Commission's April 2018 communication "Towards a common European data space"[16] that endorses the adoption of the data space concept before the data strategy formalizes it in 2020. In the meantime, the Commission has convened a dozen sectoral workshops to prefigure the corresponding data spaces. The German graft has taken hold, and data spaces have become the big thing in the European data strategy.

The European framework is therefore set. It is now a matter of ensuring the operational and industrial implementation by defining the necessary architectures and rules, especially regarding data governance, and by promoting data flow, since this is the "raison d'être" of data spaces. This is where Gaia-X comes in. It is up to the AISBL[14] association, its head office, to provide technical specifications and scheme governance. National hubs must help Gaia-X AISBL to support the emergence of sectoral data spaces, but at the same time must remain careful not to promote too much a national vision of data spaces, which would be for example a contradiction in multiple sectors such as energy, industry or finance.

1.2. EUROPEANIZING THE AMBITION VIA GAIA-X

To launch this movement, French players were asked by Germany to set up a team of "founding members" from Gaia-X and bring together large companies from both France and Germany. The Gaia-X project was initially published in French[15] at the Dortmund's Digital Gipfel in October 2019. The Franco-German team agreed upon a common position paper[16] that confirmed their partnership announced a few months earlier by the two ministers Peter Altmaier and Bruno Le Maire.

This Franco-German tandem is certainly the most obvious sign of a European commitment, and the best guarantee of its success. But it is Hubert Tardieu and Boris Otto, who were precursors of this Franco-German collaboration through the dialogue they initiated together, culminating in a common vision of Gaia-X. This collaboration was initiated, among other things, by joint programs like the one between Atos and Siemens. A real laboratory, with a budget of 320 million euros, which allowed to confront opposing visions of strategic autonomy: France favored the definition of digital sovereignty through tools, controlled by a minimum set of European players, while Germany preferred to talk about data sovereignty, without worrying about the origins of the shareholders responsible for it.

It is through these experiences that Gaia-X could be conceived, based on the consensus resulting from this work, and with the ambition to extend the operational framework to the European level for all sectors.

[11] *Presentation to Czech representatives*
[12] *Staff Working Document on the free flow of data and emerging issues of the European data economy* – January 2017
[13] *Commission Communication - Towards a European common data area* - April 2018
[14] legal form of international association for profit in Belgium
[15] *BMWi Doc : The Gaia-X Project - Network Data Infrastructure, cradle of a vital European ecosystem* – October 2019 doc in French
[16] *Franco-German Position paper on Gaia-X - 18 February 2020*
[17] *https://www.cigref.en/*

In France, since the end of 2019, Cigref[17] has been involved in mobilizing its members to participate in the constitutive Board of Directors of the Gaia-X AISBL. Major groups such as EDF and Safran have brought the needs of major French users of cloud services to this Board of Directors. In November 2020, the Secretary of State for Digital Transition and Electronic Communications, Cédric O, entrusted Cigref, by mission letter, with the task of structuring and leading the Gaia-X Hub France, in partnership with the Académie des technologies, the Systematic Paris-Région competitiveness cluster, and with the support of the DGE[18]. In 2021, the steering committee of the Gaia-X Hub France has been expanded to include INRIA[19], Numeum[20] and IMT[21]. The Gaia-X Hub France has thus organized, structured, and led 12 sectoral working groups, most of which are now part of the European dynamic supported by the Gaia-X AISBL.

On the German side, the flagship project illustrating the industrial implementation of data spaces is Catena-X[22], which gathers German car industry giants such as BMW, Mercedes, Volkswagen, and multiple subcontractors.

In six years, the initial vision of this project designed to guarantee the supremacy of the German industry in the digital era 4.0 had evolved to become a national public policy, then a European one, while leading to a project of worldwide ambition. (Japan and Korea joined Gaia-X in 2022). Today, the world's largest digital companies are members of Gaia-X. The history of this initiative remarkably illustrates the success of the German method, that of the concerted, structured, controlled construction of data spaces, under the aegis of the federal government's association with Fraunhofer, which, since 2015, has not ceased to bring together companies and associations around it, whether they are German, European or world class.

However, this European success (+20 countries) is neither shared, nor even truly recognized in France, where Gaia-X raises more questions than enthusiasm. Over the same period, a French strategy on data emerged with success, but with a different national orientation, in this case especially more oriented towards open data than data spaces. Both approaches are different but could perfectly be complementary.

[18] Direction générale des entreprises
[19] National Computer Science Research Institute
[20] These have no governance open to competition, and are not intended to be connected with other initiatives in a European policy frameworkSyntec Digital and Tech Merge in France
[21] Institut Mines-Télécom
[22] *https://catena-x.net/en*

OPEN DATA

THE FRENCH POLICY ON OPERATING AND OPENING PUBLIC DATA

Since the beginning of the year 2000, the French government has constantly promoted and made available the largest number of public data based on the remarkable wealth of data that public services continue to provide. This ambition constituted from 2014 a public policy entrusted to the State Data Administrator[23]. It was confirmed and reinforced by the Act for a Digital Republic[24] in 2016.

More recently, the report on "public data policy, algorithms, and source code" entrusted to MP Eric Bothorel, delivered at the end of 2020, revived the same policy enriched by a new chapter on the opening of code and algorithms. We can be proud of this national strategy, largely inspired by the "State as a Platform" concept[26], but that concept found little echo in other European countries such as German, since they do not have the same assets of public data. ***Highly focused on the state and its relationship with citizens/users, open data alone cannot be a lever for economic data policy for the private sector.***

Indeed, a data strategy cannot focus only on data that can be opened to all (open data does not cover personal or more sensitive data that are of great interest to economic sectors).

Although many companies joined the Gaia-X French Hub in its early days, too few are truly committed and the structure, which has no means or resources of its own, is struggling to move from a function of prefiguring sectoral data spaces to a function of supporting their operationalization. However, most of the working groups continue to work and develop their activities at the European level, thanks to the often-remarkable commitment of their initiators. However, due to a lack of resources, the Hub is struggling to reach its full potential. Moreover, this lack of resources does not allow the French Gaia-X Hub to get involved in the Gaia-X AISBL working groups, and to lead and coordinate its work at the national level.

Yet, elsewhere in Europe, Gaia-X rallies are multiplying. Nearly 20 countries have created their own national hub. The Brussels-based Gaia-X AISBL association has seen a steady increase in its membership and has been successfully growing since its creation in early 2021. Its rapprochement with the Commission has been achieved, which validates the indispensable partnership between the political level, DG Connect[27], and the operational level, provided by Gaia-X.

[23] The decree of September 16, 2014, creates a General Data Administrator (GDA) under the Prime Minister, attached to the Secretary General for the Modernization of Public Action (SGMAP).
[24] Law known as "Lemaire" named after the Secretary of State for Digital and Innovation Axelle Lemaire, promulgated on October 7, 2016
[25] https://www.government.fr/actualite/reporting-on-the-political-publicly-data-data-and-source-code-algorithms.
[26] The notion of the platform state emerged in Tim O'Reilly's essay "Government as a Platform" showing the value of the state in the digital era adopting platform strategies. Popularized concept from 2012 by Nicolas Colin and Henri Verdier, "L'Age de la multitude, entreprendre et gouverner après la révolution numérique".
[27] The Communication Networks, Content and Technology Directorate-General is the Commission service responsible for EU policy on the digital single market, internet security, and digital science and innovation.
[28] Amazon Web Service
[29] hyperscale: hyperscale is the method of sharing cloud computing server resources. A hyperscaler is an entity that offers this service like Scaleway, Switch, Alibaba, IBM, QTS, Digital Realty Trust, Equinix, Oracle, Facebook, Amazon Web Services, SAP, Microsoft or Google, etc. -

In the spring of 2022, the cloud and data news in France was dominated by the AWS[28] Summit which gathered 7000 people on April 12th. This popularity for the American hyperscalers[29] is not new. Between 2020 and 2021, many large francophone companies signed important contracts or partnerships with AWS, Microsoft Azure, or Google Cloud[30]. **Under the current legal conditions, none of them have any guarantee of any form of sovereignty over their data, nor any possibility of interoperability, reversibility, or portability. However, this is what Gaia-X[31] proposes to facilitate and secure data exchanges to create value.**

Will the French players remain permanently away from this powerful European movement of development of the data economy ?

> *"It is important that French companies, in particular the companies that own and use data, mobilize, with the support of the State, to put in place value-creating ecosystems between them, as many industries in Germany have done, like the automotive industry, among others. We call on these companies to join the Gaia-X working groups and bring their expertise in the face of very powerful companies outside Europe."*

Martine Gouriet, EDF,
Energy Data Space

[30] In 2021, Safran, Stelka-perseus-s, Total, SNCF, ENGIE entered into a partnership with AWS, Geometric Company and ArcelorMittal with Microsoft Azure, LVMH with Google Cloud. In 2020, Renault, Carrefour, Sanofi and Orange signed with Google Cloud.

[31] In 2021, under the drive of French stakeholders, AISBL Gaia-X has, for example, undertaken a process to translate its policy rules into labels that are applicable to cloud services.

NO FREEDOM
OF CIRCULATION
WITHOUT A
TRUSTWORTHY
FRAMEWORK.

II. TRUST, THE PILLAR OF OUR STRATEGIC AUTONOMY

2.1 ESSENTIAL COMPONENTS OF THE TRUSTWORTHY ECOSYSTEM

There can be no freedom of movement without a trustworthy framework

The European Union must continue its strategic effort to protect data. After personal data with the GDPR, our continent must now extend these guarantees to the protection of sensitive non-personal data in its industries and services. And this is a matter of urgency, because never in history have companies generated and manipulated so much data, raising the stakes of economic intelligence to a level never reached.

It should be recalled that the CJEU[32] ruling of July 16, 2020 on the invalidation of the Privacy Shield highlights section 702 of the Foreign Intelligence & Surveillance Act[33], which authorizes American intelligence agencies, including the NSA[34], to collect data from all "non-US persons", whether natural or legal persons, on a massive scale, a priori and without a warrant. However, a substantial part of the activity of intelligence agencies consists of carrying out intelligence of economic interest, for the benefit of American companies, within the framework of an intelligence plan coordinated at a federal level. Under this section 702 of FISA, and probably under other similar legislation such as the Patriot Act, American technology companies, including hyperscalers, are required to make available, at their own expense and in a secret manner, to intelligence agencies, the means to access the data of non-US targets of these agencies. The cloud, in its exponential dynamics on the European market, and the hegemonic capture of this market by the hyperscalers, offers to non-European intelligence agencies a means, unparalleled in the past, of gaining massive access to the sensitive data of the economy of our continent. Faced with this threat, the European Union must provide itself with mechanisms to guarantee the free circulation of sensitive data of European public and private organizations, under conditions of verified and binding trust.

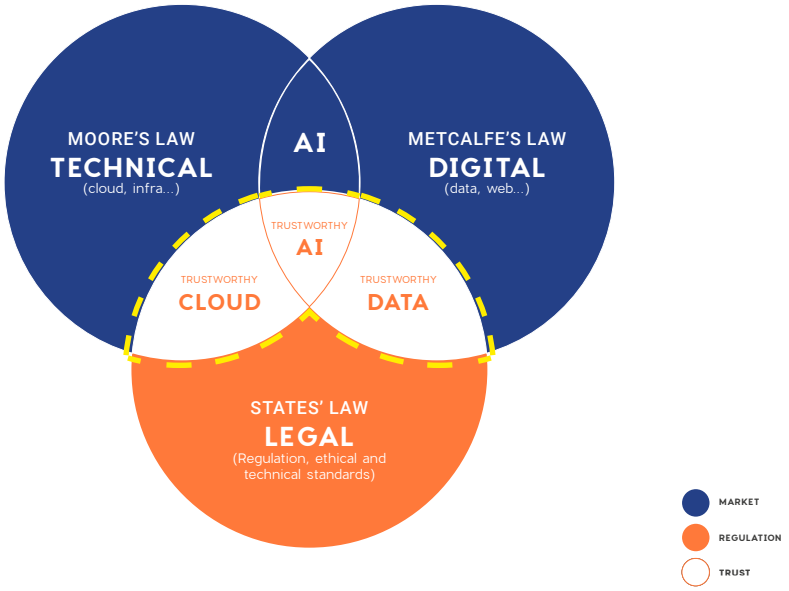
No strategic autonomy without a trustworthy ecosystem

In digital matters, as in military or energy, sovereignty is a political ambition, strategic autonomy an objective, and the level of dependence a measure. The digital sovereignty of a state can be defined in terms of its level of autonomy, its ability to choose, and not to suffer, the degree and nature of its dependencies: *"Strategic autonomy is an ability to generate and defend a trustworthy ecosystem that organizes our technological interdependencies"*[35].

There can be no digital sovereignty without trustworthy ecosystems that guarantee the strategic autonomy of our cloud infrastructures, of the data that is stored (or shared), and of the technologies that exploit them (based on AI in particular). To guarantee this technological and legal autonomy, and therefore tomorrow's political and economic autonomy, we must ensure that we control all our dependencies on the hardware and software infrastructure, which underpins the Cloud-Data-AI triptych. (cf. diagram below, from our "Trustworthy Digital" collection of publications, of which this report is a part)[36]

[32] European Court of Justice
[33] https://www.nsa.gov/Signals-Intelligence/FISA/
[34] National Security Agency (USA)
[35] *Trustworthy AI a strategic opportunity for industrial and digital sovereignty, Julien Chiarini and Arno Pons, Digital New Deal, June 2022. [company industry]*
[36] *Digital New Deal publishes a collection of four reports on the definition of "trustworthy digital" as the foundation of our strategic autonomy: Trustworthy Cloud by Laurence Houdeville and Arno Pons in 2021, then in 2022 Cybersecurity, the guardian of our strategic autonomy by Arnaud Martin and Didier Gras, Trustworthy Data by Olivier Dion and Arno Pons, and Trustworthy AI by Julien Chiaroni and Arno Pons.]*

ECOSYSTEM OF TRUST (CLOUD + DATA + AI) = STRATEGIC AUTONOMY



This trustworthy ecosystem sees its operational translation in the data spaces since they cover the scope of the three components Cloud-Data-AI (and not just data as its name might wrongly suggest). Leading the battle for trustworthy data spaces means conquering our strategic autonomy.

To set up such an infrastructure, we need generalized awareness, a comprehensive state strategy and coherent funding. For each of the three components - Cloud, Data, and AI - a program dedicated to cooperation is desirable.

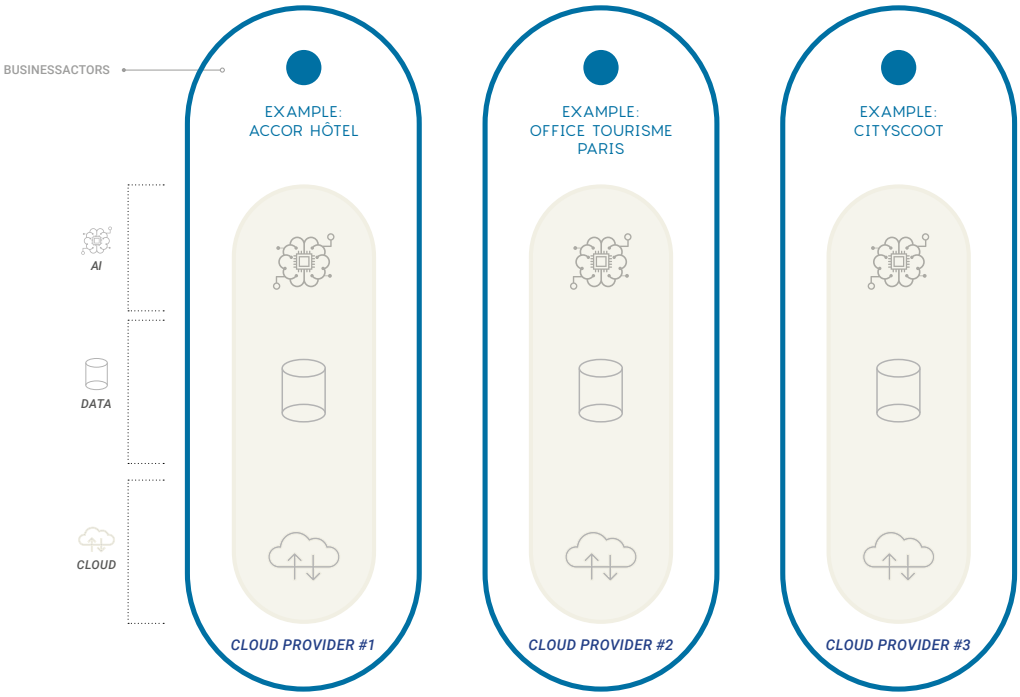
2.2. ABOUT CLOUD?

The cloud is often considered the mother of all battles. More than just a convenience, cloud has become, for public and private actors, the key to interviewing their strategic digital architectures. Hyperscalers have taken advantage of their dominant position to increase this phenomenon by making the cloud the core infrastructure of the digital world, entangling layers, and voluntarily aggregating services to the point where its technical components are no longer visible. Hyperscalers have succeeded in imposing in the collective imagination that Big Data can only be managed by their cloud services, and that any other choice would be technologically and economically unviable.

This is how **the American giants siloed all the value chains through Cloud-Data-AI packages, thereby locking their customers into proprietary systems** that cannot be easily removed (complex and expensive fees), or even mixed with competing services… This is what we were denouncing in 2021 in the "Trustworthy cloud" paper when we identified unfair competition actions via tied selling: *"By subordinating the purchase of a service to the purchase of others, some cloud providers impede the sale of related services by competing companies. These practices make the offer opaque since each service sold does not display a related price. Subordinated or so-called tied selling related obligations could begin with an individualized reading of the services and their prices included in a package. This would allow companies to choose or not the whole package, while benefiting from modular fees.*[37]*"*

[37] *Cloud de confiance*, Laurence Houdeville and Arno Pons, Digital New Deal, 2021

SILOTED VALUE CAPTURE



To get out of this "Digital Stockholm Syndrome[38]" and unlock this highly strategic business, we need to impose commercial and legal transparency, and **re-intermediate the market**. To this monolithic cluster, **we must succeed in imposing an intermediate layer that facilitates transversality.**

There is an urgent need to act because Europe's procrastination is inevitably strengthening the position acquired by the dominant cloud offers proposed by hyperscalers. In the short term, companies and public administrations will have no other choice, given their contractual deadlines, their need for competitiveness and service, to arbitrate again in favor of hyperscalers. This will reduce to a very small proportion the use cases that could benefit from a national or European offer of trustworthy clouds and control of data valuation. Under these conditions, the European Union urgently needs a legislative and operational framework to control and guarantee the confidence of the fifth freedom of data circulation on our continent. The Data Act must be designed to provide this legislative framework and Gaia-X, along with other data spaces' organizations, must develop the operational and industrial framework.

[38] *Syndrome de Stockholm numérique*, Jean-Romain Lhomme, Digital New Deal, 2018

FRENCH NATIONAL CLOUD STRATEGY

To support innovation, France had adopted by the end of 2021 a **national industrial strategy for the cloud**[38].

This third part of the national cloud acceleration strategy, after the definition of the "trustworthy cloud" and the "cloud at the at the center" doctrine for government services, must allow the mobilization of **1.8 billion euros in support of the French cloud ecosystem over four years**, or 667 million € in public funding, resulting from the Programme d'investissements d'avenir (PIA), with € 680 million in private funding and € 444 million in European funding. Among its priorities:

- bring to the forefront innovative French offers
- help French players to scale up on key critical technologies such as big data or collaborative work
- develop disruptive technologies by 2025 (edge computing,...)

This "supporting offer" policy relies:

- on the main national actors (OVHcloud, Outscale, Oodrive, Scaleway,...)
- at the European level on Gaia-X to fund the design of GXFS (Gaia-X federated services) at the heart of the emergence of trustworthy data spaces, and the implementation of a European cybersecurity certification scheme for cloud services

2.3. ABOUT AI?

By their domination over the cloud market, through which they already offer Artificial Intelligence tools (MLOps[39], etc.), **hyperscalers are perfectly positioned to take and lock the market for AI tools and methods**. As a consequence, for the trustworthy ecosystem that Europe wants to set up, **the risk is that hyperscalers make trust in AI a "private matter",** yet a major element of Europe's digital strategy[40] and its future regulation (AI Act).

To respond to this issue, France already has a national coordination program named Grand Defi IA (carried by the Secretary General for Investment - SGPI) and its operational application "confiance.ai" that specifically addresses the need for mutual investment between sub-sectors in the AI field.

EXAMPLE OF "GRAND DÉFI IA" AND "CONFIANCE.AI" PROGRAM

France's investment for a trustworthy AI is supported by the SGPI (Secretary General for Investment) through "Securing, making reliable and certifying artificial intelligence-based systems". "Grand Défi IA" is structured around 3 pillars:

- **Pillar #1 Infrastructure** (the Confiance.ai program) which will allow the development of a trustworthy AI design environment: AI algorithms, design of systems, deployment and maintenance of critical AI based systems, etc.
- **Pillar #2 Compliance** that will ensure the smooth running of trustworthy AI based systems, then define the role and the credentials of the certifiers/ trusted third parties.
- **Pillar #3 Standards** that will enable, in consultation with the different industry actors, to establish standards and certifications.

This investment allows a group of French players to develop a trustworthy AI infrastructure based on European values and regulation. We note that standardization issues are an integral part of the program, which, moreover, federates and coordinates all stakeholders (public, industrial, large groups, startups, and academics). If the trustworthy AI program currently relies mainly

[39] MLOps or ML Ops is a set of practices that aim to develop and maintain machine learning models in production in a reliable and efficient way

[40] Fostering a European approach to Artificial Intelligence - Communication from the Commission to the European Parliament, the council, the European Economic and Social Committee and the Committee of the Regions 2021

on actors in critical and high-risk systems sectors, **trustworthy AI developments, however, could be adapted to the data spaces** in all sectors (education, agriculture, mobility, tourism,...).

The Franco-German position paper *"speeding up industrial AI and trustworthiness"*[41], is the result of a partnership between the Secretary General for Investment (SGPI) and the Big Data Value Association (BDVA) in collaboration with nearly fifty academic and industrial experts. It illustrates the close connection **between trust and industrialization of AI. It underlines the industrial need for software solutions** to implement trust in AI-based processes and applications, **coupled to the availability of standards in complementarity with European initiatives such as Gaia-X** and other initiatives on AI.

2.4. ABOUT DATA ?

Data is trusted in Europe through a set of regulations (GDPR, Data Governance Act, Data Act, Digital Services Act, Digital Markets Act,...) **aimed at protecting individuals, as well as organizations.** To be truly effective, **regulation cannot be enough, it must imperatively be associated with an industrial strategy** enabling the deployment of these regulations, as well as associated standards, in business applications. **If we do not act, Europe risks seeing hyperscalers imposing de *facto* standards.**

Gaia-X has the mission to propose frameworks, standards and labels consistent with the European regulatory texts, for the data spaces. These standards are translated into digital commons, whose base is called GXFS[42] **(for Gaia-X Federation Services).** GXFS are currently being developed at European scale, they are composed of 4 bricks: Identity & Trust, Data Sovereignty Services, Federated Catalogue and Compliance. With GXFS, each Gaia-X data space will be able to build its data sharing infrastructure on top of common open source elements in all projects, thus facilitating interoperability on a large scale. GXFS will allow similar self-description for all actors in the network (identity, available data, compliance level,...).

GAIA-X FEDERATION SERVICES (GXFS)

FIRST FUNDING FOR THE DIGITAL COMMONS OF THE DATA SPACES

The French GXFS have been funded to counterbalance the German GXFS. In Germany the GXFS project is led by eco[43], in France it is the Institut Mines-Télécom (IMT) as part of the call for tender "Cloud Souverain". **The French project corresponds to a €12M funding which associates a few founding members of Gaia-X**. It extends over 24 months, with a start in July 2021.

The scope of the French project includes the initiation of open-source developments, and the writing of GXFS specifications. IMT via TeraLab[44] will serve as "testbed[45]" to validate projects compliance with Gaia-X.

GXFS are essential for the proper functioning of data spaces, they are the binder of the entire ecosystem. However, the French project covers only one part of it, and it will then be necessary to "make these GXFS live" in the data sharing infrastructures of the projects involved. They simply allow each data space to start on a common basis and ultimately represent only a small part of the volume of the necessary development required for the complete creation of the infrastructure of the data spaces; this involves the **need to encourage the development of a strong open source community around the project, and increase the number of "InfraTech" players** constituting the ecosystem that underpins the structure of Gaia-X.

[41] *"Speeding up industrial AI and trustworthiness", Secretariat General for Investment, Big Data Value Association et al., 2021*
[42] *https://www.gxfs.eu/gxfs-overview/*
[43] With over 1,000 member companies, eco is the largest association of the Internet industry in Europe. Since 1995, eco has played a key role in the evolution of the Internet, encouraging new technologies by creating framework conditions and representing its members' insights in international policy and committees. The key eco themes are the reliability and strengthening of the digital infrastructure, security and confidence in information technology, as well as ethically oriented numbering.
[44] *https://www.teralab-datascience.en/*
[45] *Test Platform*

France has therefore already laid the foundations for a potential "Trustworthy Data program" by investing in the development of GXFS, Gaia-X open source based components. We now call for a shift in scale, so that these federation efforts do not remain in vain. It is also essential to launch and animate an open source community around GXFS (France and Germany will only develop one part, we need the rest of Europe) and offer associated support, integration services and maintenance packages. Funding this "federation toolbox" is a major asset for the empowerment capacity of data spaces. **But this investment will only reach its full capacity if it is part of a complete French co-op program to create a data sharing infrastructure, and to run the data spaces, which will be based on these GXFS. Some sectoral initiatives such as education/skills, tourism, energy, mobility, etc. must be supported. Indeed, they are the ones who implement the data spaces concretely, and contribute to the building of common elements, as well as specific sectoral bricks generating value for all citizens.**

Data is the real keystone of our digital trustworthy ecosystem because it is mainly through data that value is created. There is no French "Grand Défi Data" yet, that would secure the investments made in trustworthy Cloud, trustworthy AI, and the GXFS. However, in order to have a say in Gaia-X and not be subjected to the structuring choices already underway, we need greater coordination of resources.

This "Trustworthy Data" program should be steered in coherence with current policies on Cloud and AI, in order to maintain a common roadmap on this triptych of trust "Cloud-Data-IA", while giving data a special place. This is all the more strategic as data is paradoxically the most promising area for innovation in Europe. It might be easier for us to solve data sharing internal governance issues, than to make up for our technological and commercial lag in cloud and AI when faced with hyperscalers.

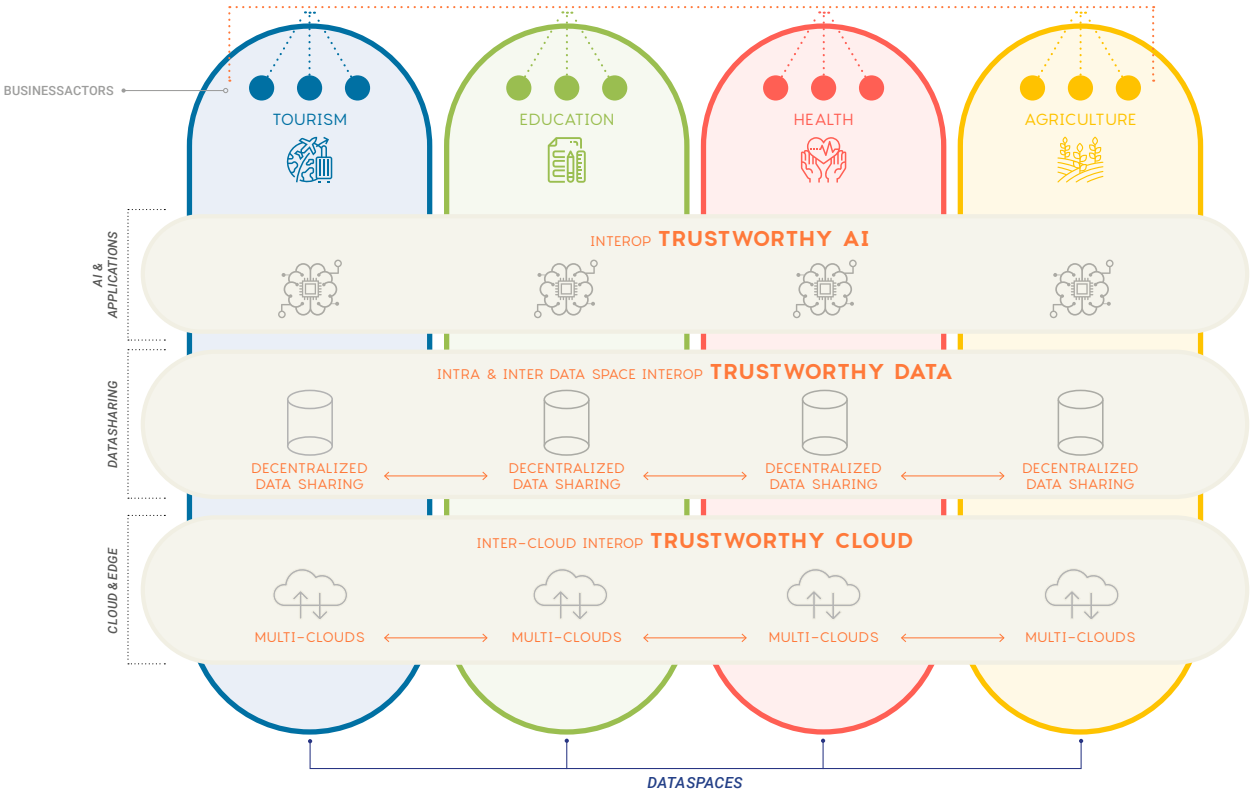
2.5. VALUE CREATION WITH TRUSTWORTHY SHARING

Data spaces must break the barriers that lock economic actors in silos, and that allow hyperscalers to be their interlocutors on the whole Cloud-Data-IA value chain, thereby capturing value in a vertical manner.

Cloud-Data-AI interoperability layers that Europe is creating, must allow for horizontalization of exchanges, and thus collectivization of value. Consider the concrete applications this represents, "Let's take health as an example. Let's imagine that Philips (client of AWS) wants to work with Sanofi (client of Google Cloud): will data interoperability be possible? And if so, will their data sharing conditions be met correctly?"[46]

[46] Hubert Tardieu, Conference on European Digital Sovereignty, Munich, May 2022

HORIZONTAL AND COLLECTIVE VALUE CREATION



“

THE LAYERS OF INTEROPERABILITY AND MUTUALIZATION CLOUD-DATA-IA MUST ALLOW A HORIZONTALIZATION OF EXCHANGES, AND THUS A COLLECTIVIZATION OF VALUE.

III. OCCUPY THE STRATEGIC DATA SPACE FIELD

3.1 EUROPEAN CONTEXT

The European Data Strategy[47] includes a set of texts that addresses the vision of President Ursula von der Leyen: "digital technologies must enrich our lives and respect European values". In this spirit, the Data Governance Act (DGA) and the Data Act (DA) establish the data sharing framework, paving the way for the development of multiple data spaces.

In 2022, the European Commission launched an investment program of almost €10 billion for the "Common European data spaces" at the heart of the data strategy, complementing investments made by Member States and the private sector. Europe's ambition is to build a "soft infrastructure" of data sharing based on European values and regulations: GDPR, Data Governance Act (DGA), Data Act (DA), Digital Services Act (DSA), Digital Markets Act (DMA), AI Act. This infrastructure should foster the emergence of an alternative model to hyperscalers', in favor of fair and human-centric data sharing ecosystems. Regarding the sharing of personal data, the EU recalls that the GDPR, with its values, its principles and rights are the constituent elements of the infrastructure that will be put in place.

The EU is promoting the emergence of data spaces projects, as well as sectoral governance bodies that will be responsible for coordinating key sectors to facilitate the flow of data based on shared standards. More than ten priority sectors have been announced including mobility, finance, health, skills, energy, Green Deal and administration.

In 2020 the EU-27 considered as the baseline the total economic value of the data economy at €325 billion (2.6% of GDP).[48]

The data spaces standardization framework is being built within organizations such as aNewGovernance[49], BDVA[50], Fiware[51], Gaia-X, IDSA, MyData[52], and others. It will be formalized via the DGA, completed by the DA, whose main objective is to "facilitate the availability of data for their use, increasing trust in data sharing and strengthening data sharing mechanisms across the EU.

The texts provide, for example, the implementation of:

- **The Data Innovation Board (DIB)** composed of representatives from all Member States, the European Data Protection Board (EDPB), some relevant data spaces and representatives of competent authorities in specific sectors. It will provide expertise when decisions on data governance will have to be taken. It will advise the Commission on cybersecurity issues for data sharing and storage, and it will push for international standardization and data portability.
- **Data Space Support Center (DSSC)** which coordinates all relevant actions on sectoral data spaces and makes available architectures and data infrastructure requirements for data spaces, including technologies, processes, possible standards, and tools that will allow data to be reused between public sector and European companies, including SMEs.

[47] https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52020DC0066&from=EN
[48] European Commission (2020): *Support Study to this Impact Assessment*, SMART 2019/0024, Deloitte
[49] https://www.anewgovernance.org/
[50] https://www.bdva.eu/
[51] https://www.fiware.org/
[52] https://www.mydata.org/

Finally, the European Commission invests in the implementation of an open source smart middleware[53], beyond GXFS, corresponding to the intermediary layer of the infrastructure common to all data spaces. Europe is clearly showing a desire to create digital commons facilitating the emergence of data spaces where coordinated action by Member States is required.

France, with its technological capabilities, large innovative groups, and startups, as well as its political will, is in a position to be one of the leaders of this process by being fully involved in the European project.

3.2. DATA SPACE CHARACTERISTICS

To better understand the issues associated with the emergence of data spaces, here are the main features of a data space:

- **A federation of secure IT infrastructure to share, access, process, use and share data** (personal, non-personal, industrial) in a decentralized and controlled manner.
- **A data governance mechanism,** including a set of legislative, administrative, technical, and contractual (business agreements) rules that determine reliable and transparent data access, processing, use, and sharing rights.
- **Control of access to data by providers of data, and citizens,** who decide for what purposes and under what conditions it may be used, including via data sharing, as described in the Data Governance Act.
- **The reuse of large quantities of data,** that can be reused under certain conditions, against compensation or free of charge, and in a controlled manner by the data provider and/or the citizen.
- **The open participation of many organizations and individuals,** that avoids the creation of cartels.

These features are implemented through four federation services:

- **The federated catalog:** which functions as a registry of all resources available on the network. Catalog entries include resource descriptions (such as data and services), as well as policies to which services adhere (e.g. GDPR), and the conditions under which resources may be used by other participants in the network.
- **Trustworthy exchange:** which allows the definition, exchange, processing, and monitoring of data usage policies. These can be understood as the terms and conditions of the data economy. Examples are restricting the number of data read actions, prohibiting the distribution of data further, the obligation to use data only in a certain geographic area, etc. This also involves managing the consent of individuals to share and process their data.
- **Identity Management:** which guarantees trust between suppliers and network users (organizations and individuals).
- **Compliance services:** which guarantees regulations (GDPR, DGA, DA, etc.) as well as user expectations translated into labels expressing levels of trust (from self-description to third-party audits)

[53] *https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-made-simple-spaces*

"Federated services facilitate the emergence and operation of data spaces. Data spaces are a concept of data integration that does not require data consolidation in a single central location, nor the use of a single database schema. On the other hand, data integration is achieved at the semantic level, and data redundancies are possible."[54]

Gaia-X AISBL makes available to its members and users a general architecture document[55] for data spaces, to coordinate the development of all the stakeholders.

3.3. BENEFITS AND USE CASES

Data spaces represent a major opportunity for the development of digital ecosystems for Europe. ***Data spaces* will streamline use and promote innovation for all actors in the economy:**

- **Private actors** will be able to (re)gain control of digital distribution channels from large platforms beyond Open Data, increase their visibility and improve their listing and that of their offerings, enable a seamless end-to-end user experience, foster greater interoperability and innovative cross-sector use cases between multiple digital services, certify data, benefit from fair redistribution of value (including the long tail of small actors) and data analytics.
- **Public actors** will be able to offer a seamless and augmented user experience for citizens, associate private services to citizens' services, reduce the administrative burden (the "Tell us once"[56] French concept avoiding data re-entry), certify data, rely on state identity services, complete B2G statistics, participate and contribute to data altruism (concept introduced by the Data Governance Act and providing a framework for data exploitation and sharing within an altruistic purpose).
- **Individuals (digital services users, Citizens)** will benefit from a seamless user experience independent from large platforms, a high degree of customization (thanks to trustworthy AI), ways to control the flow of their data (like requires by the GDPR) and services that combine public and private offerings in a seamless way according to their needs, ways to get involved in the governance of their data.

At first, data spaces projects envisioned in the Fraunhofer work, that led to the creation of Gaia-X, were rather industrial by nature. They were part of the German industrial strategy 4.0.

One of the new great promises of the data spaces is the improvement of services to citizens, in line with the precursor project X-Road , a data exchange infrastructure (Data Exchange Layer) launched in Estonia in 2001, which had enabled Estonia to digitize 99% of its public services.

Finally, a lot of data spaces under construction are investing all sectors of the economy, including the end-user-oriented sectors, raising the issue of personal data sharing, and thus the consideration of the GDPR. These topics are covered for example by MyData Global NGO[58], launched in Finland in 2018 to promote a vision of human-centric data sharing.

[54] Hubert Tardieu & Boris Otto - La Revue européenne du droit, décembre 2021, n°3
[55] https://www.gaiax.es/sites/default/files/2022-01/Gaia-X_Architecture_Document_2112.pdf
[56] https://www.numerique.gouv.fr/services/guichet-dites-le-nous-une-fois/
[57] https://x-road.global/
[58] https://www.mydata.org/

Here are some examples of use cases by market:

USE CASES FOR DATA SPACES BY MARKET

GREEN DEAL	Carbon impact monitoring for all sectors as part of the Net Zero objective
SKILLS, EDUCATION AND EMPLOYMENT	Gather skills data, personality data and preferences to help individuals in their lifelong learning with contextualized and territory-relevant training and job recommendations
TOURISM & MOBILITY	Personalized tours, augmented tourism, disability & accessibility, green alternatives, and green mobility subsidies, connected vehicles
ADMINISTRATION	'Tell us once', use of administrative data in private services (finance, insurance, health, etc.), augmented citizen service
HEALTH	Digital medical record, administrative simplification, data for research, 5P medicine (personalized, preventive, predictive, participative, proof), new medical devices (IoT), digital twin health
ENERGY	Consumption monitoring, electric vehicle charging, network flexibility, energy mix, renewables, research
AGRICULTURE AND FOOD	Traceability "from farm to fork", well-eating, sustainable food, water management, soil management, livestock genetic data, research
SMART MANUFACTURING	Supply chain, data flow between manufacturers and suppliers, industry decarbonization
SMARTBUILDING	Data capture and analysis for better environmental management
FINANCE	KYC, fraud detection, budget management, easy administration, life projects support (real estate, studies, cars,...)
CULTURE AND MEDIA	Detecting fake news, traceability of content, shared press offers, Metaverse, decentralized social profile
DIGITAL TERRITORIES	Smart City, regional data spaces, digital twins for smart cities
SPACE	Exploitation of earth observation data for environment, agriculture, security, research, digital twin earth
SEA ACTIVITIES	Permaculture, marine data for the environment, navigation, research, digital twin ocean, digital twin earth

Many use cases, planned or under development, have been described in detail and are published in Gaia-X position papers[59]. The French Gaia-X Hub working groups already cover almost all these topics.

[59] https://gaia-x.eu/use-cases
[60] https://www.data.gouv.fr/datasets/route-ministeries-on-la-politique-de-la-donnee-des-algorithmes-et-des-sources/

3.4. THE DIFFERENT TYPES OF PLAYERS IN DATA SPACES

The data spaces are composed of actors of varied natures, which can be grouped into 4 categories:

- **Institutions** (public organizations, associations and alliances/unions, regulators, etc.). They play, or can play, an essential role in many data spaces in different ways. For example:
  - The state or local governments are producers or consumers of data for many data sharing use cases, notably through the 'Tell-us-once' policy and ministerial roadmaps for data[60]
  - The concerns about sectoral (sometimes national) regulation have a direct impact on the development and governance of data spaces, for example the mobilities law[61] and its impact on the Mobility data space
  - Regarding the sharing of personal data, the CNIL (French National Commission on Informatics and Liberty), or European equivalents, can set up 'regulatory sandboxes'[62] dedicated to data sharing.
  - For specific areas, regulatory authorities, for example the CRE (French Energy Regulatory Commission) for the Energy data space, should play a major governance role.
  - Public procurement can foster the development of the data spaces and set a framework for their development that can even establish and influence certain technical choices.
  - Some use cases currently being tested such as FranceConnect[63] and the evolution of digital identity with the implementation of the European e-wallet and the eIDAS V2[64] standard as part of the French program "France Identité"[65], enhancing trust in the sharing of personal data.

- **Large groups** of all sectors are big data producers and providers, which could become the driving force in their respective ecosystems:
  - These large groups are leading the vertical industries and must be the catalysts that foster coopetition. They are already grouped in industry sector logics with the ecosystem of SMEs.
  - Very often operating on an international scale, it would be in the interest of large groups to invest more openly in data spaces, thus opening to the European market at least.

- **Innovative SMEs and startups:**
  - Startups are potentially large data consumers. Startups could, if they had more access to data, propose numerous innovative use cases in all sectors (Fintech, Edtech, Healthtech, etc.).
  - They are also potential suppliers for multiple essential components of the data spaces infrastructure. By supporting the development of infrastructure startups in the data spaces (what we call "InfraTech"), we could give birth to tomorrow's champions who would then have direct access to the single European market thanks to interoperability standards, and those new players would no longer be restricted to their smaller national market.
  - SMEs in general must also be able to use their data or add value to it, if they want to

[60] https://www.data.gouv.fr/fr/datasets/feuilles-de-route-ministerielles-sur-la-politique-de-la-donnee-des-algorithmes-et-des-codes-sources/
[61] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000039666574/
[62] Un « bac à sable » réglementaire, qui peut permettre aux acteurs de tester leur technologie ou service innovant, sans devoir nécessairement respecter l'ensemble du cadre réglementaire qui s'appliquerait normalement, dans un temps délimité.
[63] https://www.numerique.gouv.fr/espace-presse/lancement-de-lexperimentation-de-mon-franceconnect-letat-accelere-la-simplification-des-demarches-administratives/
[64] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eIDAS-Node+version+2.0
[65] https://france-identite.gouv.fr/en-savoir-plus/de-quoi-s-agit-il/

THINK-TANK DIGITAL NEW DEAL

remain competitive. Data spaces will directly contribute to the digitalization of the SME community by offering shared and inclusive digital services.

- **Research actors:**
- Researchers building innovative services and AI models will be able to gain access to more data for their research via data spaces.
- They will also be major players in the construction of the data spaces infrastructure and will ensure interdisciplinary development that respects the state of the art (technical, ethics, legal, business, social sciences, societal issues, environmental,…).

Each of these actors can carry out one or more of these different functions/roles for the data spaces:

- **Data spaces users** are the business stakeholders (e.g. EDF[66] for the Energy data space), who wish to make available and/or consume data in a controlled manner.

- **Infrastructure Providers (InfraTech)** are startups/scale-up (e.g. OKP4[67], Onecub[68], OVH, Inokufu[69], Visions[70], etc. and large groups (e.g. 3DS Outscale, Atos, Cap Gemini, Digiposte, Orange Business Services, Sopra Steria, etc.), capable of supplying the software components, open source or not, of the data space infrastructure (identity management, consents, API, equal management, business, blockchain/Web3, interoperability, etc.).

- **Operators of the data spaces** are the entities responsible for the management of the data spaces (e.g. Agdatahub[71] for the agriculture data space in France, Catena-X for the car industry data space in Germany, etc.). Some data spaces can work without formal embodiment, relying solely on interoperability standards and business agreements within a sector. Other data spaces will be run by formal organizations (associations, private companies, cooperatives, etc.). Operators may be of a public nature or operate through public service delegation contracts, public/private, or purely private. They ensure that all actors in the data space can participate in a joint governance.

- **Standards, frameworks, labels & digital commons organizations,** who build the common soft infrastructure of the data spaces, are mostly associations such as Gaia-X, IDSA, or MyData[72]. It also includes sectoral governance bodies (Skills, Tourism, Agriculture, etc.), and the DIB & DSSC at the European level. They ensure that data spaces are based on common functional specifications, blueprint, and open source code so that no actor locks the capacity to connect to a data space.

[66] https://www.gaia-x.eu/news/member-story-electricity-france
[67] https://okp4.network/
[68] https://www.onecub.com/
[69] https://www.inokufu.com/
[70] https://visionspol.eu/
[71] https://agdatahub.eu/
[72] https://www.mydata.org/

“

ADDRESSING
OUR DATA SHARING
AND GOVERNANCE
ISSUES IS SIMPLER
THAN CLOSING
A TECHNOLOGICAL GAP

3.5. A STRONG NEED FOR GOVERNANCE

Governance at the heart of data spaces

The variety of actors, their number, as well as their different stakes and levels of maturity, imply a significant coordination effort to homogenize good practices. Beyond coordination, the actors of a data space must then be able to agree on common rules, and make clear decisions, if no consensus is found.

Market players, both startups and large corporations, will be key to the success of data spaces, as they will build the infrastructure and produce most of the services of tomorrow. They must therefore embrace this new "coopetition" logic. However, **we believe that only a strategist state, fully invested and aware of the stakes, will allow data spaces projects to emerge and thrive.** In fact, through public procurement and investment, the state has the capacity to ensure the launch of data spaces on almost all sectors, and to structure the market where **private players often struggle with mutualized approaches...** By those same instruments, the state is also able to align developments to a common doctrine. Finally, a large amount of data, as well as certain decision-making processes, are the responsibility of the state, which is the only legitimate entity to ensure the protection of its citizens, since trust may not be the exclusive prerogative of private interests.

Issues by type of player

Each type of stakeholder has its own issues regarding data sharing:

- **Institutions** somehow rarely develop a global vision of the issues (e.g., vision of digitalization centered on the administration rather than the sector itself, difficulty in coordinating action between ministers) and may have some difficulty understanding other actors' problems. They also have a slower adoption of innovation curve and are often reluctant to share data as well as to collaborate with private actors.
- **Large groups** express significant strategic reserves about data sharing: unclear standards, regulatory fears, fear of opening up data to competitors, "gatekeeper" behavior, difficulty with projecting over the long term (uncertain return on investment), difficulties visualizing use cases.
- **Innovative SMEs and startups** lack resources (expensive technical developments, 'time' of startups versus 'time' of large groups and public institutions) and often find it difficult to get involved in the standardization effort, or to get involved into consortium tenders.

The intra and inter data spaces coordination needs

A coordination effort must take place within each data space, where it is necessary to facilitate dialogue between the different types of players that it is sometimes complex to make coexist, to mediate potential conflicts, to animate ecosystems, to coordinate developments, to identify and remove barriers of entry, and to build common and sustainable business models.

The effort must also take place across data spaces in order to allow the mutualization of developments (standards, Digital Commons, etc.), as well as cross-sector and cross-border interoperability.

It is indeed the joint effort of all the players in the data space that will deliver real value to end users, as well as the data providers themselves.

We believe that states must now foster the development of an overall vision (public/private) on the issue of data spaces and data sharing to help all actors work together.

EXAMPLE OF ALBERTO PALOMO-LOZANO CHIEF DATA OFFICER OF SPAIN

Just over 2 years ago, the Spanish government created this key function under the authority of the Deputy Prime Minister and the Secretary of State for Digitalization and Artificial Intelligence, with the goal of optimizing the use of data.

His principal mission is to ensure that the development of AI is done in an "aligned and coordinated" way between the different branches of government and private companies, especially those that end up participating in one of the key projects done with European funds.

Another of his tasks is to **promote the implementation of the data sharing architecture developed within Gaia-X and industry relationships in terms of data sharing and cloud services.** In collaboration with the Spanish Gaia-X Hub, the Data Office supports the development of sectoral data spaces, with an emphasis on tourism, health, agri-food and sustainable mobility.

Mr. Palomo is not a senior public servant. He holds a Ph.D. in Theoretical Physics, and is currently completing a Certificate in Management, Technology, and Innovation at MIT. He is an expert in data analytics and data strategy definition, has extensive experience in managing teams of data scientists on an international level to generate efficiencies in industry processes, and has worked on hardware design to maximize the training phase for interconnection of neural networks based on artificial intelligence.

Examples of coordination challenges

Multiple data spaces represent intra or inter data spaces challenges for which state involvement is an indispensable requirement:

- **The example of the Green Deal data space**[73]**:** Europe's Green Deal has the ambition to raise climate and environmental challenges. It integrates a growth strategy while at the same time aims to protect, conserve, and enhance EU natural capital and to protect the health and well-being of citizens from environmental risks and impacts. By allowing the flow of data linked to environmental impact, the Green Deal data space positions itself as a tool for rethinking green energy supply policies throughout the economy (industry, production & consumption, major infrastructure, transport, food, agriculture, construction, taxation, social benefits). It will also increase the protection and restoration of natural ecosystems, promote the sustainable use of resources, and improve citizens' health. By nature cross-sectoral and at the intersection of issues of a private nature (energy players, retailers, industrialists, real estate developers,...) and public (cities, regions, tax administration) the Green Deal data space is the concern of all actors in the economy, without being the "core business" of most of them. Regarding the public sector, it is noted that many ministries should be part of a joint process. Only the state is currently able to align the multiple public and private aspects of all of these actors, through vision, planning, investment and public order.

[73] https://www.cigref.fr/wp/wp-content/uploads/2021/10/English-Gaia-X-Hub-Position-Paper-GT-Green-deal-long-version-Vf.pdf

- **Example of the skills data space**[74] **:** The skills data space must allow citizens to benefit from a dynamic CV, which will follow them throughout their professional career. This CV is built with data coming from the National Education Ministry, multiple educational institutions, training agencies, training accounts[75], large companies in which the person has evolved, edtech startups that will have helped him in his orientation. Initiatives such as the Erasmus[76] program, facilitating student mobility in Europe, also naturally calls for joint initiatives at the European level. Without the active and central engagement of the state, both in the governance of data access and in the funding of a sovereign infrastructure, the skills data space just cannot thrive.

Recent feedbacks

We must rely on the feedback (education, mobility, energy, agriculture, etc.) to be able to manage the creation of data spaces in all sectors and territories.

With rare exceptions, the first data spaces projects demonstrate a real difficulty in funding public-private partnerships. The incompatibility in terms of the existing public calls for projects/tenders, compared to the inherent issues of data spaces projects, makes the task particularly tricky for the project leaders, who moreover are often entrepreneurs bringing together ecosystems of both regional and national dimensions, with very limited resources.

The most common difficulties are as follows:

- Long delays in assessing projects calls, sometimes exceeding one year, which demotivates the partners,
- High proportion of refundable advances, including to fund digital commons or tasks that can be considered as public service,
- Funding framework that is too rigid and does not consider the state of the art of the emerging market of the data spaces,
- Difficulty in valuing stakeholders input, especially for startups,
- Difficulty in identifying the calls for projects at the European level (about the amounts involved it would be necessary to have a support to detect these funding opportunities),
- National calls for projects with no direct link to the European market or the Digital Europe funding program (yet natural framework for data spaces projects),
- Lack of understanding on the part of the state of the connection between the digital commons, the data space, and the value generated *in fine* by and for the participants of the data space,
- Calls for projects pushing the players to recreate their own data sharing infrastructure. This on one hand creates an unjustified competition between one another, and above all weighs heavily on the cost of their project (thus the **ambition of this note to join investments for a single infrastructure that will allow to lighten the data space core requirements and will also prevent the states from having to fund multiple times the same parts of the soft infrastructure...)**

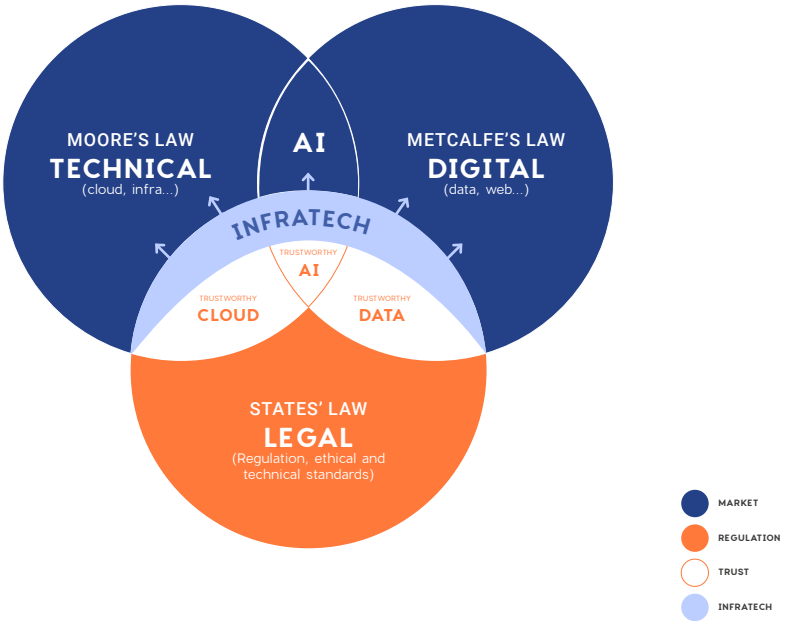
[74] https://prometheus-x.org/pages/gaia-x?locale=en-
[75] https://www.my-trainingaccount/
[76] https://www.education.gouv.fr/le-programme-erasmus-9890

We see that all of the data spaces projects being built in Europe share common data sharing infrastructure needs and building blocks (consent management, identity, legal and business models, APIs,...). **However, project developers find it very difficult to pool their developments,** including on low infrastructure layers, because they are put in competition with each other on public calls for projects. **Infrastructure actors working on their own** (education, agriculture, tourism, mobility,...) **do not naturally display a desire to fund the common layers** (nor in an independent way, nor by pooling their efforts) and prefer to direct their fundings towards use cases and business applications. Some projects though, such as Prometheus-X (skills) and Themis-X (tourism), work towards this mutualization of low-level infrastructure funding, but they encounter difficulties in their financing.

FOR THE STARTUP NATION NOT TO BECOME A MIRAGE, FRANCE MUST ALSO BECOME AN INFRASTRUCTURE NATION.

IV. THE DATA SHARING INFRATECH

ECOSYSTEM OF TRUST (CLOUD + DATA + AI) = STRATEGIC AUTONOMY



Our conviction is that France, as well as other European countries, should not be satisfied with becoming only a "startup nation", because the more it builds business-oriented startups, the more it paradoxically amplifies its dependencies (these startups being great consumers of the technical and marketing services offered by hyperscalers outside of Europe). For this brilliant tech ecosystem to be protected, for it to become truly systemic, and for each euro invested in these startups to not only benefit hyperscalers, but France, and other EU Member States, **we must also become infrastructure nations**[77].

This is the condition for our strategic autonomy, and it would be especially a major legacy to both current and future generations of citizens. Like nuclear power before it with the Messmer plan in France, digital infrastructure also deserves a state/European strategy. Especially since we are talking about an "immaterial soft infrastructure" that would cost only a couple of dozen million euros, almost a "detail" if we compare it to the 100 billion invested in the French nuclear infrastructure in the 70s.

We believe that France must contribute to the creation of what we call the "European InfraTech ecosystem", an ecosystem of European players specializing in software, composed of a multitude of start-ups, SMEs and innovators, who join forces in order to offer a qualified end-to-end competitive software infrastructure for the cloud, data or AI.

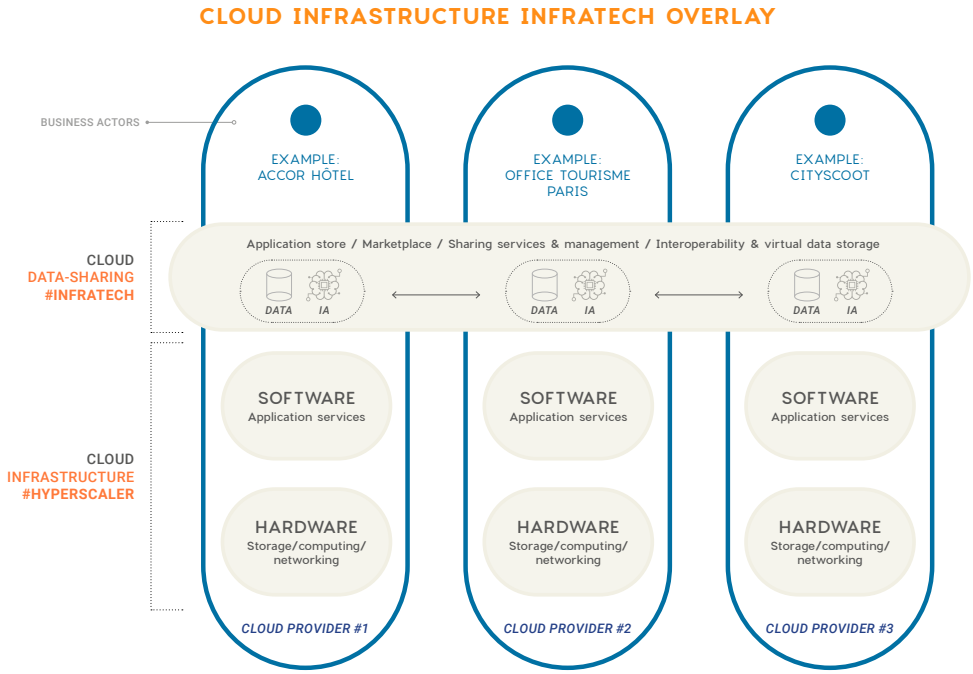
The InfraTech:

- intrinsically bears European values and regulations, as well as harmonized and shared standards.
- is based on digital commons that must be collectively exploited on the market in a horizontal strategy that will irrigate all of Europe's strategic sectors (employment, education, health, finance, tourism, mobility, administration, etc.).

[77] Tariq Krim, https://www.getrevue.co/profile/tariqkrim/issues/newsletter-4-infrastructure-nation-acte-i-923015

- requires the development of a federating common development environment to guarantee coverage of the entire value chain in the domain concerned, ensure interoperability and enable value sharing amongst all parties involved.
- requires a "business" approach within the data spaces framework in order to guarantee its deployment and adoption by the application sectors.

4.1. DECENTRALIZING THE INFRASTRUCTURE WITH AN OVERLAY OF INTEROPERABILITY



Europe must be able to define and propose a new form of data sharing architecture, which corresponds to its own objectives of sovereignty and the safeguarding of fundamental rights. The cloud definition offered by hyperscalers, or an individual packaged offer that covers the entire Cloud-Data-IA value chain, is not necessarily the one we will hold. “A fair balance between the ownership of the individual data and the community framework to use the existing data (e.g. for healthcare purposes, not only in time of pandemic) is at the focus of the debate on how to design the European data economy. The balance between individual and common interests leads to a different infrastructure design than what is available in the current offerings of U.S. or Chinese private platform providers."[78].

The European approach is more decentralized and "democratic" than hyperscalers' approaches, although the U.S. is already beginning to implement a shared infrastructure. "All stakeholders agree that Europe's data and software infrastructures must prevent the emergence of centralization of power in data." This is crucial insofar as it requires preventing network effects from leading to a "winner takes all ""[79] situation.

[78 / 79 / 80] Boris Otto & Hubert Tardieu - European Legal Journal, December 2021, #3

The European infrastructures are designed as a federation. They do not require central data storage, processing, or distribution capabilities. Instead, the infrastructure is composed of distributed nodes in a network of data and services. The nodes can provide or use resources such as data. The providers remain independent and free with respect to their data and other resources; the network is open and non-discriminatory; and users and providers trust each other providers trust each other.[80]

Developing the infrastructure (or InfraTech) of data sharing, and data spaces more generally, consists of creating a Cloud-Data-AI interoperability overlay that will be able to integrate with or add to existing cloud offerings (those of hyperscalers, as well as those of other players), to create a virtual and decentralized link between the business services of a multitude of players.

Interoperability and horizontalization work are in progress on all 3 components of the trustworthy Cloud-Data-AI system through initiatives such as Gaia-X, the French program "Grand Défi IA", and others.

	CLOUD	DATA	AI
BUSINESS ECOSYSTEMS	Cloud Services Users	Data spaces users	AI Users
INFRASTRUCTURE ECOSYSTEMS	Cloud providers (e.g. OVH, Microsoft Azure, etc.)	Structuring actors of data-sharing (e-Wallet, data-sharing intermediaries, data space infrastructure provider, etc.) (e.g. OKP4, Onecub, Visions)	Structuring actors of AI (e.g. Tools and Metrics developed as part of the Confiance. ai program)
DIGITAL COMMONS	GXFS (Gaia-X)	GXFS (Gaia-X), Simpl, Prometheus-X (and other data spaces sectoral digital commons)	Confiance.ai
NORMS	Gaia-X (Label), SecNumCloud (Fr), C5 (All), forthcoming ENISA	Gaia-X, IDSA, BDVA, MyData, etc.	Pillar #3 Upcoming standards for Grand Défi IA (SGPI), and forthcoming AI Standard Foundations

TRUSTWORTHY ECOSYSTEMS COMPONENTS

Data spaces are meta-platforms of services governed by interoperability rules, which are translated into base APIs (compliant for example to the Gaia-X standards). These meta-platforms are a set of standard and smart building blocks created by a set of actors who agree with each other through:

- intra data space interoperability: the interoperability of the building blocks within a data space, which makes it possible to create a powerful network effect, and to implement a broad and comprehensive service offer;
- inter data spaces interoperability: the alliance between data spaces (via Gaia-X, DSSC, DIB, and others), which allows to generate a critical mass at the EU level, and thus to federate all European players around the same approach.

THE DIGITAL COMMONS

The digital commons are part of a more global vision of the public goods. At the intersection of economic and political science theories, the topic was made popular by Elinor Ostrom's research in 1990 and then by her award of the Nobel Prize in Economics in 2009. The commons initially described physical resources (rivers, forests, fisheries) that could be collectively administrated.

The digital commons have as specificity: copy and distribution cost close to zero. These are for example source codes, data or databases, software or digital content (image/video/sound). In addition, these include free licenses, under which access, use, production, modification, distribution and management of these digital resources are legally organized.

Digital commons are distinct from physical commons in two main important characteristics: they are non-exclusive (no limited access if someone uses the resource), and they are non-rival (the use of the resource does not deprive other users, it remains available).

The environmental approach to data spaces will be another key element in their success.

This approach is expressed in the very principle of data spaces that do not store (or little) data, but catalog it and implement ephemeral storage means, only if necessary. In this sense, data spaces contribute to a responsible and frugal vision of data circulation. Moreover, the principles of decentralized identity (based for instance on DID protocols and blockchain technologies) and wallets (eWallet) that keep the different information necessary to track the processing and transactions performed, without storing it in its entirety, contribute to the minimization of stored data, by making each actor responsible for his data.

The infrastructure of the data spaces is "portable" and can switch cloud environments on demand. The principles of the open source "Digital Commons" and the willingness to deliver components with an orchestration layer and common documentation rules; allow for simplified deployment, and give guarantees as to the durability and resilience of these services.

4.2. INVOLVEMENT IN THE STANDARDIZATION EFFORT

Major cloud players, such as hyperscalers, do not necessarily have an interest in seeing interoperable infrastructures, that they would not control, emerge. Indeed their business models being based on vertical value capture. They have a very large market share in the cloud and data spaces on these clouds could *de facto* rely on their standards. In addition, they are already involved in creating open-source components for data spaces[81]. If we take the analogy of the autonomous vehicle, its operating system will ultimately be the one that will be able to provide all services in a packaged manner and will become the *de facto* standard of the vehicle platform. If there are no rules around this operating system, it can lock the market to any new service. A recent report from the Commission[82] highlighted that **competition from Big tech, who do not always respect European rules and core values and who place data ownership and monetization at the heart of their strategy, is a major policy challenge for Europe.**

Hyperscalers organize themselves and defend their own business interests in Gaia-X type initiatives, sometimes making it difficult to develop clear and transparent rules for a

[81] *https://projects.eclipse.org/projects/technology.dataspaceconnector*
[82] *Digital sovereignty for Europe - EPRS Ideas Paper Towards a more resilient EU*
[83] *https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-made-simple-spaces*

fair market. Furthermore, Europe has this complex nature of being composed of Member States that do not all have the same policies for technological autonomy. France has been a pioneer in these matters, **unfortunately the ecosystem is scattered.** The mobilization of certain Gaia-X actors, particularly on the French side (we can mention here the driving role of EDF) has nevertheless made it possible to lay the foundations of a Gaia-X labeling system, which takes up the ambitions of European sovereignty. The highest level of labeling includes, for example, immunity to extra-territorial laws, in accordance with the criteria established by ANSSI, and this in spite of the frontal opposition of lobbyists (some of them European, unfortunately) representing non-European interests. But all these actions remain very fragile and rely, as we said, on a few isolated good wills, which would deserve a support, notably legal, at the level of the lobbying means of the American hyperscalers.

This work comes at a cost, and no one at this point is willing to fund it. A few large groups have been able to dedicate resources to defending their interests, but this effort is far too weak compared to the forces deployed by hyperscalers. **Almost no SME or innovative startup has the means to directly invest in this work.**

In a context where **Thierry Breton,** European Commissioner for the Internal Market, **wants to make Europe a leader in standards and a "global standard-setter", we believe that it is necessary to pool the actions of French players (institutions, large groups, startups, academics) and to help them participate in the European standardization effort.**

4.3. BUILDING DATA SHARING COMMONS

The data spaces are not only going to be built on digital commons, they will be ruled by the community, through stakeholders involved in the creation of the data spaces: all aspects of data sharing related to trust (consent, data sharing contracts, business agreements, etc.), but also how we are going to describe the data or resources, how we are going to control access to the data, the identities of the people and organizations, the APIs that are going to allow data to be passed around, etc.

Some components are similar for all data spaces, and they are strategic. Europe calls them smart middleware. Some are covered by Gaia-X services (GXFS). The architectures and specifications are defined for the most part in Gaia-X (in its working groups and committees requiring significant investment and effort).

Other components can be specific to a sector (e.g. the Common data spaces for skills) and call for European cooperation.

Open source plays an active role here, it is a fundamental element for technological independence and allows to create a multi stakeholder environment. These nascent phases of open source solutions are fragile and need financial support. Hyperscalers have understood this perfectly as they are able to provide significant numbers of developers on these areas. For example, an American cloud company that is involved in open source will invest substantial funds to ensure that its components are only implemented on its solutions or are only compatible with them.

The first open source constituents of data spaces will arrive by the end of 2022. **Not only will it be a matter of developing components, but the stimulation of the open source community will also be decisive** to foster adoption and implementation of all the services that will be based

[84] *"The origins, development and role of farming cooperatives in France. Aperçus sur une histoire séculaire "* Philippe Nicolas, 1988 Philippe Nicolas, 1988

on them, such as service support for the industrial sectors that are taking them on board. The Simpl[83] initiative coming from the European Commission is fully in line with this approach.

We believe that the level of investment in data sharing's digital commons will be crucial to empower European players to expand in the data space market. We need to continue to work on identifying common needs across all data spaces currently under construction and leveraging mutual investments. **If we do not act, each data space initiative will have the only option of either redesigning its own infrastructure, which would be too expensive and would stop interoperability, or to rely on non-European solutions.**

We call upon France and Europe to take charge of these issues because we cannot resign ourselves to such a strategic impasse. **It is the role of the strategist state to ensure that the infrastructures that condition and underpin the future markets are properly funded under national or European plans.** Following the example of the agricultural cooperatives in 1880, which were financed by the French state through the Crédit Agricole Mutuel and the organization of markets , or rail transport systems such as the Freycinet plan, the data market has to rely on an infrastructure of which the cost will be largely compensated by the creation of value that it will generate. Evidently, there is no question of going as far as entirely state funding as in the case of a proxy, such as the SNCF, but rather as a public-private central association such as the GSMA in the telecom sector. Today, we are at a point in the digital world that is comparable to that of the telecom industry in the early 1990s, when Europe chose to cooperate by investing (Nokia, Ericsson,...) in common standards to confront Motorola's monopoly.

That's why we need to ensure that the public markets that aim to achieve this goal are indeed able to support our ecosystems. Because having a continuous effect on the overall progress of Gaia-X, its digital commons and standards today, is to guarantee our strategic autonomy tomorrow. Recent experiences have shown that it is impossible for sectors alone to make such investments, and as we cannot bring ourselves to select American or Chinese solutions only, which is moreover a sovereign thematic of general interest, we therefore propose another path based on the mutualization and cooperation, while also supporting projects already initiated and still in need of funding (GXFS, Skills data space, etc).

Such an effort would be ecologically beneficial and economically advantageous for the entire EU territory. All this for the same amount of money as a few hundred meters of high-speed lines...

4.4. THE INFRATECH ECOSYSTEM AS A NECESSITY

In front of the hyperscalers and in the context of the European data strategy, the data spaces represent a real opportunity for Europe to develop an ecosystem of data sharing infrastructure players. The GDPR, for example, has not allowed the emergence of such a market. The American giants have seized the opportunity with resources that no European player had, and new non-European players very well-funded (eg: OneTrust + €900M) have taken advantage of the lack of support of local players and their difficulty to scale. Let's not repeat the same mistakes and do everything we can to help our players to lead the data space market.

The state can facilitate the emergence of an InfraTech ecosystem composed of, among others, startups in the infrastructure of data spaces relying on the data spaces' digital commons. Some startups already offer high-quality expertise encompassing many of the needs of the data spaces (identity / e-Wallets, legal and business model management, data

sharing intermediaries, consent and contract management for data sharing, data pipeline, general or sectoral data management, API managers, AI, etc.). However, InfraTech startups often find it very challenging to voice their opinions to hyperscalers, large corporations or institutional players, including local ones.

In the last few years, FrenchTech and its unicorns have demonstrated France's ability to create champions in business applications (HealthTech, FinTech, EdTech, etc.), but is still struggling to develop InfraTech players, who are nevertheless located at a more fundamental level of the digital value chain, in all sectors. **If we develop it at the European level, the multitude of interoperable InfraTech players, working together along common value chains and offerings (no winner takes all model), will in return allow data spaces to offer trustworthy and more sovereign projects.**

Finally, the strategist state we call for can enhance trustworthy supplies through public procurement. Everyone will observe that the considerable power acquired by the American hyperscalers has significantly flourished, over the last ten years, by public procurement of the Federal State of the United States of America, and its major agencies. As an illustration, Amazon web services maintains a leading position in cloud services, with a 32% market share as of May 2020, a position it gained in 2013 when the company won the CIA's cloud computing contract for a reported record of USD 600 million at the time. **It therefore seems essential to us that the French government's public procurement (and Member States equivalents), but also those of local authorities and operators of public networks and services,** subject to the same regulatory obligations, **be able to access, in a simple and unobstructed manner, to a "trustworthy digital" offer, carried in large part by an ecosystem of interoperable startups, which most public players also want. Such an ecosystem, which could be operated by a European public agency or operator,** would be a strong signal towards a "trustworthy Cloud-Data-AI" market, and to its public and private customers.

The funding request that we are expressing here (see right side of the figure below) is ultimately to provide France, and Europe, with the means to have weight in the implementation of Gaia-X and thus **seize the economic opportunities that it creates for our related industries and ecosystems.**

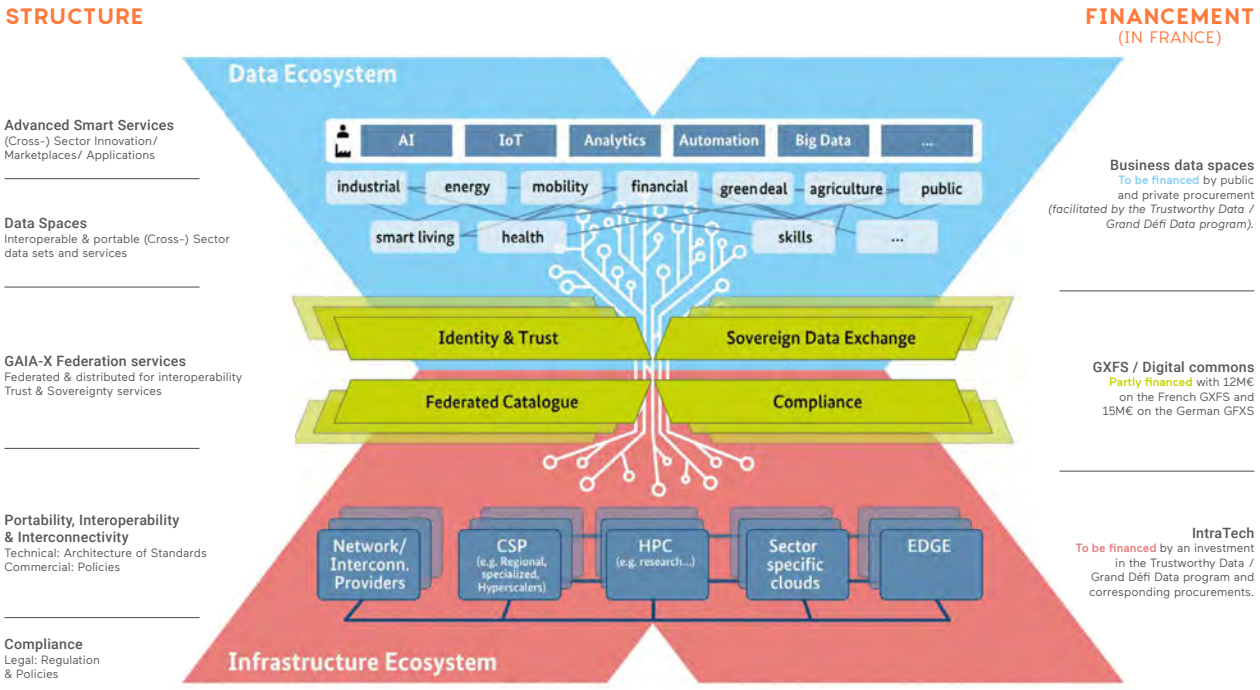
How?

• Funding an InfraTech that allows:

- To build a common cross-sector infrastructure to reduce the funding costs of the data space business
- To no longer depend on non sovereign solutions
- To enhance the digital commons

• By funding joint leadership that enables:

- To support actors for the development of standards and labels
- To create a governance structure with the resources to coordinate the entire program between the sectors and the states



"THE STRATEGIST STATE MUST ENSURE THAT THE INFRASTRUCTURES, WHICH ARE A PREREQUISITE FOR THE MARKETS OF THE FUTURE, ARE WELL FINANCED ACCORDING TO A NATIONAL PLAN."

V. PROPOSITIONS FOR A NATIONAL PROGRAM

NATIONAL TRUSTWORTHY DATA PROGRAM

This is a **Trustworthy Data Program** we call for, and for which we have identified two priority directions:

FUNDING PLAN

Funding a dedicated Research, Development and Innovation (RDI) program, supported by a full-time team, based on the 3 following principles:

PILLAR #1 STANDARDS, COMPLIANCE AND INTEROPERABILITY:

Supporting actors to participate in a coordinated way in the development of standards and compliance labels for data spaces on a European scale.

PILLAR #2 INFRATECH :

• **Digital Commons :**
- identify the digital commons data spaces needs (beyond GXFS Gaia-X), support building bricks already or under development by means of call for proposals
- Animation of the open source community
- Coordinate InfraTech and the European actors and ecosystem
- Implementation of pilots

• **InfraTech Actors:**
- Incubate/accelerate players in the infrastructure of data spaces (startups, large groups, etc.)
- Designate a place and set up online programs
- develop the use of data brokers by public actors, allowing citizens to control and share their data
- develop a "packaged" infrastructure solution based on the digital commons, the multiple solutions of InfraTech players, and stakeholders in data spaces such as the French Gaia-X Hub, TeraLab[85], competitive clusters (e.g. Cap Digital), SMEs, open source community players (e.g. Linagora) and French Tech players (Paris&Co, Lab BPI, Station F, etc.).

PILLAR #3 BUSINESS ECOSYSTEMS:

• Creation of ***data spaces*** **involving public and private** players in priority sectors in order to:
- Promote **key projects** (e.g. Energy, Green Deal,...)
- Set up **regulatory sandboxes** for these projects
- Develop a **trustworthy AI** layer for data spaces
- **Integrate pilots** of packaged infrastructure and Digital Commons.

• Facilitation and **intra and inter data spaces synergies:**
- Coordination with the **European authorities sectoral wide**
- Orchestration with the **European funding programs** (Digital Europe)
- Participation of civil society[86] (consultations, participation of institutions, ...)

Supporting data spaces through public procurement, via the incentive in tenders to:
- **Consistency with Gaia-X** (European level)
- **InfraTech actors' plurality** (startups, SMEs and large groups)
- **the reuse of the data spaces' digital commons**

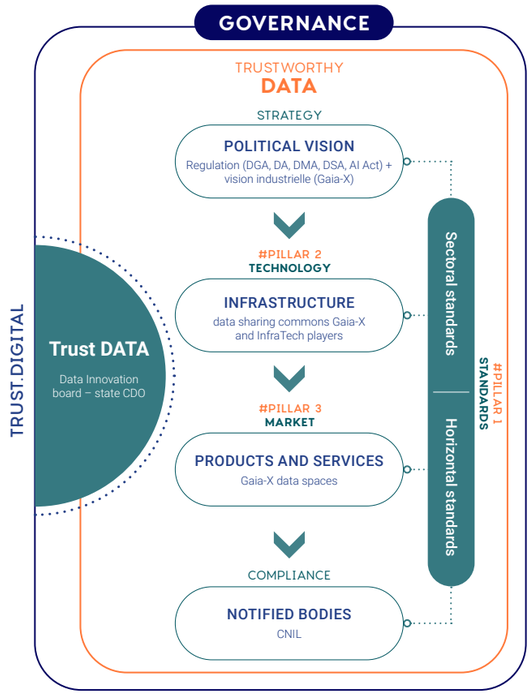
GOVERNANCE

Creation of a Trustworthy Data Governance

• **A public governance body by EU member state** is expected to be assigned to data spaces, following the example of the Spanish Data Office

• **State CDO:** Inspired by the Spanish model we propose to create the Chief Data Officer of the State (CDO) who will be responsible for:
- the orchestration of the entire trustworthy ecosystem (Cloud-Data-AI), including the supervision of data sharing and data spaces projects
- State-level public-private coordination
- representation of France in European bodies

• **Data General Assembly[87]:** In order to make people understand the opportunity for value creation in all industries, we propose to organize the Data General Assembly by industry or by theme to :
- bring together all categories of participants (institutions, large groups, SMEs/startups) to develop use cases
- Initiate the deployment of data spaces (health, mobility, government data, culture, tourism, education, etc.)



[85] https://www.teralab-datascience.en/

[86] Par example in institutions such as CNNum or CNIL, proposal from note GDPR Act II: the collective metrics of our data as impediment, Julia Roussoulie<unk> res and Jean Rérolle, Digital New Deal, 2022

[87] Like those organized by Digital New Deal and Bpifrance for loyal tourism (2 days with economic and institutional actors to arrive on a rational roadmap: White papers "Data Sharing")

CONCLUSION

This white paper is the result of multiple discussions with private and public actors involved in the creation of a sovereign and competitive French and European digital ecosystem. It attempts to express as pragmatically and as usefully as possible tech actors' expectations in this great data adventure, and trustworthy digital world in its entirety.

Stakeholders sometimes worry about our time gap, and are consequently vigilant about the technological, legal and commercial impacts that this may imply. Are we agreeing to see our strategic data, and even data of general interest (health today, education or energy tomorrow,...) entirely delegated to American or Chinese hyperscalers ? The current geopolitical turmoil invites us to avoid choosing the dangerous path of dependencies.

We are indeed at a crossroads in our history if we do not seize our opportunity now, we risk again missing the boat of the digital revolution. The European Union with its on-going regulatory package (DSA, DMA, DGA, DA, AI Act) offers the political context of liberation that we have been waiting for years, and even provides us with the operational framework, with Gaia-X, as the anchor. We won't have another momentum this favorable any time soon...

France has shown in its history that it can launch major equipment infrastructure programs (railway network for example). It is time for us **to reinstate this ambition by achieving the first major intangible infrastructure, in partnership with motivated EU players.** Therefore, we call the EU Member States to deal with the key aspects of data. From our perspective, the states should harness the emergence of a data sharing InfraTech, driven by digital commons, to support the development of the data spaces. In doing so, **to provide the technical and political foundation and impose trust as a doctrine of the digital third way.**

EU has the largest and most innovative funding instrument, with the covid crisis triggering a European Recovery and Resilience Plan of 750 billion euros, 20% of which is dedicated to a "digital dimension". This 2030 plan for "achieving digital targets" includes a new governance framework and "a facility to organize with EU Member States, projects in critical areas that are relevant to building Europe's digital transition". Now is not too late to **direct some of these billions on existing programs (Gaia-X) and missing programs (Grand Defi Data)** and create a trustworthy digital ecosystem for a stronger Europe.

CONTRIBUTORS & ACKNOWLEDGEMENTS

On the initiative of the Digital New Deal think tank, the French community, and actors of data spaces (large groups, institutions, academics, and startups), we urge a strategist state on trustworthy data sharing and data spaces issues.

AUTHORS

Olivier Dion — *Tech expert Digital New Deal, CEO Onecub, Co-founder aNewGovernance, CTO Themis -X (data space Tourism)*

Arno Pons — *General delegate Digital New Deal Think-Tank, Founder Digital New Deal Do-Tank, CEO Themis -X (data space Tourism)*

RECOMMENDATIONS AND CONTRIBUTIONS

Henri d'Agrain — *Chief Executive Cigref - Hub France Gaia-X*

Patrick Bezombes — *Chair of the IOB and Big Data Standardization Board, Co-President Workshop "Governor" - CEN CENELEC, Vice President - JTC 21 (IA) and AFNOR*

Yannick Bonhomme — *Manager valorization - IRT SystemX, Confiance.ai*

Bertrand Braunschweig — *Coordonnateur scientifique du programme Confiance.ai*

David Chassan — *Chief Strategy Officer - 3DS Outscale*

Antoine Couret — *President - Hub France IA, President - ALEIA - SaaS collaborative and sovereign AI platform*

Matthias De Bievre — *CEO Visions, President Prometheus-X, Data space Education*

Laurent Giovachini — *Chief executive of Sopra Steria*

Pierre Gronlier — *CTO Gaia-X*

Patrick Laurens-Frings — *CTO Caisse des Dépôts, Data space Finance*

Jennyfer Lecompte — *Director of Research Programs, Development & Innovation - Cap-Digital*

François Marchessaux — *Senior Data Consultant, Franz Partners*

Emmanuel Mondon — *Managing Director - Space Cooperative Europe - Space data space*

Yves Nicolas — *CTO Sopra Steria*

Michel Pauli — *Chief executive- OVHcloud*

Sébastien Picardat — *CEO - Agdatahub - Data space Agriculture*

Eric Pol — *Founder President aNewGovernance, Brussels*

Bruno Sportisse — *CEO - Inria*

Olivier Senot — *Digital Innovation Officer - Docaposte, Founding Member of Gaia-X*

Marine de Sury — *Mission Director Cigref - Hub France Gaia-X*

Anne-Sophie Taillandier — *Institut Mines Telecom - TeraLab - Founding Member of Gaia-X. Head of the Gaia-X Featured Services*

Hubert Tardieu — *Independent Administrator - Gaia-X (ex Chairman of the Board)*

Sabine Van Haecke-Lepic — *Researcher Teacher, Lawyer Web3, RegTech*

COMPLEMENTARY INTERVIEWS AND PARTICIPATIONS

Delphine Bianchini-Manno — *Tourism & Hospitality Consultant - Themis - data space Tourism*

Hervé Bonazzi — *CEO Archipels - Digital Trust & Identity, Blockchain, SelfSovereignIdentity*

Claudio Cimelli — *Direction du Numérique - National Education and Youth Ministry, coordinator DASES - Data Space Education & Skills in GAIA-X*

Jérôme Giusti — *Lawyer specialist in Intellectual Property Law, Digital Law, and Digital Common*

Tarek Habib — *CEO Murmuration - Quantification of the human impact on the environment*

Benedict Jaffeux — *Founder President - Neobridge - Training Software Editor and Training Organization*

Sylvain Le Bon — *Startin'Blox Co-founder, Solid Technology*

Arnaud Muller — *Chief executive - Cleyrop - Industrial & Modular sovereign DataHub*

Sébastien Picardat — *CEO - Agdatahub - Data space Agriculture*

Vincent Puren — *Member of Future4care, multi-disciplinary cluster of health data*

Zokama Sakanga — *Co-Founder - SMART B, blockchain data sharing*

Thomas Saint-Aubin — *CEO - Seraphin.legal - LegalTech*

Matthieu Sonnati — *President - Inokufu, co-founder - Prometheus-X - Data space Education & Competences*

Alexandre Zapolsky — *President-Founder - Linagora*

OLIVIER DION

Graduate engineer from Telecom SudParis / Hong Kong University of Science and Technology in 2007, a precursor of Open Data in France, Olivier Dion founded in 2011 the startup Onecub, and is a pioneer in the field of personal data portability and sharing. He also co-founded aNewGovernance, a key player in the data spaces, with multiple stakeholders from the European data community. He was involved in the construction of the GDPR with the French CNIL (on the right to portability), and in the European data strategy with the European Commission. He is now leading the tourism data space Themis-X.

ARNO PONS

General Delegate of the "Think-tank Digital New Deal", co-author of four notes on trustworthy digital (Trustworthy Cloud, Trustworthy Infrastructures, Trustworthy AI, and Trustworthy Data). He founded in 2021 the "Do Tank - Digital New Deal" activity dedicated to cooperation challenges and helping companies and local authorities in structuring their industries into digital ecosystems, through technological alliances. In 2022, he is launching the first initiative of the Do Tank by co-creating Themis (data space for tourism including sixty partner entities). He had previously created several start-ups (Checkfood-food waste, Medicimo in Canada, ...), and also taught at SciencesPo on the issues of digital sovereignty related to the centralization of powers by Big Tech.

DIGITAL NEW DEAL
THE THINK-TANK OF THE NEW DEAL

Digital New Deal accompanies private and public decision-makers in the creation of an Internet of the Enlightenment, European and humanistic. We are convinced that we can offer a 3rd digital way by aiming at a double objective: to defend our values by proposing a new regulation against the centralization of powers; and to defend our interests by creating the conditions of cooperation against the capture of value by the "Big Tech".

The purpose of our publication activity is to shed as much light as possible on the developments at work within the issues of "digital sovereignty", in the broadest sense of the term, and to develop concrete courses of action, even operative via the Do Tank, for economic and political organizations.

THE BOARD OF DIRECTORS

Olivier Sichel (founding president) and Arno Pons (general delegate), steer the strategic orientations of the think-tank under the supervision of the board of directors.

Strengthened by their common interest in digital issues, the members of the Board of Directors have decided to deepen their debates by formalizing a framework for production and publication within which the complementarity of their experiences can be put at the service of public and political debate. They are personally involved in the life of Digital New Deal, especially in the choice of reports and their editors. They are the guarantors of our academic and economic independence.

SÉBASTIEN BAZIN
PDG AccorHotels

NATHALIE COLLIN
General Manager, Consumer and Digital Division La Poste Group

NICOLAS DUFOURCQ
DG of Bpifrance

AXELLE LEMAIRE
Former Secretary of State for Digital Technology and Innovation

ALAIN MINC
President AM Conseil

DENIS OLIVENNES
DG Libération

YVES POILANE
DG Ionis Education Group

ARNO PONS
General Delegate of the Digital New Deal think tank

JUDITH ROCHFELD
Associate Professor of Law, Panthéon Sorbonne

OLIVIER SICHEL
President Digital New Deal
DGA Caisse des Dépôts

BRUNO SPORTISSE
PDG Inria

ROBERT ZARADER
PDG Bona fidé

OUR PUBLICATIONS

Cybersécurité, vigile de notre autonomie stratégique | Arnaud Martin, Didier Gras - *June 2022*

RGPD, acte II : la maîtrise collective de nos données comme impératif | Julia Roussoulières, Jean Rérolle – *May 2022*

Fiscalité numérique, le match retour | Vincent Renoux - *September 2021*

Défendre l'état de droit à l'ère des plateformes | Denis Olivennes and Gilles Le Chatelier - *June 2021*

Cloud de confiance : un enjeu d'autonomie stratégique pour l'Europe | Laurence Houdeville and Arno Pons - *mai 2021*

Livres blancs : Partage des données & tourisme | Fabernovel and Digital New Deal - *April 2021*

Partage de données personnelles : changer la donne par la gouvernance | Matthias de Bièvre and Olivier Dion - *September 2020*

Réflexions dans la perspective du Digital Services Act européen | Liza Bellulo - *March 2020*

Préserver notre souveraineté éducative : soutenir l'EdTech française | Marie-Christine Levet - *November 2019*

Briser le monopole des *Big Tech* : réguler pour libérer la multitude | Sébastien Soriano - *September 2019*

Sortir du syndrome de Stockholm numérique | Jean-Romain Lhomme - *October 2018*

Le Service Public Citoyen | Paul Duan - *June 2018*

L'âge du web décentralisé | Clément Jeanneau - *April 2018*

Fiscalité réelle pour un monde virtuel | Vincent Renoux - *September 2017*

Réguler le « numérique » | Joëlle Toledano - *May 2017*

Appel aux candidats à l'élection présidentielle pour un #PacteNumérique | *January 2017*

La santé face au tsunami des NBIC et aux plateformistes | Laurent Alexandre - *June 2016*

Quelle politique en matière de données personnelles ? | Judith Rochfeld - *September 2015*

Etat des lieux du numérique en Europe | Olivier Sichel - *July 2015*

contact@thedigitalnewdeal.org

october 2022

www.thedigitalnewdeal.org